A New Model For Predicting Boiling Points of Alkanes

Simon Mukwembi

University of the Witwatersrand

Farai Nyabadza (✉ fnyabadza@uj.ac.za)

University of Johannesburg

Research Article

Keywords: Graph, conduction, alkanes, Wiener index, boiling points, model fitting.

Posted Date: August 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-812041/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on December 1st, 2021.
See the published version at <https://doi.org/10.1038/s41598-021-03541-z>.

A new model for predicting boiling points of alkanes

Simon Mukwembi¹ and Farai Nyabadza^{2*}

¹School of Mathematics

University of the Witwatersrand
South Africa

²Department of Mathematics and Applied Mathematics
University of Johannesburg
South Africa

Abstract

A general perception among researchers is that boiling points, which is a key property in the optimization of lubricant performance, are difficult to predict successfully using a single-parameter model [5, 6]. In this contribution, we propose a new graph parameter which we call, for lack of better terminology, the conduction of a graph. We exploit the conduction of a graph to develop a single-parameter model for predicting the boiling point of any given alkane. The accuracy of our model compares favourably to the accuracy of experimental data in literature. Our results have significant implications on the estimation of boiling points of chemical compounds in the absence of experimental data.

Keywords: Graph, conduction, alkanes, Wiener index, boiling points, model fitting.

1 Introduction

A *graph* $G = (V, E)$ is a mathematical object which consists of a finite set V of elements called *vertices*, together with a set E , of 2-element subsets of V , called the *edges* of G . As early as 1875, Cayley (see for instance [1]), in his quest to enumerate chemical molecules called alkanes, he made an observation that molecules

*Corresponding author: fnyabadza@uj.ac.za

can be modelled by graphs where atoms are represented by vertices and two vertices are joined by an edge if the corresponding atoms are linked by bonds. This graph model became widely known as the *molecular graph*. An interesting numeric value attached to each molecule is the boiling point, i.e., the temperature at which a substance has a vapour pressure of 760 mm Hg [2]. Whilst the boiling point is a key property in classifying molecules, such as alkanes which dominate mixtures of lubricants in industry, there is generally lack of experimental boiling point data (see, for instance [1] and references cited therein). Nevertheless, although expected to be accurate - experimental boiling points are sometimes inaccurate due to the presence of impurities, and as a result, have wide discrepancies especially for higher boiling points. This has the propensity to lead to wider and elevated boiling point ranges. For instance, the boiling point of cyclooctane (an alkane) is reported to range from 120.3°C to 156.8°C , and that of methylcycloheptane ranges from 113°C to 136.8°C [2]. The lack of data, and the inaccuracies where data is available, have necessitated the development of boiling point models which can be used to estimate boiling points of chemical compounds for which no boiling point data is available or where inaccuracies exist. A natural foundation on which these models are built are graph parameters.

The Wiener index of a graph, introduced by Harold Wiener in 1947 [7], was the first graph parameter to be used in chemistry. In particular, it was used to predict boiling points of alkanes. Ever since its introduction, a lot of effort was invested in research that gave rise to the development of extensions of the Wiener index [5], and of other indices such as the Hosoya index [5], the Gutman and the Schultz indices [4], and a legion of other distance-based topological indices (see, for example, [3]), which are used for the prediction of various properties of molecules.

For alkanes, single variable models that are available are either weak [5] or they consider only a special class of the molecules [1]. Burch, Wakefield and Whitehead [1] successfully developed a single variable model to calculate boiling points of spe-

cial families of alkanes and developed multivariate models that can predict boiling points of all alkanes up to and including those of order 12. Several other models are abound in literature with the most recent one being due to Sandak and Conduit [8] who trained artificial neural networks to predict the physical properties of linear, single branched, and double branched alkanes. Clearly, as Dearden, [2], pointed out several years ago, modeling a property is easier when one is dealing with a single chemical class. However, from the point of view of an engineer concerned with a wide range of compounds, the methods of greater interest are those that can adequately model the behavior of varied data sets.

In this paper, we are concerned with relating two mathematical objects, associated with a chemical molecule, namely the molecular graph and the numeric value attached to the boiling point of the molecule. We focus on molecular graphs for all alkanes and introduce a new graph parameter which, we define as the 'conduction'. We use this new parameter, the conduction, to develop a single-variable model that predicts the boiling point of any given alkane.

This paper is organised as follows: in the next section, we introduce the new graph parameter, the conduction of a graph, and illustrate how it is computed by determining the conduction of four special classes of graphs representing some important series of alkanes. The model is developed in the Section 3. The main results are presented in Section 4 in which the model is tested. This will be followed by the discussion and the conclusion is Section 5.

2 The graph parameter: conduction

Consider a connected graph $G = (V, E)$ of order n . The *distance* $d_G(u, v)$ between vertices u and v in G is defined as the length of a shortest path joining u and v in G . Let v be a vertex of G . Denote by T_v , the breadth-first search tree of G based at v , and let S_v be the set of end vertices of T_v . To contrive the conduction parameter, we envisage heat flowing from one atom of a chemical compound to the other atoms.

We assume that in the heat conduction process between atoms (depicted by vertices of a graph), heat is transferred from vertex v outwardly to all other vertices in the graph through contact via the edges of the tree T_v until it reaches the end vertices in T_v . The speed of transfer, conceivably, is governed by the breadth of T_v which can be approximated by the square of the degree, $\deg v$, of vertex v . As a result, in the heat conduction process, let the *score* $s(v)$ of vertex v be the quantity

$$s(v) = \deg^2 v \left[\sum_{x \in S_v} d(v, x) \right].$$

Mathematically, the *conduction* of a graph G , $c(G)$, is defined as

$$c(G) = \frac{1}{n} \sum_{v \in V(G)} s(v).$$

For instance, the graph in Figure 1 has conduction $\frac{366}{8}$.

Figure 1: Molecular graph for the alkane, 2,2,3,3-tetramethylbutane.

We use four special classes of graphs. The first is the path, P_n , of order n . The *broom graph*, $B_{n,q}$, is a graph of order n obtained by taking the path P_{n-q} and attaching q end vertices to one end of P_{n-q} . The second class of graphs we will consider is $B_{n,2}$ while the third class is $B_{n,3}$. These classes of graphs, namely $P_n = B_{n,1}$, $B_{n,2}$ and $B_{n,3}$, represent the class of normal-alkanes, the 2-methyl, and the 2,2-dimethyl series of alkanes, respectively. The fourth class is that of graphs E_n of order n where E_n has fairly large conduction among all graphs of order n , and E_n consists of a path P , and having as many vertices of degree 4 on P as possible. For instance, E_8 is given in Figure 1. Table 1 shows some of the graphs, E_n .

Number	Alkane	c	Boiling point in °C
E_5	2,2-dimethylpropane	17.6	9.5
E_6	2,3-dimethylbutane	23.33	58.1
E_7	2,2,3-trimethylbutane	33.71	81
E_8	2,2,3,3-tetramethylbutane	45.75	106.5
E_9	2,3,3,4-tetramethylpentane	52	141.5
E_{10}	2,2,3,3,4-pentamethylpentane	66.3	161.1

Table 1: Some E_n graphs.

A simple calculation establishes the following formulae.

Proposition 2.1 *Let $n \geq 4$. Then*

1. $c(P_n) = 4n - 10 + \frac{6}{n}$,
2. $c(B_{n,2}) = 6n - 21 + \frac{27}{n}$,
3. $c(B_{n,3}) = 8n - 38 + \frac{78}{n}$.

Using data in Table 1, we can approximate the conduction of E_n by an equation of the form

$$c(E_n) = \alpha n^2 + \beta n + \gamma + \lambda \frac{1}{n}, \quad (1)$$

where α, β, γ and λ are constants.

Fitting data in Table 1 to equation (1), we get the following equation with a *coefficient of determination*, CoD, $R^2 = 0.9924$. We discuss in detail the concept of the CoD in subsection 3.1.

Proposition 2.2 *Let $n \geq 5$. Then*

$$c(E_n) \approx \frac{4}{9}n^2 + 3.44857532n - 14.95083692 + \frac{18.25213877}{n}.$$

3 Model development

3.1 Model quality assessment

In the development process of our model, we adopt the folklore method, goodness of fit, of determining the quality of the model. After performing a fitting process, it is important to determine the goodness of the fit. While there are many methods of determining the goodness of a fit, in this paper we evaluate the residuals and make a plot of the residuals. The evaluation of the residuals is important in determining the goodness of fit of our model to data. We use the CoD, that is used to explain the variability between our model output and the data. This coefficient is commonly known as R-squared, (R^2). The formula of coefficient of determination is given by:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}},$$

where, R^2 = Coefficient of Determination, RSS = Residuals sum of squares and TSS = Total sum of squares.

A CoD value of 1 indicates a perfect fit, and thus the model can be deemed very reliable for any future forecasts, while a value of 0 indicates that the model does not accurately model the data. Some researchers have argued that it is better to look at adjusted R-squared rather than the R-squared, but for the work presented in this paper, it suffices to use the non-adjusted R-squared.

3.2 Problem Statement

Let A_n be an alkane with n carbon atoms. Our problem is to estimate the boiling point of the alkane, A_n . We will represent this by a graph, G_n , of order n . We will often use A_n and G_n interchangeably. We will thus denote the boiling point of A_n by $b(A_n)$ or $b(G_n)$.

3.3 Model when sufficient data is available

We first consider a situation in which there is sufficient data on the boiling points of a group of alkanes. In general, to find $b(A_n)$, we propose a model of the form:

$$b(A_n) = \alpha c(G_n) + \beta, \quad (2)$$

where $c(G_n)$ is the conduction of graph G_n and α and β are constants to be determined by fitting the model to boiling points data of alkanes of order n . In many instances, the data may have an oscillatory tendencies as conduction increases. To improve the fit and accuracy, we propose a combination of linear fit with logarithmic and trigonometric functions. We thus propose a function of the form,

$$b(A_n) = \alpha c(G_n) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3 c(G_n) + \alpha_4] \sin[\alpha_5 c(G_n) + \alpha_6], \quad (3)$$

where $\alpha, \beta, \alpha_i, i = 1, 2, \dots, 6$, are constants to be determined by fitting the model to data of alkanes of order n .

To illustrate the application of (3), we consider the data set depicted in Table 2 of all alkanes of order 6.

Number	Alkane	c	Boiling point in °C
1	hexane	15	68.8
2	3-methylpentane	18.83	63.3
3	2-methylpentane	19.5	60.9
4	2,2-dimethylbutane	23	49.8
5	2,3-dimethylbutane	23.33	58.1

Table 2: Data for 5 alkanes with 6 carbon atoms.

While a linear model, (2), fits the data with $(\alpha, \beta) = (-1.8245, 96.5461)$ and a CoD value of $R^2 = 0.7903$ our model, (3), fits the data with

$$(\alpha, \beta, \alpha_i) = (2.7173, -0.2030, 24.5237, 3.5331, 0.4679, -6.7879, -0.9061, 18.1828)$$

and a CoD value of $R^2 = 1$. The graphs of the fit using the least squares (lsq) fitting method and the residuals are given in Figure 2.

Figure 2: shows the model fit with a 95% confidence interval and the residuals.

3.4 Model with insufficient data

Our model to estimate the boiling point of an alkane A_n of order n depends on the availability of data for alkanes of order n . In the absence of such data, our initial step will be to develop the data. Once we develop the data, we can then use (2) (or (3), for more accuracy) to find the equation that estimates the boiling point to a reasonable degree of accuracy. We cover this below.

3.4.1 Generating data

We need to generate data for some alkanes of order n as more data yields better accuracy during the fitting process. The graphs P_n , $B_{n,2}$, $B_{n,3}$ and E_n are all alkanes of order n we estimate their boiling points in turn below.

Boiling point for P_n

We propose the model

$$b(P_n) = \alpha c(P_n) + \beta + \alpha_1 \log_{\alpha_2}[\alpha_3 c(P_n) + \alpha_4],$$

where α , β , α_i , $i = 1, 2, 3, 4$, are constants to be determined by fitting the model to data of normal-alkanes. From Proposition 2.1, this reduces to

$$b(P_n) = \alpha \left(4n - 10 + \frac{6}{n} \right) + \beta + \alpha_1 \log_{\alpha_2} \left[\alpha_3 \left(4n - 10 + \frac{6}{n} \right) + \alpha_4 \right]. \quad (4)$$

Fitting (4) to the data of the first few normal-alkanes given in Table 3, we obtain the values of the constants as

$$(\alpha, \beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1.5911, 20.9101, 41.6488, 1.2703, 0.0429, 0.5061)$$

with a CoD value of $R^2 = 0.9999$.

Number	Alkane	c	Boiling point in °C
1	butane	7.5	0
2	pentane	11.2	36.1
3	hexane	15	68.8
4	heptane	18.85714	98.38
5	octane	22.75	125.6
6	nonane	26.66667	150.7
7	decane	30.6	174.1
8	undecane	34.54545	195
9	dodecane	38.5	216
10	tridecane	42.46154	234
11	tetradecane	46.42857	255

Table 3: Data for 11, normal-alkanes.

The graph and the plot of residuals are given in Figure 3.

Figure 3: shows the model fit and the residuals.

Boiling point for $B_{n,2}$

We propose the model

$$b(B_{n,2}) = \alpha c(B_{n,2}) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3 c(B_{n,2}) + \alpha_4],$$

where α , β , α_i , $i = 1, 2, 3, 4$, are constants to be determined by fitting the model to data of 2-methyl series of alkanes. From Proposition 2.1, this reduces to

$$b(B_{n,2}) = \alpha \left(6n - 21 + \frac{27}{n} \right) + \beta + \alpha_1 \log_{\alpha_2} \left[\alpha_3 \left(6n - 21 + \frac{27}{n} \right) + \alpha_4 \right]. \quad (5)$$

Fitting (5) to the data of the first few 2-methyl series given in Table 4, we obtain the values of the constants as

$$(\alpha, \beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1.2125, 23.5857, 40.7654, 1.3216, 0.03559, 0.3998)$$

with a CoD value of $R^2 = 0.9996$.

Number	Alkane	c	Boiling point in °C
1	2-methylbutane	14.5	27.4
2	2-methylpentane	19.5	60.9
3	2-methylhexane	24.85714	90.1
4	2-methylheptane	30.375	117.6
5	2-methyloctane	36	143
6	2-methylnonane	41.7	166.9
7	<i>2-methyldecane</i>	47.45455	189.3
8	<i>2-methylundecane</i>	53.25	208.9
9	2-methyldodecane	59.07692	229.5
10	2-methyltridecane	64.92857	247.9

Table 4: Data for 8, 2-methyl series.

The graph and the plot of residuals are given in Figure 4.

Figure 4: shows the model fit and the residuals.

Boiling point for $B_{n,3}$

We propose the model

$$b(B_{n,3}) = \alpha c(B_{n,3}) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3 c(B_{n,3}) + \alpha_4],$$

where $\alpha, \beta, \alpha_i, i = 1, 2, 3, 4$, are constants to be determined by fitting the model to data of 2-dimethyl series of alkanes. From Proposition 2.1, this reduces to

$$b(B_{n,3}) = \alpha \left(8n - 38 + \frac{78}{n} \right) + \beta + \alpha_1 \log_{\alpha_2} \left[\alpha_3 \left(8n - 38 + \frac{78}{n} \right) + \alpha_4 \right]. \quad (6)$$

Fitting (6) to the data of the first few 2-dimethyl series given in Table 5, we obtain the values of the constants as

$$(\alpha, \beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (2.2207, 15.9874, 9.8662, 1.3271, 0.0613, -0.8074)$$

with a CoD value of $R^2 = 0.9999$. The graph and the plot of residuals are given in

Number	Alkane	c	Boiling point in °C
1	2,2-dimethylpropane	17.6	9.5
2	2,2-dimethylbutane	23	49.8
3	2,2-dimethylpentane	29.14286	79.2
4	2,2-dimethylhexane	35.75	106.9
5	2,2-dimethylheptane	42.66667	131.9
6	2,2-dimethyloctane	49.8	154
7	<i>2,2-dimethylnonane</i>		
8	2,2-dimethyldecane	64.5	200.12
9	2,2-dimethylundecane	72	220
10	<i>2,2-dimethyldodecane</i>		

Table 5: Data for 8, 2,2-dimethyl series.

Figure 5.

Boiling point for E_n

We propose the model

$$b(E_n) = \alpha c(E_n) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3 c(E_n) + \alpha_4],$$

Figure 5: shows the model fit and the residuals.

where α , β , α_i , $i = 1, 2, 3, 4$, are constants to be determined by fitting the model to data given in Table 1. From Proposition 2.2, this reduces to

$$b(E_n) = \alpha \left(\frac{4}{9}n^2 + 3.44857532n - 14.95083692 + \frac{18.25213877}{n} \right) + \beta + \alpha_1 \log_{\alpha_2} \left[\alpha_3 \left(\frac{4}{9}n^2 + 3.44857532n - 14.95083692 + \frac{18.25213877}{n} \right) + \alpha_4 \right]. \quad (7)$$

Fitting (7) to the data given in Table 1, we obtain the values of the constants as

$$(\alpha, \beta, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (1.0404, -360.7375, 4.2800, 1.0802, 63.5399, -499.3487)$$

with a CoD value of $R^2 = 0.9751$. The graph and the plot of residuals are given in Figure 6.

Figure 6: shows the model fit and the residuals.

4 Main Results

From the previous sections, we have now generated data for boiling points of some alkanes with n carbon atoms. Let

$$\begin{aligned}x_1 &= 4n - 10 + \frac{6}{n}, \\x_2 &= 6n - 21 + \frac{27}{n}, \\x_3 &= 8n - 38 + \frac{78}{n},\end{aligned}$$

and

$$x_4 = \frac{4}{9}n^2 + 3.44857532n - 14.95083692 + \frac{18.25213877}{n}.$$

We present the data in Table 6.

Alkane	c	Boiling point in °C
P_n	x_1	$y_1 = \alpha(x_1) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3(x_1) + \alpha_4]$ where constants are as given in (4)
$B_{n,2}$	x_2	$y_2 = \alpha(x_2) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3(x_2) + \alpha_4]$ where constants are as given in (5)
$B_{n,3}$	x_3	$y_3 = \alpha(x_3) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3(x_3) + \alpha_4]$ where constants are as given in (6)
E_n	x_4	$y_4 = \alpha(x_4) + \beta + \alpha_1 \log_{\alpha_2} [\alpha_3(x_4) + \alpha_4]$ where constants are as given in (7)

Table 6: Data for 4 alkanes of order n .

To find $b(A_n)$, we propose the lsq model:

$$b(A_n) = \lambda \cdot c(G_n) + \gamma, \tag{8}$$

where

$$\lambda = \frac{x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4 - \frac{1}{3}(x_1 + x_2 + x_3 + x_4)(y_1 + y_2 + y_3 + y_4)}{x_1^2 + x_2^2 + x_3^2 + x_4^2 - \frac{1}{4}(x_1 + x_2 + x_3 + x_4)^2}$$

and

$$\gamma = \frac{1}{4}(y_1 + y_2 + y_3 + y_4) - \frac{1}{4}(x_1 + x_2 + x_3 + x_4)\lambda.$$

Model testing

We first test the predictive ability of our model, (8), on all alkanes of order 6. We present the results in Table 7. The CoD value is $R^2 = 0.7516$ and the graphs are

Number	Alkane	c	Bp in °C	predicted
1	2-methylpentane	19.5	60.9	59.8
2	3-methylpentane	18.8	63.3	61.1
3	2,2-dimethylbutane	23	49.8	53.1
4	2,3-dimethylbutane	23.3	58.1	52.4
5	hexane	15	68.8	68.5

Table 7: All alkanes of order 6: Predictive ability of model (8).

presented in Figure 7 below.

Figure 7: shows a comparison of the linear least squares fit to experimental data and the linear model (8) for all alkanes of order 6.

Next we test the predictive ability of our model, (8), on all alkanes of order 7. We

present the results in Table 8. The CoD value is $R^2 = 0.7898$ and the graphs are

Number	Alkane	c	Bp in °C	predicted
1	2,2,3-trimethylbutane	33.71	80.9	80.7
2	3-ethylpentane	23.14	93.5	92.1
3	3,3-dimethylpentane	28	86.1	86.9
4	2,4-dimethylpentane	30.86	80.6	83.8
5	2,3-dimethylpentane	28.57	89.8	86.2
6	2,2-dimethylpentane	29.14	79.2	85.6
7	2-methylhexane	24.86	90.1	90.3

Table 8: All alkanes of order 6: Predictive ability of model (8).

presented in Figure 8 below.

Figure 8: shows a comparison of the linear least squares fit to experimental data and the linear model (8) for all alkanes of order 7.

We now test the predictive power of our model, (8), on alkanes of order 8. We present the results in Table 9.

Number	Alkane	c	Bp in °C	predicted
1	2,2,3,3-tetramethylbutane	45.75	106.5	106.7
2	2,2,3-trimethylpentane	39.625	110	111.1
3	2,3,3-trimethylpentane	39.125	114.7	111.4
4	2,2,4-trimethylpentane	43.375	99.2	108.4
5	2,2-dimethylhexane	35.75	106.9	113.8
6	3,3-dimethylhexane	33.75	112	115.2
7	3-ethyl-3-methylpentane	32.75	118.2	115.9
8	2,3,4-trimethylpentane	40.625	113.4	110.4
9	2,3-dimethylhexane	34.5	115.6	114.7
10	3-ethyl-2-methylpentane	33.5	115.6	115.4
11	3,4-dimethylhexane	33.5	117.7	115.4
12	2,4-dimethylhexane	36	109.4	113
13	2,5-dimethylhexane	38.5	109	111.9
14	2-methylheptane	30.375	117.6	117.6
15	3-methylheptane	25.875	118	120.8
16	4-methylheptane	28.375	117.7	119
17	3-ethylhexane	27.875	118.5	119.4
18	octane	22.75	125.7	123

Table 9: All alkanes of order 8: Predictive ability of model (8).

The CoD value is $R^2 = 0.6488$ and the graphs are presented in Figure 9 below.

Figure 9: shows a comparison of the linear least squares fit to experimental data and the linear model (8) for all alkanes of order 8.

5 Discussion and conclusion

In this paper we provide a novel model for predicting the boiling points of alkanes. We consider alkanes of orders 6, 7 and 8 to test the usefulness of our model. We consider alkanes with n carbon atoms represented by a graph G_n whose boiling points may be known experimentally. In particular, the methods presented here have the ability to develop the boiling points data using the conduction parameter, which we defined as the conduction of a graph. The nature of the boiling points data necessitated the use of a combination of trigonometric and logarithmic functions in coming up with models that fit to the data. The models are then used to predict the boiling points of alkanes whose number of carbon atoms are known.

The model presented in this article is without inculpability. First we fit linear

models to experimental data that has very few number of data points in some instances. Second, some boiling points of certain compounds could not be ascertained from literature and last, the approximation of the numerical values of some parameters give rise to variability in the model predictions. Despite these shortcomings, the work presents a unique and innovative way of finding boiling points of using graph theoretic approaches with a single variable in the absence of experimental considerations.

References

- [1] K.J. Burch, D.K. Wakefield and E.G. Whitehead, Jr, Boiling Point Models of Alkanes. *MATCDY* **47** (2003) 25-52.
- [2] J.C. Dearden, Quantitative structure property relationships for prediction of boiling point, vapor pressure, and melting point. *Environmental Toxicology and Chemistry* **22**(8) (2003) 1696-1709.
- [3] P. Dankelmann and S. Mukwembi, The Distance Concept and Distance in Graphs (I. Gutman, B. Furtula (Eds.)), *Distance in Molecular Graphs Theory*, Univ. Kragujevac, Kragujevac (2012) 3-48.
- [4] I. Gutman, Selected Properties of the Schultz Molecular Topological Index. *J. Chem. Inf. Comput. Sci.* **34** (1994) 1087–1089.
- [5] S. Nikolić, N. Trinajstić and Z. Mihalić, The Wiener Index: Development and Applications. *Croatica Chemica Acta* **68**(1) (1995) 105-129.
- [6] M. Randić, Z. Mihalić, S. Nikolić, and N. Trinajstić, *Croat. Chem. Acta* **66** (1993) 411.
- [7] H. Wiener, Structural determination of paraffin boiling points. *J. Amer. Chem. Soc.* **69** (1947), 17–20.

- [8] P. Santaka and G. Conduit, Predicting physical properties of alkanes with neural networks. *Preprint submitted to Elsevier* (2019).