Contributions to Mathematics
www.shahindp.com/locate/cm

Contrib. Math. **7** (2023) 1–10
DOI: 10.47443/cm.2022.051

Research Article

The average size of independent sets in unicyclic graphs

Zuwen Luo[1,2], Kexiang Xu[1,2], Ahmet Sinan Çevik[3], Ivan Gutman[4,*]

[1] *College of Mathematics, Nanjing University of Aeronautics & Astronautics, Nanjing, China*
[2] *MIIT Key Laboratory of Mathematical Modelling and High Performance Computing of Air Vehicles, Nanjing, China*
[3] *Department of Mathematics, Faculty of Science, Konya Selcuk University, Konya, Turkey*
[4] *Faculty of Science, University of Kragujevac, Kragujevac, Serbia*

(Received: 28 September 2022. Received in revised form: 14 November 2022. Accepted: 15 November 2022. Published online: 23 November 2022.)

Abstract

The average size of independent sets of a graph (avi) is considered. It can be viewed as the logarithmic derivative of the independence polynomial at 1. Lower and upper bounds on avi for unicyclic graphs of a given order are determined, and the respective extremal graphs are given. The unicyclic graphs that maximize (minimize) avi coincide with those that maximize (minimize, respectively) the number of independent sets.

Keywords: independent set (in graphs); independence polynomial; chemical graph theory.

2020 Mathematics Subject Classification: 05C09, 05C69.

1. Introduction

In 1980s, Jamison [21,22] initiated the study of the average size of subtrees of a tree. He studied an extremal problem and proved that the path P_n has the smallest average size of subtrees among all trees of a fixed order n. However, although the average size of subtrees of a tree has been extensively studied (see [8,25,27,31,34]), the problem of describing the tree(s) of a given order with the largest average size of subtrees remains unresolved. Inspired by these works, the study of the average size of certain substructures of graphs, including subtrees [9,10], independent sets [1], matching [2], dominating sets [5,14] and connected sets [15,30], has attracted increasing attention in recent years.

An independent set of a graph G is a set of pairwise non-adjacent vertices of G. The total number of such sets has been studied extensively and is usually called Merrifield–Simmons index in honor of the work of Merrifield and Simmons in mathematical chemistry [26]. In fact it has been called the Fibonacci number of G (cf. [23]), based on the fact that in the case of a path, $i(P_n) = F_{n+2}$, where F_h is the standard h-th Fibonacci number, $F_1 = 1\,,\,, F_2 = 1\,,\, F_3 = 2\,,\, F_{h+2} = F_{h+1} + F_h$. In addition, the association of $i(G)$ with the hard core model is well established in statistical physics [7]. It is of interest to determine the graphs having extremal (maximal or minimal) $i(G)$-values in several given graph classes, especially in the context of mathematical chemistry, such as trees, unicyclic graphs, bicyclic graphs and so on (see [13,29,35,36,39,40]). For more extremal results on $i(G)$ see the surveys [32,33] and for related results see [19,20,37,38].

This paper is focused on the extremal problems of the average size of independent sets (avi), instead of their number $(i(G))$. Different authors have already studied bounds on this invariant for specific graphs, see [11,12,24]. Recently Andriantiana, Misanantenaina, and Wagner [1] studied the extreme problem of this invariant and showed that the complete and empty graphs attain the minimum and maximum among all graphs of a given order, whereas the path and the star attain the minimum and maximum among all trees of a given order. It turns out that these results parallel the extremal graph for the number of independent sets. In view of the definition of correlation, this may not be too surprising. However, the correlation is not fully understood between these two invariants (the average size of independent sets and the number of independent sets).

A natural question is whether the extremal graph for the average size of independent sets in other graph classes remains parallel to the extremal graph for the number of independent sets. A unicyclic graph, as a natural extension of a tree, is a connected graph containing exactly one cycle. Many extremal problems, considered for various classes of trees, can be extended to unicyclic graphs. For example, Pedersen and Vestergaard [29] determined sharp lower and upper bounds for the Merrifield–Simmons index in a unicyclic graph of a given order. Andriantiana and Wang [4] characterized the extremal unicyclic graphs with respect to the number of subtrees. Ou [28] described the unicyclic graphs of order n with the largest and second largest Hosoya index. Hou [17] determined the unicyclic graphs of order n with the minimal energy. Hou,

*Corresponding author (gutman@kg.ac.rs).

Gutman, and Woo [18] characterized the unicyclic bipartite graphs of order n having the maximal energy. Andriantiana and Wagner [3] also found the non-bipartite unicyclic graphs of order n with the maximum energy.

In this paper, we study the average size of independent sets in a unicyclic graph and characterized the unicyclic graphs of order n having the largest and smallest average size of independent sets. These turn out that the unicyclic graphs that maximize (minimize) the average size of independent sets are those that also maximize (minimize, respectively) the number of independent sets. Section 2 gives some auxiliary results that will be used for the subsequent proofs. In Section 3, we determine the graphs minimizing and maximizing the average size of independent sets among unicyclic graphs with the given order. We conclude in Section 4 with some questions.

2. Preliminaries

Let G be a simple graph with vertex set $V(G)$ and edge set $E(G)$. The number of vertices of G is called the *order* of G, denoted by $|G|$. For a vertex $v \in V(G)$, the *open neighborhood* of v is $N_G(v) = \{u \in V(G) : uv \in E(G)\}$ and the *closed neighborhood* of v is $N_G[v] = N_G(v) \cup \{v\}$. We use $N(v)$ and $N[v]$, respectively, if there is no ambiguity on G. The degree of $v \in V(G)$ is just $d_G(v) = |N_G(v)|$, and a vertex of degree 1 is called a leaf vertex. For a vertex subset $X \subseteq V(G)$, we write $G - X$ for the graph obtained from G by removing all the vertices in X and the edges incident with them. The distance $d(u, v)$ between two vertices u and v of G is the number of edges on a shortest path from u to v in G. Let v be a vertex of G and $S \subseteq V(G)$, the distance between v and S is $d_G(S, v) = \min\{d_G(u, v) : u \in S\}$.

As usual, we use P_n, S_n, and C_n to denote the path, star, and cycle, respectively, of order n. For $n \geq 3$, we denote by US_n the graph obtained by inserting an edge connecting two leaves of S_n, and UP_n the graph obtained by identifying one end of P_{n-2} with one of the vertices of C_3 (see Figure 1). For all other notations we refer to [6].

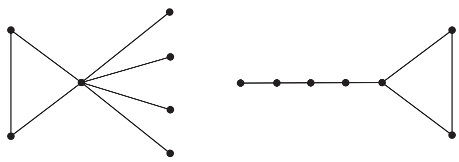

Figure 1: The unicyclic graphs US_n and UP_n for $n = 7$.

For a graph G, an independent set in G is a set of vertices with no two adjacent vertices. Let $\mathrm{i}(G, k)$ be the number of independent sets of cardinality k in G. It is both consistent and convenient to set $\mathrm{i}(G, 0) = 1$. The independence polynomial of G is defined by

$$\mathrm{I}(G, x) = \sum_{k \geq 0} \mathrm{i}(G, k) x^k.$$

One can easily see that $\mathrm{I}(G, 1) = \sum_{k\geq 0} \mathrm{i}(G, k)$ is the total number of independent sets of G, and that $\mathrm{I}'(G, 1) = \sum_{k\geq 0} k\mathrm{i}(G, k)$ is the total size of all independent sets of G. For brevity we use the notations $i(G) = \mathrm{I}(G, 1)$ and $t(G) = \mathrm{I}'(G, 1)$. Then the average size of independent sets in G is given by the logarithmic derivative

$$avi(G) = \frac{\mathrm{I}'(G, 1)}{\mathrm{I}(G, 1)} = \frac{t(G)}{i(G)}.$$

The following results are very useful for calculating the independence polynomial.

Proposition 2.1 (see [16]). *Let v be a vertex of G and $G_1, G_2, \ldots, G_k$ be the disjoint components of G. Then*

(i) $\mathrm{I}(G, x) = \mathrm{I}(G - v, x) + x\mathrm{I}(G - N[v], x)$;

(ii) $\mathrm{I}(G, x) = \prod_{i=1}^{k} \mathrm{I}(G_i, x)$.

A recursive formula for $i(G)$ and $t(G)$ is obtained as a immediate consequence.

Proposition 2.2 (see [1]). *For any graph G and $v \in V(G)$, we have*

(i) $i(G) = i(G - v) + i(G - N[v])$;

(ii) $t(G) = t(G - v) + i(G - N[v]) + t(G - N[v])$.

For the equality in Proposition 2.1 (ii), by taking the logarithm and the differentiation on both sides, letting $x = 1$, we get the next result.

Proposition 2.3. *Let $G_1, G_2, \ldots, G_k$ be the disjoint components of G. Then*

$$avi(G) = \sum_{i=1}^{k} avi(G_i).$$

The following result established in [1], will be fundamental for proving our main results.

Lemma 2.1 (see [1]). *For every tree T of order n, we have $avi(S_n) \geq avi(T) \geq avi(P_n)$. Moreover,*

(i) $avi(T) \geq avi(P_n) \geq \frac{5-\sqrt{5}}{10}n + \frac{1}{\sqrt{5}} - \frac{1}{3}$ for $n \geq 1$;

(ii) $avi(T) \geq avi(P_n) \geq \frac{5-\sqrt{5}}{10}n + \frac{2}{\sqrt{5}} - \frac{3}{4}$ for $n \geq 3$;

(iii) if $T \ncong P_n$, then $avi(T) \geq \frac{5-\sqrt{5}}{10}n + \frac{79\sqrt{5}-165}{70} > avi(P_n)$.

3. Main results

For notational simplicity, in this section we always set

$$a = \frac{5-\sqrt{5}}{10} \approx 0.27639320, \quad b = \frac{2}{\sqrt{5}} - \frac{3}{4} \approx 0.14442719,$$

$$c = \frac{1}{\sqrt{5}} - \frac{1}{3} \approx 0.11388026, \quad d = \frac{79\sqrt{5}-165}{70} \approx 0.16641957, \quad \phi = \frac{\sqrt{5}+1}{2}.$$

It is known that US_n maximizes the number of independent sets and C_n and UP_n minimizes the number of independent sets among all unicyclic graphs of order n (see [29]). We show that these results parallel the extremal graph for the average size of independent sets. We begin with some lemmas that are useful for proving the subsequent results.

Lemma 3.1.

$$avi\,(S_n) = \frac{n-1}{2} + \frac{3-n}{2^n+2}. \tag{1}$$

$$avi\,(US_n) = \frac{3n-5}{6} + \frac{11-3n}{18 \cdot 2^{n-3}+6}. \tag{2}$$

$$avi\,(C_n) = avi\,(UP_n) = an + \frac{2n}{2\sqrt{5} - (3\sqrt{5}+5)(-\phi^2)^{n-1}}. \tag{3}$$

Moreover,

$$avi\,(C_n) \leq an + \frac{9\sqrt{5}-20}{15} \approx an + 0.0083$$

for $n \geq 5$ with equality only for $n = 6$. For $n \geq 7$,

$$avi\,(C_n) \leq an + \frac{8836\sqrt{5}-19740}{11045} \approx an + 0.0016$$

with equality only for $n = 8$.

Proof. Applying Proposition 2.1, we have

$$\mathrm{I}\,(S_n, x) = (1+x)^{n-1} + x, \quad \mathrm{I}\,(US_n, x) = (1+2x)(1+x)^{n-3} + x,$$

which yields

$$i\,(S_n) = 2^{n-1} + 1 \quad \text{and} \quad i\,(US_n) = 3 \cdot 2^{n-3} + 1,$$
$$t\,(S_n) = (n-1)2^{n-2} + 1 \quad \text{and} \quad t\,(US_n) = 3(n-3) \cdot 2^{n-4} + 2^{n-2} + 1.$$

Therefore (1) and (2) follow by direct computation.

Let u be one of the vertices of degree 2 that lies on the unique cycle of UP_n and v be any vertex of C_n. By applying Proposition 2.1, we have

$$\mathrm{I}(UP_n, x) = \mathrm{I}(UP_n - u, x) + x\mathrm{I}(UP_n - N[u], x) = \mathrm{I}(P_{n-1}, x) + x\mathrm{I}(P_{n-3}, x),$$

$$\mathrm{I}(C_n, x) = \mathrm{I}(C_n - v, x) + x\mathrm{I}(C_n - N[v], x) = \mathrm{I}(P_{n-1}, x) + x\mathrm{I}(P_{n-3}, x).$$

Clearly, $\mathrm{I}(UP_n, x) = \mathrm{I}(C_n, x)$, which implies that $avi\,(UP_n) = avi\,(C_n)$.

Note that $i(P_n)$ is the Fibonacci number $F_{n+2} = \frac{1}{\sqrt{5}}\left(\phi^{n+2} - (-\phi)^{-n-2}\right)$ (see [23]). Moreover, by Proposition 2.2, we have the recursion $t(P_n) = t(P_{n-2}) + i(P_{n-2}) + t(P_{n-1})$ with $t(P_1) = 1$ and $t(P_2) = 2$. By solving this recursive equation we get

$$t(P_n) = \left(\frac{1+\sqrt{5}}{10}n + \frac{2\sqrt{5}}{25}\right)\phi^n + \left(\frac{1-\sqrt{5}}{10}n - \frac{2\sqrt{5}}{25}\right)(-\phi)^{-n}.$$

On the other hand, it follows from Proposition 2.2 that $i(C_n) = i(P_{n-1}) + i(P_{n-3})$, $t(C_n) = t(P_{n-1}) + i(P_{n-3}) + t(P_{n-3})$. Direct calculation give the formula (3) for $avi(C_n) = t(C_n)/i(C_n)$.

Let $c_n = avi(C_n) - an$. We prove that the absolute value of c_n is decreasing. When $n \geq 5$, we get

$$\begin{aligned}
\left|\frac{c_n}{c_{n-1}}\right| &= \left|\frac{2n}{2\sqrt{5} - (3\sqrt{5}+5)(-\phi^2)^{n-1}} \cdot \frac{2\sqrt{5} - (3\sqrt{5}+5)(-\phi^2)^{n-2}}{2(n-1)}\right| \\
&\leq \left(1 + \frac{1}{n-1}\right)\frac{(3\sqrt{5}+5)\phi^{2(n-2)} + 2\sqrt{5}}{(3\sqrt{5}+5)\phi^{2(n-1)} - 2\sqrt{5}} \\
&= \phi^{-2}\left(1 + \frac{1}{n-1}\right)\frac{3\sqrt{5}+5+2\sqrt{5}\phi^{-2(n-2)}}{3\sqrt{5}+5-2\sqrt{5}\phi^{-2(n-1)}} \\
&\leq \phi^{-2} \cdot \frac{5}{4} \cdot \frac{3\sqrt{5}+5+2\sqrt{5}\phi^{-6}}{3\sqrt{5}+5-2\sqrt{5}\phi^{-8}} \approx 0.4916 < 1.
\end{aligned}$$

Therefore, $|c_n|$ is decreasing for n. Meanwhile, observe that the sign of c_n alternates, i.e., $c_n < 0$ if n is odd and $c_n > 0$ if n is even. This implies that $avi(C_n)$ is alternately greater and less than an. It follows that if $n \geq 5$, the maximum of c_n is attained for $n = 6$, and if $n \geq 7$, the maximum of c_n is attained for $n = 8$. In both cases, the values of $avi(C_n)$ are easy to calculate and the two inequalities in Lemma 3.1 follow. □

Lemma 3.2. *For any graph G and $v \in V(G)$, we have*

$$\frac{1}{2} \leq \frac{i(G-v)}{i(G)} < 1.$$

Proof. Using Proposition 2.2, we get $i(G) = i(G-v) + i(G-N[v])$. Note that $i(G-N[v]) \leq i(G-v)$ because $G-N[v]$ is a subgraph of $G-v$. This establishes the left inequality. And the right inequality follows easily from the fact that $G-v$ is a proper subgraph of G. □

Lemma 3.3. *If T is a tree on n vertices, then*

(i) for any vertex v of T, we have $avi(T-v) \geq a(n-1) + c$ for $n > 1$;

(ii) for any two vertices u, v of T, we have $avi(T-u-v) \geq a(n-2) + c$ for $n \geq 3$.

Proof. We only prove (i), whereas the proofs of (ii) is similar and is omitted.

If $T-v$ is connected, the inequality follows from Lemma 2.1 (i). Consider, therefore, the case when $T-v$ be not connected. Let k be the number of components of $T-v$, where $k \geq 2$. Applying Lemma 2.1 (i) to each component of $T-v$, then by Proposition 2.3, $avi(T-v) \geq a(n-1) + kc > a(n-1) + c$ as desired. □

We first characterize the graphs minimizing the average size of independent sets among all unicyclic graphs of order n.

Theorem 3.1. *For any unicyclic graph G of order $n \geq 3$, we have*

$$avi(G) \geq an + \frac{2n}{2\sqrt{5} - (3\sqrt{5}+5)(-\phi^2)^{n-1}}$$

with equality if and only if $G \in \{C_n, UP_n\}$.

Proof. If $G \in \{C_n, UP_n\}$, then by Lemma 3.1,

$$avi(G) = an + \frac{2n}{2\sqrt{5} - (3\sqrt{5}+5)(-\phi^2)^{n-1}}.$$

Assume that $G \notin \{C_n, UP_n\}$ is a unicyclic graph of order $n \geq 3$. It suffices to prove that $avi(G) > avi(C_n)$. We write C to denote the unique cycle in G and consider a vertex v in C with $d_G(v) \geq 3$. Assume that $N_G(v) = \{v_0, v_1, \ldots, v_k\}$ with

$v_0, v_1 \in V(C)$, we use $T_1, T_2, \ldots, T_k$ to denote the components of $G-v$ such that v_0, v_1 is contained in T_1, v_i is contained in T_i for $2 \leq i \leq k$. By Propositions 2.2 and 2.3, we obtain

$$\begin{aligned}
avi(G) &= \frac{t(G-v)+i(G-N[v])+t(G-N[v])}{i(G)} \\
&= \frac{i(G-v)}{i(G)} \cdot \frac{t(G-v)}{i(G-v)} + \frac{i(G-N[v])}{i(G)} \cdot \left(1+\frac{t(G-N[v])}{i(G-N[v])}\right) \\
&= \frac{i(G-v)}{i(G)} avi(G-v) + \frac{i(G)-i(G-v)}{i(G)}(1+avi(G-N[v])) \\
&= \frac{i(G-v)}{i(G)} avi\left(T_1\right) + \left(1-\frac{i(G-v)}{i(G)}\right) avi\left(T_1-v_0-v_1\right) \\
&\quad + \frac{i(G-v)}{i(G)} \sum_{i=2}^{k} avi\left(T_i\right) + \left(1-\frac{i(G-v)}{i(G)}\right)\left(1+\sum_{i=2}^{k} avi\left(T_i-v_i\right)\right).
\end{aligned}$$

Assume that $k \geq 5$ and let $T' = G - T_1$. Repeating the above calculation yields

$$avi\left(T'\right) = \frac{i\left(T'-v\right)}{i\left(T'\right)} \sum_{i=2}^{k} avi\left(T_i\right) + \left(1-\frac{i\left(T'-v\right)}{i\left(T'\right)}\right)\left(1+\sum_{i=2}^{k} avi\left(T_i-v_i\right)\right).$$

For simplicity, we use σ and σ' to denote $\frac{i(G-v)}{i(G)}$ and $\frac{i(T'-v)}{i(T')}$, respectively, and set

$$\begin{aligned}
A &= \sigma avi\left(T_1\right) + (1-\sigma)\, avi\left(T_1-v_0-v_1\right), \\
B &= \sigma \sum_{i=2}^{k} avi\left(T_i\right) + (1-\sigma)\left(1+\sum_{i=2}^{k} avi\left(T_i-v_i\right)\right).
\end{aligned}$$

Note that

$$\begin{aligned}
\sigma = \frac{i(G-v)}{i(G)} &= \frac{\prod_{i=1}^{k} i\left(T_i\right)}{\prod_{i=1}^{k} i\left(T_i\right) + i\left(T_1-v_0-v_1\right)\prod_{i=2}^{k} i\left(T_i-v_i\right)} \\
&= \left(1+\frac{i\left(T_1-v_0-v_1\right)}{i\left(T_1\right)} \prod_{i=2}^{k} \frac{i\left(T_i-v_i\right)}{i\left(T_i\right)}\right)^{-1} \qquad (4)
\end{aligned}$$

and likewise

$$\sigma' = \left(1+\prod_{i=2}^{k} \frac{i\left(T_i-v_i\right)}{i\left(T_i\right)}\right)^{-1}. \qquad (5)$$

Applying Lemma 3.2, we deduce that

$$\frac{1}{4} \leq \frac{i\left(T_1-v_0-v_1\right)}{i\left(T_1\right)} = \frac{i\left(T_1-v_0-v_1\right)}{i\left(T_1-v_0\right)} \frac{i\left(T_1-v_0\right)}{i(T_1)} < 1. \qquad (6)$$

Combining (4), (5) and (6), we get $\sigma > \sigma'$. Now, we write

$$\begin{aligned}
B =& \sigma \sum_{i=2}^{k} avi\left(T_i\right) + (1-\sigma)\left(1+\sum_{i=2}^{k} avi\left(T_i-v_i\right)\right) \\
=& \frac{1-\sigma}{1-\sigma'}\left(\sigma' \sum_{i=2}^{k} avi\left(T_i\right) + (1-\sigma')\left(1+\sum_{i=2}^{k} avi\left(T_i-v_i\right)\right)\right) + \frac{\sigma-\sigma'}{1-\sigma'} \sum_{i=2}^{k} avi\left(T_i\right) \\
=& \frac{1-\sigma}{1-\sigma'} avi\left(T'\right) + \frac{\sigma-\sigma'}{1-\sigma'} \sum_{i=2}^{k} avi\left(T_i\right). \qquad (7)
\end{aligned}$$

Observe that T' is not a path. Using Lemma 2.1 (iii) gives $avi\left(T'\right) > a|T'|+d$. Moreover, by Lemma 2.1 (i), $avi\left(T_i\right) \geqslant a\left|T_i\right|+c$ for all $2 \leq i \leq k$. Hence

$$\begin{aligned}
\sum_{i=2}^{k} avi\left(T_i\right) \geqslant \sum_{i=2}^{k}\left(a\left|T_i\right|+c\right) &= a\left(|T'|-1\right)+c(k-1) \\
&\geqslant a\left|T'\right|+4c-a > a|T'|+d.
\end{aligned}$$

By substituting the above inequality and $avi\,(T') > a|T'| + d$ into (7), we obtain $B > a|T'| + d$. Then we consider the order of T_1.

If $|T_1| \geq 3$, then, by Lemmas 2.1 (ii), 3.3 (ii) and 3.2, we have

$$\begin{aligned} A &= \sigma avi\,(T_1) + (1-\sigma)\,avi\,(T_1 - v_0 - v_1) \\ &\geqslant \sigma\,(a\,|T_1| + b) + (1-\sigma)\,(a\,(|T_1| - 2) + c) \\ &= a|T_1| - 2a + c + \sigma\,(b - c + 2a) \\ &\geq a|T_1| - 2a + c + \frac{1}{2}\,(b - c + 2a) \quad \text{as } b - c + 2a > 0 \\ &= a|T_1| + \frac{48\sqrt{5} - 125}{120}. \end{aligned}$$

Hence $avi(G) = A + B \geq a|T_1| + \frac{48\sqrt{5}-125}{120} + a|T'| + d \approx an + 0.0192$. Note that in this case $n \geq 8$. Applying Lemma 3.1, we get $avi(G) > avi(C_n)$.

If $|T_1| = 2$, then $T_1 \cong P_2$. Thus we have $avi\,(T_1) = \frac{2}{3}$, $avi\,(T_1 - v_0 - v_1) = 0$ and $\frac{i(T_1 - v_0 - v_1)}{i(T_1)} = \frac{1}{3}$. By applying Equality (4) and Lemma 3.2, we obtain

$$\sigma = \left(1 + \frac{1}{3}\prod_{i=2}^{k}\frac{i\,(T_i - v_i)}{i\,(T_i)}\right)^{-1} > \frac{3}{4},$$

and so

$$A = \sigma avi\,(T_1) + (1-\sigma)\,avi\,(T_1 - v_0 - v_1) = \frac{2}{3}\sigma > \frac{1}{2}.$$

Observe that in this case $n \geq 7$. Therefore by Lemma 3.1,

$$avi(G) = A + B > 1/2 + a|T'| + d \approx an + 0.1136 > avi(C_n).$$

This completes the proof of $k \geq 5$.

Next we deal with the cases of $k = 4$ and $k = 3$. In each case, we divide into two subcases $|T_1| \geq 3$ and $|T_1| = 2$. Then for each subcase we continue to distinguish different cases that depend on how many $T_i's$ are a single vertex for $i \geq 2$. This yields eight cases for $k = 4$ and six cases for $k = 3$.

Now, we return to the expression

$$avi(G) = \sigma\sum_{i=1}^{k} avi(T_i) + (1-\sigma)\left(1 + avi(T_1 - v_0 - v_1) + \sum_{i=2}^{k} avi(T_i - v_i)\right). \tag{8}$$

For each of these 14 cases, we employ the following bounds and values, which can be easily obtained by Theorems 2.1 and 3.3 or simple calculations.

$$avi\,(T_i)\begin{cases} = \frac{1}{2} & |T_i| = 1, \\ = \frac{2}{3} & |T_i| = 2, \\ \geq a|T_i| + c & |T_i| > 1, \\ \geq a|T_i| + b & |T_i| \geq 3, \end{cases}$$

$$avi\,(T_i - v_i)\begin{cases} = 0 & |T_i| = 1, \\ \geq a|T_i - 1| + c & \text{otherwise}, \end{cases}$$

$$avi(T_1 - v_0 - v_1)\begin{cases} = 0 & |T_1| = 2, \\ \geq a(|T_1| - 2) + c & \text{otherwise}. \end{cases}$$

Also by Theorem 3.2 and Inequality (6), we have

$$\frac{i\,(T_i - v_i)}{i\,(T_i)}\begin{cases} = \frac{1}{2} & |T_i| = 1, \\ \in \left[\frac{1}{2}, 1\right] & \text{otherwise}, \end{cases} \qquad \frac{i\,(T_1 - v_0 - v_1)}{i\,(T_1)}\begin{cases} = \frac{1}{3} & |T_1| = 2, \\ \in \left[\frac{1}{4}, 1\right] & \text{otherwise}. \end{cases}$$

We plug these above bounds and values into (8), and use the expression (4) to obtain the maximum or minimum possible value of σ. Since the expression (8) is linear in σ, its minimum is either obtained at the maximum or minimum value of

σ. Then we get lower bounds of $avi(G)$ for the above 14 cases, which are easily calculated by computer. We consider, as an example, the case with the worst bound, which is obtained for $|T_1| \geq 3, |T_2| > 1, |T_3| > 1, |T_4| > 1$ when $k = 4$. Then

$$\sum_{i=1}^{4} avi(T_i) \geq a\,|T_1| + b + a(|T_2| + |T_3| + |T_4|) + 3c = a(n-1) + b + 3c,$$

$$\begin{aligned} avi(T_1 - v_0 - v_1) + \sum_{i=2}^{4} avi(T_i - v_i) &\geq a(|T_1| - 2) + c + a\left(\sum_{i=2}^{4} |T_i| - 3\right) + 3c \\ &= a(n-6) + 4c, \end{aligned}$$

$$\sigma = \left(1 + \frac{i\,(T_1 - v_0 - v_1)}{i\,(T_1)} \frac{i\,(T_2 - v_2)\, i\,(T_3 - v_3)\, i\,(T_4 - v_4)}{i\,(T_2)\, i\,(T_3)\, i\,(T_4)}\right)^{-1} \in \left[\frac{1}{2}, \frac{32}{33}\right].$$

Note that $n \geq 7$ in this case, combining with the Lemma 3.1, we have

$$\begin{aligned} avi(G) &= \sigma \sum_{i=1}^{4} avi(T_i) + (1-\sigma)\left(1 + avi(T_1 - v_0 - v_1) + \sum_{i=2}^{4} avi(T_i - v_i)\right) \\ &\geq \sigma(a(n-1) + b + 3c) + (1-\sigma)(1 + a(n-6) + 4c) \\ &= an + 1 - 6a + 4c + \sigma(5a + b - c - 1) \\ &\geq an + 1 - 6a + 4c + \frac{1}{2}(5a + b - c - 1) \ \ \text{as } 5a + b - c - 1 > 0 \\ &\approx an + 0.0034 > avi(C_n). \end{aligned}$$

In order to illustrate a more general process, as another example, we consider the case with the second-worst bound, which is obtained for $|T_1| = 2, |T_2| > 1, |T_3| > 1$ when $k = 3$. Then

$$\sigma = \left(1 + \frac{1}{3} \frac{i\,(T_2 - v_2)\, i\,(T_3 - v_3)}{i(T_2) i(T_3)}\right)^{-1} \in \left[\frac{3}{4}, \frac{12}{13}\right].$$

Note that $n \geq 5$ in this case. By Lemma 3.1, we have

$$\begin{aligned} avi(G) &= \sigma \sum_{i=1}^{3} avi(T_i) + (1-\sigma)\left(1 + avi(T_1 - v_0 - v_1) + \sum_{i=2}^{3} avi(T_i - v_i)\right) \\ &\geq \sigma\left(\frac{2}{3} + a(|T_2| + |T_3|) + 2c\right) + (1-\sigma)\,(1 + a(|T_2| + |T_3| - 2) + 2c) \\ &= \sigma\left(\frac{2}{3} + a(n-3) + 2c\right) + (1-\sigma)\,(1 + a(n-5) + 2c) \\ &= an + 1 - 5a + 2c + \sigma\left(2a - \frac{1}{3}\right) \\ &\geq an + 1 - 5a + 2c + \frac{3}{4}\left(2a - \frac{1}{3}\right) \ \ \text{as } 2a - \frac{1}{3} > 0 \\ &\approx an + 0.0104 > avi(C_n). \end{aligned}$$

The other 12 cases can be treated analogously. Each time we obtain $avi(G) \geq an + \varepsilon$ where $\varepsilon > 0.0104$ is a constant. Note that we have $n \geq 5$ for all these 12 cases. Therefore for each of these we have $avi(G) \geq an + \varepsilon > avi(C_n)$ by Lemma 3.1.

Now, only the case $k = 2$ remains. Consider the following subcases.

If $|T_1| \geq 3$, $|T_2| \geq 3$, then $n \geq 7$. By Lemmas 2.1, 3.1, 3.2, 3.3, we have

$$\begin{aligned} avi(G) &= \sigma\,(avi(T_1) + avi(T_2)) + (1-\sigma)\,(1 + avi(T_1 - v_0 - v_1) + avi(T_2 - v_2)) \\ &\geq \sigma\,(a\,|T_1| + b + a|T_2| + b) + (1-\sigma)\,(1 + a(|T_1| - 2) + c + a(|T_2| - 1) + c) \\ &= an + 1 - 4a + 2c + \sigma\,(3a + 2b - 2c - 1) \\ &\geq an + 1 - 4a + 2c + (3a + 2b - 2c - 1) \ \ \text{as } 3a + 2b - 2c - 1 < 0 \\ &\approx an + 0.0125 > avi(C_n). \end{aligned}$$

If $|T_1| \geq 3$, $|T_2| = 2$, then $avi(T_2 - v_2) = \frac{1}{2}$, $\frac{i(T_2-v_2)}{i(T_2)} = \frac{2}{3}$. Moreover, by the inequality in (6), we have

$$\sigma = \left(1 + \frac{2}{3}\frac{i\,(T_1 - v_0 - v_1)}{i\,(T_1)}\right)^{-1} \in \left[\frac{3}{5}, \frac{6}{7}\right].$$

Note that $n \geq 6$ in this case. It follows from Lemmas 2.1, 3.1 and 3.3 that

$$\begin{aligned}
avi(G) &= \sigma\,(avi(T_1) + avi(T_2)) + (1-\sigma)\,(1 + avi(T_1 - v_0 - v_1) + avi(T_2 - v_2)) \\
&\geq \sigma\left(a\,|T_1| + b + \frac{2}{3}\right) + (1-\sigma)\left(1 + a(|T_1| - 2) + c + \frac{1}{2}\right) \\
&= an + 1 - 5a + c + \frac{1}{2} + \sigma\left(2a + b - c - \frac{5}{6}\right) \\
&\geq an + 1 - 5a + c + \frac{1}{2} + \frac{6}{7}\left(2a + b - c - \frac{5}{6}\right) \quad \text{as } 2a + b - c - \frac{5}{6} < 0 \\
&\approx an + 0.0176 > avi(C_n).
\end{aligned}$$

If $|T_1| \geq 3$, $|T_2| = 1$, then $avi(T_2 - v_2) = 0$. Note that $n \geq 5$ in this case. It follows from Lemmas 2.1, 3.1, 3.2 and 3.3 that

$$\begin{aligned}
avi(G) &= \sigma\,(avi(T_1) + avi(T_2)) + (1-\sigma)\,(1 + avi(T_1 - v_0 - v_1) + avi(T_2 - v_2)) \\
&\geq \sigma\left(a\,|T_1| + b + \frac{1}{2}\right) + (1-\sigma)\,(1 + a(|T_1| - 2) + c) \\
&= an + 1 - 4a + c + \sigma\left(2a + b - c - \frac{1}{2}\right) \\
&\geq an + 1 - 4a + c + \frac{1}{2}\left(2a + b - c - \frac{1}{2}\right) \quad \text{as } 2a + b - c - \frac{1}{2} > 0 \\
&\approx an + 0.0500 > avi(C_n).
\end{aligned}$$

If $|T_1| = 2$, $|T_2| \geq 4$, then $n \geq 7$ in this case. We consider the structure of T_2. Suppose first that T_2 is not a path. Then by Lemma 2.1 (iii), $avi(T_2) \geq a\,|T_2| + d$. It follows from Lemmas 3.1 3.2 and 3.3 that

$$\begin{aligned}
avi(G) &= \sigma\,(avi(T_1) + avi(T_2)) + (1-\sigma)\,(1 + avi(T_1 - v_0 - v_1) + avi(T_2 - v_2)) \\
&\geq \sigma\left(\frac{2}{3} + a\,|T_2| + d\right) + (1-\sigma)\,(1 + a(|T_2| - 1) + c) \\
&= an + 1 - 4a + c + \sigma\left(a + d - c - \frac{1}{3}\right) \\
&\geq an + 1 - 4a + c + \left(a + d - c - \frac{1}{3}\right) \quad \text{as } a + d - c - \frac{1}{3} < 0 \\
&\approx an + 0.0039 > avi(C_n).
\end{aligned}$$

Suppose now that T_2 is a path. Since $G \notin \{C_n, UP_n\}$, $T_2 - v_2$ is not connected. Then by applying Lemma 2.1 (i) to each component of $T_2 - v_2$, we obtain $avi(T_2 - v_2) \geq a(|T_2| - 1) + 2c$. Moreover, by Lemma 3.2,

$$\sigma = \left(1 + \frac{1}{3}\frac{i\,(T_2 - v_2)}{i\,(T_2)}\right)^{-1} \in \left[\frac{3}{4}, \frac{6}{7}\right].$$

It follows from Lemmas 2.1 and 3.1 that

$$\begin{aligned}
avi(G) &= \sigma\,(avi(T_1) + avi(T_2)) + (1-\sigma)\,(1 + avi(T_1 - v_0 - v_1) + avi(T_2 - v_2)) \\
&\geq \sigma\left(\frac{2}{3} + a\,|T_2| + b\right) + (1-\sigma)\,(1 + a(|T_2| - 1) + 2c) \\
&= an + 1 - 4a + 2c + \sigma\left(a + b - 2c - \frac{1}{3}\right) \\
&\geq an + 1 - 4a + 2c + \frac{6}{7}\left(a + b - 2c - \frac{1}{3}\right) \quad \text{as } a + b - 2c - \frac{1}{3} < 0 \\
&\approx an + 0.0020 > avi(C_n).
\end{aligned}$$

If $|T_1| = 2$, $|T_2| \leq 3$, then G must be the graph obtained by identifying one of vertices of C_3 with a leaf of S_4 since $G \notin \{C_n, UP_n\}$. Simple computations show that $avi(G) = \frac{33}{19}$, $avi(C_6) = \frac{5}{3}$. Thus $avi(G) > avi(C_n)$.

We have thus examined all cases and so these complete the proof. □

Now we turn to considering the graphs maximizing the average size of independent sets.

Theorem 3.2. *For any unicyclic graph G of order $n \geq 3$, we have*

$$avi\,(G) \leq \frac{3n-5}{6} + \frac{11-3n}{18 \cdot 2^{n-3} + 6}$$

with equality if and only if $G \cong US_n$.

Proof. If $G \cong US_n$, then

$$avi(G) = \frac{3n-5}{6} + \frac{11-3n}{18 \cdot 2^{n-3} + 6}$$

by Equality (2). If $G \cong C_n$, then by Theorem 3.1, we have

$$avi(G) < avi(US_n) = \frac{3n-5}{6} + \frac{11-3n}{18 \cdot 2^{n-3} + 6}.$$

Assume that $G \notin \{C_n, US_n\}$ is a unicyclic graph. It suffices to prove that $avi(G) < avi(US_n)$. Then we proceed by induction on n (note that $n \geq 5$ since $G \notin \{C_n, US_n\}$). For $n = 5$, we easily see that $G \cong UP_5$, and thus by Theorem 3.1, $avi(G) < avi(US_n)$. We write C for the unique cycle of G and let $v \in V(G)$ that has the maximum distance from C. Clearly, v is a leaf. Assume that $vu \in E(G)$ with $d_G(u) = k$. We first prove the following inequality:

$$avi(G - N[v]) < \frac{3n-8}{6}. \tag{9}$$

Consider the distance between v and C. If $d_G(C, v) = 1$, then the maximality of $d_G(C, v)$ implies that $G - N[v]$ consists of a tree of order $n - k + 1$, say T', and $k - 3$ singletons. Moreover, since $G \notin \{C_n, US_n\}$, we must have $n - k \geq 2$. By Proposition 2.3, Lemma 2.1 and Equality (1), deduce that

$$\begin{aligned}
avi(G - N[v]) &= 1/2(k-3) + avi(T') \\
&\leq 1/2(k-3) + avi(S_{n-k+1}) \\
&= \frac{n-3}{2} + \frac{2-(n-k)}{2^{n-k+1}+2} \\
&\leq \frac{n-3}{2} < \frac{3n-8}{6} \quad as \quad n-k \geq 2.
\end{aligned}$$

Suppose now that $d_G(C, v) \geq 2$. Then the maximality of $d_G(C, v)$ implies that $G - N[v]$ consists of a unicyclic graph of order $n - k$, say G', and $k - 2$ singletons, where $n - k \geq 3$. By Proposition 2.3, induction hypothesis and Equality (2), we have

$$\begin{aligned}
avi(G - N[v]) &= 1/2(k-2) + avi(G') \\
&\leq 1/2(k-2) + avi(US_{n-k}) \\
&= \frac{3n-11}{6} + \frac{11-3(n-k)}{18 \cdot 2^{n-k-3} + 6} \\
&\leq \frac{3n-11}{6} + \frac{1}{12} < \frac{3n-8}{6}.
\end{aligned}$$

Therefore, the inequality (9) is proved.

Observe next that $i(G-v) - i(G-N[v]) = i(G-v) - i(G-v-u) \geq 1$. Since US_n maximizes the number of independent sets among all unicyclic graphs of order n ([29]), we have $i(G-v) \leq i\,(US_{n-1}) = 3 \cdot 2^{n-4} + 1$. It follows that

$$\frac{i(G-N[v])}{i(G-v)} = 1 - \frac{i(G-v) - i(G-N[v])}{i(G-v)} \leq 1 - \frac{1}{i(G-v)} \leq 1 - \frac{1}{3 \cdot 2^{n-4} + 1}. \tag{10}$$

Using Proposition 2.2 and Inequality (9), we get

$$\begin{aligned}
avi(G) = \frac{t(G)}{i(G)} &= \frac{t(G-v) + i(G-N[v]) + t(G-N[v])}{i(G-v) + i(G-N[v])} \\
&= \frac{i(G-v)avi(G-v) + i(G-N[v])(1 + avi(G-N[v]))}{i(G-v) + i(G-N[v])} \\
&< \frac{avi\,(US_{n-1})\, i(G-v)/i(G-N[v]) + (3n-8)/6 + 1}{i(G-v)/i(G-N[v]) + 1}.
\end{aligned}$$

By Equality (2), we can easily check that $avi\,(US_{n-1}) < (3n-8)/6+1$, and thus

$$\frac{avi\,(US_{n-1})\,x + (3n-8)/6+1}{x+1}$$

is decreasing for $x \geq 0$. This together with the inequality in (10) yields

$$avi(G) < \frac{avi\,(US_{n-1})\left(3\cdot 2^{n-4}+1\right)/(3\cdot 2^{n-4}) + (3n-8)/6+1}{(3\cdot 2^{n-4}+1)\,/\,(3\cdot 2^{n-4})+1} = avi\,(US_n)\,,$$

which completes the proof. □

4. Concluding remarks

Considering the family of unicyclic graphs of order $n \geq 3$, we have proved that the minimum average size of independent sets is attained by C_n and UP_n, while the graph US_n resulting from adding one edge between two leaves of the star of order n uniquely realizes the maximum. These results parallel the extremal graphs for the number of independent sets.

This leaves some natural questions. First, we can consider the average size of independent sets by choosing any different graph instead of unicyclic graph, and see if there are similar parallel conclusions. Secondly, it is well-known that there is a negative correlation between Merrifield–Simmons index and Hosoya index, which are the number of independent sets and the number of matchings, respectively. Recently, Andriantiana et al. [2] also considered the extreme problem on the average size of matching of graphs. It was shown that in all the instances they dealt with, including general graphs and trees of given order, graphs that minimize the average size of the independent set also maximize the average size of the matching, and vice versa. So, it is also interesting to consider whether the average size of independent set and the average size of matching also have a negative correlation in other graph classes.

References

[1] E. O. D. Andriantiana, V. Razanajatovo Misanantenaina, S. Wagner, The average size of independent sets of graphs, *Eur. J. Math.* **6** (2020) 561–576.

[2] E. O. D. Andriantiana, V. Razanajatovo Misanantenaina, S. Wagner, The average size of matchings in graphs, *Graphs Combin.* **36** (2020) 539–560.

[3] E. O. D. Andriantiana, S. Wagner, Unicyclic graphs with large energy, *Linear Algebra Appl.* **435** (2011) 1399–1414.

[4] E. O. D. Andriantiana, H. Wang, Subtrees and independent subsets in unicyclic graphs and unicyclic graphs with fixed segment sequence, *MATCH Commun. Math. Comput. Chem.* **84** (2020) 537–566.

[5] I. Beaton, J. I. Brown, The average order of dominating sets of a graph, *Discrete Math.* **344** (2021) #112595.

[6] J. A. Bondy, U. S. R. Murty, *Graph Theory with Applications*, Macmillan Press, New York, 1976.

[7] G. R. Brightwell, P. Winkler, Hard constraints and the Bethe lattice: adventures at the interface of combinatorics and statistical physics, *Proceedings of the International Congress of Mathematicians*, Vol. III, Higher Ed. Press, Beijing, 2002, 605–624.

[8] S. Cambie, S. Wagner, H. Wang, On the maximum mean subtree order of trees, *European J. Combin.* **97** (2021) #103388.

[9] B. Cameron, L. Mol, On the mean subtree order of graphs under edge addition, *J. Graph Theory* **96** (2021) 403–413.

[10] A. J. Chin, G. Gordon, K. J. MacPhee, C. Vincent, Subtrees of graphs, *J. Graph Theory* **89** (2018) 413–438.

[11] E. Davies, M. Jenssen, W. Perkins, B. Roberts, Independent sets, matchings, and occupancy fractions, *J. Lond. Math. Soc.* **96** (2017) 47–66.

[12] E. Davies, M. Jenssen, W. Perkins, B. Roberts, On the average size of independent sets in triangle-free graphs, *Proc. Amer. Math. Soc.* **146** (2018) 111–124.

[13] H. Deng, S. Chen, The extremal unicyclic graphs with respect to Hosoya index and Merrifield–Simmons index. *MATCH Commun. Math. Comput. Chem.* **59** (2008) 171–190.

[14] A. Erey, On the average order of a dominating set of a forest, *Discrete Math.* **346** (2023) #113127.

[15] J. Haslegrave, The path minimises the average size of a connected induced subgraph, *Discrete Math.* **345** (2022) #112799.

[16] C. Hoede, X. Li, Clique polynomials and independent set polynomials of graphs, *Discrete Math.* **125** (1994) 219–228.

[17] Y. Hou. Unicyclic graphs with minimal energy, *J. Math. Chem.* **29** (2001) 163–168.

[18] Y. Hou, I. Gutman, C. W. Woo, Unicyclic graphs with maximal energy, *Linear Algebra Appl.* **356** (2002) 27–36.

[19] H. Hua, X. Hua, H. Wang, Further results on the Merrifield–Simmons index, *Discrete Appl. Math.* **283** (2020) 231–241.

[20] H. Hua, M. Wang, On the Merrifield–Simmons index and some Wiener-type indices, *MATCH Commun. Math. Comput. Chem.* **85** (2021) 131–146.

[21] R. Jamison, On the average number of nodes in a subtree of a tree, *J. Combin. Theory Ser. B* **35** (1983) 207–223.

[22] R. Jamison, Monotonicity of the mean order of subtrees, *J. Combin. Theory Ser. B* **37** (1984) 70–78.

[23] P. Kirschenhofer, H. Prodinger, R. F. Tichy, Fibonacci numbers of graphs. II, *Fibonacci Quart.* **21** (1983) 219–229.

[24] N. Linial, M. Saks, Every poset has a central element, *J. Combin. Theory Ser. A* **40** (1985) 195–210.

[25] Z. Luo, K. Xu, S. Wagner, H. Wang, On the mean subtree order of trees under edge contraction, *J. Graph Theory* (2022) DOI: 10.1002/jgt.22885, In press.

[26] R. E. Merrifield, H. E. Simmons, *Topological Methods in Chemistry*, Wiley, New York, 1989.

[27] L. Mol, O. R. Oellermann, Maximizing the mean subtree order, *J. Graph Theory* **91** (2019) 326–352.

[28] J. Ou. On extremal unicyclic molecular graphs with maximal Hosoya index, *Discrete Appl. Math.* **157** (2009) 391–397.

[29] A. S. Pedersen, P. D. Vestergaard, The number of independent sets in unicyclic graphs, *Discrete Appl. Math.* **152** (2005) 246–256.

[30] A. Vince, A lower bound on the average size of a connected vertex set of a graph, *J. Combin. Theory Ser. B* **152** (2022) 153–170.

[31] A. Vince, H. Wang, The average order of a subtree of a tree, *J. Combin. Theory Ser. B* **100** (2010) 161–170.

[32] S. Wagner, Upper and lower bounds for Merrifield–Simmons index and Hosoya index, In: I. Gutman, B. Furtula, K. C. Das, E. Milovanović, I. Milovanović, (Eds.), *Bounds in Chemical Graph Theory – Basics*, Univ. Kragujevac, Kragujevac, 2017, 155–187.

[33] S. Wagner I. Gutman, Maxima and minima of the Hosoya index and the Merrifield–Simmons index, *Acta Appl. Math.* **112** (2010) 323–346.

[34] S. Wagner, H. Wang, On the local and global means of subtree orders, *J. Graph Theory* **81** (2016) 154–166.

[35] H. Wang, H. Hua, Unicycle graphs with extremal Merrifield–Simmons index, *J. Math. Chem.* **43** (2008) 202–209.

[36] B. Wang, C. Ye, H. Zhao, Extremal unicyclic graphs with respect to Merrifield–Simmons index, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 203–216.

[37] K. Xu, K. C. Das, I. Gutman, M. Wang, Comparison between Merrifield–Simmons index and Wiener index of graphs, *Acta Math. Sin. (Engl. Ser.)* (2022) DOI: 10.1007/s10114-022-0540-9, In press.

[38] K. Xu, M. Wang, J. Tian, Relations between Merrifield–Simmons and Wiener indices, *MATCH Commun. Math. Comput. Chem.* **85** (2021) 147–160.

[39] K. Xu, B. Xu, Some extremal unicyclic graphs with respect to Hosoya index and Merrifield–Simmons index, *MATCH Commun. Math. Comput. Chem.* **62** (2009) 629–648.

[40] Y. Ye, X. F. Pan, H. Liu, Ordering unicyclic graphs with respect to Hosoya indices and Merrifield–Simmons indices, *MATCH Commun. Math. Comput. Chem.* **59** (2008) 191–202.