A Unified Approach to the Extremal Trees for Different Indices *

Xueliang Li and Jie Zheng

Center for Combinatorics and LPMC

Nankai University, Tianjin 300071, P.R. China

Email: lxl@nankai.edu.cn; jzheng@eyou.com

Abstract

Many chemical indices have been invented in theoretical chemistry, such as the Wiener index, Merrifield and Simmons index, Hosoya index, Randić index and several kinds of Zagreb indices, etc. The extremal trees for these chemical indices are interested in existing literature. Though the definitions of these indices are quite different, the extremal trees have the same or very similar structures, i.e., stars or paths, but the proofs are quite diverse. This paper presents a unified and simple approach to these results by introducing two kinds of uniform transformations. Our transformations very clearly show that all the chemical indices mentioned above are indices acceptable as measures of branching.

1 Introduction

Half a century ago, in 1947, Harold Wiener introduced the first chemical index, now called the *Wiener index*. He published a series of papers [15, 16, 17, 18, 19] to

*Supported by National Science Foundation of China.

show that there are excellent correlations between the Wiener index of the molecular graph of an organic compound and a variety of physical and chemical properties of the organic compound. In the past fifty years, a large number of other chemical indices of molecular graphs, including *Merrifield and Simmons index, Hosoya index, Randić* index and *Zagreb indices*, have been proposed and widely used in chemistry. There have been many publications on these chemical indices, see [1, 3, 4, 7, 8], etc.

For convenience of our discussion, we first recall some relevant terminology and notations. For other definitions and notations not defined here, we refer to [2] and [14].

Let $G = (V(G), E(G))$ denote a molecular graph with $V(G)$ as the set of vertices and $E(G)$ the set of edges. Two vertices u and v of G are said to be adjacent if $uv \in E(G)$. We denote by $N_G(u)$ the set of vertices adjacent to u in G. A subset of $V(G)$ containing no two mutually adjacent vertices is called an *independent set*. We say that two edges of G are independent if they share no vertex in common, otherwise they are called adjacent. An *independent edge set*, also called a *match*, is defined to be a subset of $E(G)$ containing no two mutually adjacent edges.

We next list the definitions of some topological indices with which we are particularly concerned in this paper.

(1) The *Wiener index* of G, defined in [15], is

$$W(G) = \sum_{\{u,v\}} d_G(u, v), \tag{1}$$

where $d_G(u, v)$ denotes the distance between u and v in G and the sum goes over all of the unordered pairs of vertices.

(2) The *Merrifield and Simmons index* of G is defined to be the total number of independent sets of G including the empty set, denoted by $\sigma(G)$, i.e.,

$$\sigma(G) = \sum_{k \geq 0} i(G, k), \tag{2}$$

where $i(G, k)$ denotes the number of k-independent sets of G. Note that $i(G, 0) = 1$.

(3) The *Hosoya index* of G, defined in [8], is the total number of independent edge sets of G, also including the empty set, usually denoted by $Z(G)$, i.e.,

$$Z(G) = \sum_{k \geq 0} m(G, k), \tag{3}$$

where $m(G, k)$ denotes the number of k-independent edge sets. Similarly, $m(G, 0) = 1$.

(4) The *Randić index* of G is defined by

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d_G(u) d_G(v)}}, \tag{4}$$

where $d_G(u)$ denotes the degree of u in G. The *general Randić index* of a molecular graph G is defined by

$$R_\alpha(G) = \sum_{uv \in E(G)} (d_G(u) d_G(v))^\alpha, \tag{5}$$

where $\alpha \in \mathbb{R}$ and $\alpha \neq 0$.

(5) The *first Zagreb index* $M_1(G)$ and the *second Zagreb index* $M_2(G)$, defined in [1], are

$$M_1(G) = \sum_{u \in V(G)} (d_G(u))^2 \tag{6}$$

and

$$M_2(G) = \sum_{uv \in E(G)} d_G(u) d_G(v), \tag{7}$$

respectively. Motivated by the first Zagreb index and the general Randić index, as well as the generalized topological index introduced in [12], we define an index in a more general setting, called the *first general Zagreb index* and defined as follows:

$$M_\alpha(G) = \sum_{u \in V(G)} (d_G(u))^\alpha, \tag{8}$$

where $\alpha \in \mathbb{R}$, $\alpha \neq 0$ and $\alpha \neq 1$.

It is not difficult to see that there are close relations between these topological indices. For example, we can check that $R_1(G) = M_2(G)$, $m(G, 2) = \frac{1}{2}|E(G)|(|E(G)| + 1) - \frac{1}{2}M_1(G)$, and so on. Nevertheless, it is not difficult to observe from [3, 5, 6, 7, 13]

that among all the trees with n vertices, the extremal structures of different chemical indices mentioned above are in most cases S_n and P_n, the star and the path with n vertices, respectively. These results were proved by different means. Some proofs were simple but some were considerably complicated. Some authors used inductive methods but some did not. A natural question to ask is whether we can use a unified approach to find the extremal trees for different chemical indices instead of dealing with the indices separately.

The aim of the article is trying to answer the above question. Our main idea is as follows: To avoid handling the indices one by one we try to discover a transformation from a tree T to another tree T' which makes the values of these indices of the two trees monotonically change, and then by iteratively using the transformation step by step, finally we obtain the extremal tree for different indices. In Section 2, a transformation f is defined which changes a tree $T \neq S_n$ to another tree $T' = f(T)$ such that from T to T', the Wiener index decreases, the Merrifield and Simmons index increases, the Hosoya index decreases, the Randić index decreases, the second Zagreb index increases and the first general Zagreb index increases when $\alpha < 0$ or $\alpha > 1$ but decreases when $0 < \alpha < 1$. Using a similar argument, in Section 3 we find another transformation g from a tree $T \neq P_n$ to another tree $T'' = g(T)$ such that the Wiener index increases, the Merrifield and Simmons index decreases, the Hosoya index increases, the second Zagreb index decreases and the first general Zagreb index decreases when $\alpha < 0$ or $\alpha > 1$ but increases when $0 < \alpha < 1$.

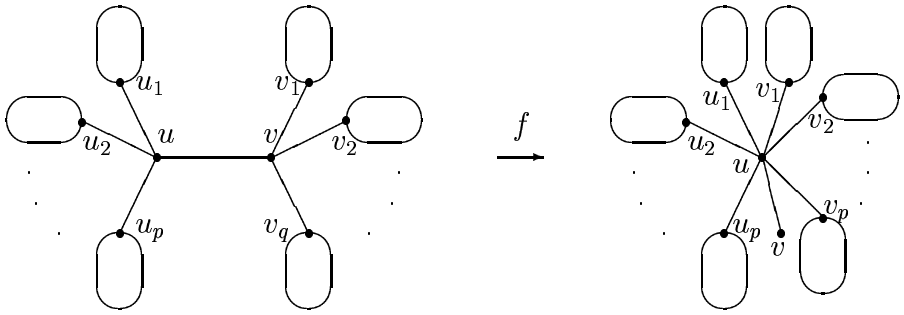
2 The star S_n as the extremal tree

Let $T \neq S_n$ be a tree with n vertices and u be a vertex with the maximum degree. Then there is an element v in $N_T(u)$ satisfying $d_T(v) > 1$. Without loss of generality, assume

$$N_T(u) = \{v, u_1, u_2, \cdots, u_p\} \text{ and } N_T(v) = \{u, v_1, v_2, \cdots, v_q\},$$

where $1 \leq p \leq n-3$ and $1 \leq q \leq p$.

We make the following changes on T: delete the edges $vv_i (i = 1, 2, \cdots, q)$ and connect $v_i (i = 1, 2, \cdots, q)$ to u. Then we get a tree T'. The following figure explains the changes. In fact, $f : T \to T'$ is a transformation on trees that are not S_n.



T $\qquad\qquad$ T'

It is not hard to check that for a given tree $T \neq S_n$, we can iterate the transformation f step by step until T is changed into S_n. The following fact is easily seen:

Theorem 2.1 *Let $I(G)$ be any index of a graph G. If $I(T) < I(f(T))$ $(I(T) > I(f(T)))$, then among all trees with n vertices S_n is the unique extremal structure with the largest (smallest) I-index.*

Our aim in this section is to prove the following crucial lemma.

Lemma 2.2 *For two trees $T (T \neq S_n)$ and $T' = f(T)$, we have*

(1) $W(T) > W(T')$.

(2) $\sigma(T) < \sigma(T')$.

(3) $Z(T) > Z(T')$.

(4) $R(T) > R(T')$.

(5) $M_2(T) < M_2(T')$.

(6) $M_\alpha(T) < M_\alpha(T')$ if $\alpha < 0$ or $\alpha > 1$; $M_\alpha(T) > M_\alpha(T')$ if $0 < \alpha < 1$.

Proof. (1) Color the vertices of T with two colors, red and blue, by the following method: Delete the edge uv and denote the remaining subgraph by $T\backslash uv$. The elements in $R = \{w|w \text{ and } u \text{ are in a same component of } T \setminus uv\}$ are colored red, while the elements in $B = \{w \neq v|w \text{ and } v \text{ are in a same component of } T \setminus uv\}$ are colored blue, and finally v is colored red. By straightforward observation, we have that from T to T' the distance between any pair of the vertices that have the same color does not change, and the distance between any pair of vertices having form $w_1w_2, w_1 \in R, w_2 \in B$ decreases 1. There are altogether $|R||B|$ such pairs. On the contrary, the distance between the pair of vertices that have form $vw_2, w_2 \in B$ increases 1. The number of such pairs is $|B|$. Since u is a vertex that has the maximum degree in T, we have $|R| > 1$. Then

$$W(T) - W(T') = |R||B| - |B| > 0.$$

(2) From T to T', an essential observation is that the independent sets in T that contain u and some elements of $\{v_1, v_2, \cdots, v_q\}$ disappear, while the independent sets that contain v and some elements of $\{v_1, v_2, \cdots, v_q\}$ appear in T'. Suppose $\{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$ is a nonempty subset of $\{v_1, v_2, \cdots, v_q\}$. Delete the vertices $u, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ and the vertices adjacent to them from T, we get $\acute{T}$. Obviously, $\sigma(\acute{T})$ is the number of independent sets in T that contain $u, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$. Similarly, we can delete the vertices $v, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ and their neighbors from T' to get $\grave{T}$ and the number of independent sets in T' that contain $v, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ is exactly $\sigma(\grave{T})$. By comparison, it is easy to get $\sigma(\acute{T}) < \sigma(\grave{T})$. Since $v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ are chosen arbitrary, the following assertion is true:

$$\sigma(T) < \sigma(T').$$

(3) From T to T', the matchings in T that contain uu_i and vv_j for some $1 \leq i \leq p$ and some $1 \leq j \leq q$ disappear and no new matchings are produced in T'. Therefore,

$$Z(T) > Z(T').$$

(4) See [9] for a detailed proof.

(5) Since $d_T(u_i) \geq 1 (i = 1, 2, \cdots, p)$ and $d_T(v_j) \geq 1 (j = 1, 2, \cdots, q)$, this allows the following argument.

$$\begin{aligned}
& M_2(T') - M_2(T) \\
= & \sum_{i=1}^{p}(p+q+1)d_T(u_i) + \sum_{j=1}^{q}(p+q+1)d_T(v_j) + (p+q+1) \\
& - \sum_{i+1}^{p}(p+1)d_T(u_i) - \sum_{j=1}^{q}(q+1)d_T(v_j) - (p+1)(q+1) \\
= & \; q\sum_{i=1}^{p} d_T(u_i) + p\sum_{j=1}^{q} d_T(v_j) - pq \\
\geq & \; pq + qp - pq = pq > 0.
\end{aligned}$$

(6) Comparing T and T', we have

$$\begin{aligned}
& M_\alpha(T') - M_\alpha(T) \\
= & \; (p+q+1)^\alpha + 1 - (p+1)^\alpha - (q+1)^\alpha \\
= & \; [(p+q+1)^\alpha - (p+1)^\alpha] - [(q+1)^\alpha - 1].
\end{aligned}$$

Using Lagrange's mean-value theorem, we conclude that there is number $y \in (p+1, p+q+1)$ such that $(p+q+1)^\alpha - (p+1)^\alpha = [(p+q+1) - (p+1)]\alpha y^{\alpha-1} = q\alpha y^{\alpha-1}$. Analogously, there is number $x \in (1, q+1)$ such that $(q+1)^\alpha - 1 = [(q+1) - 1]\alpha x^{\alpha-1} = q\alpha x^{\alpha-1}$. Hence, $M_\alpha(T') - M_\alpha(T)$ equals $q\alpha(y^{\alpha-1} - x^{\alpha-1})$. Since $q \leq p$, we have $1 < x < y$. Again, using Lagrange's mean-value theorem, we conclude that there is number $z \in (x, y)$ such that $q\alpha(y^{\alpha-1} - x^{\alpha-1}) = q\alpha(\alpha-1)(y-x)z^{\alpha-2}$. Note that $q(y-x)z^{\alpha-2}$ is a positive number and that $\alpha(\alpha-1)$ is positive if $\alpha \in (-\infty, 0) \cup (1, +\infty)$ and negative if $\alpha \in (0, 1)$, which proves the claim. ∎

From Theorem 2.1 and Lemma 2.2 we can get the following conclusion:

Corollary 2.3 *Among all trees with n vertices, S_n is the unique extremal structure with the smallest Wiener index [5], the largest Merrifield and Simmons index [13], the smallest Hosoya index [6, 7], the smallest Randić index [3], the largest second Zagreb*

index [11] and the largest general Zagreb index when $\alpha < 0$ *or* $\alpha > 1$*, the smallest general Zagreb index when* $0 < \alpha < 1$.

3 The path P_n as the extremal tree

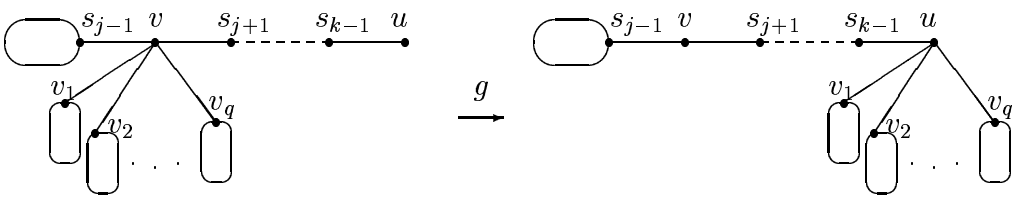
Now we continue to study another extremal tree of the chemical indices mentioned above. The result is completely similar with that in Section 2, but unfortunately, our approach cannot unify the result for the Randić index. We will demonstrate the reason in next section.

Let $T \neq P_n$ be a tree with n vertices and $(s_1, s_2, \cdots, s_k = u)$ be a longest path in T. Assume that $s_j = v$ is the vertex satisfying the following two conditions: (1) $d_T(s_j) > 2$, and (2) for any integer l such that $j < l < k$, we have $d_T(s_l) \leq 2$. Without loss of generality, suppose

$$N_T(v) = \{s_{j-1}, s_{j+1}, v_1, v_2, \cdots, v_q\},$$

where $1 \leq q \leq n-3$.

We do the following transformation g on T: Delete the edges $vv_i (i = 1, 2, \cdots, q)$ and connect every v_i to u. Then we get a new tree $T'' = g(T)$. The following figure is an illustration of g.



T $\qquad\qquad$ T''

For a given tree $T \neq P_n$, we can apply the transformation g on T step by step until it is changed into P_n. The following fact is easily seen:

Theorem 3.1 *Let $I(G)$ be any index of a graph G. If $I(T) < I(g(T))$ $(I(T) > I(g(T)))$, then among all trees with n vertices P_n is the unique extremal structure with the largest (smallest) I-index.*

The following lemma is crucial.

Lemma 3.2 *For two trees $T \neq P_n$ and $T'' = g(T)$, we have*

(1) $W(T) < W(T'')$.

(2) $\sigma(T) > \sigma(T'')$.

(3) $Z(T) < Z(T'')$.

(4) $M_2(T) > M_2(T'')$.

(5) $M_\alpha(T) > M_\alpha(T'')$ if $\alpha < 0$ or $\alpha > 1$; $M_\alpha(T) < M_\alpha(T'')$ if $0 < \alpha < 1$.

Proof. (1) Use the following method to color the vertices of T with three colors, red, blue and yellow: Delete the edges $vv_i (i = 1, 2, \cdots, q)$ and vs_{j-1}, then several components appear. Color the vertices in the same component with v red, the vertices in the same component with s_{j-1} blue and the other vertices yellow. One can observe that, from T to T'', the distance between a blue vertex and a yellow one increases. The sum of the distances between pairs of red vertices and yellow ones does not change. Furthermore, the distances between other vertices do not change, either. This complete the proof.

(2) From T to T'', the independent sets in T that contain u and some elements in the set $\{v_1, v_2, \cdots, v_q\}$ disappear, and conversely, another kind of independent sets that contain v and some elements in $\{v_1, v_2, \cdots, v_q\}$ appear in T''. Choose $\{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$ from $\{v_1, v_2, \cdots, v_q\}$, randomly. Delete $u, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ and the vertices adjacent to them from T, then we get $\dot{T}$. Clearly, $\sigma(\dot{T})$ is the number of independent sets in T containing u and $\{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$. Similarly, delete $v, v_{j_1}, v_{j_2}, \cdots, v_{j_k}$ and the vertices adjacent to them from T''. We get $\ddot{T}$, and $\sigma(\ddot{T})$ is the number of independent sets in T' containing v and $\{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$. Note that $s_{j+1}, s_{j+2}, \cdots, s_{k-1}$ are all of degree

2. By carefully comparing, we can draw the conclusion that $\sigma(\dot{T}) > \sigma(\ddot{T})$. Since $\{v_{j_1}, v_{j_2}, \cdots, v_{j_k}\}$ is chosen arbitrarily, we have

$$\sigma(T) > \sigma(T'').$$

(3) From T to T'', the matches in T that contain the edge $s_{k-1}u$ and the edge vv_j for some $1 \leq j \leq q$ disappear and two other kinds of matches appear in T''. One is the matches that contain vs_{j+1} and uv_j, and the other is the matches that contain vs_{j-1} and uv_j. Delete $s_{k-1}u$, vv_j and the adjacent edges with them from T to get the graph $\hat{T}$. Obviously, $Z(\hat{T})$ is the number of matches in T that contain $s_{k-1}u$ and vv_j. Again, delete the edges vs_{j+1}, uv_j and the adjacent edges with them from T'' to obtain $\breve{T}$, and $Z(\breve{T})$ is equal to the number of matches in T'' that contain vs_{j+1} and uv_j. Note that all the vertices between u and v are of degree 2. It is not difficult to conclude that $Z(\hat{T}) = Z(\breve{T})$. Because there actually exist matches in T'' that contain vs_{j-1} and uv_j, the following assertion is true:

$$Z(T) < Z(T'').$$

(4) Since $d_T(s_{j-1}) \geq 1$, $d_T(s_{j+1}) \geq 1$ and $d_T(v_j) \geq 1 (j = 1, 2, \cdots, q)$, we have

$$\begin{aligned}
& M_2(T) - M_2(T'') \\
= & \sum_{i=1}^{q}(q+2)d_T(v_i) + (q+2)d_T(s_{j-1}) + (q+2)d_T(s_{j+1}) + 2 \\
& - \sum_{i=1}^{q}(q+1)d_T(v_i) - 2d_T(s_{j-1}) - 2d_T(s_{j+1}) - 2(q+1) \\
= & qd_T(s_{j-1}) + qd_T(s_{j+1}) + \sum_{i=1}^{q} d_T(v_i) - 2q \\
\geq & q + q + q - 2q = q > 0
\end{aligned}$$

(5) The proof is essentially the same as that of Lemma 2.2 (6). ∎

From Theorem 3.1 and Lemma 3.2, we can get the following conclusion:

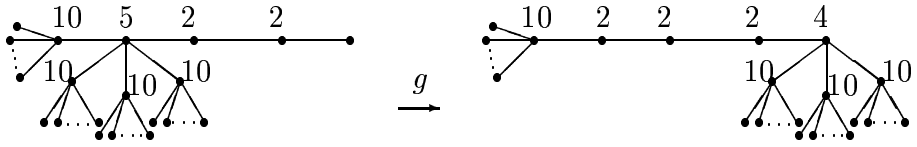
Corollary 3.3 *Among all trees with n vertices, P_n is the unique extremal structure with the largest Wiener index [5], the smallest Merrifield and Simmons index [13], the*

largest Hosoya index [6, 7], the smallest second Zagreb index [11] and the smallest general Zagreb index $M_\alpha(T)$ when $\alpha < 0$ or $\alpha > 1$, the largest general Zagreb index $M_\alpha(T)$ when $0 < \alpha < 1$.

Remark 3.4 *It is easy to see that the transformation f is to increase the number of branchings of trees, while the transformation g is to decrease the number of branchings of trees. Meanwhile, from Lemma 2.2 and Lemma 3.2, we can see that as the number of branchings changes, the values of those chemical indices change monotonically. So, all the chemical indices mentioned above are really indices acceptable as measures of branching.*

4 Further Discussion

One can notice that although the extremal tree with the maximum Randić index is P_n [10], we cannot unify the proof in Lemma 3.2. In fact, The assertion that $R(T) < R(g(T))$ is not always true. Here is a counterexample. T is the left tree in the following figure. Apply g on T to get T'', the right one in the figure. The numbers in the figure are the degrees of the non-leaf vertices.

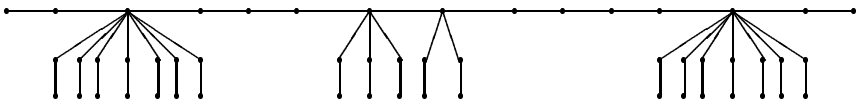


T T''

By computation, we obtain

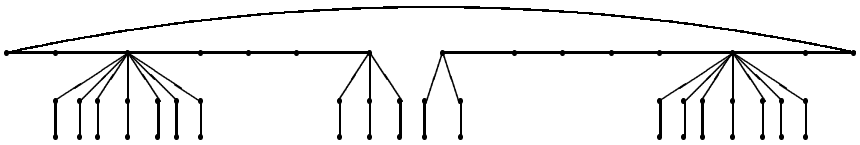
$$R(T) = 14.7381 > R(T'') = 14.7006.$$

Though if g is replaced by another transformation in [10] we can show that P_n has the maximum Randić index, we can not get the desired results for the other chemical indices in Lemma 3.2. For example, let T be the following tree:



T

By the method in [10], we change T into the following tree $\widetilde{T}$:



$\widetilde{T}$

By calculation, we have

$$R(T) < R(\widetilde{T}) \text{ but } W(T) > W(\widetilde{T}).$$

Therefore, it is natural to ask for a more unified approach that can cover extremal results for as many chemical indices as possible.

References

[1] A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Topological indices for struture-activity correlations, Topics Curr. Chem. **114**(1983), 21-55.

[2] B. Bollobás, Graph Theory: An Introductory Course, Spring-Verlag, 1979.

[3] B. Bollobás and Erdös, Graphs with extremal weights, Ars Combin. **50**(1998), 225-233.

[4] O. Chan, I. Gutman, T.K. Lam and R. Merris, Algebraic connections between topological indices, J. Chem. Inform Comput. Sci. **38**(1998), 62-65.

[5] R.C. Entringer, D.E. Jackson and D.A. Snyder, Distance in graphs, Czechoslovak Math. J. **26**(1976), 283-296.

[6] I. Gutman, Acyclic systems with extremal Hückel π-electron, Theor. Chim. Acta **45**(1977), 79-87.

[7] I. Gutman and O.E. Polansky, Mathematical Concepts in Organic Chemistry, Spring, Berlin, 1986.

[8] H. Hosoya, Topological index, a newly proposed quantity characterizing the topological nature of structure isomers of saturated hydrocarbons, Bull. Chem. Soc. Japan. **44**(1971), 2332-2339.

[9] Y. Hu, X. Li and Y. Yuan, Trees with minimum general Randic index, MATCH Commun. Math. Comput. Chem. **52**(2004), 119-128.

[10] Y. Hu, X. Li and Y. Yuan, Trees with maximum general Randic index, MATCH Commun. Math. Comput. Chem. **52**(2004), 129-146.

[11] R. Lang, X. Li and S. Zhang, Inverse problem for Zagreb index of molecular graphs, Appl. Math. J. Chinese Univ. Ser.A **18(4)**(2003), 487-493.

[12] X. Li and H. Zhao, Trees with the first three smallest and largest generalized topological indices, MATCH Commun. Math. Comput. Chem. **50**(2004), 57-62.

[13] X. Li, Z. Li and L. Wang, The inverse problems for some topological indices in combinatorial chemistry, J. Comput. Biology **10(1)**(2003), 47-55.

[14] N. Trinajstić, Chemical Graph Theory, CRC Press, 1992.

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

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

[17] H. Wiener, Influence of interatomic forces on paraffin properties, J. Chem. Phys. **15** (1947), 766-766.

[18] H. Wiener, Vapor pressure-temperature relationships among the branched paraffin hydrocarbons, J. Phys. Chem. **52** (1948), 425-430.

[19] H. Wiener, Relation of physical properties of the isomeric alkanes to molecular structure, J. Phys. Chem. **52** (1948), 1082-1089.