MATCH
Communications in Mathematical and in Computer Chemistry

no. 51, April 2004
MATCDY (51) 205-210 (2004)

ISSN 0340 - 6253

The Edge Split Reconstruction Problem for Chemical Trees is NP-Complete *

Xueliang Li and Xiaoyan Zhang

Center for Combinatorics and LPMC
Nankai University
Tianjin 300071, P.R. China
E-mail: x.li@eyou.com; zhxyy@eyou.com

(Received June 30, 2003)

Abstract

A chemical tree is a tree in which there is no vertex with degree larger than 4. In this paper, we show that the *edge split reconstruction problem* for chemical trees is NP-complete. An exhaustive algorithm for the reconstruction problem is described, though it does not run in polynomial time in general.

Keywords: Chemical tree; Wiener index; Edge split; Reconstruction; NP-Complete

1 Introduction

Wiener index is one of the most popular topological indices in the quantitative structure-property relationships (QSPR) [3,8,11,12], which can be used to predict physical parameters such as boiling points, molecular volumes, heats of vaporization, etc [8,9,11,12]. One of the important problems in the theory of Wiener index is to design molecular structures with given Wiener index. As known, in the development of the theory, the study of Wiener index for trees forms a single subject of special interests. Given an integer W, is there any tree such that the Wiener index of the tree is equal to W ? This problem is called the *decision problem*. Closely related to this problem, Lepović and Gutman [6] proposed the conjecture: there is a large integer N such that for any $W \geq N$ there always exists a tree T the Wiener index of which is equal to W. If the decision problem can be solved in polynomial time, then there is an efficient (algorithmic) way to address Lepović and Gutman's conjecture. Of course, if Lepović and Gutman's conjecture can be solved, then the decision problem can be answered trivially. Unfortunately, both the decision problem and Lepović and Gutman's

*Supported by NSFC.

conjecture are still there unsolved. It seems that both of them are very difficult to solve.

For trees, the Wiener index can be calculated by its edge splits [1,7,10]. In fact, for a tree T of order n if we know the $n-1$ edge splits $s(e_1), s(e_2), ..., s(e_{n-1})$, we get the load $l(e) = s(e)(n - s(e))$ for every edge. Then the Wiener index $W(T)$ of T is the sum of all the $n-1$ edge loads, see [1,7,10]. Because of the relation between the splits and Wiener index, the decision problem naturally reminds us to ask another question: for some given integers $s(e_1), s(e_2), ..., s(e_{n-1})$, is there any tree such that the edge splits of which are exactly these $n-1$ integers ? This problem is called the *edge split reconstruction problem.*

In real world, the maximum degree of a chemical graph cannot be too large. For instance, the maximum degree of a chemical tree is at most 4 [4,9]. Chemical trees are the graph representations of alkanes, or more precisely, of the carbon atom skeleton of the molecules of alkanes. Because of this reason, throughout this paper we deal with chemical trees only, i.e., trees without any vertex with degree larger than 4. We show that the edge split reconstruction problem for chemical trees is NP-complete. In the end, an exhaustive algorithm for the reconstruction problem is described, though not efficient theoretically but useful in practice.

Before proceeding, we first give some terminology and notations which will be used in the sequel. For those not defined here, we refer to [9].

Let $G = (V, E)$ be a connected graph with vertex set $V = \{v_1, v_2, \cdots, v_n\}$. $d(v_i, v_j)$ denotes the distance between v_i and v_j in G, i.e., the length (number of edges) of a shortest path between v_i and v_j. The Wiener index $W(G)$ of G is defined as the sum of the distances of any two vertices in G, i.e., $W(G) = \sum_{i<j} d(v_i, v_j)$.

For a tree T of order n, define the split of each edge e of T as follows [10]: the removal of e results in two subtrees T_1 and T_2 of T. Then the smaller integer of the two orders of T_1 and T_2 is called the *split* of the edge e, denoted by $s(e)$. As pointed out in the above, we can calculate the Wiener index of T from the splits of its edges. First, we have a load $l(e) = s(e)(n - s(e))$ for every edge of T. Then, the Wiener index of T is equal to the sum of all the edge loads.

The notion of a famous NP-complete problem, the 3-PARTITION problem will be used in next section, which is described as follows [2]:

INSTANCE: A set A of $3m$ elements $a_1, a_2, \cdots, a_{3m}$, a bound $B \in Z^+$, and a size $s(a) \in Z^+$ for each $a \in A$ such that $B/4 < s(a) < B/2$ and such that $\sum_{a \in A} s(a) = mB$.

QUESTION: Can A be partitioned into m disjoint sets $A_1, A_2, \cdots, A_m$ such that, for $1 \leq i \leq m, \sum_{a \in A_i} s(a) = B$ (note that each A_i must therefore contain exactly 3 elements from A) ?

Now we are at the point to give our main results.

2 Main Results

In this section we show that the edge split reconstruction problem for chemical trees is NP-complete. We also describe an exhaustive algorithm to construct chemical trees for given integers as its possible edge splits, or to claim that such chemical tree does not exist.

Theorem 1 *The edge split reconstruction problem for chemical trees is NP-Complete.*

Proof. We shall use an NP-complete problem, the 3-PARTITION problem, to show that the edge split reconstruction problem for chemical trees is NP-complete.

For any "yes" instance of 3-PARTITION problem described above, we reduce it to a chemical tree with its edge splits as some integers constructed from the integers given in the 3-PARTITION problem. Without loss of generality, we can assume that $m \geq 4$, since for $m \leq 3$ the 3-PARTITION problem can be easily solved by exhausting all the possibilities of 3 element partitions. We only deal with the case for m even. The case of m odd can be dealt with in a similar way, and the details are omitted.

Let m be even, and let $A_1, A_2, \cdots, A_m$ be the 3 element sets partitioning A. We construct a chemical tree shown in Figure 1 with the following constructed integers as its edge splits: m integers of $s(e) = B + 1$, 2 integers of $s(e) = 2B + 3$, 2 integers of $s(e) = 3B + 5$, $\cdots$, 2 integers of $s(e) = (m/2 - 1)B + 2(m/2 - 1) - 1$, 1 integer of $s(e) = (m/2)B + m - 1$, 1 integer of $s(e) = s(a_1), s(e) = s(a_1) - 1, \cdots, s(e) = 1$, respectively, 1 integer of $s(e) = s(a_2), s(e) = s(a_2) - 1, \cdots, s(e) = 1$, respectively, $\cdots$, $\cdots$, 1 integer of $s(e) = s(a_{3m}), s(e) = s(a_{3m}) - 1, \cdots, s(e) = 1$, respectively. Altogether, we have constructed $2m + mB - 3$ integers from the $3m + 1$ integers $s(a_1), s(a_2), \cdots, s(a_{3m})$ and B in a 3-PARTITION problem.

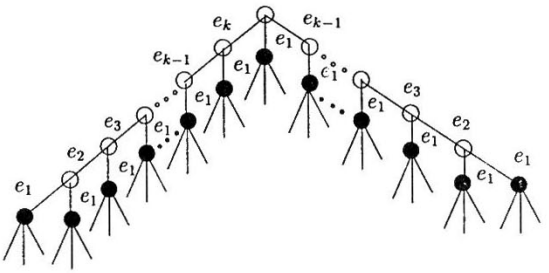

Figure 1

In Figure 1 we see the following facts:

F_1. $s(e_1) = B + 1, s(e_2) = 2B + 3, s(e_3) = 3B + 5, \cdots, s(e_{k-1}) = (m/2 - 1)B + 2(m/2 - 1) - 1, s(e_k) = (m/2)B + m - 1$, where $k = m/2$.

F_2. There are m dark vertices corresponding to the m disjoint 3 element sets $A_1, A_2, \cdots, A_m$ such that for $1 \leq i \leq m, \sum_{a \in A} s(a) = mB$.

F_3. There are 3 vertices corresponding to a_{i1}, a_{i2}, a_{i3} as the children for each dark vertex corresponding to A_i.

F_4. Starting from each vertex corresponding to $a_{ij}, j \in \{1, 2, 3\}$, a path is generated with $s(a_{ij})$ vertices.

Obviously, each of the vertices in the tree shown in Figure 1 but the dark ones has a degree smaller than 4, while each of the dark vertices has a degree equal to 4. So the tree shown in Figure 1 is a chemical tree. It is easy to see that the reduction can be done in polynomial time, and that the chemical tree shown in Figure 1 has the constructed integers as its edge splits.

Conversely, if there is a chemical tree with the above constructed integers as its edge splits, then we show that the tree must be constructed as shown in Figure 1. First, it is a crucial observation that every edge e_{it} with a value of split $s(e_{it}) := s(a_i) - t$ $(1 \leq i \leq 3m, 0 < t < s(a_i) - 1)$ must be adjacent to a unique edge e' with a value of split $s(e') = s(a_i) - t + 1$ at one end, and another unique edge e'' with a value of split $s(e'') = s(a_i) - t - 1$ at the other end, on a path departing from the edge e_i with a value of split $s(e_i) = s(a_i)$. The number of such paths must be $3m$, because there are $3m$ edges having values of splits $s(a_i), 1 \leq i \leq 3m$, respectively. Since for each $a \in A$, we have that $B/4 < s(a) < B/2$. So for any two elements a_i, a_j, we have that $s(a_i) + s(a_j) < B$, and for any four elements a_i, a_j, a_k, a_l, we have that $B < s(a_i) + s(a_j) + s(a_k) + s(a_l)$. The constraints imply that there must be 3 paths with $s(a_i) + s(a_j) + s(a_k)$ vertices departing from the edge e with a value of split $s(e) = B + 1$ such that e is incident to a dark vertex. Then from the definition of the split of $s(e)$, we have that $s(a_i) + s(a_j) + s(a_k) = B$. This indicates that every dark vertex corresponds to a 3-element set and the sum of the 3 weights is equal to B. So, a "yes" instance of 3-PARTITION problem for A is obtained.

Next, consider the edges whose splits are $B + 1, 2B + 3, 3B + 5, \cdots, (m/2)B + m - 1$. For simplicity, we call them $(B + 1), (2B + 3), (3B + 5), \cdots, [(m/2)B + m - 1]$-edges, respectively. First, it is easy to see that there must exist two $(B + 1)$-edges departing from a $(2B + 3)$-edge, since $(B + 1) + (B + 1) + 1 = 2B + 3$ is the only possible choice. For a $(3B+5)$-edge, suppose there is no $(2B+3)$-edge departing from it. Then, only $(B+1)$-edges can depart from a $(3B + 5)$-edge. There are at most three $(B + 1)$-edges that can depart from a $(3B+5)$-edge. However, $3(B+1)+1 = 3B+4 < 3B+5$, a contradiction. So, we get that there must exist a $(2B+3)$-edge departing from a $(3B+5)$-edge. Then, there is no doubt that the other edge can only be a $(B+1)$-edge since $(2B+3)+(B+1)+1 = 3B+5$. Similarly, from any edge with a value of split $2B + 3, \cdots, (m/2)B + m - 1$, respectively, we can get that every $[iB+2i-1]$-edge is adjacent to a $[(i-1)B+2(i-1)-1]$-edge and a $(B+1)$-edge.

From the above arguments, it is easily seen that the chemical tree must be reconstructed as shown in Figure 1. So, for each dark vertex of the chemical tree, there is a set of 3 splits $s(a_i), s(a_j)$ and $s(a_k)$ of the 3 edges incident to a $(B+1)$-edge, and all the 3-element sets form a 3-partition of A, or provide a solution to the "yes" instance of 3-PARTITION PROBLEM. The proof is now complete. □

Although the above theorem tells us that it is not easy to design an algorithm to decide in polynomial time whether there is a chemical tree with some given integers as its edge splits, or at least no one knows whether there is such an efficient algorithm, in real world we still hope that there is a way to settle this problem. In the following, we describe an exhaustive algorithm for constructing chemical trees with given integers as their edge splits.

Description of the Algorithm: The problem for constructing a chemical tree of order n from some given integers $s(e_1), s(e_2), \cdots, s(e_{n-1})$ as its edge splits can be solved by the following exhaustive algorithm. First, sort the integers such that $s(e_1) \geq s(e_2) \geq \cdots \geq s(e_{n-1}) = 1$. Then, starting from a tree with only one vertex of weight n, we insert the edges one at a time, ending up with a chemical tree on n vertices, each of which has a weight equal to 1. At step k we

1. look exhaustively for a vertex i with degree smaller than 4 and with a weight w_i larger than $s(e_k)$;

2. augment: attach to the vertex i a new vertex j, setting $w_j := s(e_k)$ and decreasing w_i to $w_i - s(e_k)$.

Note that at Step 1 we may have to break ties. The presence of these ties is what makes the algorithm exponential, since we may have to backtrack from a wrong choice. It is not immediately seen that this algorithm does indeed work. For instance, the sorting of the $s(e_i)$ is crucial, as the following example shows: Take $n = 4$ and $s(e_1) = s(e_2) = 1$ and $s(e_3) = 2$. Then there is no way for placing the split 2 after having placed the two splits 1. So we need to show that it is sufficient to consider the sorted permutation of the splits out of the $(n-1)!$ possibilities.

Theorem 2 *If $s(e_1) \geq s(e_2) \geq \cdots \geq s(e_{n-1})$ is a "yes" instance of the edge split reconstruction problem, then the algorithm ends up with a chemical tree with the $n-1$ integers as its edge splits.*

Proof. We may reason backwards by starting from the chemical tree and finding the correct sequence of vertices to augment. Let T be a chemical tree with edge splits $s(e_1), s(e_2), \cdots, s(e_{n-1})$. Then, give weight 1 to each vertex of T and repeat the following operation: for $k = 1$ to $n-1$, until T has only one vertex. Take a leaf i of minimum weight among the leaves. Let $(i, j(k))$ be the unique edge out of i. Delete the vertex i and increase $w_{j(k)}$ as $w_{j(k)} := w_{j(k)} + w_i$. By looking backwards at the sequence of chemical trees thus obtained, we see a possible run of the algorithm which augments on $j(n-1), \cdots, j(1)$ creating edges of decreasing splits. □

This argument also implies that for "yes" instances there always exits a choice of vertices to augment which requires no backtrack.

From Step 1 one can see that the algorithm exhausts all the possibilities to construct a chemical tree for the given integers as its edge splits. Therefore, if, at some stage, Step 1 can not find a candidate vertex i to augment, then we can definitely claim that there does not exist any chemical tree with the given integers as its edge splits. Although in general the algorithm does not run in polynomial time, it is better than nothing. For some small examples or special cases, it runs efficiently, indeed.

References

[1] A.A. Dobrynin, R. Entringer and I. Gutman, Wiener index of trees: theory and applications, Acta Appl. Math. 66(2001) 211-249.

[2] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, 1979.

[3] I. Gutman, S. Klavzar and B. Mohar eds., Fiftieth Anniversary of the Wiener Index, Discrete Applied Mathematics, Special Issue, 80(1)(1997), 122 pages.

[4] I. Gutman and O.E. Polansky, Mathematical Concepts of Organic Chemistry, Springer-Verlag, Berlin, 1986.

[5] H. Hosoya, Topological index. Anewly proposed quantity charaterizing the topological nature of structural isomers of saturated hydrocarbons, Bull. Chem. Soc. Japan 44(1971) 2332-2339.

[6] M. Lepović and I. Gutman, A collective property of trees and chemical trees, J. Chem. Inf. Comput. Sci. 38(5)(1998) 823-826.

[7] I. Lukovits and I. Gutman, Edge-decomposition of the Wiener number, MATCH Commun. Math. Comput. Chem. 31(1994) 133-144.

[8] D.E. Needham, I.C. Wei and P. G. Seybold, Molecular modelling of the physical properties of alkanes, J. Am. Chem. Soc. 110(1998) 4186-4194.

[9] N. Trinajstić, Chemical Graph Thoery, CRC Press, 1992.

[10] M.S. Waterman, Introduction to Computational Biology, Chapman & Hall, London, 1995.

[11] H. Wiener, Structural determination of pataffin boiling points, J. Am. Chem. Soc. 69(1947) 17-20.

[12] H. Wiener, Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the pataffin hydrocarbons, J. Am. Chem. Soc. 69(1947) 2636-2638.