The Average Size of Matchings in Graphs

Eric O. D. Andriantiana[1] · Valisoa Razanajatovo Misanantenaina[2] · Stephan Wagner[2]

Abstract

In this paper, we consider the average size of independent edge sets, also called matchings, in a graph. We characterize the extremal graphs for the average size of matchings in general graphs and trees. In addition, we obtain inequalities between the average size of matchings and the number of matchings as well as the matching energy, which is defined as the sum of the absolute values of the zeros of the matching polynomial.

1 Introduction

An independent vertex set in a graph is a set of vertices such that no two vertices are adjacent. An independent edge set, also called a matching, is a set of edges such that no two edges are adjacent. It is not surprising that these two concepts are closely

✉ Valisoa Razanajatovo Misanantenaina
   valisoa@sun.ac.za

   Eric O. D. Andriantiana
   E.Andriantiana@ru.ac.za

   Stephan Wagner
   swagner@sun.ac.za

[1] Department of Mathematics (Pure and Applied), Rhodes University, PO Box 94, Grahamstown 6140, South Africa

[2] Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

related, an elementary example being the fact that a matching in a graph is an independent set in the corresponding line graph. Two popular graph invariants associated with these concepts are the Merrifield–Simmons index and the Hosoya index, which are the total number of independent sets and the total number of matchings respectively. Extremal problems, where one is looking for the maximum or minimum of an invariant in a specified class of graphs, have been studied quite thoroughly for both the Merrifield–Simmons index and the Hosoya index. It is straightforward that among all n-vertex graphs, the complete graph has the maximum Hosoya index and the minimum Merrifield–Simmons index, while on the other hand the empty graph has the minimum Hosoya index and the maximum Merrifield–Simmons index. Among n-vertex trees, the path and the star are extremal, and there are numerous other examples of graph classes where the graphs that minimize the Merrifield–Simmons index also maximize the Hosoya index, and vice versa [13].

In a recent paper [1], we were interested in extremal questions for the average size of independent sets of graphs rather than their number. This was partly inspired by the work of Jamison [6, 7] and later authors [5, 10, 12, 14] on the average size of subtrees of trees. In the present paper, which complements our paper [1], we are concerned with the study of the average size of matchings in a graph. In view of the aforementioned relation between independent sets and matchings, we expect to get similar results as for the average size of independent sets. Indeed, we find that the graphs that minimize the average size of independent sets are also those that maximize the average size of matchings and vice versa in all instances that we treat. Specifically, it holds true for arbitrary graphs and trees of a prescribed size.

Finally, we also prove inequalities between the average size of matchings and the number of matchings as well as the matching energy of a graph, an invariant introduced in [4].

2 Preliminaries

Let G be a graph. A subset A of $E(G)$ is called a matching of G if no pair of edges of A share a common vertex. Let $\mathrm{m}(G, k)$ be the number of matchings of cardinality k (also called k-matchings) in G. We use the following notation for the total number of matchings in G, the sum of the sizes of all matchings in G and the average size of matchings in G:

$$\mathrm{M}(G) = \sum_{k \geq 0} \mathrm{m}(G, k),$$

$$\mathrm{S}(G) = \sum_{k \geq 0} k \, \mathrm{m}(G, k),$$

$$\mathrm{avm}(G) = \frac{\mathrm{S}(G)}{\mathrm{M}(G)}.$$

The greatest cardinality of a matching in G is called the matching number of G and denoted by $\mu(G)$.

As examples, let us consider the n-vertex edgeless graph E_n and the star S_n. We have

$$M(E_n) = 1, \quad M(S_n) = n, \quad S(E_n) = 0, \quad S(S_n) = n - 1$$

and hence

$$\mathrm{avm}(E_n) = 0, \quad \mathrm{avm}(S_n) = \frac{n-1}{n}.$$

The following standard and well-known proposition gives us a recursion for the total number and size of matchings.

Proposition 1 *If $e = uv$ is an edge of G, then*

$$M(G) = M(G - e) + M(G - v - u) \tag{1}$$

and

$$S(G) = S(G - e) + S(G - v - u) + M(G - v - u). \tag{2}$$

Similarly, if v is a vertex of G, then

$$M(G) = M(G - v) + \sum_{u:uv \in E(G)} M(G - v - u) \tag{3}$$

and

$$S(G) = S(G - v) + \sum_{u:uv \in E(G)} (S(G - v - u) + M(G - v - u)). \tag{4}$$

Proof A matching in G either contains the edge e or not. The number of matchings containing e is $M(G - v - u)$, and the number of those not containing e is $M(G - e)$. Hence, the first equation holds. The argument for the second equation is similar, with the last term taking the edge e itself into account.

Using similar reasoning, the last two equations are obtained by distinguishing between matchings that do not contain an edge with v as an endpoint and those that do contain such an edge. $\qquad\square$

Remark 1 In particular, if v is a leaf of a tree and w its unique neighbor, we obtain the relations

$$M(G) = M(G - v) + M(G - v - w)$$

and

$$S(G) = S(G - v) + S(G - v - w) + M(G - v - w).$$

Moreover, we have the following basic result on disjoint unions:

Proposition 2 *Let $G_1, G_2, \ldots, G_k$ be the connected components of a graph G. Then we have*

$$\mathrm{M}(G) = \prod_{j=1}^{k} \mathrm{M}(G_j)$$

and

$$\mathrm{S}(G) = \sum_{i=1}^{k} \mathrm{S}(G_i) \prod_{\substack{j=1 \\ j \neq i}}^{k} \mathrm{M}(G_j) = \mathrm{M}(G) \sum_{i=1}^{k} \frac{\mathrm{S}(G_i)}{\mathrm{M}(G_i)},$$

thus

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

Proof This follows easily from the fact that every matching of G decomposes uniquely into matchings of its connected components. $\qquad\square$

3 General Graphs

Unlike the total number of matchings M, the average size of matchings avm is not always a monotone function under addition of edges to the graph. For example, consider the tree in Fig. 1. We have

$$\mathrm{avm}(T - e_1) = \frac{7}{6} > \frac{8}{7} = \mathrm{avm}(T), \text{ but } \mathrm{avm}(T - e_2) = \frac{3}{4} < \frac{8}{7} = \mathrm{avm}(T).$$
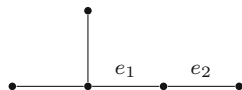
However, we can make use of the following result obtained in [1]:

Theorem 1 *Let X be a nonempty finite set, and $\mathcal{P}(X)$ its powerset. For a set $\mathcal{A} \subseteq \mathcal{P}(X)$, we define*

$$av(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{A \in \mathcal{A}} |A|.$$

Let $\mathcal{B} \subseteq \mathcal{P}(X)$, such that the cardinalities of the elements of $\mathcal{B}$ are not all the same and for every $x \in X$ there exists $B \in \mathcal{B}$ with $x \in B$. Then there exists $x_0 \in X$ such that

Fig. 1 A tree T and two of its edges

$$av(\mathcal{B}) > av(\mathcal{B} \cap \mathcal{P}(X - \{x_0\})).$$

Applying Theorem 1, with $\mathcal{B}$ being the set of matchings of G, we immediately obtain the following theorem.

Theorem 2 *If G is a nonempty graph, then there exists an edge e in $E(G)$ such that*

$$\mathrm{avm}(G - e) < \mathrm{avm}(G).$$

As an immediate consequence, we have the following corollary (which of course is also rather trivial without Theorem 2).

Corollary 1 *For every n-vertex graph G that is not the edgeless graph E_n,* $0 = \mathrm{avm}(E_n) < \mathrm{avm}(G).$

One might wonder whether there is an analogous statement for adding edges. If it was possible to add an edge to every non-complete graph in such a way that the average matching size increases, it would follow immediately that complete graphs maximize the invariant avm. While the latter is true (as will be shown in the following), the analogue of Theorem 1 fails, as the example of a four-vertex cycle shows: when an edge e is added to the cycle C_4, we have

$$\mathrm{avm}(C_4) = \frac{8}{7} > \frac{9}{8} = \mathrm{avm}(C_4 + e).$$

Thus we need another approach to show that the complete graph is still extremal. For this purpose, we first introduce some notation.

In analogy to $\mathrm{M}(G)$, $\mathrm{S}(G)$ and $\mathrm{avm}(G)$, we define the following partial quantities for every nonnegative integer k:

$$\mathrm{M}_k(G) = \sum_{i=0}^{k} \mathrm{m}(G, i),$$

$$\mathrm{S}_k(G) = \sum_{i=0}^{k} i\,\mathrm{m}(G, i),$$

$$\mathrm{avm}_k(G) = \frac{\mathrm{S}_k(G)}{\mathrm{M}_k(G)}.$$

We have the following lemmas.

Lemma 1 *For every nonnegative integer k and every graph G, we have*

$$\mathrm{avm}_{k+1}(G) \geq \mathrm{avm}_k(G).$$

If $k \geq \mu(G)$, then

$$\mathrm{avm}_{k+1}(G) = \mathrm{avm}_k(G) = \mathrm{avm}(G).$$

Proof This is straightforward from the definition of avm_k. $\square$

Lemma 2 *For every n-vertex graph G and every nonnegative integer k such that $k < \mu(G)$, we have*

$$\frac{\mathrm{m}(K_n, k)}{\mathrm{m}(K_n, k+1)} \le \frac{\mathrm{m}(G, k)}{\mathrm{m}(G, k+1)}.$$

Proof Let N be any k-matching of the complete graph K_n. When the $2k$ vertices that are covered by N are removed, a complete graph on $n - 2k$ vertices remains. Thus there are $\mathrm{m}(K_{n-2k}, 1) = \dbinom{n - 2k}{2}$ possible ways to extend N to a $(k+1)$-matching. Conversely, every $(k+1)$-matching can be obtained as an extension of $k + 1$ different k-matchings. It follows that

$$\mathrm{m}(K_n, k+1) = \mathrm{m}(K_n, k) \cdot \frac{\mathrm{m}(K_{n-2k}, 1)}{k+1}. \tag{5}$$

Likewise, if N is a k-matching of G and $v(N)$ the set of vertices covered by N in G, then there are $\mathrm{m}(G - v(N))$ ways to extend N to a $(k+1)$-matching of G. So by the same double-counting argument, we have

$$\mathrm{m}(G, k+1) = \frac{1}{k+1} \sum_{N:\, k-\mathrm{matching\, of\, }G} \mathrm{m}(G - v(N), 1).$$

Clearly, $\mathrm{m}(G - v(N), 1) \le \mathrm{m}(K_{n-2k}, 1)$ for all k-matchings N (with equality if and only if $G - v(N)$ is complete), thus

$$\mathrm{m}(G, k+1) \le \frac{1}{k+1} \cdot \mathrm{m}(G, k) \cdot \mathrm{m}(K_{n-2k}, 1), \tag{6}$$

and the desired inequality results from (5) and (6). $\square$

Remark 2 Equality in Lemma 2 may hold for some (but not all) k even if G is not complete: for example, for the 4-cycle C_4, we have

$$\frac{\mathrm{m}(K_4, 2)}{\mathrm{m}(K_4, 1)} = \frac{3}{6} = \frac{2}{4} = \frac{\mathrm{m}(C_4, 2)}{\mathrm{m}(C_4, 1)}.$$

Lemma 2 can easily be extended to the following lemma by induction:

Lemma 3 *For every n-vertex graph G and for every pair of integers k, l with $\mu(G) \ge k \ge l \ge 0$, we have*

$$\frac{\mathrm{m}(K_n, l)}{\mathrm{m}(K_n, k)} \le \frac{\mathrm{m}(G, l)}{\mathrm{m}(G, k)}$$

and thus

$$\frac{M_l(K_n)}{m(K_n,k)} = \frac{\sum_{i=0}^{l} m(K_n,i)}{m(K_n,k)} \leq \frac{\sum_{i=0}^{l} m(G,i)}{m(G,k)} = \frac{M_l(G)}{m(G,k)}.$$

Theorem 3 *For every n-vertex graph G and every integer k with $\mu(G) \geq k > 0$, we have*

$$\mathrm{avm}_k(K_n) \geq \mathrm{avm}_k(G),$$

with equality if and only if G is a complete graph.

Proof We only need to consider the case that G is not complete.

We proceed by induction on k. For $k = 1$ we have

$$\mathrm{avm}_1(K_n) = \frac{|E(K_n)|}{|E(K_n)| + 1} > \frac{|E(G)|}{|E(G)| + 1} = \mathrm{avm}_1(G).$$

The inequality holds because $\frac{x}{x+1}$ is an increasing function of x on the interval $[0, \infty)$.

Assume that $\mathrm{avm}_k(K_n) > \mathrm{avm}_k(G)$ for some positive integer k, where $k < \mu(G)$. Then we have $m(k+1, G) \neq 0$ and

$$
\begin{aligned}
\mathrm{avm}_{k+1}(K_n) &= \frac{(k+1)m(K_n, k+1) + \sum_{i=0}^{k} i\, m(K_n, i)}{m(K_n, k+1) + \sum_{i=0}^{k} m(K_n, i)} \\
&= \frac{(k+1)m(K_n, k+1) + S_k(K_n)}{m(K_n, k+1) + M_k(K_n)} \\
&= \frac{(k+1)m(K_n, k+1) + \mathrm{avm}_k(K_n)M_k(K_n)}{m(K_n, k+1) + M_k(K_n)} \qquad (7) \\
&= \frac{(k+1) + \mathrm{avm}_k(K_n)\frac{M_k(K_n)}{m(K_n, k+1)}}{1 + \frac{M_k(K_n)}{m(K_n, k+1)}}.
\end{aligned}
$$

Since $k + 1 > \mathrm{avm}_k(K_n)$, we know that $\frac{(k+1) + \mathrm{avm}_k(K_n)x}{1+x}$ is decreasing as a function of x on the interval $[0, \infty)$, so Lemma 3 and (7) imply that

$$\mathrm{avm}_{k+1}(K_n) \geq \frac{(k+1) + \mathrm{avm}_k(K_n)\frac{M_k(G)}{m(G, k+1)}}{1 + \frac{M_k(G)}{m(G, k+1)}}. \qquad (8)$$

Finally, using the induction hypothesis $\mathrm{avm}_k(K_n) > \mathrm{avm}_k(G)$, we obtain

$$\mathrm{avm}_{k+1}(K_n) > \frac{(k+1) + \mathrm{avm}_k(G)\frac{M_k(G)}{m(G, k+1)}}{1 + \frac{M_k(G)}{m(G, k+1)}} = \mathrm{avm}_{k+1}(G). \qquad (9)$$

□

Corollary 2 *For every n-vertex graph G we have* $\mathrm{avm}(K_n) \geq \mathrm{avm}(G)$, *with equality only if G is a complete graph.*

Proof Theorem 3 and Lemma 1 give us

$$\mathrm{avm}(K_n) = \mathrm{avm}_{\lfloor n/2 \rfloor}(K_n) \geq \mathrm{avm}_{\mu(G)}(K_n) \geq \mathrm{avm}_{\mu(G)}(G) = \mathrm{avm}(G).$$

$\square$

Remark 3 While there is no simple explicit formula for $\mathrm{avm}(K_n)$, it can be expressed in terms of the number of matchings in complete graphs. Every edge of the complete graph K_n is contained in $\mathrm{M}(K_{n-2})$ matchings, thus we have $\mathrm{S}(K_n) = \binom{n}{2} \mathrm{M}(K_{n-2})$ and consequently

$$\mathrm{avm}(K_n) = \frac{\mathrm{S}(K_n)}{\mathrm{M}(K_n)} = \binom{n}{2} \frac{\mathrm{M}(K_{n-2})}{\mathrm{M}(K_n)}.$$

A relatively simple asymptotic formula can be provided as well. There is a straightforward bijection between matchings of K_n and involutions of an n-element set (a permutation is called an involution if it is equal to its own inverse, or equivalently if all cycles are of length 1 or 2). Thus the number of matchings of K_n is the same as the number of involutions of an n-element set, for which there is a well-known asymptotic formula (see [3, Proposition VIII.2]):

$$\mathrm{M}(K_n) \sim \frac{1}{\sqrt{2}} n^{n/2} e^{-n/2+\sqrt{n}-1/4}.$$

It follows that

$$\mathrm{avm}(K_n) \sim \frac{n}{2}$$

as $n \to \infty$.

4 Trees

In this section, we will be concerned with trees. Our main goal is to determine the maximum and minimum of $\mathrm{avm}(T)$ when T is a tree with n vertices. Let us first consider the problem of minimizing the average size of matchings. As it turns out, the minimum for trees is also the minimum for connected graphs in general.

Theorem 4 *For every connected n-vertex graph,* $\mathrm{avm}(S_n) \leq \mathrm{avm}(G)$, *with equality only if G is a star.*

Proof We have shown earlier that $\mathrm{avm}(S_n) = \frac{n-1}{n} < 1$. However any other connected graph G (except for the complete graph K_3, for which

$\mathrm{avm}(K_3) = \frac{3}{4} > \frac{2}{3})$ on n vertices satisfies $\mathrm{avm}(G) \geq 1$, since it possesses matchings of size greater than 1, which make up for the empty set. $\qquad\square$

The maximization problem requires more effort. Note that the line graph of the n-vertex path P_n is the $(n-1)$-vertex path P_{n-1}. This implies that the matchings of P_n can be identified with the independent sets of P_{n-1}. Thus, the average size of matchings of P_n is the same as the average size of the independent sets of P_{n-1}. A formula for this average size was determined in [1], where it was also shown that this average is in fact the minimum among trees of the same size.

Lemma 4 *The average size of matchings of the n-vertex path P_n is*

$$\mathrm{avm}(P_n) = \frac{5-\sqrt{5}}{10} n + \frac{1-\sqrt{5}}{10} - \frac{n+1}{\sqrt{5}((-\phi^2)^{n+1} - 1)}, \qquad (10)$$

where $\phi = \frac{\sqrt{5}+1}{2}$ is the golden ratio. In particular,

(a) $\lim_{n\to\infty} \mathrm{avm}(P_n) - \frac{5-\sqrt{5}}{10}n = \frac{1-\sqrt{5}}{10}$,

(b) $\mathrm{avm}(P_n) \leq \dfrac{5-\sqrt{5}}{10} n + \dfrac{1}{\sqrt{5}} - \dfrac{1}{2}$, *with equality only for $n = 2$. For all positive*

  integers $n \neq 2$, we even have $\mathrm{avm}(P_n) \leq \dfrac{5-\sqrt{5}}{10} n + \dfrac{2}{\sqrt{5}} - 1$.

Proof The formula for $\mathrm{avm}(P_n)$ is taken from [1] (using the aforementioned correspondence between matchings of P_n and independent sets of P_{n-1}). The limit in (a) is a straightforward consequence. For (b), one only needs to note that the sign of the final term in (10) alternates, and that its absolute value is decreasing in n (see also [1]). It follows that its maximum is attained for $n = 2$, and the second largest value for $n = 4$. $\qquad\square$

For ease of notation, we set $a = \frac{5-\sqrt{5}}{10} \approx 0.27639320$ and $c_n = \mathrm{avm}(P_n) - an$. As mentioned above, the final in (10) has alternating sign, and its absolute value is decreasing. Thus we have

$$c_1 < c_3 < c_5 < \cdots < c_6 < c_4 < c_2. \qquad (11)$$

Table 1 gives values of c_n for small n.

Before we prove the main result of this section, we require one more lemma:

Lemma 5 *For every graph G and every vertex v of G, we have*

$$\frac{1}{1+d(v)} \leq \frac{\mathrm{M}(G-v)}{\mathrm{M}(G)} \leq 1,$$

where $d(v)$ denotes the degree of v.

Proof Note first that $\mathrm{M}(G) = \mathrm{M}(G-v) + \sum_{u:uv\in E(G)} \mathrm{M}(G-v-u)$. Since $G - v - u$ is a subgraph of $G - v$, we have $\mathrm{M}(G-v-u) \leq \mathrm{M}(G-v)$, hence $\mathrm{M}(G) \leq (1 + d(v))\mathrm{M}(G-v)$, which proves the first inequality. The second

Table 1 Values of $c_1, c_2, \ldots, c_5$

n	1	2	3
c_n	$\frac{\sqrt{5}}{10} - \frac{1}{2} \approx -0.2764$	$\frac{1}{\sqrt{5}} - \frac{1}{2} \approx -0.0528$	$\frac{3}{2\sqrt{5}} - \frac{5}{6} \approx -0.1625$

n	4	5	6
c_n	$\frac{2}{\sqrt{5}} - 1 \approx -0.1056$	$\frac{\sqrt{5}}{2} - \frac{5}{4} \approx -0.1320$	$\frac{3}{\sqrt{5}} - \frac{19}{13} \approx -0.1199$

inequality simply follows from the fact that $G - v$ is a subgraph of G, so matchings of $G - v$ are also matchings of G.   □

Theorem 5 *For every tree T of order n that is not a path, we have the inequality* $\mathrm{avm}(T) \leq an + b$, *where* $b = (7\sqrt{5} - 17)/10 \approx -0.13475241$. *Consequently, the path maximizes the value of* $\mathrm{avm}(T)$ *among all trees of order n.*

Proof We prove the inequality by induction on n. For $n \leq 3$, there is nothing to prove since the only trees with three or fewer vertices are paths. Thus assume now that $n \geq 4$, and consider a vertex v of the tree T whose degree is at least 3 (which must exist if T is not a path). Denote the neighbors of v by $v_1, v_2, \ldots, v_k$ and the components of $T - v$ by $T_1, T_2, \ldots, T_k$ (in such a way that v_j is contained in T_j). Let e be the edge between v and v_k, and $T' = T - T_k$ be the tree obtained by removing T_k from T. We have

$$
\begin{aligned}
\mathrm{avm}(T) &= \frac{\mathrm{S}(T)}{\mathrm{M}(T)} = \frac{\mathrm{S}(T-e) + \mathrm{S}(T-v-v_k) + \mathrm{M}(T-v-v_k)}{\mathrm{M}(T)} \\
&= \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)} \cdot \frac{\mathrm{S}(T-e)}{\mathrm{M}(T-e)} + \frac{\mathrm{M}(T-v-v_k)}{\mathrm{M}(T)} \cdot \left(1 + \frac{\mathrm{S}(T-v-v_k)}{\mathrm{M}(T-v-v_k)} \right) \\
&= \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)} \mathrm{avm}(T-e) + \frac{\mathrm{M}(T) - \mathrm{M}(T-e)}{\mathrm{M}(T)} (1 + \mathrm{avm}(T-v-v_k)) \\
&= \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)} (\mathrm{avm}(T') + \mathrm{avm}(T_k)) \\
&\quad + \left(1 - \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)} \right) \left(1 + \sum_{j=1}^{k-1} \mathrm{avm}(T_j) + \mathrm{avm}(T_k - v_k) \right).
\end{aligned}
\tag{12}
$$

Set $A = \mathrm{avm}(T') + \mathrm{avm}(T_k)$ and $B = 1 + \sum_{j=1}^{k-1} \mathrm{avm}(T_j) + \mathrm{avm}(T_k - v_k)$.

Assume first that $k \geq 4$. By Lemma 4 and the induction hypothesis, we have $\mathrm{avm}(T_j) \leq a|T_j| + \frac{1}{\sqrt{5}} - \frac{1}{2}$ for all j and $\mathrm{avm}(T') \leq a|T'| + b$. It follows that

$$
A \leq a(|T'| + |T_k|) + b + \frac{1}{\sqrt{5}} - \frac{1}{2} = a|T| + b + \frac{1}{\sqrt{5}} - \frac{1}{2} < a|T| + b.
$$

If $B \leq a|T| + b$, then we are done immediately. Hence we can assume that $A < a|T| + b \leq B$. This implies that the expression for $\mathrm{avm}(T)$ in (12) is decreasing

regarded as a function of $\frac{M_{(T-e)}}{M_{(T)}}$, which means that we will need lower bounds for this quotient. So let us first find a formula for $\frac{M_{(T-e)}}{M_{(T)}}$. We observe that

$$\frac{M(T - e)}{M(T)} = \frac{M(T')M(T_k)}{M(T')M(T_k) + M(T' - v)M(T_k - v_k)},$$

thus

$$\frac{M(T - e)}{M(T)} = \left(1 + \frac{M(T' - v)}{M(T')} \cdot \frac{M(T_k - v_k)}{M(T_k)}\right)^{-1}. \tag{13}$$

Let us also find an expression for $\frac{M_{(T'-v)}}{M_{(T')}}$:

$$\frac{M(T' - v)}{M(T')} = \frac{\prod_{j=1}^{k-1} M(T_j)}{\prod_{j=1}^{k-1} M(T_j) + \sum_{j=1}^{k-1} M(T_j - v_j) \prod_{\substack{i=1 \\ i \neq j}}^{k-1} M(T_i)} \tag{14}$$

$$= \left(1 + \sum_{j=1}^{k-1} \frac{M(T_j - v_j)}{M(T_j)}\right)^{-1}.$$

Moreover, by Lemma 5, we have $\frac{M_{(T'-v)}}{M_{(T')}} \leq 1$, and plugging this estimate into (13), we obtain

$$\frac{M(T - e)}{M(T)} \geq \left(1 + \frac{M(T_k - v_k)}{M(T_k)}\right)^{-1}. \tag{15}$$

We have to consider two different cases:

Case 1: One of the T_j's is the two-vertex path P_2. Then we can without loss of generality assume that $T_k = P_2$, so that $\text{avm}(T_k) = \frac{1}{2}$ and $\text{avm}(T_k - v_k) = 0$. Let us distinguish two subcases depending on the number of other branches T_j that are isomorphic to P_2.

– At least one of the T_j's is different from P_2. We have

$$A \leq a|T| + b + \frac{1}{\sqrt{5}} - \frac{1}{2},$$

as it was established earlier. Moreover, by Lemma 4 and the induction hypothesis,

$$\text{avm}(T_j) \leq a|T_j| + \frac{1}{\sqrt{5}} - \frac{1}{2}$$

for all j, and

$$\text{avm}(T_j) \le a|T_j| + \frac{2}{\sqrt{5}} - 1$$

if T_j is different from P_2. Since this is the case for at least one index j, it follows that

$$B = 1 + \sum_{j=1}^{k-1} \text{avm}(T_j)$$

$$\le 1 + \sum_{j=1}^{k-1} a|T_j| + (k-2)\left(\frac{1}{\sqrt{5}} - \frac{1}{2}\right) + \frac{2}{\sqrt{5}} - 1$$

$$= a(|T| - 3) + (k-2)\left(\frac{1}{\sqrt{5}} - \frac{1}{2}\right) + \frac{2}{\sqrt{5}}$$

$$\le a(|T| - 3) + 2\left(\frac{1}{\sqrt{5}} - \frac{1}{2}\right) + \frac{2}{\sqrt{5}}$$

$$= a|T| - 3a + \frac{4}{\sqrt{5}} - 1.$$

Since we are assuming that T_k is a two-vertex path, we have $\frac{M_{(T_k - v_k)}}{M_{(T_k)}} = \frac{1}{2}$. Plugging this into (15), we obtain $\frac{M_{(T-e)}}{M_{(T)}} \ge \frac{2}{3}$. Hence, (12) gives us

$$\text{avm}(T) \le a|T| + \frac{2}{3}\left(b + \frac{1}{\sqrt{5}} - \frac{1}{2}\right) + \frac{1}{3}\left(-3a + \frac{4}{\sqrt{5}} - 1\right)$$

$$= a|T| + \frac{29}{6\sqrt{5}} - \frac{23}{10} \approx a|T| - 0.13847 < a|T| + b.$$

– All of the T_j's are equal to P_2. In this case, we can determine $M(T)$ and $S(T)$ explicitly (as functions of k only) by means of Proposition 1:

$$M(T) = 2^k + k2^{k-1}$$

and

$$S(T) = k2^{k-1} + k(k+1)2^{k-2},$$

thus

$$\text{avm}(T) = \frac{k^2 + 3k}{2k + 4}.$$

Now one verifies easily that

$$\text{avm}(T) = \frac{k^2 + 3k}{2k + 4} \le a(2k + 1) + b = a|T| + b$$

holds for all $k \ge 4$, completing the proof in Case 1.

Case 2: None of the T_j's is a 2-vertex path P_2.

Let us distinguish different cases depending on the shape of T_k. We may assume that T_k is the smallest branch, i.e. $|T_k| = \min_{1 \le j \le k} |T_j|$.

– If $|T_k| = 1$, then $\text{avm}(T_k) = \text{avm}(T_k - v_k) = 0$. It follows that

$$A = \text{avm}(T') \le a|T'| + b = a|T| + b - a.$$

Moreover, since $\text{avm}(T_j) \le |T_j| + \frac{2}{\sqrt{5}} - 1$ for every j by the induction hypothesis and Lemma 4 (and the assumption that none of the T_j is a 2-vertex path), we have

$$B = 1 + \sum_{j=1}^{k-1} \text{avm}(T_j)$$

$$\le 1 + a \sum_{j=1}^{k-1} |T_j| + (k-1)\left(\frac{2}{\sqrt{5}} - 1\right)$$

$$\le 1 + a(|T| - 2) + 3\left(\frac{2}{\sqrt{5}} - 1\right)$$

$$= a|T| - 2a + \frac{6}{\sqrt{5}} - 2.$$

Since $\frac{M_{(T_k - v_k)}}{M_{(T_k)}} = 1$, (15) gives us $\frac{M_{(T-e)}}{M_{(T)}} \ge \frac{1}{2}$. Thus,

$$\text{avm}(T) \le \frac{1}{2}(a|T| + b - a) + \frac{1}{2}\left(a|T| - 2a + \frac{6}{\sqrt{5}} - 2\right)$$

$$= a|T| + \frac{11}{2\sqrt{5}} - \frac{13}{5} \approx a|T| - 0.14033 < a|T| + b.$$

– If $|T_k| = 3$, then $\text{avm}(T_k) = a|T_k| + \frac{3}{2\sqrt{5}} - \frac{5}{6}$ and $\text{avm}(T_k - v_k) \le a(|T_k| - 1) + \frac{1}{\sqrt{5}} - \frac{1}{2}$. In the same way as in the previous case, it follows that

$$A \le a|T| + b + \frac{3}{2\sqrt{5}} - \frac{5}{6},$$

$$B \le 1 + a(|T| - 2) + 3\left(\frac{2}{\sqrt{5}} - 1\right) + \frac{1}{\sqrt{5}} - \frac{1}{2} = a|T| - 2a + \frac{7}{\sqrt{5}} - \frac{5}{2}.$$

Since $\frac{M_{(T_k - v_k)}}{M_{(T_k)}} \le \frac{2}{3}$, in this case (15) gives us $\frac{M_{(T-e)}}{M_{(T)}} \ge \frac{3}{5}$. We obtain

$$\text{avm}(T) \le a|T| + \frac{3}{5}\left(b + \frac{3}{2\sqrt{5}} - \frac{5}{6}\right) + \frac{2}{5}\left(-2a + \frac{7}{\sqrt{5}} - \frac{5}{2}\right)$$

$$= a|T| + \frac{31}{5\sqrt{5}} - \frac{73}{25} \approx a|T| - 0.14728 < a|T| + b.$$

- If $|T_k| = 4$, then $\mathrm{avm}(T_k) \leq a|T_k| + \frac{2}{\sqrt{5}} - 1$ and $\mathrm{avm}(T_k - v_k) \leq a(|T_k| - 1)$ $+ \frac{3}{2\sqrt{5}} - \frac{5}{6}$. In the same way as before, it follows that

$$A \leq a|T| + b + \frac{2}{\sqrt{5}} - 1,$$

$$B \leq 1 + a(|T| - 2) + 3\left(\frac{2}{\sqrt{5}} - 1\right) + \frac{3}{2\sqrt{5}} - \frac{5}{6} = a|T| - 2a + \frac{3\sqrt{5}}{2} - \frac{17}{6}.$$

Moreover, $\frac{\mathrm{M}(T_k - v_k)}{\mathrm{M}(T_k)} \leq \frac{3}{4}$ in this case, so using (15) again, we get $\frac{\mathrm{M}(T - e)}{\mathrm{M}(T)} \geq \frac{4}{7}$. Hence

$$\mathrm{avm}(T) \leq a|T| + \frac{4}{7}\left(b + \frac{2}{\sqrt{5}} - 1\right) + \frac{3}{7}\left(-2a + \frac{3\sqrt{5}}{2} - \frac{17}{6}\right)$$

$$= a|T| + \frac{19\sqrt{5}}{14} - \frac{223}{70} \approx a|T| - 0.15105 < a|T| + b.$$

- If $|T_k| \geq 5$, then $\mathrm{avm}(T_j) \leq a|T_j| + \frac{3}{\sqrt{5}} - \frac{19}{13}$ for all j (by the induction hypothesis and (11), we have $\mathrm{avm}(T_j) \leq a|T_j| + c_6 = a|T_j| + \frac{3}{\sqrt{5}} - \frac{19}{13}$ if T_j is a path, and $\mathrm{avm}(T_j) \leq a|T_j| + b \leq a|T_j| + \frac{3}{\sqrt{5}} - \frac{19}{13}$ otherwise) and $\mathrm{avm}(T_k - v_k) \leq a(|T_k| - 1) + \frac{2}{\sqrt{5}} - 1$. So it follows now that

$$A \leq a|T| + b + \frac{3}{\sqrt{5}} - \frac{19}{13},$$

$$B \leq 1 + a(|T| - 2) + 3\left(\frac{3}{\sqrt{5}} - \frac{19}{13}\right) + \frac{2}{\sqrt{5}} - 1 = a|T| - 2a + \frac{11}{\sqrt{5}} - \frac{57}{13}.$$

Since $\frac{\mathrm{M}(T_k - v_k)}{\mathrm{M}(T_k)} \leq 1$, we have $\frac{\mathrm{M}(T - e)}{\mathrm{M}(T)} \geq \frac{1}{2}$ by (15). Thus,

$$\mathrm{avm}(T) \leq a|T| + \frac{1}{2}\left(b + \frac{3}{\sqrt{5}} - \frac{19}{13} - 2a + \frac{11}{\sqrt{5}} - \frac{57}{13}\right)$$

$$= a|T| + \frac{37}{4\sqrt{5}} - \frac{1111}{260} \approx a|T| - 0.13635 < a|T| + b.$$

This completes the proof in the case that $k \geq 4$, so we are left with the case $k = 3$. We return to the representation

$$\mathrm{avm}(T) = \frac{\mathrm{M}(T - e)}{\mathrm{M}(T)}(\mathrm{avm}(T') + \mathrm{avm}(T_k))$$

$$+ \left(1 - \frac{\mathrm{M}(T - e)}{\mathrm{M}(T)}\right)\left(1 + \sum_{j=1}^{k-1} \mathrm{avm}(T_j) + \mathrm{avm}(T_k - v_k)\right). \tag{16}$$

Plugging (14) into (13), we obtain

$$\frac{M(T - e)}{M(T)} = \left(1 + \frac{1}{1 + \sum_{j=1}^{k-1} \frac{M(T_j - v_j)}{M(T_j)}} \cdot \frac{M(T_k - v_k)}{M(T_k)} \right)^{-1}. \tag{17}$$

Now we distinguish different cases depending on how many of the branches T_j have one, two, three, four and five or more vertices respectively. This gives us a total of 35 cases corresponding to the solutions of

$$x_1 + x_2 + x_3 + x_4 + x_5 = 3.$$

Here, x_1, x_2, x_3, x_4 stand for the number of T_j's with one, two, three, and four vertices respectively, and x_5 is the number of T_j's with five or more vertices. In each of the cases, we use the following explicit values and bounds. The bounds and explicit values for $|T_j| \leq 4$ are obtained by an exhaustive case check, while the bounds for $|T_j| > 4$ follow from the induction hypothesis and Lemma 4.

$$\mathrm{avm}(T_j) \begin{cases} = a|T_j| - a & |T_j| = 1, \\ = a|T_j| + c_2 & |T_j| = 2, \\ = a|T_j| + c_3 & |T_j| = 3, \\ \leq a|T_j| + c_4 & |T_j| = 4, \\ \leq a|T_j| + c_6 & \text{otherwise,} \end{cases}$$

$$\mathrm{avm}(T_j - v_j) \begin{cases} = a|T_j| - a & |T_j| = 1, \\ = a|T_j| - 2a & |T_j| = 2, \\ \leq a(|T_j| - 1) + c_2 & |T_j| = 3, \\ \leq a(|T_j| - 1) + c_3 & |T_j| = 4, \\ \leq a(|T_j| - 1) + c_4 & \text{otherwise.} \end{cases}$$

We can assume that the degree of v_j is at most 3 for every j, since otherwise we can go back to the case that $k \geq 4$. Using this assumption, we have

$$\frac{M(T_j - v_j)}{M(T_j)} \begin{cases} = 1 & |T_j| = 1, \\ = \dfrac{1}{2} & |T_j| = 2, \\ \in \left[\dfrac{1}{3}, \dfrac{2}{3} \right] & |T_j| = 3, \\ \in \left[\dfrac{2}{5}, \dfrac{3}{4} \right] & |T_j| = 4, \\ \in \left[\dfrac{4}{11}, \dfrac{3}{4} \right] & \text{otherwise.} \end{cases}$$

The first four statements are obtained by checking all possible cases. For the last one, we use a similar recursion to (14) combined with Lemma 5 as follows: note first that v_j has at most two neighbors in T_j, since its degree in T is at most 3. If there is only one neighbor, let w be this neighbor, and set $S = T_j - v_j$. We have

$$\frac{M(T_j - v_j)}{M(T_j)} = \left(1 + \frac{M(S - w)}{M(S)}\right)^{-1}.$$

Applying Lemma 5 to S and w yields $\frac{1}{3} \leq \frac{M(S-w)}{M(S)} \leq 1$ (if the degree of w was greater than 2, we could go back to the case $k \geq 4$ again), thus $\frac{M(T_j-v_j)}{M(T_j)} \in [\frac{1}{2}, \frac{3}{4}]$. If there are two neighbors w_1 and w_2, let S_1 and S_2 be the respective components of $T_j - v_j$. Since $\frac{1}{3} \leq \frac{M(S_i-w_i)}{M(S_i)} \leq 1$, we obtain

$$\frac{M(T_j - v_j)}{M(T_j)} = \left(1 + \frac{M(S_1 - w_1)}{M(S_1)} + \frac{M(S_2 - w_2)}{M(S_2)}\right)^{-1} \leq \frac{1}{1 + \frac{1}{3} + \frac{1}{3}} = \frac{3}{5}$$

in this case, which readily proves the upper bound of $\frac{3}{4}$ in all cases. To improve the lower bound even further, we can note that one of the two trees S_1 and S_2 has more than one vertex; without loss of generality, let this be S_1. Applying the same argument to S_1 as to T_j, we find $\frac{M(S_1-w_1)}{M(S_1)} \leq \frac{3}{4}$. Thus

$$\frac{M(T_j - v_j)}{M(T_j)} = \left(1 + \frac{M(S_1 - w_1)}{M(S_1)} + \frac{M(S_2 - w_2)}{M(S_2)}\right)^{-1} \geq \frac{1}{1 + \frac{3}{4} + 1} = \frac{4}{11},$$

and we have also established the lower bound.

Next we return to the representation (16). Note that $|T'| \geq 3$, so the induction hypothesis combined with (11) yields

$$\mathrm{avm}(T') \leq \mathrm{avm}(P_{|T'|}) = a|T'| + c_{|T'|} \leq a|T'| + c_4.$$

The same reasoning shows that $\mathrm{avm}(T_k) \leq a|T_k| + c_2$, thus $\mathrm{avm}(T') + \mathrm{avm}(T_k) \leq (a|T'| + c_4) + (a|T_k| + c_2) < a|T| + b$. As before, if $1 + \sum_{j=1}^{k-1} \mathrm{avm}(T_j) + \mathrm{avm}(T_j - v_j) \leq a|T| + b$, then we are done. So we may assume that

$$\mathrm{avm}(T') + \mathrm{avm}(T_k) < a|T| + b \leq 1 + \sum_{j=1}^{k-1} \mathrm{avm}(T_j) + \mathrm{avm}(T_j - v_j).$$

Hence the expression (16) is linear and decreasing in $\frac{M(T-e)}{M(T)}$, and thus its maximum is attained for the smallest possible value of $\frac{M(T-e)}{M(T)}$.

By the induction hypothesis, $\mathrm{avm}(T') \leq \mathrm{avm}(P_{|T'|}) = a|T'| + c_{|T'|}$. This inequality is plugged into (16) along with the bounds for $\mathrm{avm}(T_j)$ and $\mathrm{avm}(T_j - v_j)$. The identity (17) is used to obtain a lower bound on the quotient $\frac{\mathrm{M}(T-e)}{\mathrm{M}(T)}$. All this gives us an upper bound for $\mathrm{avm}(T)$ in each of the aforementioned 35 cases, which can all be checked easily with a computer. The worst case happens when $x_1 = x_3 = x_4 = x_5 = 0$ and $x_2 = 3$, where we have the equality $\mathrm{avm}(T) = a|T| + b$. As another example to illustrate the general procedure, let us consider the case that gives us the second worst estimate: it is obtained for $x_1 = x_3 = x_5 = 0$, $x_2 = 2$ and $x_4 = 1$. Let T_2 and T_3 both have two vertices, so that the first branch T_1 consists of four vertices. Note that T_3 is isomorphic to P_2 and $\mathrm{avm}(T_1) \leq \mathrm{avm}(P_4)$ from the first part of the proof. We then have

$$\mathrm{avm}(T_3) = a|T_3| + c_2, \ \mathrm{avm}(T') \leq a|T'| + c_7,$$

thus

$$\mathrm{avm}(T') + \mathrm{avm}(T_3) \leq a|T| + c_2 + c_7 = a|T| + \frac{9}{2\sqrt{5}} - \frac{46}{21}.$$

Moreover,

$$\mathrm{avm}(T_3 - v_3) = 0, \ \mathrm{avm}(T_1) \leq 4a + c_4, \ \mathrm{avm}(T_2) = \frac{1}{2},$$

and thus

$$1 + \sum_{j=1}^{2} \mathrm{avm}(T_j) + \mathrm{avm}(T_3 - v_3) \leq 1 + 4a + c_4 + \frac{1}{2} = a|T| + \frac{9}{2\sqrt{5}} - 2.$$

Finally, we have

$$\frac{\mathrm{M}(T-e)}{\mathrm{M}(T)} = \left(1 + \frac{\mathrm{M}(T_3 - v_3)}{\mathrm{M}(T_3)} \cdot \frac{1}{1 + \frac{\mathrm{M}(T_1 - v_1)}{\mathrm{M}(T_1)} + \frac{\mathrm{M}(T_2 - v_2)}{\mathrm{M}(T_2)}}\right)^{-1}$$

$$\geq \left(1 + \frac{1}{2} \cdot \frac{1}{1 + \frac{1}{2} + \frac{2}{5}}\right)^{-1} = \frac{19}{24}.$$

Putting everything together, we obtain

$$\mathrm{avm}(T) = \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)}(\mathrm{avm}(T') + \mathrm{avm}(T_3))$$

$$+ \left(1 - \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)}\right)\left(1 + \sum_{j=1}^{2}\mathrm{avm}(T_j) + \mathrm{avm}(T_3 - v_3)\right)$$

$$\leq \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)}\left(a|T| + \frac{9}{2\sqrt{5}} - \frac{46}{21}\right)$$

$$+ \left(1 - \frac{\mathrm{M}(T-e)}{\mathrm{M}(T)}\right)\left(a|T| + \frac{9}{2\sqrt{5}} - 2\right)$$

$$\leq \frac{19}{24}\left(a|T| + \frac{9}{2\sqrt{5}} - \frac{46}{21}\right) + \frac{5}{24}\left(a|T| + \frac{9}{2\sqrt{5}} - 2\right)$$

$$= a|T| + \frac{9}{2\sqrt{5}} - \frac{271}{126} \approx a|T| - 0.13833 < a|T| + b.$$

The other cases are treated in the same fashion and give upper bounds with smaller constant terms. Thus the induction is complete. In order to complete the proof of the theorem, we only need a lower bound on $\mathrm{avm}(P_n)$. However, (11) shows that

$$\mathrm{avm}(P_n) = an + c_n \geq an + c_5$$

for $n > 3$, and $c_5 = \frac{\sqrt{5}}{2} - \frac{5}{4} \approx -0.131966 > b$. Thus $\mathrm{avm}(P_n) > an + b \geq \mathrm{avm}(T)$ for every tree T with n vertices other than P_n. This completes the proof. □

5 Relations to Other Invariants

In this section, we will prove inequalities between the average matching size and other matching-related quantities associated with a graph. Let G be an n-vertex graph. The matching polynomial and the matching generating polynomial are defined as follows:

$$\Phi(G, x) = \sum_{k \geq 0} \mathrm{m}(G, k)(-1)^k x^{n-2k},$$

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

Note that the average size of matchings in G can be expressed as

$$\mathrm{avm}(G) = \frac{\sum_{k \geq 0} k\,\mathrm{m}(G, k)}{\sum_{k \geq 0}\mathrm{m}(G, k)} = \frac{\mathrm{M}'(G, 1)}{\mathrm{M}(G, 1)},$$

where $\mathrm{M}'(G, x)$ is the first derivative of $\mathrm{M}(G, x)$ with respect to x.

It is easy to see that $\Phi(G, x) = x^n \mathrm{M}(G, -\frac{1}{x^2})$. Using this relation, we can write the derivative of Φ in terms of M and its derivative as follows:

$$\Phi'(G, x) = nx^{n-1}\mathrm{M}\left(G, -\frac{1}{x^2}\right) + 2x^{n-3}\mathrm{M}'\left(G, -\frac{1}{x^2}\right).$$

This gives us

$$\frac{\Phi'(G,x)}{\Phi(G,x)} = \frac{n}{x} + \frac{2}{x^3} \frac{\mathrm{M}'\left(G, -\frac{1}{x^2}\right)}{\mathrm{M}\left(G, -\frac{1}{x^2}\right)}. \tag{18}$$

Let $\mu_1, \mu_2, \ldots, \mu_n$ be the zeros of the matching polynomial $\Phi(G,x)$; it is well-known that these zeros are real, see for example Section 8.5 in [9]. Now we can express Φ and Φ' in terms of the zeros as follows:

$$\Phi(G,x) = \prod_{j=1}^{n}(x - \mu_j),$$

$$\Phi'(G,x) = \sum_{k=1}^{n} \frac{\prod_{j=1}^{n}(x - \mu_j)}{x - \mu_k}.$$

Therefore,

$$\frac{\Phi'(G,x)}{\Phi(G,x)} = \sum_{k=1}^{n} \frac{1}{x - \mu_k}. \tag{19}$$

Now, we can establish a relation between the average size of matchings of G and the zeros of its matching polynomial.

Lemma 6 *Let G be an n-vertex graph and $\mu_1, \ldots, \mu_n$ be the zeros of the matching polynomial of G. Then*

$$\mathrm{avm}(G) = \frac{1}{2} \sum_{j=1}^{n} \frac{\mu_j^2}{\mu_j^2 + 1}.$$

Proof Using (18) and (19), and plugging in $x = i$, we obtain

$$\sum_{j=1}^{n} \frac{1}{i - \mu_j} = \frac{n}{i} + \frac{2}{i^3} \frac{\mathrm{M}'(G,1)}{\mathrm{M}(G,1)},$$

and this simplifies to

$$\sum_{j=1}^{n} \frac{\mu_j}{\mu_j - i} = 2\,\mathrm{avm}(G). \tag{20}$$

Let us rearrange the left hand side of (20). We have

$$\sum_{j=1}^{n} \frac{\mu_j}{\mu_j - i} = \sum_{j=1}^{n} \frac{\mu_j(\mu_j + i)}{(\mu_j - i)(\mu_j + i)} = \sum_{j=1}^{n} \frac{\mu_j^2 + i\mu_j}{\mu_j^2 + 1}.$$

Since the imaginary part must be 0, we get the desired result.   □

Having established this relation, we can now prove two inequalities. The first relates the average matching size with the total number of matchings. Note that the latter is $M(G) = M(G, 1)$, which can be expressed in terms of the zeros $\mu_1, \ldots, \mu_n$ as well:

$$M(G) = M(G, 1) = |M(G, 1)| = |i^{-n}\Phi(G, i)| = \left|\prod_{j=1}^{n}(i - \mu_j)\right| = \prod_{j=1}^{n}\sqrt{1 + \mu_j^2}.$$

It is not difficult to verify that the inequality

$$\frac{x}{1 + x} \leq \beta \log(1 + x) + 1 - \beta + \beta \log \beta$$

holds for all positive real numbers β and x. Plugging in μ_j^2 for x and summing over all j yields the following result:

Proposition 3 *For every positive real number β and every n-vertex graph G, we have*

$$\mathrm{avm}(G) \leq \beta \log M(G) + (1 - \beta + \beta \log \beta)\frac{n}{2}.$$

In particular,

$$\mathrm{avm}(G) \leq \log M(G).$$

We can still choose β arbitrarily. Differentiating with respect to β, we find that the optimal value for β (that minimizes the upper bound) is $\beta = M(G)^{-2/n}$. Plugging this back into the inequality, we obtain the following theorem:

Theorem 6 *For every n-vertex graph, we have*

$$\mathrm{avm}(G) \leq \frac{n}{2}\left(1 - M(G)^{-2/n}\right).$$

An alternative way to prove this theorem is to apply the inequality between the arithmetic and the geometric mean.

We conclude this section with a similar inequality involving the matching energy. This invariant is defined as follows [4]:

$$ME(G) = \sum_{j=1}^{n}|\mu_j|.$$

Following an analogous approach, we can prove a relation between the average size of matchings in G and the matching energy of G.

Theorem 7 *For every graph G,*

$$ME(G) \geq 4\,\mathrm{avm}(G).$$

Proof For all nonnegative real x, we have $\frac{x^2}{1+x^2} \leq \frac{x}{2}$. Therefore, by Lemma 6,

$$\text{avm}(G) = \frac{1}{2}\sum_{j=1}^{n}\frac{\mu_j^2}{1+\mu_j^2} \leq \frac{1}{2}\sum_{j=1}^{n}\frac{|\mu_j|}{2} = \frac{1}{4}\text{ME}(G).$$

□

Remark 4 Note that in the case of trees, the matching polynomial coincides with the characteristic polynomial. So we have a correspondence between the average size of matchings of a tree and the classical energy of a tree, which is the sum of the absolute values of the eigenvalues, see [8].

6 The Weighted Average Size of Matchings in a Graph

In the context of the monomer-dimer model from statistical physics, one often considers a probability distribution on the set of matchings where the probability of a k-matching is proportional to α^k for some constant α, see for example [2]. This provides the motivation to study the weighted average size of matchings. We consider a random matching according to the aforementioned probability distribution, where α is a fixed positive number. We define the weighted total number of matchings in G, the weighted total size of G and the weighted average size of matchings in G as follows:

$$M^\alpha(G) = \sum_{k \geq 0} m(G,k)\alpha^k,$$

$$S^\alpha(G) = \sum_{k \geq 0} k\,m(G,k)\alpha^k,$$

$$\text{avm}^\alpha(G) = \frac{S^\alpha(G)}{M^\alpha(G)}.$$

Following a similar reasoning as in the special case where $\alpha = 1$, it is still possible to prove the following inequalities.

Theorem 8 *For every fixed positive real number α and every n-vertex graph G, we have*

$$\text{avm}^\alpha(E_n) \leq \text{avm}^\alpha(G) \leq \text{avm}^\alpha(K_n).$$

Moreover, for every real number $\alpha \in (0,1]$ and every n-vertex tree T, we have

$$\text{avm}^\alpha(S_n) \leq \text{avm}^\alpha(T) \leq \text{avm}^\alpha(P_n).$$

We refer to [11] for more details on the proof. Note that the final inequality $(\text{avm}^\alpha(T) \leq \text{avm}^\alpha(P_n))$ is not generally true for all values of α. One can also express the weighted average matching size in terms of the zeros of the matching polynomial:

Lemma 7 *Let G be an n-vertex graph and $\mu_1, \ldots, \mu_n$ be the zeros of the matching polynomial of G. Then*

$$\mathrm{avm}^\alpha(G) = \frac{1}{2} \sum_{j=1}^{n} \frac{\alpha \mu_j^2}{\alpha \mu_j^2 + 1}.$$

Finally, it is also possible again to prove inequalities that relate $\mathrm{avm}^\alpha(G)$ to other invariants. Specifically, we have the following straightforward generalization of Theorem 7:

Theorem 9 *For every graph G and every positive real number α,*

$$\mathrm{ME}(G) \geq \frac{4}{\sqrt{\alpha}} \mathrm{avm}^\alpha(G).$$

References

1. Andriantiana, E.O.D., Razanajatovo Misanantenaina, V., Wagner, S.: The average size of independent sets in a graph. Eur. J. Math. (2019). https://doi.org/10.1007/s40879-019-00333-8
2. Davies, E., Jenssen, M., Perkins, W., Roberts, B.: Independent sets, matchings, and occupancy fractions. J. Lond. Math. Soc. **96**(1), 47–66 (2017)
3. Flajolet, P., Sedgewick, R.: Analytic Combinatorics. Cambridge University Press, Cambridge (2009)
4. Gutman, I., Wagner, S.: The matching energy of a graph. Discrete Appl. Math. **160**(15), 2177–2187 (2012)
5. Haslegrave, J.: Extremal results on average subtree density of series-reduced trees. J. Comb. Theory Ser. B **107**, 26–41 (2014)
6. Jamison, R.E.: On the average number of nodes in a subtree of a tree. J. Comb. Theory Ser. B **35**(3), 207–223 (1983)
7. Jamison, R.E.: Monotonicity of the mean order of subtrees. J. Comb. Theory Ser. B **37**(1), 70–78 (1984)
8. Li, X., Shi, Y., Gutman, I.: Graph Energy. Springer Science & Business Media, Berlin (2012)
9. Lovász, L., Plummer, M.D.: Matching Theory, volume 121 of North-Holland Mathematics Studies. Annals of Discrete Mathematics, vol. 29. North-Holland Publishing Co., Amsterdam (1986)
10. Mol, L., Oellermann, O.: Maximizing the mean subtree order. J. Graph Theory (**To appear**)
11. Razanajatovo Misanantenaina, V.: Properties of graph polynomials and related parameters. Ph.D. thesis, Stellenbosch University (2017)
12. Vince, A., Wang, H.: The average order of a subtree of a tree. J. Comb. Theory Ser. B **100**(2), 161–170 (2010)
13. Wagner, S.: Upper and lower bounds for Merrifield–Simmons index and Hosoya index. In: Gutman, I., Furtula, B., Das, K.C., Milovanović, E., Milovanović, I. (eds.) Bounds in Chemical Graph Theory—basics, volume 19 of Mathematical Chemistry Monographs, pp. 155–187. University of Kragujevac and Faculty of Science, Kragujevac (2017)
14. Wagner, S., Wang, H.: On the local and global means of subtree orders. J. Graph Theory **81**(2), 154–166 (2016)