RELATION BETWEEN RANDIĆ INDEX
AND AVERAGE DISTANCE OF TREES

Marek Cygan   Michał Pilipczuk

Riste Škrekovski

Ljubljana, September 17, 2010

Relation between Randić index and average distance of trees[*]

Marek Cygan, Michał Pilipczuk[†] and Riste Škrekovski[‡]

Abstract

The Randić index $R(G)$ of a graph G is the sum of weights $(\deg(u)\deg(v))^{-0.5}$ over all edges uv of G, where $\deg(v)$ denotes the degree of a vertex v. We prove that for any tree T with n_1 leaves $R(T) \geq \mathrm{ad}(T) + \max(0, \sqrt{n_1} - 2)$, where $\mathrm{ad}(T)$ is the average distance between vertices of T. As a consequence we resolve the conjecture $R(G) \geq \mathrm{ad}(G)$ given by Fajtlowicz in 1988 for the case when G is a tree.

1 Introduction

In chemical graph theory topological indices belong to the set of molecular descriptors that are calculated based on the molecular graph of a chemical compound. In 1975 Milan Randić [11] introduced a topological connectivity index $R(G)$ of a graph G defined as the sum of weights $(\deg(u)\deg(v))^{-0.5}$ over all edges uv of G, i.e.,

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

where $\deg(v)$ is the degree of a vertex v. Randić has shown that there exists a correlation of the Randić index with several physico-chemical properties of alkanes such as boiling points, chromatographic retention times, enthalpies of formation, parameters in the Antoine equation for vapor pressure, surface areas and others. More information about Randić index can be found in a survey [7] by Li and Shi or in a book [8] by Li and Gutman.

The Wiener index is a distance-based topological index defined as the sum of distances between all pairs of vertices in a graph and is denoted by $W(G)$. It was the first topological index in chemistry [13], introduced in 1945. Since then, the Wiener index has been used to explain various chemical and physical properties of molecules. However the Wiener index can be cubic in the number of vertices of a graph, hence for the sake of simplicity we use the average distance instead

$$\mathrm{ad}(G) = W(G)/\binom{n}{2}.$$

For a one vertex graph K_1 we assume $R(K_1) = \mathrm{ad}(K_1) = 0$.

For the last two decades researchers are investigating extremal values and relations between topological indices. In 1988 Fajtlowicz [5] stated the following conjecture based on the computer program Graffiti.

[†]Department of Mathematics, Computer Science and Mechanics, University of Warsaw, Warsaw, Poland, {cygan@,michal.pilipczuk@students}.mimuw.edu.pl

[‡]Department of Mathematics, University of Ljubljana, Jadranska 19, 1111 Ljubljana, Slovenia, skrekovski@gmail.com

Conjecture 1.1. *For all connected graphs G, $R(G) \geq \mathrm{ad}(G)$.*

Very recently Li and Shi [9] have proven Conjecture 1.1 when $\delta(G) \geq n/5$ and $n \geq 15$. More recent results related to extremal values of the Randić index can be found in [1, 3, 6, 14, 15].

In order to state the results of this paper let us introduce some notation. By $\mathrm{dist}_G(u, v)$ we denote the length of the shortest path between vertices u and v in a connected graph G. The length $\max_{u,v} \mathrm{dist}_G(u, v)$ is a diameter of a graph G, which we denote by $\mathrm{diam}(G)$. In particular for trees the diameter is the length of the longest simple path.

A *star* is a tree where at most one vertex is of degree greater than one. Note that a tree with at most three vertices is a star. Also notice that stars are precisely the trees with diameter at most two.

The main result of this paper is the following theorem.

Theorem 1.2. *For any tree T with n vertices and n_1 leaves the following inequality holds:*

$$R(T) \geq \mathrm{ad}(T) + \max(0, \sqrt{n_1} - 2).$$

Moreover if we consider the limit when n goes to infinity, the inequality is sharp for stars.

As a consequence we prove Conjecture 1.1 for the case when G is a tree.

2 Proof of Theorem 1.2

A vertex of degree one is called a *pendant vertex* or a *leaf*. By $\mathcal{T}_{n,n_1}$ we denote the set of all trees with exactly n nodes and n_1 pendant vertices.

Firstly let us define a special class of trees, which will be frequently used in our paper. For $a, b \geq 1$, $n \geq a+b+2$ by $\mathrm{DC}(n, a, b)$ we denote a *double comet*, which is a tree composed of a path containing $n - a - b$ vertices with a pendant vertices attached to one of the ends of the path and b pendant vertices attached to the other end of the path. Thus, $\mathrm{DC}(n, a, b)$ has n vertices and $a + b$ leaves, i.e. $\mathrm{DC}(n, a, b) \in \mathcal{T}_{n,a+b}$.
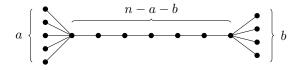

Figure 1: A double comet $\mathrm{DC}(n, a, b)$ for $n = 16, a = 5, b = 4$.

We begin our proof with a lemma proving Theorem 1.2 for stars and paths.

Lemma 2.1. *Let $T \in \mathcal{T}_{n,n_1}$ be a tree such that $n_1 \leq 2$ (a path) or $n_1 = n - 1$ (a star). Then $R(T) \geq \mathrm{ad}(T) + \max(0, \sqrt{n_1} - 2)$.*

Proof. If $n \leq 2$ then the inequality trivially holds, hence we assume $n \geq 3$. If T is a star then by direct calculations we obtain $\mathrm{ad}(T) = 2 - \frac{2}{n}$ and $R(T) = \sqrt{n_1} = \sqrt{n - 1}$, whereas if T is a path then $\mathrm{ad}(T) = \frac{n+1}{3}$ and $R(T) = \sqrt{2} + \frac{n-3}{2}$. In both cases the lemma holds. □

Consequently from now on we assume that $T \in \mathcal{T}_{n,n_1}$ where $3 \leq n_1 \leq n - 2$, which means that we may use the following result proved by Liu et al. in Theorem 5 from [10].

Theorem 2.2. *Let $T \in \mathcal{T}_{n,n_1}$, where $3 \leq n_1 \leq n - 2$. Then*

$$R(T) \geq \frac{n - n_1}{2} + \sqrt{n_1} + (1 - \sqrt{2})(\frac{1}{\sqrt{2n_1}} + \frac{1}{\sqrt{2}}).$$

Corollary 2.3. *Using the assumption $n - 2 \geq n_1 \geq 3$ we obtain:*

$$R(T) \geq \frac{n - n_1}{2} + \sqrt{n_1} - 0.462\,.$$

The flow of the proof is as follows. Firstly we show that each tree can be transformed into a double comet without decreasing the average distance and simultaneously preserving the number of vertices and leaves. Since in Corollary 2.3 we only use n and n_1 disregarding the actual structure of a tree it is enough to show that for double comets we have $\frac{n-n_1}{2} + \sqrt{n_1} - 0.462 \geq \mathrm{ad}(T) + \max(0, \sqrt{n_1} - 2)\,.$

Lemma 2.4. *Let $T \in \mathfrak{T}_{n,n_1}$, where $3 \leq n_1 \leq n - 2$. There exists a double comet $T' = \mathrm{DC}(n, a, b)$ for some $a, b \geq 1$, $a + b = n_1$ such that $\mathrm{ad}(T) \leq \mathrm{ad}(T')\,.$*

Proof. Let us assume that contrary. Let $T \in \mathfrak{T}_{n,n_1}$ be a non-double-commet tree such that for each double comet $T' = \mathrm{DC}(n, a, b)$, where $a, b \geq 1, a + b = n_1$ we have $\mathrm{ad}(T') < \mathrm{ad}(T)$. Among all such trees from $\mathfrak{T}_{n,n_1}$ we consider a tree with the largest diameter $\mathrm{diam}(T)$.

Let $d = \mathrm{diam}(T)$ and $D = v_1 v_2 \ldots v_d$ be a diameter of the tree T. Since T is neither a star nor a double comet we have $d \geq 5$. Moreover, we know that v_1, v_d are leaves and all neighbours of v_2, v_{d-1} outside D are pendants, because otherwise D would not be a diameter. Furthermore, since T is not a double comet there exists a vertex outside the diameter $x \notin D$ that is a neighbour of some vertex v_i for $3 \leq i \leq d - 2$. Observe that after a removal of the edge xv_i the tree T decomposes into exactly two parts. Let us denote by T_x the part containing the vertex x and by T_D the other one. Similarly after removal of the edge $v_{i-1}v_i$ the tree T_D decomposes into two parts T_{D_1}, T_{D_2}, where $v_{i-1} \in V(T_{D_1})$ and $v_i \in V(T_{D_2})$. Without loss of generality we may assume, that $|V(T_{D_1})| \leq |V(T_{D_2})|$, since otherwise we can reverse the diameter D. Let T' be a tree T with the edge xv_i removed and xv_2 inserted (as in Fig. 2). We claim $\mathrm{ad}(T') \geq \mathrm{ad}(T)$. Observe that distances between vertices within $V(T_D)$ and within $V(T_x)$ in both trees T, T' are the same. Now we consider distances between vertices in $V(T_D)$ and $V(T_x)$. Let $a \in V(T_D), b \in V(T_x)$. Consider two cases:

- $a \in V(T_{D_1})$, then $\mathrm{dist}_{T'}(a, b) \geq \mathrm{dist}_T(a, b) - (i - 2)$,

- $a \in V(T_{D_2})$, then $\mathrm{dist}_{T'}(a, b) \geq \mathrm{dist}_T(a, b) + (i - 2)$.

Since $|V(T_{D_1})| \leq |V(T_{D_2})|$, we have $W(T') \geq W(T)$ and so $\mathrm{ad}(T') \geq \mathrm{ad}(T)$. If $|V(T_x)| > 1$ then $\mathrm{diam}(T') > \mathrm{diam}(T)$, hence the lemma holds for T' because $\mathrm{diam}(T')$ is larger. Which means that there exists a double comet $T'' = DC(n, a, b)$, where $a + b = n_1$, such that $\mathrm{ad}(T'') \geq \mathrm{ad}(T') \geq \mathrm{ad}(T)$, a contradiction. Thus we know that $|V(T_x)| = 1$. However we can continue in this manner moving pendant neighbours of vertices v_i for $3 \leq i \leq d - 2$ ending up with a double comet, thus completing the proof. $\qquad \square$

Obviously, the biggest value of $\mathrm{ad}(\mathrm{DC}(n, a, b))$, where $a + b = n_1$, is obtained when a, b are as close as possible to $n_1/2$, i.e. $a = \lfloor \frac{n_1}{2} \rfloor$ and $b = \lceil \frac{n_1}{2} \rceil$. These bounds are also observed in [4, 12] (see Theorem 31 in the survey [2]) but as the papers [4, 12] may not be easy to access and as our proof is short, for the sake of completness we decide to include it.

Lemma 2.5. *Let $T \in \mathfrak{T}_{n,k}$ be a double comet $\mathrm{DC}(n, a, b)$ for $a, b \geq 1$, where $3 \leq a + b = n_1 \leq n - 2$. Then*

$$\frac{n - n_1}{2} + \sqrt{n_1} - 0.462 \geq \mathrm{ad}(T) + \max(0, \sqrt{n_1} - 2)\,.$$

Proof. Firstly we calculate the average distance for the double comet T:

$$\binom{n}{2}\mathrm{ad}(T) = W(T)$$

$$= ab(n - n_1 + 1) + 2\binom{a}{2} + 2\binom{b}{2} + (a + b)(n - n_1)(n - n_1 + 1)/2 + \binom{n - n_1 + 1}{3},$$
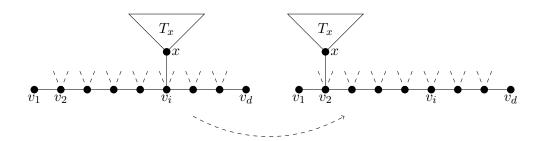
Figure 2: Ilustration of the transformation used in the proof of Lemma 2.4.

where $ab(n-n_1+1)+2\binom{a}{2}+2\binom{b}{2}$ comes from distances between leaves, $(a+b)(n-n_1)(n-n_1+1)/2$ is the sum of distances between leaves and inner vertices and $\binom{n-n_1+1}{3}$ is the sum of distances between inner vertices. Using following inequalities:

$$2ab + 2\binom{a}{2} + 2\binom{b}{2} \leq (a+b)^2 = n_1^2$$

$$ab \leq \frac{n_1^2}{4}$$

$$\binom{n-n_1+1}{3} \leq \frac{(n-n_1)^3}{6}$$

we obtain

$$\mathrm{ad}(T)\binom{n}{2} \leq \frac{n_1^2(n-n_1-1)}{4} + n_1^2 + \frac{n_1(n-n_1)(n-n_1+1)}{2} + \frac{(n-n_1)^3}{6}$$

$$= \frac{n_1^2(n-n_1)}{4} + \frac{3n_1^2}{4} + \frac{n_1(n-n_1)^2}{2} + \frac{n_1(n-n_1)}{2} + \frac{(n-n_1)^3}{6}.$$

For the sake of simplicity we put $x = n_1, y = n - n_1$.

$$\mathrm{ad}(T)\binom{n}{2} \leq \frac{x^2y}{4} + \frac{3x^2}{4} + \frac{xy^2}{2} + \frac{xy}{2} + \frac{y^3}{6}$$

In order to prove the lemma it is enough to show the following inequality:

$$\binom{n}{2}\left(\frac{n-n_1}{2} + \sqrt{n_1} - 0.462 - \max(0, \sqrt{n_1} - 2) - \mathrm{ad}(T)\right) \geq 0.$$

We multiply the expression by 4, put x, y instead of $n_1, n - n_1$ and use the previously obtained inequality for $\mathrm{ad}(T)\binom{n}{2}$.

$$4\binom{n}{2}\left(\frac{n-n_1}{2} + \sqrt{n_1} - 0.462 - \max(0, \sqrt{n_1} - 2) - \mathrm{ad}(T)\right) =$$

$$2(x+y)(x+y-1)\left(\frac{y}{2} + \min(\sqrt{x} - 0.462, 1.538)\right) - 4\mathrm{ad}(T)\binom{n}{2} \geq$$

$$2(x^2 + y^2 + 2xy - x - y)\min(\sqrt{x} - 0.462, 1.538) + y^3/3 - 3xy - y^2 - 3x^2$$

4

Now we consider two cases. Either $n_1 = x = 3$ or $n_1 = x \geq 4$. When $x = 3$ from the above expression we obtain:

$$
\begin{aligned}
2(9 + y^2 + 6y - 3 - y)\min(\sqrt{3} - 0.462, 1.538) + y^3/3 - 9y - y^2 - 27 &\geq \\
2.5(y^2 + 5y + 6) + y^3/3 - y^2 - 9y - 27 &= \\
y^3/3 + 1.5y^2 + 3.5y - 12 \, .
\end{aligned}
$$

Using the assumption $n_1 \leq n - 2$ which is equivalent to $y \geq 2$ we obtain $y^3/3 + 1.5y^2 + 3.5y - 12 \geq 11/3 \geq 0$.

Now we assume that $x \geq 4$, hence $\min(\sqrt{x} - 0.462, 1.538) \geq 1.5$. Since $x, y \geq 2$ then $xy \geq x + y$, and so we have:

$$
2(x^2 + y^2 + 2xy - x - y)1.5 + y^3/3 - 3xy - y^2 - 3x^2 = y^3/3 + 2y^2 + 3xy - 3(x + y) \geq 0 \, ,
$$

what ends the proof of the lemma. $\qquad\square$

In Lemma 2.1 we prove Theorem 1.2 for paths and stars whereas by Lemmas 2.4 and 2.5 together with Corollary 2.3 we obtain the inequality for all other trees, thus completing the proof of Theorem 1.2.

References

[1] M. Aouchiche, P. Hansen, M. Zheng, *Variable neighborhood search for extremal graphs 18: conjectures and results about Randić index*, MATCH Commun. Math. Comput. Chem. **56** (2007) 541–550.

[2] A. Dobrynin, R. Entringer, I. Gutman, *Wiener Index of Trees: Theory and Applications*, Acta Applicandae Mathematicae **66** (3)(2001) 211–249.

[3] Z. Du, B. Zhou, N. Trinajstic, *On Randic Indices of Chemical Trees and Chemical Unicyclic Graphs*, MATCH Commun. Math. Comput. Chem. **62** (1)(2009) 131–142.

[4] R. Entringer, *Bounds for the average distance-inverse degree product in trees*, In: Y. Alavi, D. Lick and A. Schwenk (Eds), *Combinatorics, Graph Theory, and Algorithms*, New Issues Press, Kalamazoo, (1999) 335–352.

[5] S. Fajtlowicz, *On conjectures of Graffiti*, Discrete Math. **72** (1988) 113–118.

[6] J. Gao, M. Lu, *On the Randić index of unicyclic graphs*, MATCH Commun. Math. Comput. Chem. **53** (2005) 377–384.

[7] X. Li, Y. Shi, *A survey on the Randić index*, MATCH Commun. Math. Comput. Chem. **59** (1)(2008) 127–156.

[8] X. Li, I. Gutman, *Mathematical Aspects of Randić-Type Molecular Structure Descriptors*, Mathematical Chemistry Monographs No.1, Kragujevac, 2006.

[9] X. Li, Y. Shi, *Randic index, diameter and average distance*, MATCH Commun. Math. Comput. Chem. **64** (2)(2010) 425–431.

[10] H. Liu, M. Liu, F. Tian, *Trees of extremal connectivity index*, Discrete Applied Math. **154** (2006) 106–119.

[11] M. Randic, *On characterization of molecular branching*, J. Amer. Chem. Soc. **97** (1975) 6609–6615.

[12] R. Shi, *The average distance of trees*, Systems Sci. Math. Sci. **6** (1993) 18–24.

[13] H. Wiener, *Correlation of Heats of Isomerization and Differences in Heats of Vaporization of Isomers among the Paraffin Hydrocarbons*, J. Amer. Chem. Soc. **69** (1947) 2636–2638.

[14] L. Zhang, M. Lu, F. Tian, *Maximum Randić index on trees with k pendent vertices*, J. Math. Chem. **41** (2007) 161–171.

[15] M. Zhang, B. Liu, *On a Conjecture about the Randic Index and Diameter*, MATCH Commun. Math. Comput. Chem. **64** (2)(2010) 433–442.