Mathematical model for determining octanol-water partition coefficient of alkanes

Jelena Sedlar

University of Split, Croatia

2nd Croatian Combinatorial Days,
Zagreb, Croatia

September 27-28, 2018

Introduction

Chemical molecules are the subject of the study of chemistry and they consist of:

- atoms,
- bonds.

Introduction

Chemical molecules are the subject of the study of chemistry and they consist of:

- atoms,
- bonds.

Graph theoretical model of a molecule is a graph in which:

- atoms are represented by vertices,
- bonds are represented by edges.

Introduction

The example of molecules represented by a graph can be:

Introduction

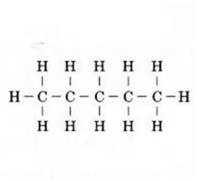
The example of molecules represented by a graph can be:

- isomers of pentane,

Introduction

The example of molecules represented by a graph can be:

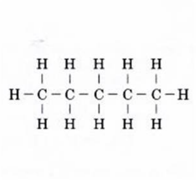
- isomers of pentane,

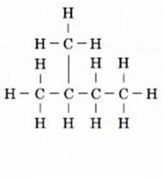


Introduction

The example of molecules represented by a graph can be:

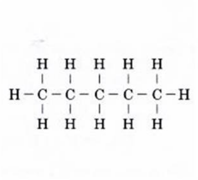
- isomers of pentane,

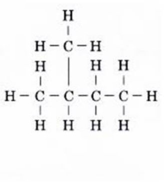


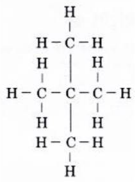


The example of molecules represented by a graph can be:

- isomers of pentane,



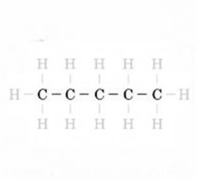


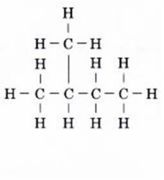


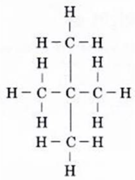
Introduction

The example of molecules represented by a graph can be:

- isomers of pentane,



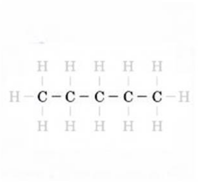


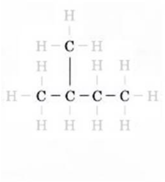


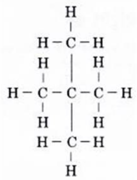
Introduction

The example of molecules represented by a graph can be:

- isomers of pentane,



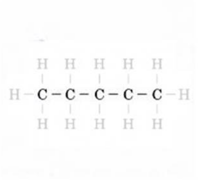


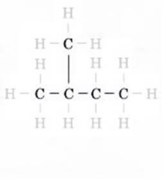


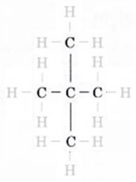
Introduction

The example of molecules represented by a graph can be:

- isomers of pentane,



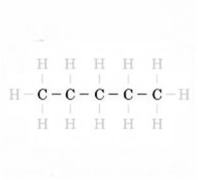


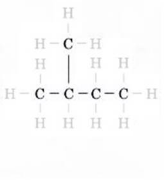


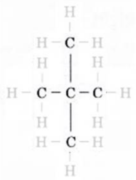
Introduction

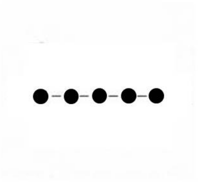
The example of molecules represented by a graph can be:

- isomers of pentane,





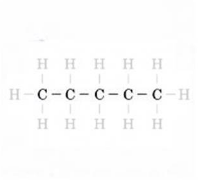


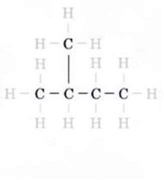


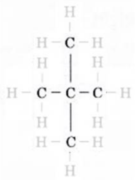
Introduction

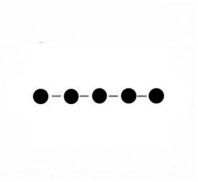
The example of molecules represented by a graph can be:

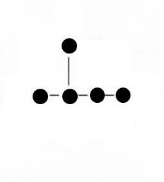
- isomers of pentane,







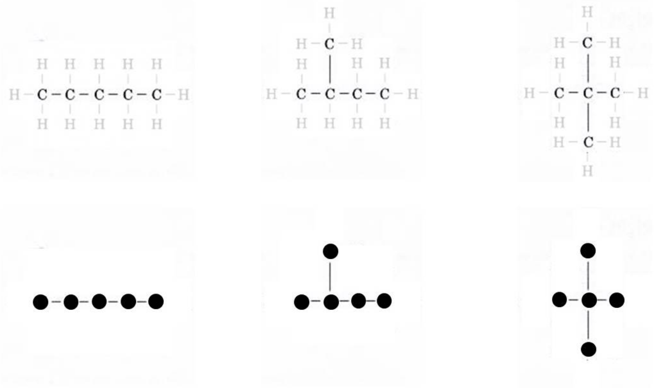




Introduction

The example of molecules represented by a graph can be:

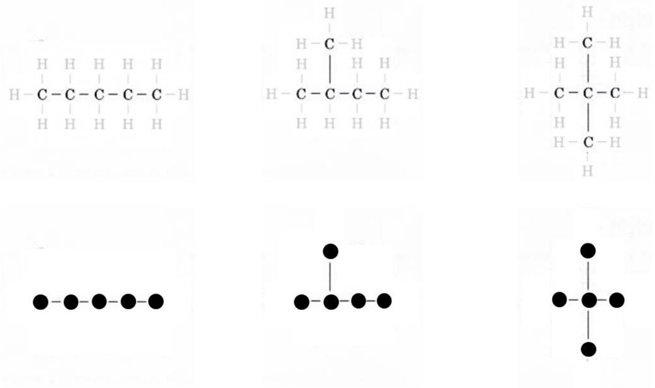
- isomers of pentane,



Introduction

The example of molecules represented by a graph can be:

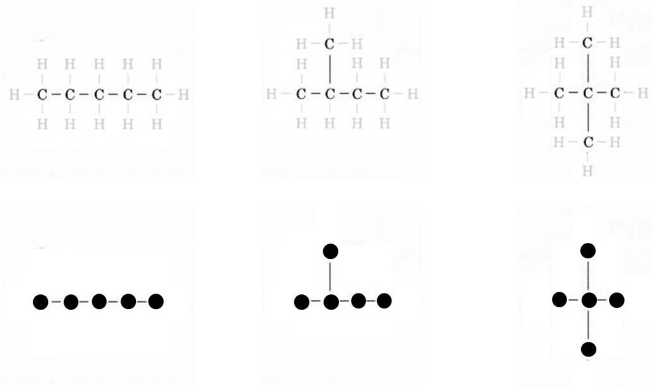
- isomers of pentane,
- isomers of octanes,



Introduction

The example of molecules represented by a graph can be:

- isomers of pentane,
- isomers of octanes,
- isomers of nonanes,...



Introduction

Chemical properties of the molecules which are of interest to people are various, for example:

- boiling point,
- melting point,
- heat capacities,
- density,
- heat of vaporization,
- enthalpy of formation,
- motor octane number,
- molar refraction,
- total surface area,
- octanol-water partition coefficient,
- molar volume,
- entropy,
- acentric factor.

Introduction

Chemical properties of molecules are usually presented in a table.

Molecule	Entropy	AcenFac	HVAP	DHVAP
n-octane	111.70	0.39790	73.19	9.915
2-methyl-heptane	109.80	0.37792	70.30	9.484
3-methyl-heptane	111.30	0.37100	71.30	9.521
4-methyl-heptane	109.30	0.37150	70.91	9.483
3-ethyl-hexane	109.40	0.36247	71.70	9.476
2,2-dimethyl-hexane	103.40	0.33943	67.70	8.915
2,3-dimethyl-hexane	108.00	0.34825	70.20	9.272
2,4-dimethyl-hexane	107.00	0.34422	68.50	9.029
2,5-dimethyl-hexane	105.70	0.35683	68.60	9.051
3,3-dimethyl-hexane	104.70	0.32260	68.50	8.973
3,4-dimethyl-hexane	106.60	0.34035	70.20	9.316
2-methyl-3-ethyl-pentane	106.10	0.33243	69.70	9.209
3-methyl-3-ethyl-pentane	101.50	0.30690	69.30	9.081
2,2,3-trimethyl-pentane	101.30	0.30082	67.30	8.826
2,2,4-trimethyl-pentane	104.10	0.30537	64.87	8.402
2,3,3-trimethyl-pentane	102.10	0.29318	68.10	8.897
2,3,4-trimethyl-pentane	102.40	0.31742	68.37	9.014
2,2,3,3-tetramethylbutane	93.06	0.25529	66.20	8.410

Obtaining chemical properties of the molecules can be done:

Introduction

Obtaining chemical properties of the molecules can be done:

- experimentally in a laboratory - expensive;

Introduction

Obtaining chemical properties of the molecules can be done:

- experimentally in a laboratory - expensive;
- mathematically - cheap.

Introduction

Obtaining chemical properties of the molecules can be done:

- experimentally in a laboratory - expensive;
- mathematically - cheap.

Mathematical model for obtaining chemical properties is based on:

- deriving topological index of a graph which correlates well with the chemical property of a molecule.

Introduction

Obtaining chemical properties of the molecules can be done:

- experimentally in a laboratory - expensive;
- mathematically - cheap.

Mathematical model for obtaining chemical properties is based on:

- deriving topological index of a graph which correlates well with the chemical property of a molecule.

Wiener index of a graph is the first and maybe the best studied example, it is defined as

$$W(G) = \sum_{u,v \in V(G)} d(u,v)$$

and correlates well with a boiling point of paraffines.

Introduction

In 1972 a paper [1] was published, in which approximate formulas for the total $\pi-$electron energy (E) were derived.

[1] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total $\pi-$electron energy of alternant hydrocarbons, Chem. Phys. Lett. 17 (1972) 535-538.

In 1972 a paper [1] was published, in which approximate formulas for the total $\pi-$electron energy (E) were derived.

[1] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total $\pi-$electron energy of alternant hydrocarbons, Chem. Phys. Lett. 17 (1972) 535-538.

In the formulas for the total $\pi-$electron energy the folowing terms ocured:

Introduction

In 1972 a paper [1] was published, in which approximate formulas for the total $\pi-$electron energy (E) were derived.

[1] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total $\pi-$electron energy of alternant hydrocarbons, Chem. Phys. Lett. 17 (1972) 535-538.

In the formulas for the total $\pi-$electron energy the folowing terms ocured:

- the first Zagreb index

$$M_1(G) = \sum_{v \in V(G)} \deg(v)^2$$

Introduction

In 1972 a paper [1] was published, in which approximate formulas for the total $\pi-$electron energy (E) were derived.

[1] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total $\pi-$electron energy of alternant hydrocarbons, Chem. Phys. Lett. 17 (1972) 535-538.

In the formulas for the total $\pi-$electron energy the folowing terms ocured:

- the first Zagreb index

$$M_1(G) = \sum_{v \in V(G)} \deg(v)^2$$

- the Forgotten index

$$F(G) = \sum_{v \in V(G)} \deg(v)^3$$

The predictive ability of the $F-$index was tested in [2]:

- using a dataset of octane isomers,
- found at http://www.moleculardescriptors.eu/dataset/dataset.htm,
- concurring with a suggestion of the International Academy of Mathematical Chemistry.

[2] B. Furtula, I. Gutman, A forgotten topological index, J. Math. Chem. 53 (2015) 1184-1190.

Introduction

The predictive ability of the $F-$index was tested in [2]:

- using a dataset of octane isomers,
- found at http://www.moleculardescriptors.eu/dataset/dataset.htm,
- concurring with a suggestion of the International Academy of Mathematical Chemistry.

[2] B. Furtula, I. Gutman, A forgotten topological index, J. Math. Chem. 53 (2015) 1184-1190.

The result of the test showed that both F and M_1 have similar predictive ability, i.e. both:

Introduction

The predictive ability of the $F-$index was tested in [2]:

- using a dataset of octane isomers,
- found at http://www.moleculardescriptors.eu/dataset/dataset.htm,
- concurring with a suggestion of the International Academy of Mathematical Chemistry.

[2] B. Furtula, I. Gutman, A forgotten topological index, J. Math. Chem. 53 (2015) 1184-1190.

The result of the test showed that both F and M_1 have similar predictive ability, i.e. both:

- correlate well with entropy and acentric factor;

The predictive ability of the $F-$index was tested in [2]:

- using a dataset of octane isomers,
- found at http://www.moleculardescriptors.eu/dataset/dataset.htm,
- concurring with a suggestion of the International Academy of Mathematical Chemistry.

[2] B. Furtula, I. Gutman, A forgotten topological index, J. Math. Chem. 53 (2015) 1184-1190.

The result of the test showed that both F and M_1 have similar predictive ability, i.e. both:

- correlate well with entropy and acentric factor;
- correlate badly with all other properties.

Introduction

The example of correlation of F and M_1 with (log of) octanol-water partition coefficient is presented below (from [2]).

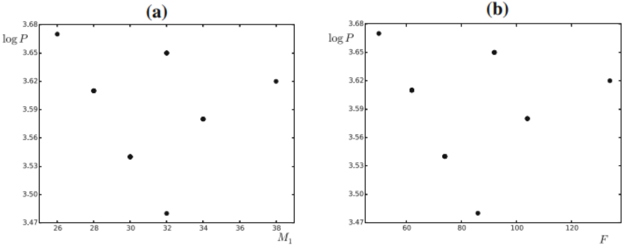


Fig. 1 Logarithm of the octanol-water partition coefficient (P) plotted versus the first Zagreb and the forgotten indices, **a** $\log P$ versus M_1; correlation coefficient is 0.079, **b** $\log P$ versus F; correlation coefficient is 0.0055

Introduction

In order to improve the predictive ability of these indices, a simple linear model is devised (in [2]):

$$M_1 + \lambda F$$

where λ ($-20 \leq \lambda \leq 20$) is fitting parameter.

Introduction

In order to improve the predictive ability of these indices, a simple linear model is devised (in [2]):

$$M_1 + \lambda F$$

where λ $(-20 \leq \lambda \leq 20)$ is fitting parameter.

The model gave:

Introduction

In order to improve the predictive ability of these indices, a simple linear model is devised (in [2]):

$$M_1 + \lambda F$$

where λ $(-20 \leq \lambda \leq 20)$ is fitting parameter.

The model gave:

- insignificant improvement for all physico-chemical properties but the octanol-water partition coefficient;

Introduction

In order to improve the predictive ability of these indices, a simple linear model is devised (in [2]):

$$M_1 + \lambda F$$

where λ $(-20 \leq \lambda \leq 20)$ is fitting parameter.

The model gave:

- insignificant improvement for all physico-chemical properties but the octanol-water partition coefficient;
- a major improvement for the octanol-water partition coefficient.

The correlation coefficient is the best for $\lambda = -0.14$.

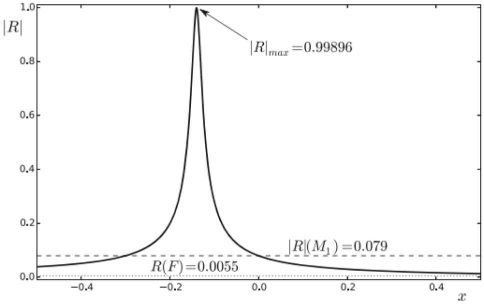


Fig. 2 *Solid line* shows the variation of the absolute value of the correlation coefficient with λ; the *dashed* and *dotted lines* pertain to the correlation coefficients of $\log P$ versus M_1 and F, respectively

The experimental and calculated values of (log of) the octanol-water partition coefficient for $\lambda = -0.14$.

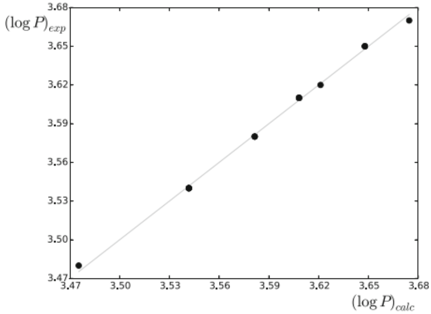


Fig. 3 Experimental versus calculated log P of octanes

Introduction

The questions that arise are the following:

Introduction

The questions that arise are the following:

- is

$$M_1 - 0.14\ F$$

a good model for (log of) the octanol-water partition coefficient of all alkanes (and not just octanes);

Introduction

The questions that arise are the following:

- is
$$M_1 - 0.14\ F$$
  a good model for (log of) the octanol-water partition coefficient of all alkanes (and not just octanes);
- can a coefficient $\lambda = -0.14$ be derived from the structure of the molecule?

Main results

The elimination of a free factor λ.

Main results

The elimination of a free factor λ. Note that for octanes ($n = 8$) we have

$$\lambda = -0.14 =$$

Main results

The elimination of a free factor λ. Note that for octanes ($n = 8$) we have

$$\lambda = -0.14 = \{n = 8\} \approx$$

Main results

The elimination of a free factor λ. Note that for octanes ($n = 8$) we have

$$\lambda = -0.14 = \{n = 8\} \approx -\frac{1}{n-1}$$

Main results

The elimination of a free factor λ. Note that for octanes $(n = 8)$ we have

$$\lambda = -0.14 = \{n = 8\} \approx -\frac{1}{n-1}$$

so maybe

$$M_1(G) - \frac{1}{n-1} F(G)$$

would be a more appropriate term for modeling the octanol-water partition coefficient.

Main results

Lanzhou index $Lzh(G)$ of a graph G is defined by

$$\begin{aligned} Lz(G) &= (n-1)(M_1(G) - \frac{1}{n-1}F(G)) = \\ &= (n-1)M_1(G) - F(G). \end{aligned}$$

Main results

Lanzhou index $Lzh(G)$ of a graph G is defined by

$$\begin{aligned} Lz(G) &= (n-1)(M_1(G) - \frac{1}{n-1}F(G)) = \\ &= (n-1)M_1(G) - F(G). \end{aligned}$$

Practical formula for Lanzhou index obtained by inserting the formulas for $M_1(G)$ and $F(G)$ is

$$Lz(G) = \sum_{u \in V} d^2(u)(n-1-d(u)) =$$

Main results

Lanzhou index $Lzh(G)$ of a graph G is defined by

$$\begin{aligned} Lz(G) &= (n-1)(M_1(G) - \frac{1}{n-1}F(G)) = \\ &= (n-1)M_1(G) - F(G). \end{aligned}$$

Practical formula for Lanzhou index obtained by inserting the formulas for $M_1(G)$ and $F(G)$ is

$$Lz(G) = \sum_{u \in V} d^2(u)(n-1-d(u)) = \sum_{u \in V} d^2(u)\bar{d}(u).$$

Main results

Lanzhou index $Lzh(G)$ of a graph G is defined by

$$\begin{aligned} Lz(G) &= (n-1)(M_1(G) - \frac{1}{n-1}F(G)) = \\ &= (n-1)M_1(G) - F(G). \end{aligned}$$

Practical formula for Lanzhou index obtained by inserting the formulas for $M_1(G)$ and $F(G)$ is

$$Lz(G) = \sum_{u \in V} d^2(u)(n-1-d(u)) = \sum_{u \in V} d^2(u)\bar{d}(u).$$

We also define

$$\overline{Lz}(G) = \sum_{u \in V} \bar{d}^2(u)\bar{d}(u).$$

Main results

Justification of the introduced index.

Main results

Justification of the introduced index. Note that:

Main results

Justification of the introduced index. Note that:

- for octanes models $M_1 - 0.14\ F$ and $M_1 - \frac{1}{n-1}\ F$ are equivalent,

Main results

Justification of the introduced index. Note that:

- for octanes models $M_1 - 0.14\ F$ and $M_1 - \frac{1}{n-1}\ F$ are equivalent,
- the performance has to be checked on another group of alkanes (nonanes, dekanes,...).

Main results

Justification of the introduced index. Note that:

- for octanes models $M_1 - 0.14\ F$ and $M_1 - \frac{1}{n-1}\ F$ are equivalent,
- the performance has to be checked on another group of alkanes (nonanes, dekanes,...).

This was done for nonanes aaaand...

Main results

Justification of the introduced index. Note that:

- for octanes models $M_1 - 0.14\ F$ and $M_1 - \frac{1}{n-1}\ F$ are equivalent,
- the performance has to be checked on another group of alkanes (nonanes, dekanes,...).

This was done for nonanes aaaand...

	M_1	F	$M_1 - 0.140F$	Lz	$\overline{Lz}$
octanes	-0.07933	0.00550	-0.99876	-0.96693	0.16616
nonanes	-0.78729	-0.73379	-0.65970	-0.98869	0.82598

Table 1. Comparison of correlation coefficients of five indices for the octanol-water partition coefficient of octane and nonane isomers.

Main results

Remark. Now that the new index is chemically justified, we can gladly do the math work. :-)

Main results

Remark. Now that the new index is chemically justified, we can gladly do the math work. :-)

The math work is finding:

- extremal graphs,
- extremal trees,
- extremal unicyclic graphs,
- ...

Main results

General graps

Main results

General graps

Proposition. Let G be a graph on n vertices. Then

$$0 \leq Lz(G) \leq \frac{4}{27} n(n-1)^3$$

The left bound is obtained if and only if G is either complete or empty graph. The right bound is obtained if and only if $n \equiv 1 \pmod 3$ and G is $r-$regular with $r = \frac{2}{3}(n-1)$.

Proposition. Let G be a graph on n vertices. Then

$$0 \leq Lz(G) \leq \frac{4}{27}n(n-1)^3$$

The left bound is obtained if and only if G is either complete or empty graph. The right bound is obtained if and only if $n \equiv 1 \pmod 3$ and G is $r-$regular with $r = \frac{2}{3}(n-1)$.

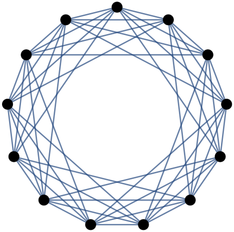
$Lz(G) = \sum_{u \in V} d^2(u)(n-1-d(u))$

Proposition. Let G be a graph on n vertices. Then

$$0 \leq Lz(G) \leq \frac{4}{27}n(n-1)^3$$

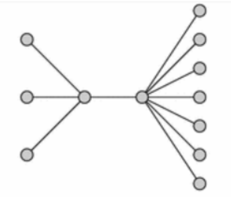
The left bound is obtained if and only if G is either complete or empty graph. The right bound is obtained if and only if $n \equiv 1 \pmod 3$ and G is $r-$regular with $r = \frac{2}{3}(n-1)$.

$Lz(G) = \sum_{u \in V} d^2(u)(n-1-d(u))$

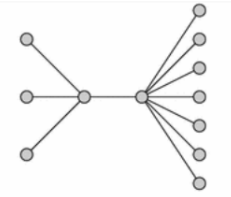


A **double star** $S_{k,l}$ is a tree obtained from K_2 by attaching $k-1$ leaves to one of its vertices and $l-1$ leaves to the other one.

A **double star** $S_{k,l}$ is a tree obtained from K_2 by attaching $k-1$ leaves to one of its vertices and $l-1$ leaves to the other one.

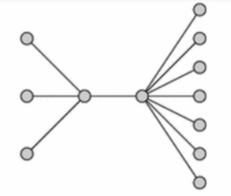


A **double star** $S_{k,l}$ is a tree obtained from K_2 by attaching $k-1$ leaves to one of its vertices and $l-1$ leaves to the other one.



A **balanced double star** $BDS(n)$ on n vertices is a double star in which the diference between k and l is the smallest possible,

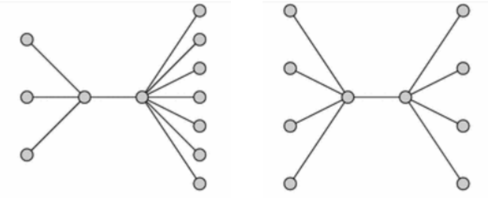
A **double star** $S_{k,l}$ is a tree obtained from K_2 by attaching $k-1$ leaves to one of its vertices and $l-1$ leaves to the other one.



A **balanced double star** $BDS(n)$ on n vertices is a double star in which the diference between k and l is the smallest possible, hence

$$BDS(n) = S_{n/2,n/2} \text{ or } BDS(n) = S_{(n-1)/2,(n+1)/2}.$$

A **double star** $S_{k,l}$ is a tree obtained from K_2 by attaching $k-1$ leaves to one of its vertices and $l-1$ leaves to the other one.



A **balanced double star** $BDS(n)$ on n vertices is a double star in which the diference between k and l is the smallest possible, hence

$$BDS(n) = S_{n/2,n/2} \text{ or } BDS(n) = S_{(n-1)/2,(n+1)/2}.$$

Proposition. Let T_n be a tree on $n \geq 15$ vertices. Then

$$Lz(S_n) \leq Lz(T_n) \leq Lz(BDS(n)).$$

The lower bound is achieved if and only if $T_n = S_n$, while the upper bound is achieved if and only if $T_n = BDS(n)$.

It remains to consider the trees with less than 15 vertices.

It remains to consider the trees with less than 15 vertices.

We say that a tree $T_n \neq BDS(n)$ is **excessive** if $Lz(T_n) > Lz(BDS(n))$.

It remains to consider the trees with less than 15 vertices.

We say that a tree $T_n \neq BDS(n)$ is **excessive** if $Lz(T_n) > Lz(BDS(n))$.

If $Lz(Tn) \geq Lz(BDS(n))$, we say that T_n is **weakly excessive**.

It remains to consider the trees with less than 15 vertices.

We say that a tree $T_n \neq BDS(n)$ is **excessive** if $Lz(T_n) > Lz(BDS(n))$.

If $Lz(Tn) \geq Lz(BDS(n))$, we say that T_n is **weakly excessive**.

It can be proved that:

It remains to consider the trees with less than 15 vertices.

We say that a tree $T_n \neq BDS(n)$ is **excessive** if $Lz(T_n) > Lz(BDS(n))$.

If $Lz(Tn) \geq Lz(BDS(n))$, we say that T_n is **weakly excessive**.

It can be proved that:

- number of excessive and of weakly excessive trees is finite,

It remains to consider the trees with less than 15 vertices.

We say that a tree $T_n \neq BDS(n)$ is **excessive** if $Lz(T_n) > Lz(BDS(n))$.

If $Lz(Tn) \geq Lz(BDS(n))$, we say that T_n is **weakly excessive**.

It can be proved that:

- number of excessive and of weakly excessive trees is finite,
- there are no excessive nor weakly excessive trees on more than 14 vertices.

Main results

n	$t(n)$	$t_{we}(n)$	$t_{ex}(n)$	$t_{max}(n)$	$Lz_{max}(n)$	$Lz(BDS(n))$
4	2	0	0	0	12	12
5	3	1	1	1	30	26
6	6	2	2	1	56	52
7	11	7	6	6	90	84
8	23	16	12	1	138	132
9	47	29	24	2	196	188
10	106	21	2	2	270	264
11	235	14	6	1	360	350
12	551	4	4	2	464	460
13	1301	9	2	2	588	582
14	3159	1	0	1	732	732

Table 2. Statistics of weakly excessive, excessive, and extremal trees, and maximum value of Lanzhou index for trees on $4 \leq n \leq 14$ vertices.

Main results

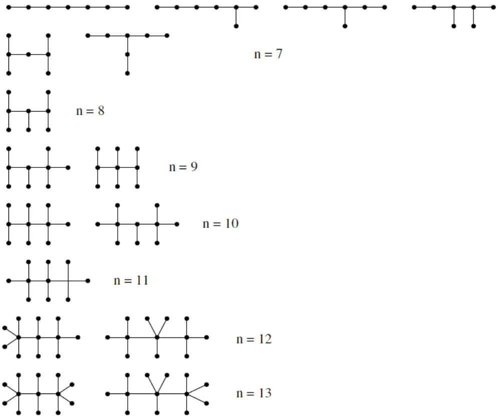


Figure 1. Excessive trees maximizing Lanzhou index.

Main results

Degree bound graphs

Let $\mathcal{T}_n^{\Delta}$ denote the set of all trees on n vertices with maximum degree at most Δ.

Let $\mathcal{T}_n^{\Delta}$ denote the set of all trees on n vertices with maximum degree at most Δ.

Note that:

Let $\mathcal{T}_n^{\Delta}$ denote the set of all trees on n vertices with maximum degree at most Δ.

Note that:

- $\Delta = 3$ is the smallest interesting case,

Let $\mathcal{T}_n^{\Delta}$ denote the set of all trees on n vertices with maximum degree at most Δ.

Note that:

- $\Delta = 3$ is the smallest interesting case,
- for $\Delta = 4$ we obtain the class of chemical trees.

Main results

Proposition. Let $n \geq 8$ be an integer and $T_n \in \mathcal{T}_n^3$. Then

$$4n^2 - 18n + 20 \leq Lz(T_n) \leq 5n^2 - 27n - (n-7)\frac{1-(-1)^n}{2}.$$

The left inequality is sattisfied if and only if $T_n = P_n$. The right inequality is sattisfied for any tree without vertices of degree 2 if n is even, and for any tree with exactly one vertex of degree 2 if n is odd.

Main results

Proposition. Let $n \geq 8$ be an integer and $T_n \in \mathcal{T}_n^4$. Then

$$4n^2 - 18n + 20 \leq Lz(T_n) \leq 6n^2 + O(n).$$

The left inequality is sattisfied if and only if $T_n = P_n$. The maximum value of $Lz(T_n)$ is achieved for any tree having the largest possible number of vertices of degree 4 for a given n.

Main results

Proposition. Let $n \geq 8$ be an integer and $T_n \in \mathcal{T}_n^4$. Then

$$4n^2 - 18n + 20 \leq Lz(T_n) \leq 6n^2 + O(n).$$

The left inequality is sattisfied if and only if $T_n = P_n$. The maximum value of $Lz(T_n)$ is achieved for any tree having the largest possible number of vertices of degree 4 for a given n.

The details of the proofs can be seen in [3].

[3] D. Vukičević, Q. Li, J. Sedlar, T. Došlić: Lanzhou Index, MATCH Commun. Math. Comput. Chem. 80 (2018) 863-876.

In the end

Thank you for your attention!