

# Chemical applications of graph theory

# 8

Kimberly Jordan Burch

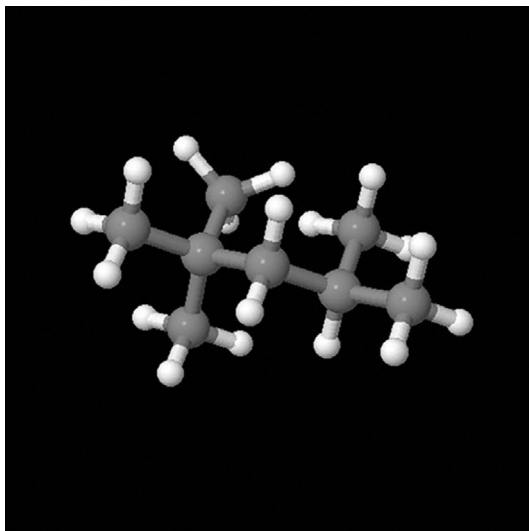
Department of Mathematics, Indiana University of Pennsylvania, Indiana, PA, United States

## 1 INTRODUCTION

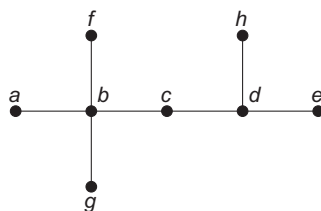
Chemical graph theory applies this branch of mathematics to model molecules in order to study their various physical properties. A *graph*  $G = (V, E)$  consists of a set  $V$  of vertices (or nodes) and a set  $E$  of unordered pairs of distinct elements of  $V$ , which are the edges. In chemistry, the atoms of a molecule are represented by the vertices and the chemical bonds are represented by the edges. The resulting graph is often called a *chemical graph*. When studying alkanes which have the chemical formula  $C_nH_{2n+2}$ , the hydrogen atoms are removed from the graph resulting in what is known as a *hydrogen-depleted molecular graph* or a *carbon tree*. Since each carbon has four bonds and each hydrogen has one bond, no information about the molecule is lost by removing the hydrogen atoms. The resulting graph is in fact easier to study since the geometric structure of the alkane is more apparent. For example, a rendering of 2,2,4-trimethylpentane is given in Fig. 1. The chemical graph of 2,2,4-trimethylpentane is given in Fig. 2.

The vertex set of the graph in Fig. 2 is  $V = \{a, b, c, d, e, f, g, h\}$ , and the edge set of this graph is given by  $E = \{ab, bc, cd, de, bf, bg, dh\}$ . The *degree* of a vertex  $v$  is the number of edges attached to vertex  $v$  and is denoted  $\deg(v)$ . In a molecular compound, this is the number of bonds an atom has and is defined as the *valency* of an atom. In Fig. 2,  $\deg(a) = \deg(f) = \deg(g) = \deg(h) = \deg(e) = 1$ ,  $\deg(b) = 4$ ,  $\deg(c) = 2$ , and  $\deg(d) = 3$ .

Given a graph  $G$  with vertices  $u, v \in V(G)$ , we define a  *$u$ - $v$  path* as a sequence of edges from  $G$  of the form  $uv_1, v_1v_2, v_2v_3, \dots, v_nv$  where each vertex  $v_i \in V(G)$  for  $1 \leq i \leq n$  is distinct. The *length* of the given  $u$ - $v$  path is  $n$ , the number of edges in the path. The *shortest path* from vertex  $u$  to  $v$  is the path for which  $n$ , the number of edges in the path, is minimal. A path from vertex  $g$  to  $h$  in Fig. 2 is given by  $gb, bc, cd, dh$  and has length four. This is also the shortest path from  $g$  to  $h$  since this is the only  $g$ - $h$  path. Given that 2,2,4-trimethylpentane is an acyclic hydrocarbon, its chemical graph contains no cycles. A *cycle* in a graph is a path that begins and ends at the

**FIG. 1**

Rendering of 2,2,4-trimethylpentane molecule.

**FIG. 2**

Chemical graph of 2,2,4-Trimethylpentane.

same vertex. Since the graph of 2,2,4-trimethylpentane contains no cycles, the path between any pair of vertices in the graph will always be the shortest path between the two vertices.

In graph theory, vertices are traditionally labeled with letters so as not to confuse the label of a vertex with the degree of a vertex. However, when the compounds studied are alkanes (also known as acyclic saturated hydrocarbons, paraffins, and the methane series), it is traditional to label the “straight” chain of the hydrocarbon with the numbers 1 to  $n$  and all branching carbons with the labels  $n+1$ ,  $n+2$ , etc. as needed in the chemical graph. For instance, the chemical graph of 2,2,4-trimethylpentane would be given as seen in Fig. 3. All chemical graphs of alkanes will be labeled in this manner from this point on in this chapter.

A graph is said to be *connected* if there is a path between every pair of vertices in the graph. Otherwise, the graph is said to be *disconnected*. If a graph is connected, the

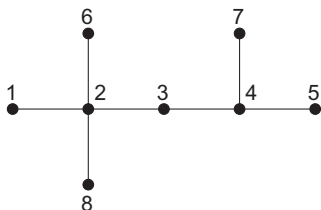


FIG. 3

2,2,4-Trimethylpentane with traditional labeling.

*distance*  $d(u, v)$  from  $u$  to  $v$  is the length of the shortest path from  $u$  to  $v$ . The *distance matrix* of a graph  $G$  having  $n$  vertices is the  $n \times n$  matrix whose  $i$ - $j$ th entry is  $d(i, j)$  and is denoted  $D(G)$ . For example, the distance matrix for 2,2,4-trimethylpentane is given by

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 2 & 2 & 4 \\ 1 & 0 & 1 & 2 & 3 & 1 & 1 & 3 \\ 2 & 1 & 0 & 1 & 2 & 2 & 2 & 2 \\ 3 & 2 & 1 & 0 & 1 & 3 & 3 & 1 \\ 4 & 3 & 2 & 1 & 0 & 4 & 4 & 2 \\ 2 & 1 & 2 & 3 & 4 & 0 & 2 & 4 \\ 2 & 1 & 2 & 3 & 4 & 2 & 0 & 4 \\ 4 & 3 & 2 & 1 & 2 & 4 & 4 & 0 \end{bmatrix}$$

When the molecular compounds being studied are alkanes as is the case with this example, the entries are the lengths of the shortest paths between any two pairs of carbons. The *adjacency matrix*,  $A(G)$ , of a labeled graph  $G$  having  $n$  vertices is the  $n \times n$  vertex whose entries are given by

$$A_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

For example, the adjacency matrix for 2,2,4-trimethylpentane is given by

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Edges are said to be *independent* if they do not share a common vertex. A *matching* of a graph  $G$  is a (possibly empty) set of independent edges of  $G$ . A  $k$ -*matching* is a matching which has exactly  $k$  edges. Letting  $a_k$  denote the

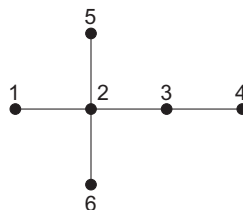


FIG. 4

2,2-Dimethylbutane.

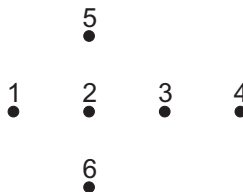


FIG. 5

0-Matching for 2,2-dimethylbutane.

number of  $k$ -matchings, the *simple matching polynomial* of a graph with  $p$  vertices is given by the formula  $\sum_{k=0}^{\lfloor p/2 \rfloor} a_k \omega^{p-k}$ . We will demonstrate how to compute the simple matching polynomial of 2,2-dimethylbutane using the chemical graph of this molecule as seen in Fig. 4.

There is only one way to choose zero edges for a matching ( $a_0 = 1$ ) as seen in Fig. 5. There are five ways to form a 1-matching since the graph of 2,2-dimethylbutane has five edges ( $a_1 = 5$ ). These 1-matchings are seen in Fig. 6. We exhibit the three possible 2-matchings of 2,2-dimethylbutane in Fig. 7. These are the only possible 2-matchings of the graph of 2,2-dimethylbutane since any other choice of two edges will not be independent ( $a_2 = 3$ ). There are no  $k$ -matchings for  $k \geq 3$  since any choice of 3, 4, 5, etc. edges will not be independent. Using the values of  $a_k$  as explained previously in the formula for the simple matching formula, the simple matching polynomial for 2,2-dimethylbutane is  $\omega^6 + 5\omega^5 + 3\omega^4$ .

## 2 TOPOLOGICAL INDICES

Physical properties of chemical compounds are highly correlated to the geometric structure of the compounds. Using indices which can accurately depict the shape of the molecules allows one to model these physical properties with surprising precision. Topological indices are graph invariant in which they do not change if the graph is drawn or labeled differently. Several topological indices are given with corresponding examples. The notation used for the indices is in agreement with the notation used by Todeschini and Consonni [1].

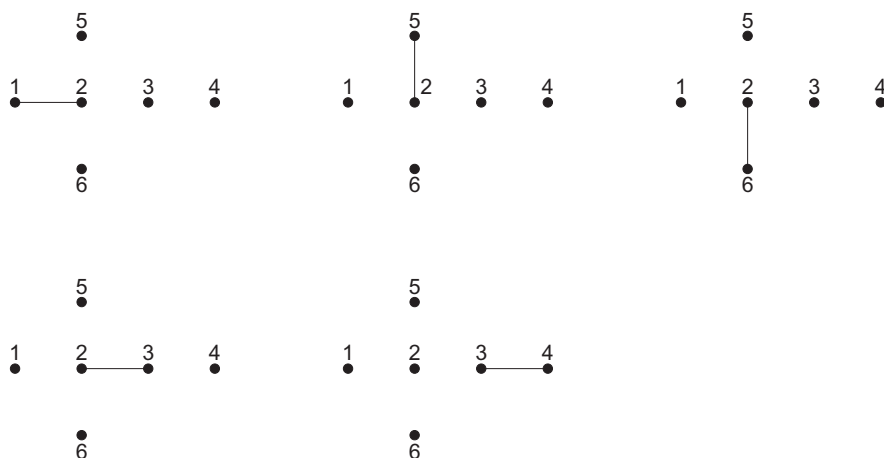


FIG. 6

1-Matchings for 2,2-dimethylbutane.

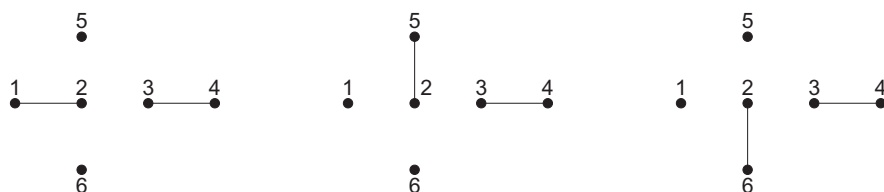


FIG. 7

2-Matchings for 2,2-dimethylbutane.

**Relative molecular mass** ( $M_r$ ) is defined as the total mass of a compound and is equal to the sum of the individual atomic masses of each atom in the molecule. Let  $n$  be the number of carbons in an alkane. The relative molecular mass of an alkane is given by the formula

$$M_r(n) = 12.01115n + 1.00797(2n + 2)$$

For example, the relative molecular mass of 2,2-dimethylbutane is 86.17848.

Perhaps the best known topological index is the relative molecular mass. For alkanes with similar geometric structure, the relative molecular mass alone may be used to determine a model of the boiling points of these alkanes.

**Hosoya index** ( $Z$ ) is the sum of the coefficients of the simple matching polynomial. The Hosoya index was introduced by Haruo Hosoya in 1971 [2] as he studied the correlation between the boiling points of alkanes and the molecular descriptor which he defined. Recall that the chemical graph of 2,2-dimethylbutane, as

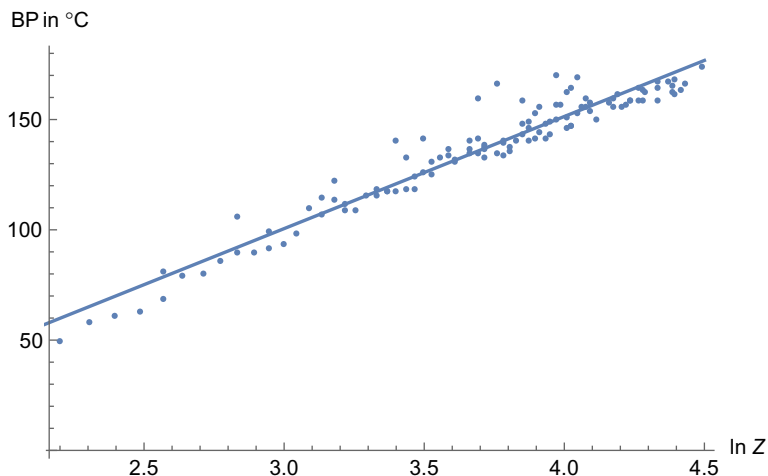
seen in Fig. 4, has exactly one matching with no edges, five 1-matchings, and three 2-matchings. These matchings were shown in Figs. 5–7, respectively. Hence, the simple matching polynomial of 2,2-dimethylbutane is  $\omega^6 + 5\omega^5 + 3\omega^4$  so that the Hosoya index of 2,2-dimethylbutane is  $1 + 3 + 5 = 9$ .

In [3] the authors present a graph of the boiling points of alkanes having six, seven, and eight carbon atoms versus the natural logarithm of the Hosoya index. This particular graph was the inspiration for the paper [4] which further investigated modeling the boiling points of alkanes. Fig. 8 is the graph of the alkanes having 6 to 10 carbon atoms versus the natural logarithm of the Hosoya index. The best linear fit is given by the equation  $f(\ln Z) = 50.8287 \ln Z - 51.9346$ . The coefficient of determination for this function, denoted  $R^2$ , is  $R^2 = 0.933748$ . The standard deviation is  $7.1^\circ\text{C}$ .

**Wiener number** ( $W$ ) is the sum of the distances between all pairs of vertices in a graph. The index was introduced in 1947 by Wiener [5] and is the oldest of the topological indices associated to chemical graph theory. Many subsequent topological indices were defined based on Wiener's work. The Wiener number is given by the formula

$$W = \frac{1}{2} \sum_i \sum_j d_{ij}$$

where  $d_{ij}$  is the  $ij$ th entry in the distance matrix,  $D(G)$  [6].



**FIG. 8**

$\ln Z$  versus BP of alkanes having 6–10 carbons.

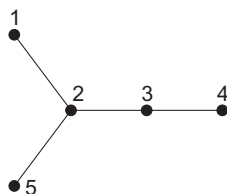


FIG. 9

2-Methylbutane.

We consider the alkane 2-methylbutane whose chemical graph is given in Fig. 9. The distance matrix of 2-methylbutane is given by

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 2 & 1 & 2 & 3 & 0 \end{bmatrix}$$

Therefore, the Wiener number for 2,2-dimethylbutane is

$$W = \frac{8(1) + 8(2) + 4(3)}{2} = 18$$

**Mean Wiener index** ( $\bar{W}$ ) is the average of the distances between all pairs of vertices in a graph. For a graph with  $n$  vertices,

$$\bar{W} = \frac{W}{\binom{n}{2}}$$

For 2,2-dimethylbutane,  $\bar{W} = \frac{18}{10} = 1.8$ .

**Wiener Path numbers** ( ${}^1P, {}^2P, \dots, {}^9P$ ) are the numbers  ${}^iP$  where  ${}^iP$  is the number of pairs of vertices in the graph separated by  $i$  edges (chemical bonds). Again using the distance matrix,  ${}^iP$  can be obtained by counting the number of times  $i$  appears in the upper triangular part of the matrix. For example, 2-methylbutane has

$${}^1P = 4, {}^2P = 4, {}^3P = 2, {}^4P = 0, {}^5P = 0, \dots, {}^9P = 0$$

When studying alkanes,  ${}^iP$  is the number of shortest paths of length  $i$  that occur in the alkane. This means that  ${}^1P$  always equals the number of bonds in an alkane.

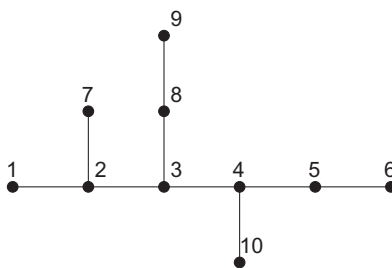


FIG. 10

3-Ethyl-2,4-dimethylheptane.

**Methyl** (*Mth*) is defined to be the number of degree one vertices, which are adjacent to a vertex of degree three or greater. This is not equivalent to the number of methyls in the IUPAC name.

Consider the chemical graph of 3-ethyl-2,4-dimethylheptane as seen in Fig. 10. We see that 3-ethyl-2,4-dimethylheptane has  $Mth = 3$  which differs from the two methyls given in the IUPAC name. The three methyls that are used in the computation of the methyl index are the edges  $\{1, 2\}$ ,  $\{2, 7\}$ , and  $\{4, 10\}$  as seen in Fig. 10.

The methyl index and the ethyl index, which will be defined shortly, were introduced in Burch et al. [4]. Since methyl is an electron donor group, the hydrogen atoms have a slightly positive charge which causes the carbon atoms where the methyl group attaches to be slightly negative. This change in polarity will have an effect on certain physical properties of the alkanes such as the boiling point.

**Ethyl** (*Eth*) is defined to be the number of degree one vertices that are adjacent to a degree two vertex, which is adjacent to a vertex of degree three or higher. This is not equivalent to the number of ethyls in the IUPAC name.

For example, 3-ethyl-2,4-dimethylheptane has  $Eth = 2$ . The first ethyl is the path  $\{3, 8\}$ ,  $\{8, 9\}$  and the second is the path  $\{4, 5\}$ ,  $\{5, 6\}$  as seen in Fig. 10.

**Randić index** ( $\chi$ ) is defined for a graph  $G$  with  $n - 1$  edges by the formula

$$\chi = \sum_{q \in E(G)} \frac{1}{\text{Geometric mean of } q}$$

where the geometric mean of an edge  $q$  joining vertices  $v_i$  and  $v_j$  is  $\sqrt{\deg(v_i) * \deg(v_j)}$  [7]. The Randić index was originally introduced as the *connectivity index* in 1975 by the chemist Milan Randić [8]. Randić noticed that this index was highly correlated with several physical properties of the



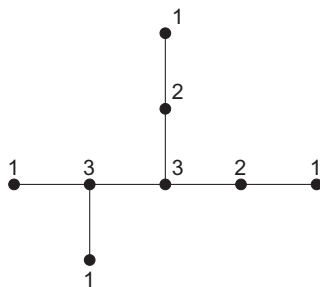


FIG. 11

3-Ethyl-2-methylpentane.

alkanes, including their boiling points. Extensive research was done in the following 25 years based on the Randić index.

We compute the Randić index for the alkane 3-ethyl-2-methylpentane. The chemical graph of 3-ethyl-2-methylpentane is given in Fig. 11 where the vertices are labeled with their corresponding degrees. Therefore, the Randić index for 3-ethyl-2-methylpentane is given by

$$\chi = 2\frac{1}{\sqrt{3}} + \frac{1}{3} + 2\frac{1}{\sqrt{6}} + 2\frac{1}{\sqrt{2}} = 3.7187$$

**Balaban index** ( $J$ ) is defined for a graph  $G$  with  $n$  vertices,  $m$  edges, and  $c$  connected components by

$$J = \frac{m}{\gamma + 1} \sum_{q \in E(G)} (D_i D_j)^{-1/2}$$

where  $\gamma = m - n + c$  is the *cyclomatic number* of the graph  $G$  and  $D_i$  is the sum of all the entries in the  $i$ th row of the distance matrix of  $G$ ,  $D(G)$ . The cyclomatic number of a graph is the minimum number of edges that must be removed from the graph in order for the graph to be acyclic. Since the alkanes are all represented by acyclic carbon trees, their corresponding cyclomatic number is zero.

Professor Alexandru Balaban introduced this index in 1982 in order to differentiate between isomers, since indices such as the Randić index failed to do so [9]. The index was originally called the *average distance sum connectivity index* and was based on the Randić index. Where the Randić index uses the degrees of the vertices, the Balaban index uses the distance between the vertices. Hence, this “highly discriminating distance-based topological index” as defined in Balaban’s paper of the same name, was critical in the advancement of chemical graph theory.

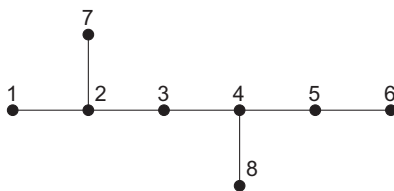


FIG. 12

2,4-Dimethylhexane.

We compute the Balaban index for the alkane 2,4-dimethylhexane whose chemical graph is given in Fig. 12. The distance matrix for the chemical graph of 2,4-dimethylhexane and the corresponding  $D_i$  are

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 2 & 4 \\ 1 & 0 & 1 & 2 & 3 & 4 & 1 & 3 \\ 2 & 1 & 0 & 1 & 2 & 3 & 2 & 2 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 & 1 \\ 4 & 3 & 2 & 1 & 0 & 1 & 4 & 2 \\ 5 & 4 & 3 & 2 & 1 & 0 & 5 & 3 \\ 2 & 1 & 2 & 3 & 4 & 5 & 0 & 4 \\ 4 & 3 & 2 & 1 & 2 & 3 & 4 & 0 \end{bmatrix} \begin{array}{l} D_1 = 21 \\ D_2 = 15 \\ D_3 = 13 \\ D_4 = 13 \\ D_5 = 17 \\ D_6 = 23 \\ D_7 = 21 \\ D_8 = 19 \end{array}$$

Using the fact that  $\gamma = 7 - 8 + 1 = 0$  and  $m = 7$  for the graph of 2,4-dimethylhexane, we compute that the Balaban index for 2,4-dimethylhexane is given by

$$J = 7 \left( \frac{2}{\sqrt{15 \cdot 21}} + \frac{2}{\sqrt{15 \cdot 13}} + \frac{1}{\sqrt{13 \cdot 19}} + \frac{1}{\sqrt{13 \cdot 27}} + \frac{1}{\sqrt{17 \cdot 23}} \right) = 3.0988$$

**Harary index** ( $H$ ) is the sum of the inverse of the distance between all pairs of vertices in a graph. The index was introduced by Plavšić et al. in 1991 at a conference held in honor of Professor Frank Harary's 70th birthday and was thus named the Harary index [10]. This index attempts to model the fact that atoms closer together in a molecule should have a greater influence on each other than atoms that are separated by a large distance. The formula for the Harary index of a graph  $G$  having  $n$  vertices is

$$H = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (D')_{ij}$$

where  $D'$  is the *reciprocal distance matrix* of a graph  $G$  defined by

$$(D')_{ij} = \frac{1}{D_{ij}}; \quad i \neq j$$

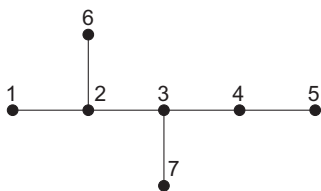


FIG. 13

2,3-Dimethylpentane.

We compute the Harary index for the alkane 2,3-dimethylpentane whose chemical graph is given in Fig. 13.

The distance matrix and the reciprocal distance matrix for the chemical graph of 2,3-dimethylpentane are as follows:

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 2 & 3 \\ 1 & 0 & 1 & 2 & 3 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 & 2 & 1 \\ 3 & 2 & 1 & 0 & 1 & 3 & 2 \\ 4 & 3 & 2 & 1 & 0 & 4 & 3 \\ 2 & 1 & 2 & 3 & 4 & 0 & 3 \\ 3 & 2 & 1 & 2 & 3 & 3 & 0 \end{bmatrix} \quad D'(G) = \begin{bmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{2} & \frac{1}{3} \\ 1 & 0 & 1 & \frac{1}{2} & \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 0 & 1 & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{2} & 1 & 0 & 1 & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{3} & 1 & 0 & 1 & \frac{1}{4} & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}$$

Therefore, the Harary index for 2,3-dimethylpentane is

$$H = \frac{12(1) + 14\left(\frac{1}{2}\right) + 12\left(\frac{1}{3}\right) + 4\left(\frac{1}{4}\right)}{2} = 12$$

**Schultz index or Molecular Topological index (MTI)** is defined for a graph  $G$  having  $n$  vertices by

$$MTI = \sum_{j=1}^n e_j$$

where

$$e_j = \sum_{i=1}^n v_i(A_{ij} + D_{ij})$$

where  $v_i$  is the degree (valence) of vertex  $i$  in the graph  $G$ .

We compute the Schultz index for the alkane 2,3-dimethylbutane whose chemical graph is given in Fig. 14. The adjacency matrix for the chemical graph of 2,3-dimethylbutane is

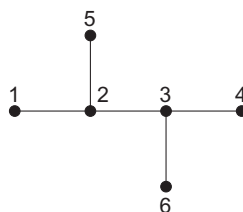


FIG. 14

2,3-Dimethylbutane.

$$A(G) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The distance matrix for 2,3-dimethylbutane is

$$D(G) = \begin{bmatrix} 0 & 1 & 2 & 3 & 2 & 3 \\ 1 & 0 & 1 & 2 & 1 & 2 \\ 2 & 1 & 0 & 1 & 2 & 1 \\ 3 & 2 & 1 & 0 & 3 & 2 \\ 2 & 1 & 2 & 3 & 0 & 3 \\ 3 & 2 & 1 & 2 & 3 & 0 \end{bmatrix}$$

The valency vector for 2,3-dimethylbutane is  $[1 \ 3 \ 3 \ 1 \ 1 \ 1]$ . To compute each of the  $e_i$  from earlier we take

$$\begin{aligned} v(G)[A + D](G) &= [1 \ 3 \ 3 \ 1 \ 1 \ 1] \begin{bmatrix} 0 & 2 & 2 & 3 & 2 & 3 \\ 2 & 0 & 2 & 2 & 2 & 2 \\ 2 & 2 & 0 & 2 & 2 & 2 \\ 3 & 2 & 2 & 0 & 3 & 2 \\ 2 & 2 & 2 & 3 & 0 & 3 \\ 3 & 2 & 2 & 2 & 3 & 0 \end{bmatrix} \\ &= [20 \ 14 \ 14 \ 20 \ 20 \ 20] \end{aligned}$$

Hence, the *MTI* for 2,3-dimethylbutane is  $4 \cdot 20 + 2 \cdot 14 = 108$ .

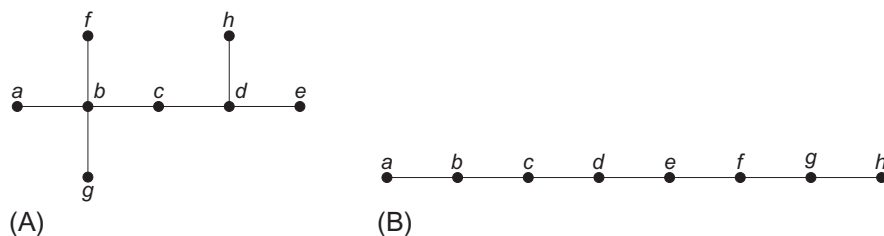
**Methyl Locant** (*loc*) is defined for alkanes having one and only one methyl branch. It is the number of the carbon to which the methyl branch is attached to the straight chain whose carbons are numbered 1 to  $n - 1$  consistent with IUPAC nomenclature. The methyl locant number is defined as 0 for the normal alkanes. For example, 4-methylheptane has *loc* = 4. This index was defined to model the melting points of a set of alkanes having a single methyl group which are used in synthetic fuel as seen in [Section 3](#).

### 3 MODELS

Extensive research had been conducted using the topological indices described here to model physical properties of many different types of chemical molecules. Using combinations of these indices has produced multivariable models of physical properties which may be used to determine said physical properties which are not available by direct measurement.

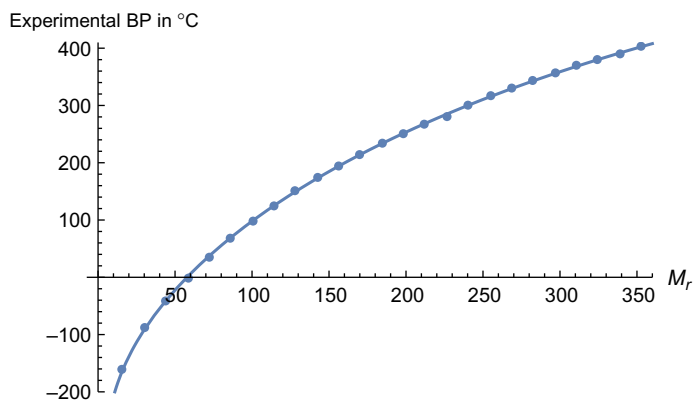
Experimental boiling point data are available for most alkanes having less than or equal to 10 carbon atoms [11,12]. It is unlikely that comprehensive boiling point data will be available for all alkanes having more than 10 carbon atoms. This has motivated several studies to determine successful models for the boiling points which may be used to predict the boiling points for which no data are available. The boiling point of an alkane depends strongly on the geometric structure of the alkane since boiling points are a measure of the forces of attraction between individual atoms or molecules. Given that alkanes are essentially nonpolar compounds, these forces are London dispersion forces due to instantaneous dipole-induced dipole attractions. Since the forces are weak, they depend on contact between molecules. This means that the forces increase with surface area and therefore the boiling point of an alkane will increase as the relative molecular mass increases. The geometry of the alkane also plays a pivotal role in determining its boiling point. The dependence is complex, but the boiling point of alkanes with the same relative molecular weight will generally decrease as the compactness of the molecules increases. Balaban noted in his 2001 paper that boiling points decreased for alkanes with increased branching for the same relative molecular mass [13].

We examine this phenomenon by viewing the structure of 2,2,4-trimethylpentane and octane. Both acyclic hydrocarbons are composed of eight carbons and therefore have the same relative molecular mass. The geometric structure of these two alkanes differ greatly as seen by their chemical graphs in Fig. 15. We note that 2,2,4-trimethylpentane is clearly a more compact molecule than the straight chain octane and we therefore expect the boiling point of 2,2,4-trimethylpentane to be lower than that of octane. This is indeed the case since the boiling point of 2,2,4-trimethylpentane is 99.3°C whereas the boiling point of octane is 125.6°C.



**FIG. 15**

Comparing the geometric structure of 2,2,4-trimethylpentane and octane.  
(A) 2,2,4-Trimethylpentane; (B) octane.



**FIG. 16**

$M_r$  versus BP of  $n$ -alkanes for  $1 \leq n \leq 25$ .

When studying a family of alkanes with similar geometric structure, a single topological index can be used to model the boiling point of the alkanes. The normal alkanes or  $n$ -alkanes are one such family in which their molecular graphs are the path graphs having  $n$  vertices. Although these molecules are simple, their boiling points do not increase in a linear fashion with respect to their length. The boiling points of the first few normal alkanes increase rapidly because the London forces increase with the relative molecular weight which are the forces holding the alkane together in liquid form. As the number of carbons increase, the chain will fold back on itself causing the London forces to be directed toward themselves and therefore diminish the dependence of the boiling point on the relative molecular weight. Therefore, a model using the relative molecular mass fits the boiling points of the  $n$ -alkanes for  $1 \leq n \leq 25$  quite well as seen in Fig. 16. This model was developed in Burch et al. [4] using Mathematica. The equation of the model is  $f(M_r) = -240.103 - 330.713M_r^{-0.797084} + 43.9249M_r^{0.793242} - 26.3154M_r^{0.854845}$  where  $R^2 = 0.999925$  and the standard deviation is  $1.2^\circ\text{C}$ .

Developing models to predict the physical properties of all the alkanes necessitates using a combination of the topological indices described in order to effectively represent the geometric structures of the different alkanes. Multivariable models were developed to model the boiling points of alkanes having 6–12 carbon atoms. These models were then used to predict the boiling points of alkanes having 13–22 carbon atoms. Similar research was conducted using neural nets to predict the boiling points of the alkanes having 6–10 carbon atoms [14]. Using neural nets to predict physical properties does not provide equations for the models. Therefore, models were developed in Burch et al. [4] using the NEOS solver FilterSQP available online from Argonne National Laboratory [15]. FilterSQP is a sequential quadratic programming algorithm that optimizes the nonlinear function by using a “filter” to promote global convergence. FilterSQP uses second derivative information and the

constraints in such a way that it is less prone to converge to shallow local minima. By using FilterSQP, we were able to construct equations for our boiling point models. These models can be used to predict the boiling points of alkanes for which no experimental boiling point data exist.

The first boiling point model was developed using all the alkanes having 1–12 carbon atoms for which boiling point data were available and the normal alkanes having 13–22 carbons. This range of normal alkanes was included in the model's dataset because the models were used to predict the boiling points of a set of miscellaneous alkanes having 13–22 carbon atoms. Table 1 gives the boiling point data and corresponding source for the alkanes having 1–12 carbon atoms. Model 1 was developed using the Wiener path numbers from one to eight and the methyl indices. The number of alkanes in the dataset for Model 1 is 195 and  $R^2 = 0.995322$ . The standard deviation is 4.3°C.

**Model 1.** This model uses the Wiener  $iP$  numbers ( $1 \leq i \leq 8$ ) and the Methyl index. The formula is

$$\begin{aligned} f_{\text{Model 1}}({}^1P, {}^2P, \dots, {}^8P, Mth) \\ = -167.49997 + 84.28344({}^1P)^{0.46072} + 15.94534({}^2P)^{0.00348} \\ + 17.42198({}^3P)^{0.53517} + 11.16515({}^4P)^{0.00164} + 4.74582({}^5P)^{0.00089} \\ + 5.23270({}^6P)^{0.00143} + 6.67018({}^7P)^{0.14687} + 5.27829({}^8P)^{0.96677} \\ - 5.53723Mth^{0.00072} \end{aligned}$$

The absolute boiling point deviation, denoted BP dev., is given by the formula BP dev. = |Model BP – Experimental BP|. Table 2 gives the number and percentage of alkanes whose absolute boiling point deviations determined by Model 1 are within the specified ranges.

By adding the Hosoya index as one of the variables in the boiling point model, the standard deviation improved by over a degree. Model 2 uses the same 195 alkanes as its dataset and yields  $R^2 = 0.997648$ . The standard deviation of Model 2 shows an improvement by 3.0°C.

**Model 2.** This model uses the Wiener  $iP$  numbers ( $1 \leq i \leq 8$ ), the Methyl index, and the Hosoya index. The formula is

$$\begin{aligned} f_{\text{Model 2}}({}^1P, {}^2P, \dots, {}^8P, Mth, Z) \\ = -157.39801 - 1.71578({}^1P)^{2.06741} + 0.38684({}^2P)^{1.98738} \\ - 3.53937({}^3P)^{0.73904} - 4.69814({}^4P)^{0.04224} - 2.31936({}^5P)^{0.85500} \\ - 0.00004({}^6P)^{5.13758} + 1.57097({}^7P)^{0.57937} \\ + 0.23048({}^8P)^{2.70854} + 1.20993Mth^{1.32159} + 108.36449(\ln Z)^{1.00972} \end{aligned}$$

Table 3 gives the number and percentage of alkanes whose absolute boiling point deviations determined by Model 2 are within the specified ranges.

Table 1 Boiling Point Data

Alkane	BP (C)	Source	Alkane	BP (C)	Source	Alkane	BP (C)	Source
Methane	-162.15	NIST	<i>n</i> -Decane	174.15	NIST	<i>n</i> -Dodecane	215.85	NIST
Ethane	-88.55	NIST	2-Methylnonane	166.94	NIST	2,2,4,6,6-Pentamethylheptane	177.85	NIST
Propane	-42.15	NIST	3-Methylnonane (DL)	167.75	NIST	2-Methylundecane	210.05	NIST
Butane	-0.15	NIST	4-Methylnonane (DL)	166.85	NIST	3,3,6,6-Tetramethyloctane	189.75	NIST
2-Methylpropane	-11.15	NIST	5-Methylnonane	165.11	NIST	2,2,4,7-Tetramethyloctane	184.7	Beilstein
<i>n</i> -Pentane	36.05	NIST	3-Ethyloctane	166	Beilstein	2,2,7,7-Tetramethyloctane	186.85	NIST
2-Methylbutane	27.03	Beilstein	4-Ethyloctane v168	163.36	NIST	2,2,4,4,6-Pentamethylheptane	185.05	NIST
2,2-Dimethylpropane	9.45	NIST	2,2-Dimethyloctane	154	Beilstein	3-Ethyl-2,3,5-trimethylheptane	197.8	Beilstein
<i>n</i> -Hexane	68.75	NIST	2,3-Dimethyloctane	164.24	NIST	3-Ethyl decane	208.85	Beilstein
2-Methylpentane	60.85	NIST	2,4-Dimethyloctane	155.9	NIST	4-Ethyldecane	204.5	Beilstein
3-Methylpentane	63.25	NIST	2,5-Dimethyloctane	156.93	NIST	3-Methylundecane	211.2	Beilstein
2,2-Dimethylbutane	49.75	NIST	2,6-Dimethyloctane	158.85	NIST	4-Methylundecane	207.1	Beilstein
2,3-Dimethylbutane	58.05	NIST	2,7-Dimethyloctane	160.05	NIST	5-Methylundecane	204.6	Beilstein
<i>n</i> -Heptane	98.45	NIST	3,3-Dimethyloctane	161.2	NIST	4,5-Diethyloctane	200.5	Beilstein
3-Ethylpentane	93.45	NIST	3,5-Dimethyloctane	158.85	NIST	2,4,5,7-Tetramethyloctane	186.25	NIST
3-Methylhexane	91.82	Beilstein	4,4-Dimethyloctane	157.5	NIST	2,2-Dimethyldecane	199	Beilstein
2-Methylhexane	90.05	NIST	4,5-Dimethyloctane v162.13	162.42	NIST	2,3-Dimethyldecane	203.4	Beilstein
2,3-Dimethylpentane	89.75	NIST	4-Propylheptane	161.85	NIST	2,4-Dimethyldecane	198.5	Beilstein
3,3-Dimethylpentane	86.05	NIST	4-Isopropylheptane	158.8	Beilstein	5,6-Dimethyldecane	197.5	Beilstein



2,4-Dimethylpentane	80.55	NIST	3-Ethyl-3-methylheptane	163.85	NIST	3-Ethyldecane	208.8	Beilstein
2,2-Dimethylpentane	79.15	NIST	5-Ethyl-3-methylheptane	158.3	Beilstein	2,6-Dimethyl-3-isopropylheptane	186.85	NIST
2,2,3-Trimethylbutane	80.95	NIST	4-Ethyl-4-methylheptane	159	Beilstein	2,2,3,5,6-Pentamethylheptane	188.8	NIST
<i>n</i> -Octane	125.55	NIST	2,2,4-Trimethylheptane	148.85	NIST	2,2,6,6-Tetramethyloctane	182.75	NIST
3-Methylheptane	118.85	NIST	2,2,5-Trimethylheptane	146	Beilstein	2,4,4,7-Tetramethyloctane	183.85	NIST
3-Ethylhexane	118.55	NIST	2,2,6-Trimethylheptane	148.33	NIST	2,2,3,4,5,5-Hexamethylhexane	191.85	NIST
3-Ethyl-3-methylpentane	118.35	NIST	2,3,3-Trimethylheptane	160.05	NIST	2-Methyl-5-propyloctane	188.85	NIST
3,4-Dimethylhexane	117.85	NIST	2,3,4-Trimethylheptane	155.85	NIST	2-Methyl-3-isopropyloctane	197.7	Beilstein
4-Methylheptane	117.75	NIST	2,3,6-Trimethylheptane	149.85	NIST	2,2,4-Trimethylnonane	191.6	Beilstein
2-Methylheptane	117.55	NIST	2,4,4-Trimethylheptane	149.55	NIST	2,4,6-Trimethylnonane	186	Beilstein
3-Ethyl-2-methylpentane	115.65	NIST	2,4,6-Trimethylheptane v144.8	146.75	NIST	2,5,8-Trimethylnonane	189.85	NIST
2,3-Dimethylhexane	115.85	NIST	2,5,5-Trimethylheptane	152.8	NIST	3,6-Diethyloctane	202.85	NIST
2,3,3-Trimethylpentane	114.65	NIST	3,3,5-Trimethylheptane	155.67	NIST	3,3-Diethyloctane	204.85	NIST
2,3,4-Trimethylpentane	113.65	NIST	3,3-Diethylhexane	164.48	NIST	4-Ethyl-5-methylnonane	194.5	Beilstein
3,3-Dimethylhexane	111.95	NIST	3,4-Diethylhexane	162.75	NIST	5-Ethyl-5-methylnonane	196	Beilstein

(Continued)

Table 1 Boiling Point Data—Cont'd

Alkane	BP (C)	Source	Alkane	BP (C)	Source	Alkane	BP (C)	Source
2,2,3-Trimethylpentane	109.95	NIST	2,2-Dimethyl-4-ethylhexane	147.05	NIST	4-Propylnonane	202.5	Beilstein
2,4-Dimethylhexane	108.85	NIST	2,4-Dimethyl-4-ethylhexane	158	Beilstein	5-Propylnonane	204.5	Beilstein
2,5-Dimethylhexane	108.95	NIST	2,2,3,3-Tetramethylhexane	158.85	NIST			
2,2-Dimethylhexane	106.85	NIST	2,2,3,4-Tetramethylhexane	156.65	NIST			
2,2,3,3-Tetramethylbutane	106.45	NIST	2,2,3,5-Tetramethylhexane	148.65	NIST			
2,2,4-Trimethylpentane	99.25	NIST	2,2,4,5-Tetramethylhexane	148.04	NIST			
<i>n</i> -Nonane	150.65	NIST	2,2,5,5-Tetramethylhexane	137.17	NIST			
3-Ethylheptane	143.27	NIST	2,3,3,4-Tetramethylhexane	164.59	NIST			
4-Ethylheptane	141.3	NIST	2,3,3,5-Tetramethylhexane	152.85	NIST			
3-Methyloctane	143.85	NIST	2,3,4,4-Tetramethylhexane	162.25	NIST			
4-Methyloctane	141.85	NIST	2,3,4,5-Tetramethylhexane	155.45	NIST			
3,3-Diethylpentane	146.15	NIST	3,3,4,4-Tetramethylhexane	170.32	NIST			
4-Ethyl-3-methylhexane	140.4	NIST	3-Isopropyl-2,4-dimethylpentane	157.04	NIST			
2-Methyloctane	142.95	NIST	2,2,4-Trimethyl-3-ethylpentane	155.3	NIST			

3,4-Dimethylheptane	140.45	NIST	2,3,4-Trimethyl-3-ethylpentane	169.44	NIST		
3-Ethyl-2-methylhexane	138	NIST	2,2,3,3,4-Pentamethylpentane	166.05	NIST		
3,5-Dimethylheptane	135.61	NIST	2,2,3,4,4-Pentamethylpentane	159.29	NIST		
3-Ethyl-3-methylhexane	140.6	NIST	Undecane	194.85	NIST		
2,3-Dimethylheptane	139.85	NIST	2-Methyldecane	189.15	NIST		
4-Ethyl-2-methylhexane	133.8	NIST	3-Methyldecane	190.9	Beilstein		
2,5-Dimethylheptane	134.85	NIST	4-Methyldecane	186.95	NIST		
2,3,4-Trimethylhexane	138.85	NIST	5-Methyldecane	186.8	Beilstein		
3,3-Dimethylheptane	136.85	NIST	4,4-Diethylheptane	183.35	NIST		
2,4-Dimethylheptane	132.85	NIST	4-Ethyl-2,4-dimethylheptane	148.85	NIST		
3-Ethyl-2,3-dimethylpentane	141.6	NIST	2,2,3,5,5-Pentamethylhexane	165.15	NIST		
2,6-Dimethylheptane	134.85	NIST	2,2,4,6-Tetramethylheptane	162.05	NIST		
3,3,4-Trimethylhexane	140.16	NIST	2,2,6,6-Tetramethylheptane	160.3	Beilstein		
3-Ethyl-2,4-dimethylpentane	136.66	NIST	2,2,3,3,5-Pentamethylhexane	173.05	NIST		
4,4-Dimethylheptane	134.93	NIST	2,2,4-Trimethyloctane	170.05	NIST		
2,2-Dimethylheptane	131.85	NIST	2,4,6-Trimethyloctane	168	Beilstein		

(Continued)

Table 1 Boiling Point Data—Cont'd

Alkane	BP (C)	Source	Alkane	BP (C)	Source	Alkane	BP (C)	Source
2,3,5-Trimethylhexane	130.85	NIST	2,2-Dimethylnonane,	186	Beilstein			
2,3,3-Trimethylhexane	137.08	NIST	2,3-Dimethylnonane	186.8	Beilstein			
3-Ethyl-2,2-dimethylpentane	133.75	NIST	2,4-Dimethylnonane	178	Beilstein			
2,2,3-Trimethylhexane	133.12	NIST	3-Isopropyl-2-methyl- heptane	178.1	Beilstein			
2,4,4-Trimethylhexane	130.85	NIST	2,3,4,6- Tetramethylheptane	173.55	NIST			
2,3,3,4-Tetramethylpentane	141.5	NIST	2,2,3,3- Tetramethylheptane	179.65	NIST			
2,2,4-Trimethylhexane	126.45	NIST	4-Ethyl-3,4- dimethylheptane	180.85	NIST			
2,2,5-Trimethylhexane	123.85	NIST	3-Ethyl-3- methyloctane	185.75	NIST			
2,2,3,4-Tetramethylpentane	133	NIST	3,3-Diethylheptane	186.75	NIST			
2,2,3,3-Tetramethylpentane	140.27	NIST	4-Propyloctane	182.6	Beilstein			
2,2,4,4-Tetramethylpentane	122.25	NIST						

**Table 2** Alkanes and Their Corresponding Absolute BP Deviations for **Model 1**

BP Dev. (degrees)	No. of Alkanes	Alkanes (%)
0–1	49	25.1
1–2	35	17.9
2–4	52	26.7
4–6	32	16.4
6–9	22	11.3
9–11	2	1.0
>11	3	1.5

**Table 3** Alkanes and Their Corresponding Absolute BP Deviations for **Model 2**

BP Dev. (degrees)	No. of Alkanes	Alkanes (%)
0–1	65	33.3
1–2	54	27.7
2–4	53	27.2
4–6	18	9.2
6–9	4	2.1
9–10	0	0
>10	1	0.5

In examining **Models 1** and **2**, we found that the first eight alkanes having one to five carbon atoms greatly constrained the intercept and the coefficients and exponents of some of the indices. This was especially true with the indices  ${}^1P$ ,  ${}^2P$ , and  $Z$ . This motivated us to exclude these eight alkanes from our data set and to construct **Model 3**. The dataset for this model is therefore the set of alkanes having 6–12 carbon atoms for which boiling point data are available and consists of 187 alkanes. The model converged with zero coefficients for the last two Wiener path numbers used in the previous models,  ${}^7P$ , and  ${}^8P$ , so these two indices were eliminated from **Model 3**. The coefficient of determination for **Model 3** is  $R^2 = 0.997068$  and the further improved standard deviation is  $2.1^\circ\text{C}$ .

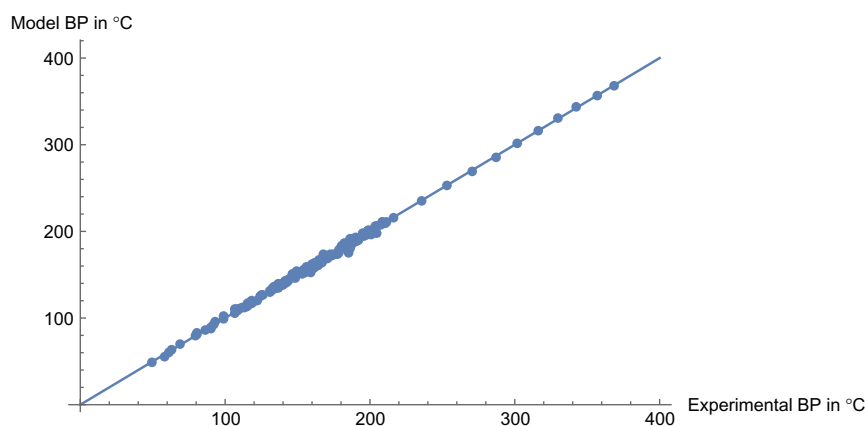
**Model 3.** This model uses the Wiener  ${}^iP$  numbers ( $1 \leq i \leq 6$ ), the Methyl index, and the Hosoya index. The formula is

$$\begin{aligned}
 f_{\text{Model 3}}({}^1P, {}^2P, \dots, {}^6P, Mth, Z) = & 847.41474 + 221.61698({}^1P)^{0.49420} \\
 & - 1182.20853({}^2P)^{0.03689} + 0.00125({}^3P)^{3.39724} - 3.02445({}^4P)^{0.93751} \\
 & - 2.16070({}^5P)^{1.01631} - 0.56366({}^6P)^{1.38233} - 2.10575Mth^{0.5695} - 9.61075Z^{0.19907}
 \end{aligned}$$

**Table 4** gives the number and percentage of alkanes whose absolute boiling point deviations determined by **Model 3** are within the specified ranges. The improvement

**Table 4** Alkanes and Their Corresponding Absolute BP Deviations for [Model 3](#)

BP Dev. (degrees)	No. of Alkanes	Alkanes (%)
0–1	84	44.9
1–2	48	25.7
2–4	46	24.6
4–6	5	2.7
6–9	3	1.6
>9	1	0.5

**FIG. 17**

Experimental BP versus model BP for [Model 3](#).

in the boiling point estimation of [Model 3](#) is demonstrated by the fact that 70% of the alkanes have an error less than 2°C.

[Fig. 17](#) shows a plot of the experimental boiling points versus the model boiling points determined by [Model 3](#). The straight line represents the exact values.

These multivariable models were used to predict the boiling points of a set of 52 alkanes having 13–22 carbon atoms. [Table 5](#) gives the experimental boiling point data for these 52 alkanes, the boiling points predicted by [Models 2](#) and [3](#), and the corresponding boiling point deviations for each model. [Model 3](#) performs particularly well in predicting the boiling points of these 52 alkanes in that 24 of the 52 alkanes have an absolute boiling point deviation less than 2°C. [Fig. 18](#) shows a plot of the experimental boiling point versus the boiling point predicted by [Model 3](#) for the additional set of 52 alkanes. The straight line again represents the exact values.

Other physiochemical properties can also be modeled using similar topological indices presented in this chapter. Melting point models of the alkanes were developed using the same techniques that were used with the previous boiling point models [16]. Modeling the melting points of alkanes is a more challenging task than that

Table 5 Predictive Ability of Models 2 and 3

Alkane	Exp. BP	Model 2		Model 3	
		Calc. BP	BP Dev.	Calc. BP	BP Dev.
2,5,9-Trimethyldecane	211.1	210.2	0.9	212.7	1.6
2,5-Dimethylundecane	215.0	214.5	0.5	218.4	3.4
3,3-Diethylnonane	222.1	222.5	0.4	228.9	6.8
2-Methyldodecane	229.5	227.2	2.3	228.2	1.3
2,3-Dimethylundecane	226.3	228.3	2.0	226.7	0.4
3-Methyldodecane	230.2	231.0	0.8	230.0	0.2
2,2-Dimethylundecane	220.0	222.3	2.3	221.4	1.4
4,5-Diethyl-2,7-dimethyloctane	215.9	202.2	13.6	212.9	2.9
2,4,4,5,5,7-Hexamethyloctane	216.9	216.2	0.6	228.1	11.2
4-Propylundecane	231.0	242.8	11.9	239.0	8.0
2-Methyltridecane	247.9	245.5	2.4	246.6	1.3
3-Methyltridecane	248.4	249.3	0.9	248.7	0.3
4-Methyltridecane	246.5	246.7	0.2	246.4	0.1
2,3-Dimethyl-dodecane	245.9	247.4	1.5	246.0	0.1
2,6,10-Trimethyldodecane	252.0	234.7	17.2	251.2	0.7
4-Methyl-6-propylundecane	242.6	220.7	21.9	239.8	2.8
2,3-Dimethyltridecane	263.7	266.0	2.3	264.5	0.8
4-Methyltetradecane	262.7	264.5	1.8	264.2	1.5
3-Methyltetradecane	265.0	267.2	2.2	266.5	1.5
2-Methyltetradecane	265.4	263.3	2.1	264.1	1.3
7,8-Dimethyltetradecane	269.9	273.0	3.2	272.3	2.5
3-Methylpentadecane	282.0	284.5	2.5	283.4	1.4
2,4-Dimethyltetradecane	266.0	278.5	12.5	275.1	9.1
2-Methylpentadecane	281.9	280.6	1.3	280.7	1.2
4-Methylpentadecane	279.7	281.8	2.1	281.1	1.4
2,3-Dimethyltetradecane	280.2	284.1	3.9	282.1	1.9
2,3-Dimethylpentadecane	295.5	301.7	6.2	298.9	3.4
4-Methylhexadecane	294.3	298.7	4.4	297.3	3.0
3-Methylhexadecane	294.0	301.4	7.4	299.5	5.5
2-Methylhexadecane	297.7	297.4	0.3	296.4	1.3
2-Methylheptadecane	312.0	313.8	1.8	311.5	0.5
3-Methylheptadecane	310.0	317.7	7.7	314.9	4.9
4-Methylheptadecane	307.1	315.1	8.0	312.8	5.7
2,3-Dimethyl-hexadecane	310.0	318.8	8.8	315.2	5.2
2,3-Dimethyl-heptadecane	324.9	335.2	10.3	330.8	5.9
4-Methyloctadecane	322.5	330.7	8.2	327.6	5.1
3-Methyloctadecane	323.5	333.4	9.9	329.8	6.3
2-Methyloctadecane	327.0	329.4	2.4	325.9	1.1
3-Ethyloctadecane	340.9	349.6	8.8	347.9	7.1

(Continued)

Table 5 Predictive Ability of Models 2 and 3—Cont'd

Alkane	Exp. BP	Model 2		Model 3	
		Calc. BP	BP Dev.	Calc. BP	BP Dev.
2,6,11,15-Tetramethylhexadecane	318.9	257.2	61.6	324.6	5.7
2,2-Dimethyloctadecane	330.0	349.1	19.1	336.7	6.7
2,3-Dimethyloctadecane	338.8	350.5	11.7	345.9	7.1
4-Methylnonadecane	336.0	345.2	9.2	341.9	5.9
3-Methylnonadecane	336.0	347.9	11.9	344.0	8.0
2-Methylnonadecane	341.0	343.9	2.9	339.6	1.4
2-Methyleicosane	354.0	356.9	2.9	352.8	1.2
3-Methyleicosane	347.0	360.9	13.9	357.6	10.6
4-Methyleicosane	348.0	358.3	10.3	355.6	7.6
2,3-Dimethylnonadecane	351.8	364.2	12.4	360.4	8.6
2-Methylheneicosane	367.0	367.8	0.8	365.4	1.6
3-Methylheneicosane	358.0	371.8	13.8	370.8	12.8
4-Methylheneicosane	359.0	369.2	10.2	368.8	9.8

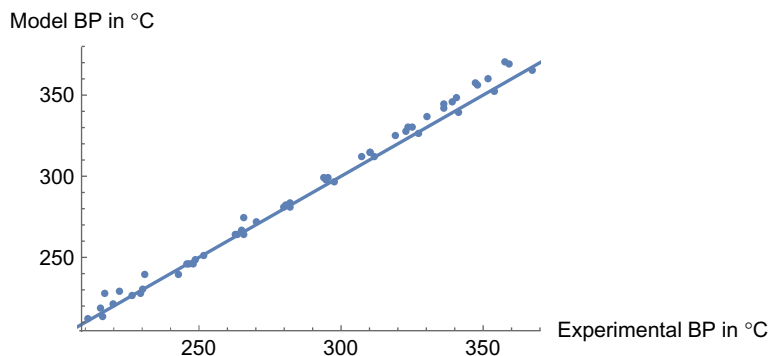


FIG. 18

Experimental BP versus model BP for Model 3 on the 52 additional alkanes.

of boiling points. As previously mentioned, boiling points are mainly determined by the London dispersion forces between the molecule in liquid form. When studying melting points, the molecules are in a solid state which introduces many additional factors in determining the melting point since the alkane is a rigid 3D structure. Attempting to model all the alkanes as was done with the boiling points produces models with very high error rates. Therefore, a select group of alkanes was chosen in order to more effectively model the melting points and use the models to predict the melting points of alkanes for which no experimental data exist.





FIG. 19

2-Methylheptane.

Professors Dr. John Tierney and Dr. Irving Wender of the Chemical and Petroleum Engineering Department at the University of Pittsburgh first approached us asking if we could model a select set of alkanes having between 10 and 20 carbon atoms. These alkanes were of particular interest to them since they comprise most of the molecules in jet and diesel fuels. Normal alkanes have a high cetane number which is environmentally desirable but they also have a high melting point. Since jet and diesel fuels are often exposed to cold temperatures, the high melting points can be problematic in which the fuels produced with these alkanes could possibly freeze. Methyl-substituted alkanes have a lower, but still acceptable, cetane number and have lower melting points than their straight chain counterparts. The methyl-substituted alkanes specifically studied in our research were alkanes that had one and only one methyl group attached by a single bond to one of the carbons in the straight chain of the alkane. An example of the chemical graph of a methyl-substituted alkane used in this study is given in Fig. 19.

Fuels containing methyl-substituted alkanes have lower pour points than fuels based on normal alkanes. Therefore, the fuels remain in liquid form even when subjected to cold temperatures. This is a distinct advantage for jet fuel since it is exposed to very low temperatures at high altitudes. Vehicular transportation also benefits from such improvements to the pour point of diesel fuels in colder climates.

While this improvement in melting points using methyl-substituted alkanes is known, there are several methyl-substituted alkanes for which no melting point data exist. Therefore, we again used the NEOS solver FilterSQP from Argonne National Laboratory to develop nonlinear models for the melting points of the alkanes having between 10 and 20 carbon atoms and a single methyl group. Experimental melting point data are available for 69 of the 80 alkanes in this selected group [11,12]. The melting points of these single methyl group alkanes are given in Table 6. All data are given in Kelvin (K).

The nonlinear models were constructed using the single methyl group alkanes presented with the exception of small group of data that were withheld to test the predictive ability of the models. This predictive set was chosen using a random number generator for each of the models we developed. We then used the models to predict the melting points of the 11 alkanes for which no experimental melting point data existed.

Model 4 was developed using 62 of the 69 single methyl alkanes with the remaining seven withheld to serve as a predictive dataset. The indices used for this model include the Wiener  $iP$  numbers ( $1 \leq i \leq 4$ ) and the mean Wiener index.

Table 6 Melting Point Data

Alkane	$T_{\text{fus}}/\text{K}$	Source	Data Points	Average	Std. Dev.
Decane	243.7	Beilstein	13 <sup>a</sup>	243.4	0.7
2-Methylnonane	198.6	Beilstein	4	198.6	0.2
3-Methylnonane	186.6	NIST	1	NA	NA
4-Methylnonane	174.5	Beilstein	1	NA	NA
5-Methylnonane	186.2	Beilstein	3	186.2	0.6
Undecane	247.4	Beilstein	7	247.4	0.4
2-Methyldecane	224.0	Beilstein	2	224.0	0.5
3-Methyldecane	189.9	Beilstein	2	189.9	5.3
4-Methyldecane	180.3	Beilstein	1	NA	NA
5-Methyldecane	183.2	Beilstein	1	NA	NA
Dodecane	263.6	Beilstein	15 <sup>a</sup>	263.4	0.6
2-Methylundecane	227.4	Beilstein	2	227.4	0.3
3-Methylundecane	216.9	Beilstein	2	216.8	2.3
4-Methylundecane	205.3	Beilstein	2	205.3	1.6
5-Methylundecane	204.3	Beilstein	2	204.3	1.6
6-Methylundecane	209.3	Beilstein	1	NA	NA
Tridecane	267.6	Beilstein	9	267.6	0.3
2-Methyldodecane	246.4	Beilstein	2	246.4	1.1
3-Methyldodecane	213.0	Beilstein	2	213.0	1.0
4-Methyldodecane	222.0	Beilstein	2	222.0	2.3
5-Methyldodecane	203.9	Beilstein	2	203.9	0.4
6-Methyldodecane	215.0	Beilstein	1	NA	NA
Tetradecane	279.0	Beilstein	13	278.9	0.2
2-Methyltridecane	247.0	Beilstein	2	247.0	0.4
3-Methyltridecane	232.4	Beilstein	2	232.4	4.6
4-Methyltridecane	228.5	Beilstein	2	228.5	1.6
5-Methyltridecane	222.1	Beilstein	2	222.1	3.6
6-Methyltridecane	222.6	Beilstein	1	NA	NA
7-Methyltridecane	234.5	Beilstein	2	234.5	2.1
Pentadecane	282.9	Beilstein	11 <sup>a</sup>	282.9	0.2
2-Methyltetradecane	264.6	Beilstein	2 <sup>a</sup>	264.6	0.4
3-Methyltetradecane	237.2	Beilstein	1	NA	NA
4-Methyltetradecane	241.5	Beilstein	2	241.5	1.7
5-Methyltetradecane	226.5	Beilstein	2	226.5	3.7
6-Methyltetradecane	237.8	Beilstein	1	NA	NA
7-Methyltetradecane	225.2	Beilstein	1	NA	NA
Hexadecane	291.7	Beilstein	30 <sup>b</sup>	291.4	0.9
2-Methylpentadecane	263.3	Beilstein	3	263.3	0.9
3-Methylpentadecane	251.8	Beilstein	4	251.8	2.7
4-Methylpentadecane	250.9	Beilstein	3	250.9	1.8
5-Methylpentadecane	242.1	Beilstein	3	242.1	2.3

(Continued)

Table 6 Melting Point Data–Cont'd

Alkane	$T_{\text{fus}}/\text{K}$	Source	Data Points	Average	Std. Dev.
6-Methylpentadecane	241.8	Beilstein	2	241.8	1.1
7-Methylpentadecane	243.1	Beilstein	2	243.1	1.0
8-Methylpentadecane	245.8	Beilstein	2 <sup>a</sup>	245.7	2.7
Heptadecane	295.2	Beilstein	18 <sup>a</sup>	294.9	0.9
2-Methylhexadecane	278.1	Beilstein	1	NA	NA
3-Methylhexadecane	258.5	Beilstein	1	NA	NA
4-Methylhexadecane	258.4	Beilstein	1	NA	NA
5-Methylhexadecane	241.3	Beilstein	1	NA	NA
Octadecane	301.7	Beilstein	32 <sup>a</sup>	301.3	0.8
2-Methylheptadecane	278.5	Beilstein	2	278.5	0.5
3-Methylheptadecane	267.0	Beilstein	1	NA	NA
4-Methylheptadecane	264.9	Beilstein	1	NA	NA
5-Methylheptadecane	253.3	Beilstein	1	NA	NA
9-Methylheptadecane	265.0	NIST	1	NA	NA
Nonadecane	305.1	Beilstein	23	304.9	0.7
2-Methyloctadecane	284.4	Beilstein	2	285.2	1.2
3-Methyloctadecane	274.5	Beilstein	2	274.5	1.2
4-Methyloctadecane	270.4	Beilstein	2	270.4	2.5
5-Methyloctadecane	257.4	Beilstein	2	258.5	1.6
6-Methyloctadecane	269.2	Beilstein	1	NA	NA
7-Methyloctadecane	256.9	Beilstein	1	NA	NA
8-Methyloctadecane	263.2	Beilstein	1	NA	NA
9-Methyloctadecane	256.7	Beilstein	1	NA	NA
Eicosane	311.0	Beilstein	40	309.8	1
2-Methylnonadecane	291.9	Beilstein	3	291.9	1.1
3-Methylnonadecane	280.8	Beilstein	1	NA	NA
4-Methylnonadecane	271.9	Beilstein	1	NA	NA
5-Methylnonadecane	266.2	Beilstein	1	NA	NA

<sup>a</sup>One additional bad data point omitted in average and standard deviation calculations.

<sup>b</sup>Two additional bad data points omitted in average and standard deviation calculations.

The coefficient of determination for Model 4 is  $R^2 = 0.97659$  and the standard deviation is 5.1 K.

#### Model 4.

$$f_{\text{Model 4}}({}^1P, {}^2P, {}^3P, {}^4P, \bar{W}) = 1625 + 2971({}^1P)^{0.1428} - 4460({}^2P)^{0.05842} \\ - 236.7({}^3P)^{0.3554} - 0.01762({}^4P)^{2.727} + 16.11\bar{W}^{1.047}$$

Fig. 20 shows a plot of the experimental melting points versus the model melting points determined by Model 4. The straight line represents an exact prediction.

Table 7 gives the number and percentage of alkanes whose absolute melting point deviations determined by Model 4 are within the specified ranges. Table 8 illustrates the predictive ability of Model 4 for the given set of seven withheld alkanes.

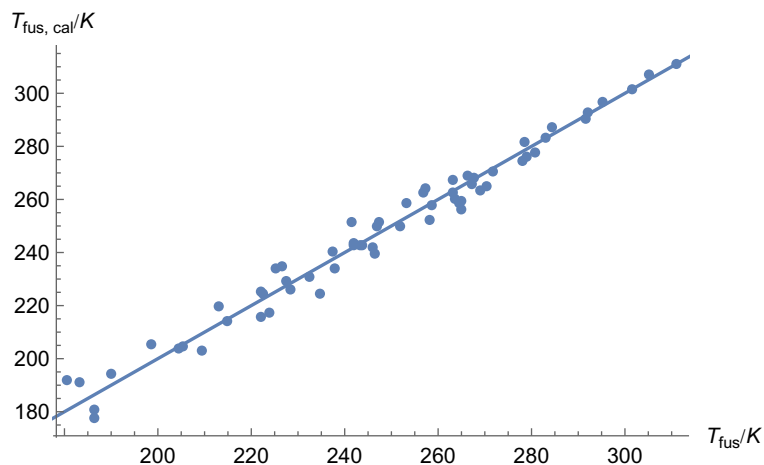


FIG. 20

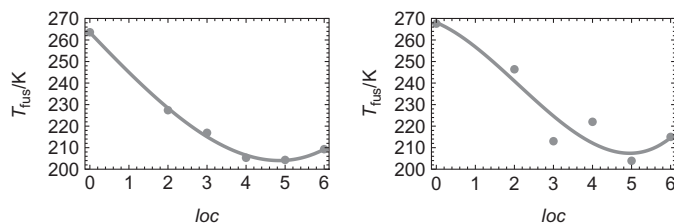
Experimental MP versus model MP for Model 4.

**Table 7** Alkanes and Their Corresponding Absolute MP Deviations for Model 4

$\Delta T/K$	No. of Alkanes	Alkanes (%)
0–1	11	17.7
1–2	9	14.5
2–4	18	29.0
4–6	6	9.7
6–9	14	22.6
>9	4	6.5

**Table 8** Predictive Ability of Model 4

Alkane	$T_{\text{fus}}/K$	$T_{\text{fus,cal}}/K$	$\Delta T/K$
4-Methylnonane	174.5	178.0	3.5
3-Methylundecane	216.9	207.6	9.3
5-Methyldodecane	203.9	214.9	11.1
4-Methyltetradecane	241.5	236.1	5.4
4-Methylpentadecane	250.9	244.7	6.2
3-Methyloctadecane	274.5	272.1	2.4
9-Methyloctadecane	256.7	262.1	5.4

**FIG. 21**

(A) MP of the single branch dodecanes. (B) MP of the single branch tridecanes.

Two of the alkanes have a high melting point deviation which made us examine the data more closely. Upon analyzing graphs of the melting points of the single methyl alkanes versus their methyl locant indices, we found that the melting point of alkanes having an odd number of carbons oscillated along the cubic regression as the methyl locant changed. In contrast, the alkanes having an even number of carbons closely followed the cubic regression. Fig. 21 shows this phenomenon with the graphs of the melting points of the single branch dodecanes and single branch tridecanes versus their respective methyl locant indices. This difference in the behavior of the melting points based on the parity of the number of carbons is a well-known fact dating back to the 19th century [17].

Based on these findings, the melting point data were split based on the parity of the number of carbon atoms. New predictive sets using a random number generator were chosen for each model. For the model based on the single methyl alkanes having an even number of carbons, 31 of the 37 alkanes were used to develop the model while the remaining 6 were withheld to serve as a predictive data. Since this data does not oscillate, the same indices used in Model 4 were used to develop Model 5. The coefficient of determination for Model 5 is  $R^2 = 0.988475$  and the standard deviation is 3.8 K.

#### Model 5.

$$f_{\text{Model 5}}({}^1P, {}^2P, {}^3P, {}^4P, \bar{W}) = -552 + 1731({}^1P)^{0.2345} - 1090.5({}^2P)^{0.2120} \\ - 4.866({}^3P)^{1.245} - 254.0({}^4P)^{0.1865} + 0.006393\bar{W}^{4.092}$$

Table 9 gives the number and percentage of alkanes whose absolute melting point deviations determined by Model 5 are within the specified ranges. Table 10 illustrates the predictive ability of Model 5 for the given set of six withheld alkanes. The standard deviation of Model 5 was notably improved over that of Model 4 since four of the six predicted melting points were within 2 K of the experimental melting points.

As seen in Fig. 21, the melting points of single methyl alkanes having an odd number of alkanes oscillate as the methyl locant index varies. Therefore, we used both the number of carbon atoms and the methyl locant index with the cosine function

**Table 9** Alkanes and Their Corresponding Absolute MP Deviations for **Model 5**

$\Delta T/K$	No. of Alkanes	Alkanes (%)
0–1	7	22.6
1–2	7	22.6
2–4	8	25.8
4–6	6	19.4
6–9	3	9.7

**Table 10** Predictive Ability of **Model 5**

Alkane	$T_{fus}/K$	$T_{fus,cal}/K$	$\Delta T/K$
Dodecane	263.6	261.9	1.7
2-Methylundecane	227.4	225.3	2.0
4-Methylundecane	205.3	206.1	0.8
2-Methyltridecane	247.0	246.0	0.9
4-Methylpentadecane	250.9	244.8	6.1
Eicosane	311.0	315.1	4.0

as one of the terms in **Model 6**. **Model 6** is fit to 27 of the 32 single methyl alkanes using the Wiener  $iP$  numbers ( $1 \leq i \leq 4$ ), the mean Wiener index, the number of carbon atoms ( $n$ ), and the methyl locant index. Five data points were randomly chosen to serve as predictive data and were therefore not used to form the model. **Model 6** has a coefficient of determination of  $R^2 = 0.99241$  and the standard deviation is 3.0 K.

**Model 6.**

$$f_{\text{Model 6}}({}^1P, {}^2P, {}^3P, {}^4P, \bar{W}, n, loc) = -6695 + 2159({}^1P)^{0.16275} - 240.5({}^2P)^{0.3518} \\ + 1076({}^3P)^{-0.4243} - 2.537({}^4P)^{1.407} \\ + 3750\bar{W}^{0.03078} + 1.203n^{0.5095} \cos(loc \pi)$$

**Table 11** gives the number and percentage of alkanes whose absolute melting point deviations determined by **Model 6** are within the specified ranges. **Table 12** illustrates the predictive ability of **Model 6** for the given set of five withheld alkanes.

**Table 11** Alkanes and Their Corresponding Absolute MP Deviations for **Model 6**

$\Delta T/K$	No. of Alkanes	Alkanes (%)
0–1	14	51.9
1–2	4	14.8
2–4	6	22.2
4–6	2	7.4
6–9	1	3.7

**Table 12** Predictive Ability of [Model 6](#)

Alkane	$T_{\text{fus}}/\text{K}$	$T_{\text{fus,cal}}/\text{K}$	$\Delta T/\text{K}$
3-Methyldodecane	213.0	217.9	4.9
4-Methyltetradecane	241.5	238.3	3.1
4-Methylhexadecane	258.4	256.1	2.2
Nonadecane	305.1	306.9	1.8
2-Methyloctadecane	284.4	291.9	7.5

**Table 13** Eleven Predicted Melting Points

Alkane	$T_{\text{fus,cal}}/\text{K}$
6-Methylhexadecane	253.9
7-Methylhexadecane	243.1
8-Methylhexadecane	253.0
6-Methylheptadecane	258.8
7-Methylheptadecane	258.6
8-Methylheptadecane	258.4
6-Methylnonadecane	271.4
7-Methylnonadecane	271.1
8-Methylnonadecane	270.8
9-Methylnonadecane	270.7
10-Methylnonadecane	270.7

Given the improved standard deviation of both [Models 5](#) and [6](#), these models were used to predict the melting points of the 11 single methyl alkanes for which no experimental data exists in the set of 80 single methyl alkanes having 10–20 carbon atoms. [Model 5](#) was used to predict the melting point if the alkane had an even number of carbon atoms and [Model 6](#) was used for the alkanes having an odd number of carbons. [Table 13](#) gives the predicted melting points of these 11 single methyl alkanes. [Fig. 22](#) gives the plot of the predicted melting points of the octadecanes versus their methyl locant indices. Existing melting point data for the octadecanes are also included. [Fig. 23](#) shows the predicted melting points of the octadecanes with available existing experimental data and [Fig. 24](#) shows the predicted melting points of the eicosanes with existing experimental data.

## 4 CONCLUSIONS

Chemical graph theory is a productive application of mathematics to the physical properties of chemical molecules. Chemists have used graph theory to model the molecules they studied long before graph theory was established as an area of mathematics. Chemists played a major role in the growth of chemical graph

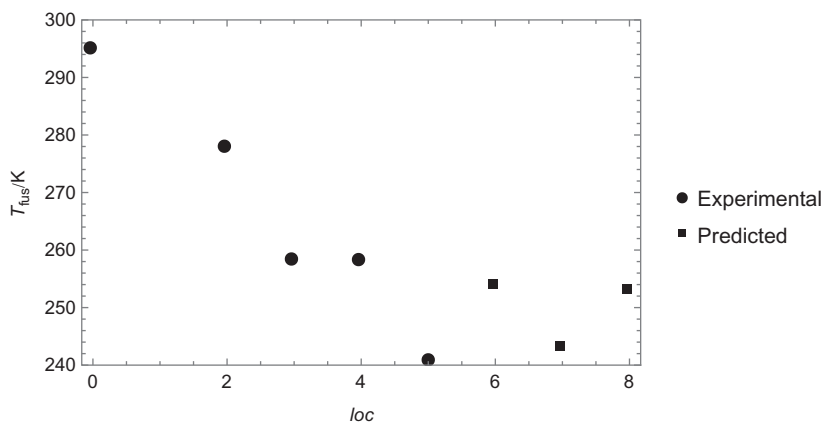


FIG. 22

Predicted melting points of the missing heptadecane methyl substituted alkanes.

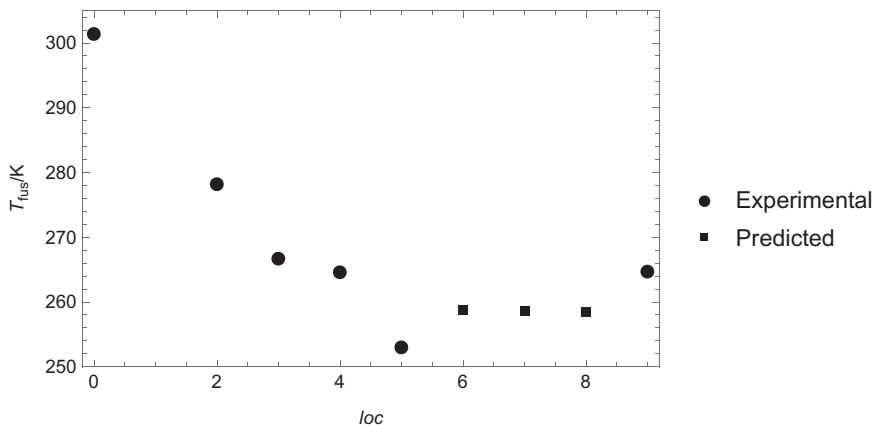


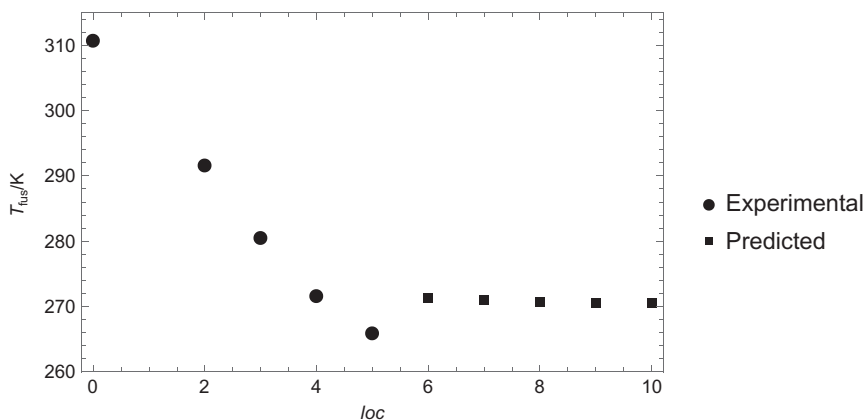
FIG. 23

Predicted melting points of the missing octadecane methyl substituted alkanes.

theory as they realized that mathematics was more than just a tool for crunching numerical data.

Using several topological indices to construct models of the physical properties of alkanes has proven successful as seen in this chapter. [Model 3](#) does extremely well at predicting the boiling points of the additional 52 alkanes as seen in [Table 5](#). The boiling point deviations for 24 of the 52 alkanes were less than  $2^{\circ}\text{C}$ . Given this is often an acceptable error for experimental measurement of physical properties, this model could be used to predict the boiling points of alkanes for which no experimental boiling point data exist. Although modeling the melting points of





**FIG. 24**

Predicted melting points of the missing eicosane methyl substituted alkanes.

alkanes proved to be more difficult, splitting the data by the parity of the number of carbons resulted in greatly improved models. **Model 4**, which attempted to model the melting points of all the methyl substituted alkanes having between 10 and 20 carbon atoms, had 11.1 K as its worst predictive error. In contrast, **Model 5**, which modeled the methyl-substituted alkanes having an even number of carbons, had 6.1 K as its worst predictive error. The worst predictive error for **Model 6**, which modeled the alkanes having an odd number of carbons, was 7.5 K. Using models similar to those presented in this chapter would allow chemists to predict physical properties of molecules which may never be measurable in a traditional lab setting.

## REFERENCES

- [1] Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim: Wiley-VCH; 2000.
- [2] Hosoya H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. Bull Chem Soc Jpn 1971;44(9): 2332–9.
- [3] Gutman I, Polansky OE. Mathematical concepts in organic chemistry. Berlin: Springer-Verlag; 1986.
- [4] Burch KJ, Wakefield DK, Whitehead EGJ. Boiling point models of alkanes. Match 2003; 47:25–52.
- [5] Wiener H. Structural determination of paraffin boiling points. J Am Chem Soc 1947; 1(69):17–20.
- [6] Rouvray D. The rich legacy of half a century of the Wiener index. In: Topology in chemistry: discrete mathematics of molecules. Horwood Publishing; 2002. p. 295–306.
- [7] Gutman I, Miljković O. Molecules with smallest connectivity indices. Match 2000;41: 57–70.

- [8] Randić M. On characterization of molecular branching. *J Am Chem Soc* 1975;97:6609–15.
- [9] Balaban AT. Highly discriminating distance-based topological index. *Chem Phys Lett* 1982;89(5):399–404.
- [10] Plavšić D, Nikolić S, Trinajstić N, Mihalić Z. On the Harary index for the characterization of chemical graphs. *J Math Chem* 1993;12:235–50.
- [11] Database BC; 2001. Available from: <http://www.beilstein.com/products/xfire/>.
- [12] Frenkel MD. Thermodynamics Research Center, NIST Boulder Laboratories. In: Linstrom PJ, Mallard WG, editors. “Thermodynamics Source Database” in NIST Chemistry WebBook, NIST Standard Reference Database Number 69. Gaithersburg MD: National Institute of Standards and Technology; 2003. p. 20899. Available from: <http://webbook.nist.gov> [Accessed October 2001].
- [13] Balaban AT. A personal view about topological indices for QSAR/QSPR. In: QSPR/QSAR studies by molecular descriptors. Huntington, NY: Nova Science Publishers; 2001. p. 1–30.
- [14] Espinosa G, Yaffe D, Cohen Y, Arenas A, Giralt F. Neural network based quantitative structural property relations (QSPRs) for predicting boiling points of aliphatic hydrocarbons. *J Chem Inf Comput Sci* 2000;40:859–79.
- [15] Fletcher R, Leyffer S. Nonlinear programming without a penalty function. *Math Program* 2002;91(2):239–69.
- [16] Burch KJ, Whitehead EGJ. Melting point models of alkanes. *J Chem Eng Data* 2004;49:858–63.
- [17] Baeyer A. Ueber Regelmässigkeiten im Schmelzpunkt homologer Verbindungen. *Ber Chem Ges* 1877;10:1286.
- [18] Mihalić Z, Trinajstić N. A graph-theoretical approach to structure-property relationships. *J Chem Educ* 1992;69(9):701–12.