Information Sciences 260 (2014) 200–208

Contents lists available at ScienceDirect

Information Sciences

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

A computational approach to construct a multivariate complete graph invariant

Matthias Dehmer [a,*], Frank Emmert-Streib [b], Martin Grabner [a]

[a] Institute for Bioinformatics and Translational Research, UMIT, Eduard Wallnoefer Zentrum 1, 6060 Hall in Tyrol, Austria
[b] 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, United Kingdom

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 29 August 2013
Accepted 1 November 2013
Available online 19 November 2013

Keywords:
Quantitative graph theory
Information inequality
Statistics
Random network model

ABSTRACT

In this paper, we present a computational approach for finding complete graph invariants. Specifically, we generate exhaustive sets of connected, non-isomorphic graphs with 9 and 10 vertices and demonstrate that a 97-dimensional multivariate graph invariant is capable to distinguish each of the non-isomorphic graphs. Furthermore, in order to tame the computational complexity of the problem caused by the vast number of graphs, e.g., involving over 10 million networks with 10 vertices, we suggest a low-dimensional, iterative procedure that is based on highly discriminative individual graph invariants. We show that also this computational approach leads to a perfect discrimination. Overall, our numerical results prove the existence of such graph invariants for networks with 9 and 10 vertices. Furthermore, we show that our iterative approach has a polynomial time complexity.

1. Introduction

Graph invariants [7,13,30,54] are quantitative measures that preserve the structural properties of graphs under isomorphism. This implies that two networks with a different structure can be distinguished from each other by such invariants unambiguously. A large number of such measures have been extensively explored and applied in various disciplines such as computational physics [17,58], structural chemistry [25,24,54], ecology [55,56] and computational linguistics [43]. In practice, graph invariants have been developed to characterize the structure of graphs by using specific structural features such as distances, vertex degrees, eigenvalues and other invariants. Also, several graph-theoretical matrices [32] such as the adjacency matrix, the distance matrix or the Laplacian matrix, to name just a few, have been utilized to derive graph invariants [9,10]. An example of graph characterization is to determine its structural complexity by using the invariants [9,10,23]. Hence, it is natural to ask for graphs that minimize or maximize an invariant. As recent work demonstrates, this problem turned out to be quite challenging when using information-theoretic graph invariants (i.e., graph entropies) [22]. Another application of graph invariants is to examine graph isomorphism [39]. However, due to the complexity of this problem, unique graph invariants for general graphs have yet not been found. For practical applications, detecting a graph isomorphism is important, e.g., for finding structurally identical molecules in chemical databases [24,26] or to examine the equivalence of electronic circuits represented by graphs [37].

A severe problem of many graph invariants is that they do not allow to discriminate graphs uniquely. That means, applying such a measure to non-isomorphic graphs, i.e., to networks that cannot be mapped to each other by application of a bijective function, may nevertheless result in the same value. This property of a graph invariant is also referred to as

* Corresponding author. Tel.: +43 (0)50 8648 3851.
E-mail addresses: matthias.dehmer@umit.at (M. Dehmer), v@bio-complexity.com (F. Emmert-Streib), martin.grabner@umit.at (M. Grabner).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.11.008

'degeneracy', 'discrimination power' or 'uniqueness' and has been investigated since decades [12,20,21,31,36,34]. Most of the existing graph invariants are so-called 'numerical graph invariants', i.e., they map graphs to real numbers by taking structural features into account. Another problem is that the degeneracy of many graph invariants has been only explored for special graph classes, for instance chemical graphs or Erdös-Rényi random networks [12,7,36,34]. Here by a special graph class, we mean a set of graphs for which each network is conform with a specific structural constraint that limits the structural complexity of all these networks within this class.

Moreover we also sketch some more up-to-date contributions dealing with graph invariants [38,46,57,59]. For instance, Liu and Liu [38] developed a structural measure based on the known graph energy formula [27] by using the Laplacian of a graph. Wang and Luo [57] proved properties thereof such as new lower bounds depending on the maximum degree. Noy [46] surveyed results for determining graphs that can be characterized by polynomials uniquely. Wu [59] explored properties of normalized Laplacian matrices which are invariant under graph isomorphism. As a result, Wu [59] found that the desired set of graph only contains the $K_{2,2}$, its complement and all complete graphs. Finally, Dehmer et al. [21] developed highly discriminating graph measures based on Shannon's entropy and presented a measure that discriminates about 98% out of nearly 12 million of graphs with 10 vertices uniquely.

However, this implies that one cannot generalize such findings to general (unconstraint) graphs, because the graph class under consideration has a considerable influence on the results [12,20,21]. For instance, graph invariants such as the well-known Balaban J index [6] and the local entropies of Konstantinova [35], possess a high uniqueness (i.e., little degeneracy) for trees, but they fail to discriminate more general networks, see [21].

Ultimately, it would be desirable to find graph invariants that describe all non-isomorphic graphs uniquely. Such a measure is called *complete graph invariant* (see Section 2). Extensive research has been performed to derive such invariants [13], however, until now, no efficiently computable complete graph invariants have been found for general (unconstraint) graphs. Finding complete graph invariants that are computable in a polynomial time complexity would be a breakthrough, as it would solve the graph isomorphism problem, whose complexity has to date still not been determined [39,51].

In this paper, we present a computational approach for finding complete multivariate graph invariants. Our method is based on applying highly discriminating topological graph invariants (numerical graph invariants) [24,30,54] to general graphs iteratively. A major contribution of this paper is that we prove the existence of complete multivariate graph invariants for graphs consisting of 9 and 10 vertices. To demonstrate the feasibility of our approach, we apply (composed) numerical graph invariants to exhaustively generated graphs that are non-isomorphic and connected with 9 and 10 vertices. We would like to note that the vast number of those graph sets (see Section 2) posses an enormous computational challenge for the discrimination power of the used numerical graph invariants, as Dehmer et al. [20] found that many invariants show a strong dependency between their discrimination power and the size of the graph set under consideration.

2. Constructing a multivariate graph invariant

2.1. Preliminaries

Let $\mathcal{G}$ be a class of graphs, i.e., $\mathcal{G}$ is a set and its elements, $G_j \in \mathcal{G}$, for $j \in I$ from an index set, are networks. A graph invariant [7,13,30,28], we call ψ, is a mapping from $\mathcal{G}$ to a d-dimensional vector of real numbers, i.e., $\psi : \mathcal{G} \longrightarrow I\!R^d, d \geqslant 1$ where from $G \cong H$ (G is isomorph to H) $\Rightarrow \psi(G) = \psi(H), G, H \in \mathcal{G}$. Here, for $d > 1$ the equality of $\psi(G) = \psi(H)$ is assessed componentwise. Furthermore, we call a graph invariant 'complete' if from $\psi(G) = \psi(H) \Rightarrow G \cong H$, for all $G, H \in \mathcal{G}$, see [13]. Note that existing topological indices [54,25] for characterizing graphs are also graph invariants. The reason for this is that these measures utilize only the underlying graph topology and do not depend on any vertex or edge labels [54]. Well known examples thereof are measures that are, e.g., based on distances, vertex degrees and eigenvalues [18,20,24].

Some failures in the scientific literature underpin the complexity of the problem to characterize graphs uniquely by using graph invariants. For instance, Spialter [52] assumed mistakenly that a polynomial closely related to the characteristic polynomial of a graph is sufficient to discriminate molecules (represented by graphs) uniquely [49]. However, this was disproven by Balaban and Harary [8]. It is interesting to note that the problem of deriving a complete graph invariant is computationally equivalent to a canonical labeling of a graph [33]. Due to the computational complexity of this problem, this implies that no polynomial-time algorithm has been found to solve this problem and, thus, the problem to find complete graph invariants with reasonable time complexity remains quite intricate.

2.2. Definition of a multivariate graph invariant

In the following, we construct a multivariate graph invariant from existing one-dimensional graph invariants and demonstrate that it is actually a complete graph invariant. We call this measure GI. Specifically, we define GI as an m-dimensional function by,

$$\mathrm{GI}[G] = (\psi_1(G), \ldots (G), \psi_m(G)), \qquad G \in \mathcal{G}, \tag{1}$$

that is a mapping from $\mathcal{G}$ to $I\!R^m$. Here, each ψ_i ($i \in [1, \ldots, m]$) is a one-dimensional graph invariant. In Tables 4 and 5, we list the graph invariants we use in this study. Due to the fact that for two graphs, G_j and G_k, the equality of $\mathrm{GI}[G_j] = \mathrm{GI}[G_k]$ is