Applied Mathematics and Computation 291 (2016) 115–121

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

A novel method for analyzing inverse problem of topological indices of graphs using competitive agglomeration

Rongling Lang[a], Tao Li[b], Desen Mo[a], Yongtang Shi[c,*]

[a] School of Electronics and Information Engineering, Beihang University, Beijing 100191, China
[b] College of Computer and Control Engineering, Nankai University, Tianjin 300071, China
[c] Center for Combinatorics and LPMC, Nankai University, Tianjin 300071, China

ARTICLE INFO

Keywords:
Graph theory
Topological index
Competitive agglomeration
Data mining

ABSTRACT

A novel method for analyzing the inverse problem of topological indices of graph is proposed in this paper, which applies the data mining technique to classification of graphs according to topological indices. Differing from the way of using topological indices to categorize graphs into different isomorphic classes, this method aims at dividing molecular graphs with n vertices into several compact classes according to their topological indices. Then the chemical and physical features of each kind of graph can be researched and new compounds can be found, which has great significance on bio-medical filed. In the experiment, three classes of simple connected graphs with 5, 6 and 7 vertices, respectively, are investigated and analyzed using the proposed method, the experimental results show the validity of it. However, there are still some problems needed to be researched further, such as selecting the vector of topological indices for clustering, selecting the distance between the vectors of topological indices, choosing the aspects to analyze the properties of each kind of graph after clustering, etc.

1. Introduction

The *molecular descriptor*, which is also called the *topological index*, is the representation of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment. The molecular descriptors is much related to the physical-chemical properties of molecules. The field of molecular descriptors is strongly interdisciplinary and involves a mass of different theories.

A topological index is a graph invariant, defined as many graph invariants are well-known, such as the number of vertices, the number of edges, vertex degree, degree sequence, matching number, independence number, chromatic number, domination number, etc [1]. In the fields of bioinformatics, molecular topology, chemical graph theory and mathematical chemistry, a topological index is a type of molecular descriptors that are calculated based on the molecular graph of a chemical compound. For the definition of molecular descriptors, a knowledge of algebra, graph theory, information theory, computational chemistry, theories of organic reactivity and physical chemistry is usually required, although at different levels. For the use of the molecular descriptors, a knowledge of statistics, chemometrics, and the principles of the QSAR/QSPR

* Corresponding author at: Center for Combinatorics and LPMC, Nankai University, Tianjin 300071, China. Tel.: +862223503682; fax: +862223509272.
E-mail addresses: ronglinglang@163.com (R. Lang), litao@nankai.edu.cn (T. Li), dason_mo@163.com (D. Mo), shi@nankai.edu.cn (Y. Shi).

http://dx.doi.org/10.1016/j.amc.2016.06.048
0096-3003/

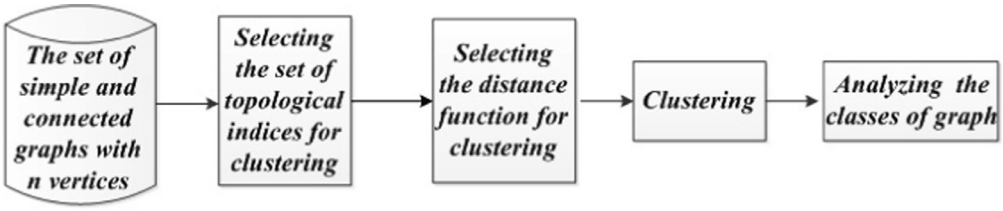

Fig. 1. Process of cluster analysis with inverse problem of topological indices.

approaches is necessary in addition to the specific knowledge of the problem. Moreover, programming, sophisticated software and hardware are often inseparable fellow-travelers of the researcher in this field. Topological indices are numerical parameters of a graph which characterize its topology and are usually graph invariants. Topological indices are used for example in the development of quantitative structure-activity relationships (QSARs) in which the biological activity or other properties of molecules are correlated with their chemical structures. The field of molecular descriptors is strongly interdisciplinary and involves a mass of different theories. We also can find some applications of topological indices in complex networks and game theory [2–5]. For more results and introductions on molecular descriptors, we refer to [1].

In the field of chemistry, it is common to calculate topological indices of the molecular graph in order to figure out the physical or chemical properties of a molecule statistically. Sometimes, the inverse problem of topological index is inevitable. For example, it happens when we expect to synthesize a compound which has a certain kind of activity. Since a molecule can be represented by its molecular graph in structure and the corresponding topological indices can reflect its physical-chemical properties, graph theory has natural advantages in studying it. Consequently, the study may narrow the scope of molecular structures, resulting in a reduced amount of experiments. In the first stage of the research on inverse problem of topological indices, the extreme value problem of topological indices of a given class of graph had been investigated. With the development of research, some researchers attempt to use some kind of topological index or a few topological indices for classification based on isomorphism [6–8], which means that the topological indices are different between non-isomorphic graphs within a certain kind of graphs. But there has been no ideal results. The main problem of classification based on isomorphism is that the topological indices may be identical even for two or several non-isomorphic graphs and it becomes worse with the increment of vertices of a graph [9]. Therefore, it is difficult to find out a set of indices to implement the classification of graphs based on isomorphism.

Given the original intention of researching, inverse problem of topological indices is to search out a molecular graph which possesses some particular kind of chemical or physical property, there is no need to obtain all the isomorphic graphs. When synthesizing a compound that has a certain boiling point or certain molecule orbital energy, only the topological indices related to these characters shall be considered. For instance, Randic index is strongly connected with the boiling point of a molecule [10,11] and energy of a graph gives a great characterization of molecule orbital energy [12,13]. Thus, the problem can be studied from a new perspective: applying the data mining technique to classification of graphs according to topological indices, and the chemical and physical features of each kind of graph can be researched and new compounds can be found, which has great significance on bio-medical filed. Differing from the way using topological indices to categorize graphs into different isomorphic classes, this method aims at dividing molecular graphs with n vertices into several compact classes according to their topological indices and then analyzes the peculiarity of each class of graphs concretely. The process is shown in Fig. 1.

There is a key issue that needs to be resolved in this process: how many classes should the simple connected graphs with n vertices be classified into? Currently, the common used methods for classification are based on a known number of classes and follow certain strategies, such as Clustering [14], Support Vector Machine [15] and Neural Network [16], etc. However, the amount of classes is unknown in this case. In order to solve this problem, a classification approach using competitive agglomeration [17,18] is presented in this paper. By taking this approach, classes with relatively small cardinality will be eliminated through competitions during the clustering process and the simple connected graphs with n vertices will be divided into C classes eventually, where C is unknown in advance.

2. Competitive agglomeration

Let x_i be a vector of topological indices of a graph G_i, $i = 1, 2, \ldots, \ell$, where ℓ is the number of non-isomorphic simple connected graphs with n vertices. For example, if the selected indices are graph entropy, Randic index and energy, then $x_i \in R^3$. The purpose of cluster is to classify $X = \{x_1, x_2, \ldots, x_\ell\}$ into C classes, where C is unknown in advance. According, the corresponding graphs G_i, $i = 1, 2, \ldots, \ell$ are classified into C classes. Each class represents an instance of graph.

Since $X = \{x_1, x_2, \ldots, x_\ell\}$ does not contain decision-making attributes, C cannot be known in advance. It is necessary to determine C during the process of classification. We exploit competitive agglomeration method to classify X and fix C, which is shown in Fig. 2.