MEASURING ASSOCIATIVITY: GRAPH ALGEBRAS OF UNDIRECTED GRAPHS

KAMILLA KÁTAI-URBÁN AND TAMÁS WALDHAUSER

Dedicated to Reinhard Pöschel on the occasion of his 75th birthday.

Abstract. We study two measures of associativity for graph algebras of finite undirected graphs: the index of nonassociativity and (a variant of) the semigroup distance. We determine "almost associative" and "antiassociative" graphs with respect to both measures. It turns out that the antiassociative graphs are exactly the balanced complete bipartite graphs, no matter which of the two measures we consider. In the class of connected graphs the two notions of almost associativity are also equivalent.

1. Introduction

There are several ways to measure the (non)associativity of a given binary operation $\circ$. The *index of nonassociativity* counts the number of triples (a, b, c) such that $(a \circ b) \circ c \neq a \circ (b \circ c)$. The *semigroup distance* counts the minimum number of changes we need to perform in the Cayley table of $\circ$ in order to make it associative. The *associative spectrum* counts the number of term functions arising from different bracketings of $x_1 \circ \cdots \circ x_n$, thereby giving information about the consequences of the associative identity that are (not) satisfied by $\circ$.

We define these notions more precisely in Section 2, and we give several examples as well as a brief overview of earlier research in this area. The examples will show that there is little relationship among these measures of associativity: it is possible that one of them is very small, while another one is very large, for the same binary operation.

We can define a binary operation on the vertices of a graph by

$$x \circ y = \begin{cases} x, & \text{if there is an edge from } x \text{ to } y; \\ 0, & \text{otherwise;} \end{cases}$$

where 0 is an external zero element. The main topic of this paper is the study of associativity measures of these *graph algebras*. Again, the more formal definition is given in Section 2, and we also recall some notions of graph theory there.

We will only consider undirected graphs, and these have been completely classified with respect to their associative spectra in [16] (we state this classification in Theorem 5.1 in Section 5). Therefore, we focus on the semigroup distance (more precisely, a "graph-algebraic" version of the semigroup distance)

Date: August 30, 2023.

This research was partially supported by the National Research, Development and Innovation Office of Hungary under grant no. K138892, and by project TKP2021-NVA-09, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

and on the index of nonassociativity in sections 3 and 4. In particular, we determine the minimal and maximal values of these measures of associativity, and we characterize graphs corresponding to these extremal values.

We will conclude in Section 5 that—as opposed to the case of arbitrary binary operations—the semigroup distance and the index of nonassociativity are closely related: for connected undirected graphs, the notions of "antiassociativity" and "almost associativity" coincide for these two measures of associativity (but not for the associative spectrum).

2. Preliminaries

2.1. Measures of (non)associativity.

Definition 2.1 ([2])**.** The *index of nonassociativity* of a finite groupoid $\mathbb{A} = (A; \circ)$ is the number of nonassociative triples in $\mathbb{A}$:

$$\operatorname{ns}(\mathbb{A}) = \left|\left\{(a,b,c) \in A^3 : (a \circ b) \circ c \neq a \circ (b \circ c)\right\}\right|.$$

The index of nonassociativity was defined by Climescu [2], and he proved that all values between 1 and n^3 are possible for n-element groupoids if $n \geq 3$. (For 2-element groupoids the possible values are 0, 2, 4 and 8.) Clearly, $\operatorname{ns}(\mathbb{A}) = 0$ if and only if $\mathbb{A}$ is associative (or, more precisely, the binary operation of $\mathbb{A}$ is associative), and we may say that $\mathbb{A}$ is almost associative if $\operatorname{ns}(\mathbb{A}) = 1$, while $\mathbb{A}$ can be regarded as antiassociative if $\operatorname{ns}(\mathbb{A}) = n^3$. Groupoids with $\operatorname{ns}(\mathbb{A}) = 1$ are also called *Szász–Hájek-groupoids*, because their structure was first studied by Szász [23] and Hájek [5], and later by Kepka and Trch in a long series of papers [8–15].

To introduce our second measure of associativity, we need to define the distance of two groupoids. Let $\mathbb{A}_1 = (A; \circ)$ and $\mathbb{A}_2 = (A; *)$ be two groupoids on the same finite set A. The distance of $\mathbb{A}_1$ and $\mathbb{A}_2$ is the Hamming distance of their operation tables, i.e., the number of positions where the operation tables differ:

$$\operatorname{dist}(\mathbb{A}_1, \mathbb{A}_2) = |\{(x,y) : x \circ y \neq x * y\}|.$$

The set of all groupoids on A is a metric space with the above defined distance. The semigroup distance a groupoid $\mathbb{A}$, introduced by Kepka and Trch [7], is simply the distance of $\mathbb{A}$ to the set of all semigroups in this metric space.

Definition 2.2 ([7])**.** The *semigroup distance* of a finite groupoid $\mathbb{A} = (A; \circ)$ is defined by

$$\operatorname{sdist}(\mathbb{A}) = \min\left\{\operatorname{dist}((A;\circ),(A;*)) : * \text{ is an associative operation on } A\right\}.$$

Informally, $\operatorname{sdist}(\mathbb{A})$ is the least number of changes one has to perform in the operation table of $\mathbb{A}$ to make it associative. We have $\operatorname{sdist}(\mathbb{A}) = 0$ if and only if $\mathbb{A}$ is associative, and we can say that $\mathbb{A}$ is almost associative if $\operatorname{sdist}(\mathbb{A}) = 1$. Antiassociativity for n-element groupoids could be defined by $\operatorname{sdist}(\mathbb{A}) = \operatorname{maxdist}(n)$, where

$$\operatorname{maxdist}(n) = \max\left\{\operatorname{sdist}(\mathbb{A}) : \mathbb{A} \text{ is an } n\text{-element groupoid}\right\}.$$

However, the value of $\operatorname{maxdist}(n)$ is not known. It is clear that $\operatorname{maxdist}(n) \leq n^2 - n$, as we can make any groupoid associative by changing each entry in the operation table to the most frequently occurring element. As a lower bound, we have $\operatorname{maxdist}(n) \geq n^2/4$, as shown by the following example.

Example 2.3. [7] Let A be an n element set, and let $x \circ y = f(x)$, where f is a permutation of A that has no fixed points. Then we have $\mathrm{ns}(\mathbb{A}) = n^3$ (this is easy to verify) and $\mathrm{sdist}(\mathbb{A}) \geq n^2/4$ (this was proved in [7]).

Let us now describe a third way to measure associativity, which was proposed by Csákány [3].

Definition 2.4 ([3])**.** For a groupoid $\mathbb{A} = (A; \circ)$, let $s_n(\mathbb{A})$ denote the number of term operations induced by bracketings of the "product" $x_1 \circ \cdots \circ x_n$. The sequence $\mathrm{spec}(\mathbb{A}) = (s_1(\mathbb{A}), s_2(\mathbb{A}), s_3(\mathbb{A}), \dots)$ is called the *associative spectrum* of $\mathbb{A}$.

Clearly, $s_1(\mathbb{A}) = s_2(\mathbb{A}) = 1$ for all groupoids, and $s_3(\mathbb{A}) = 1$ if $\mathbb{A}$ is associative, while $s_3(\mathbb{A}) = 2$ if $\mathbb{A}$ is not associative. Moreover, by the generalized associative law, $s_3(\mathbb{A}) = 1$ implies $s_n(\mathbb{A}) = 1$ for all $n \in \mathbb{N}$, thus the associative spectrum of a semigroup is $(1, 1, 1, \dots)$. The associative spectrum measures associativity by its consequences: a nonassociative groupoid $\mathbb{A}$ may still satisfy some identities that are consequences of the associative law, and a relatively small spectrum indicates that $\mathbb{A}$ satisfies relatively many of these identities. On the other hand, if $\mathbb{A}$ satisfies no nontrivial "bracketing identities", then $s_n(\mathbb{A})$ equals the number of formally different bracketings of $x_1 \circ \cdots \circ x_n$, which is the $(n-1)$-st Catalan number $C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1}$. In the latter case $\mathbb{A}$ can be regarded as antiassociative, and almost associativity could be defined by $\mathrm{spec}(\mathbb{A}) = (1, 1, 2, 1, 1, \dots)$, as this is the least nonassociative associative spectrum(!). For more background about associative spectra, we refer the reader to [3, 18].

2.2. **Comparisons.** As the following example illustrates, $\mathrm{ns}(\mathbb{A})$ can be arbitrarily large for a groupoid with $\mathrm{sdist}(\mathbb{A}) = 1$, and, at the same time, $\mathbb{A}$ can have a Catalan spectrum.

Example 2.5. Let us define a binary operation $\circ$ on $A = \{0, 1, \dots, n-1\}$ by

$$x \circ y = \begin{cases} 1, & \text{if } x = y = 0; \\ 0, & \text{otherwise.} \end{cases}$$

It is clear that $\mathrm{sdist}(\mathbb{A}) = 1$ (we get a zero semigroup by setting $0 \circ 0 = 0$), and it is not hard to verify that $\mathrm{ns}(\mathbb{A}) = 2n(n-1)$ [7]. Restricting $\circ$ to the subuniverse $\{0, 1\}$, we get the Sheffer operation (NOR operation), which does not satisfy any nontrivial bracketing identities [3], hence the same is true for $\mathbb{A}$, i.e., $\mathbb{A}$ has a Catalan spectrum.

In some sense, the example above is the worst possible: if $|A| = n$, then $\mathrm{sdist}(\mathbb{A}) = 1$ implies $\mathrm{ns}(\mathbb{A}) \leq 2n(n-1)$ [7] (and, of course, no associative spectrum can exceed the Catalan numbers). Analogously to Example 2.5, a groupoid with $\mathrm{ns}(\mathbb{A}) = 1$ can have arbitrarily large semigroup distance [1, 14]. However, here we do not know the maximal value of $\mathrm{sdist}(\mathbb{A})$ among n-element groupoids with $\mathrm{ns}(\mathbb{A}) = 1$; the authors of [1] only mention that the size of the groupoids they have constructed grow quickly.

In the next example we present groupoids that are almost associative with respect to both the index of nonassociativity and the semigroup distance but are antiassociative in the "spectral" sense.

Example 2.6. Let us consider the groupoid $\mathbb{A} = (A; \circ)$ defined by the following operation table:

$\circ$	0	1	2
0	0	0	0
1	0	1	0
2	0	1	2

We have $\operatorname{sdist}(\mathbb{A}) = 1$ (we get a zero semigroup by setting $1 \circ 2 = 2$) and $\operatorname{ns}(\mathbb{A}) = 1$ (the only nonassociative triple is $(1, 2, 1)$). On the other hand, $\mathbb{A}$ has a Catalan spectrum [3]. Repeatedly extending the groupoid by a zero element, we can construct arbitrarily large groupoids with these properties.

Let us now see how small can be the semigroup distance of an n-element groupoid $\mathbb{A}$ with $\operatorname{ns}(\mathbb{A}) = n^3$.

Proposition 2.7. *If* $\mathbb{A}$ *is an* n*-element groupoid and* $\operatorname{ns}(\mathbb{A}) = n^3$*, then* $\operatorname{sdist}(\mathbb{A}) \geq n$.

Proof. Assume for contradiction that there is an n-element groupoid $\mathbb{A} = (A; \circ)$ such that $\operatorname{ns}(\mathbb{A}) = n^3$ and $\operatorname{sdist}(\mathbb{A}) < n$. Then there is a semigroup $\mathbb{A}^* = (A; *)$ with $\operatorname{dist}(\mathbb{A}, \mathbb{A}^*) < n$. By the pigeonhole principle, there is an element $a \in A$ such that the row of a in the operation table of $\mathbb{A}$ is identical to the row of a in the operation table of $\mathbb{A}^*$, and similarly, there exist $c \in A$ such that the column of c looks the same in the two operation tables:

$$\forall x \in A\colon a \circ x = a * x \text{ and } x \circ c = x * c.$$

This implies that (a, b, c) is an associative triple in $\mathbb{A}$ for any $b \in A$:

$$(a \circ b) \circ c = (a * b) * c = a * (b * c) = a \circ (b \circ c).$$

However, this contradicts the assumption $\operatorname{ns}(\mathbb{A}) = n^3$. □

The estimate in the proposition above is sharp, as shown by the following example

Example 2.8. Let us define a binary operation $\circ$ on $A = \{0, 1, \dots, n-1\}$ by $x \circ y = f(x)$, where the map $f\colon A \to A$ is given by

$$f(x) = \begin{cases} 1, & \text{if } x = 0; \\ 0, & \text{otherwise.} \end{cases}$$

(Compare this with Example 2.3.) Then we have $(a \circ b) \circ c = f^2(a)$ and $a \circ (b \circ c) = f(a)$; therefore, $\operatorname{ns}(\mathbb{A}) = n^3$. This implies $\operatorname{sdist}(\mathbb{A}) \geq n$ by Proposition 2.7. Actually, we have $\operatorname{sdist}(\mathbb{A}) = n$, as $\mathbb{A}$ is of distance n to a zero semigroup. Considering the associative spectrum, note that evaluating $x_1 \circ \cdots \circ x_n$ over $\mathbb{A}$, we get either $f(x_1)$ or $f^2(x_1)$, depending on the bracketing. Thus the associative spectrum of $\mathbb{A}$ is quite small: $\operatorname{spec}(\mathbb{A}) = (2, 2, 2, \dots)$.

Summarizing our observations, we can say that the three measures of associativity behave quite differently, but there seems to be some weak connection between the index of nonassociativity and the semigroup distance. The latter is also illustrated by the following inequality.

Theorem 2.9 ([7])**.** *If* $|A| = n$*, then* $\operatorname{ns}(\mathbb{A}) \leq (2n^2 + 2n) \cdot \operatorname{sdist}(\mathbb{A})$.

2.3. **Graphs.** A *directed graph* is a pair $G = (V; \rho)$, where $V = V(G)$ is a nonempty set (vertices) and $\rho \subseteq V \times V$ is a relation on V. We consider finite undirected graphs, i.e., we always assume that the relation ρ is symmetric. (Sometimes we drop the adjective "undirected": by default, a graph always means an undirected graph in this paper.) If G is such a graph and $(x, y) \in \rho$ for two distinct vertices $x, y \in V(G)$, then $e = \{x, y\}$ is an (undirected) *edge* of G, which we will simply write as $e = xy$ (of course, yx is the same edge). If $(x, x) \in \rho$, then there is a *loop* on the vertex x. It will be convenient to treat loops and "real" edges separately. Therefore, by a slight abuse of terminology, in the following we say that $e = xy$ is an edge only in the case $x \neq y$. The set of "loopy" vertices is denoted by $L(G)$, thus we write $x \in L(G)$ to indicate that there is a loop on the vertex x. We denote by $E(G)$ the set of edges (loops are not included!), and we partition this set into three parts according to the number of loops at the endpoints of the edges:

$$E_0(G) = \{xy \in E(G) : x \notin L(G) \text{ and } y \notin L(G)\};$$
$$E_1(G) = \{xy \in E(G) : \text{exactly one of } x \in L(G), y \in L(G) \text{ holds}\};$$
$$E_2(G) = \{xy \in E(G) : x \in L(G) \text{ and } y \in L(G)\}.$$

(Thus $e \in E_i$ means that exactly i of the two endpoints of the edge e have a loop.) We say that G is a *reflexive* graph if the underlying relation is reflexive, i.e., $L(G) = V(G)$. Similarly, we say that G is *irreflexive* if there are no loops, i.e., if $L(G) = \emptyset$. We use the notation $G^{\circlearrowleft}$ for the graph that we obtain from G by adding loops to all vertices.

The *neighborhood* of x in G is the set $N(x) = \{y \in V(G) : xy \in E(G)\}$. Since loops and edges are treated separately, we have $x \notin N(x)$ even if there is a loop on x. The *degree* of a vertex x is the size of it neighborhood: $d(x) = |N(x)|$. Note that if there is a loop on x, it is not taken into account in $d(x)$. We say that x is an *isolated vertex* if $d(x) = 0$. Again, loops do not matter: a vertex with a loop that is not connected to any other vertices is considered isolated. A connected component of G is said to be *nontrivial* if it has at least two vertices; the trivial connected components are just the isolated vertices. For $A \subseteq V$, we denote by $G|_A$ the induced subgraph of G on the vertex set A, and $G \setminus A$ stands for $G|_{V \setminus A}$.

By a *cherry* in G, we mean a three-vertex induced subgraph in which exactly two edges are present (loops do not matter): . The set of cherries in G will be denoted by $Ch(G)$. Thus $xyz \in Ch(G)$ if x, y, z are three distinct vertices, and exactly two of $xy \in E(G)$, $yz \in E(G)$ and $xz \in E(G)$ holds. Here we use xyz as a shorthand for $\{x, y, z\}$ (just as xy is a shorthand for $\{x, y\}$ when we speak about (non)edges).

An irreflexive graph G is *bipartite* if $V(G) = A \cup B$, where A and B are disjoint sets (called the two *color classes*) such that every edge of G has one of its endpoints in A and the other endpoint in B. (We usually draw bipartite graphs in such a way that A is **a**bove B and B is **b**elow A.) For natural numbers n, m, let K_n denote the *complete graph* on n vertices, and let $K_{n,m}$ denote the complete bipartite graph with n and m vertices in the two color classes. We will often need *balanced complete bipartite graphs*, where the sizes of the two color classes differ by at most one. Up to isomorphism there is only one such graph on n vertices, namely $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. These graphs are irreflexive by definition; but we will also frequently encounter the *reflexive complete graph* $K_n^{\circlearrowleft}$. We use the notation $K_n^{\circlearrowleft} - \circlearrowleft$ for the graph that we obtain from $K_n^{\circlearrowleft}$ by removing one loop (of course, up to isomorphism, it does not matter which loop is removed). Similarly, $K_n - \updownarrow$ and $K_n^{\circlearrowleft} - \updownarrow$ indicate the removal of one edge from K_n and $K_n^{\circlearrowleft}$, respectively (loops are retained in the second case).

We will need the following classical result of extremal graph theory, which is a special case of Turán's theorem about K_r-free graphs (see, e.g., [4, Theorem 7.1.1]).

Theorem 2.10. *If G is a triangle-free irreflexive undirected graph on n vertices, then $|E(G)| \leq \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil = \lfloor n^2/4 \rfloor$. Equality holds here if and only if $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

2.4. **Graph algebras.** Shallon [22] proposed a construction of an algebra associated to a directed graph. These graph algebras have a binary and a nullary operation, but here we only consider their groupoid reducts (for simplicity, we do not introduce a new name, and we just call these reducts graph algebras).

Definition 2.11 ([22])**.** The *graph algebra* of a directed graph $G = (V; \rho)$ is the groupoid $\mathbb{A}(G) = (V \mathbin{\dot{\cup}} \{0\}; \circ)$, where

$$x \circ y = \begin{cases} x, & \text{if } x, y \in V \text{ and } (x, y) \in \rho; \\ 0, & \text{if } x, y \in V \text{ and } (x, y) \notin \rho; \\ 0, & \text{if } x = 0 \text{ or } y = 0. \end{cases}$$

Poomsa-ard [19] characterized directed graphs with associative graph algebras.

Theorem 2.12 ([19])**.** *The graph algebra of a directed graph is associative if and only if the edge relation is transitive and the outneighborhood of each vertex is a reflexive complete graph.*

For undirected graphs, the characterization takes the following simple form.

Corollary 2.13. *An undirected graph has an associative graph algebra if and only if all of its nontrivial connected components are reflexive complete graphs.*

Associative spectra of graph algebras were investigated in [16, 17]. The main tool in that study was a result of Pöschel and Wessel [20] describing satisfaction of identities in terms of graph homomorphisms. Since associative spectra of undirected graphs were completely described in [16] (we state this result in Theorem 5.1), we consider $\mathrm{ns}(\mathbb{A}(G))$ and $\mathrm{sdist}(\mathbb{A}(G))$.

More precisely, we consider a variant of the semigroup distance that seems more relevant to our setting, and, admittedly, is easier to handle. We restrict our attention to the metric space of graph algebras of undirected graphs with a fixed vertex set V, and we measure associativity by the distance to the set $\mathrm{AssGr}(V)$ of associative graph algebras in this metric space.

Definition 2.14. For a finite nonempty set V, let $\mathrm{AssGr}(V)$ denote the set of all undirected graphs H with $V(H) = V$ such that $\mathbb{A}(H)$ is associative. For a finite undirected graph G with $V(G) = V$, we define $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ as follows:

$$\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = \min \left\{ \mathrm{dist}(\mathbb{A}(G), \mathbb{A}(H)) : H \in \mathrm{AssGr}(V) \right\}.$$

Clearly, $\mathrm{sdist}(\mathbb{A}(G)) \leq \mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$, and sometimes we have equality here (e.g., when $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$), but it may also happen that the inequality is strict (e.g., for $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$; see Section 5). Observe that deleting or adding an edge requires two changes in the operation table, while deleting or adding a loop requires only one change. Thus, if G and H are two graphs on the same finite vertex set, then

$$\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H)) = 2 \cdot |E(G) \triangle E(H)| + |L(G) \triangle L(H)|. \tag{1}$$

Tables 1, 2 and 3 in Appendix C show the results of a brute force computer exploration of graph algebras of undirected graphs of sizes 3, 4 and 5 (we leave the case of 2-vertex graphs as an exercise to the interested reader). Note that the size of the graph algebra is one more than the size of the graph, due to the external zero element. We will always refer to the size of the graph; in particular, n denotes the number of vertices of the graph under consideration throughout the paper (hence the corresponding graph algebra has $n+1$ elements).

We can conjecture from the tables that the range of $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ for n-vertex graphs is an interval. This is indeed the case: we prove in Theorem 3.5 that the possible values are $0, 1, \ldots, \lfloor n^2/2 \rfloor$, and we characterize graphs with the maximal value $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = \lfloor n^2/2 \rfloor$ in Theorem 3.6. The behavior of the index of nonassociativity seems more complicated. It will be clear that $\operatorname{ns}(\mathbb{A}(G))$ is always even (see Proposition 4.1), but we see "gaps" in the range of $\operatorname{ns}(\mathbb{A}(G))$ even if we disregard odd numbers (these missing values are indicated by gray color in tables 2 and 3). We find the maximal value of $\operatorname{ns}(\mathbb{A}(G))$ and the corresponding graphs in Theorem 4.6, and we describe some gaps at the top of the range in Corollary 4.8. In Theorem 4.10 we prove that the bottom of the range contains an interval of even numbers that is asymptotically as long as the whole range. This means that the "chaotic" part at the top is very small, but it remains an open problem to complete the description of the range of $\operatorname{ns}(\mathbb{A}(G))$.

3. Semigroup distance of graph algebras

Our first result is an upper estimate of $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ in terms of the number of edges. We prove this inequality in the next lemma, then we prove that the inequality is strict in some special cases (Lemma 3.2), and in Proposition 3.3 we characterize graphs for which the estimate is sharp.

Lemma 3.1. *For any finite undirected graph G, we have* $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \leq 2 \cdot |E(G)|$.

Proof. Consider the graph H such that $V(H) = V(G)$, $E(H) = \emptyset$ and $L(H) = L(G)$, i.e., H is obtained from G by deleting all edges (but keeping the loops). By Corollary 2.13, H has an associative graph algebra, since it has no nontrivial component. Clearly, $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H)) = 2 \cdot |E(G)|$, thus $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \leq 2 \cdot |E(G)|$. ∎

Lemma 3.2. *If G is a finite undirected graph that has a loop on a non-isolated vertex, then* $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) < 2 \cdot |E(G)|$.

Proof. Assume that $x \in L(G)$ and $xy \in E(G)$. We consider the graph H such that $V(H) = V(G)$, $E(H) = \{xy\}$ and $L(H) = L(G) \cup \{y\}$, i.e., H is obtained from G by deleting all edges except xy and adding a loop to y (if there was no loop there). By Corollary 2.13, H has an associative graph algebra, since its only nontrivial component is isomorphic to $K_2^{\circlearrowleft}$. According to (1), we have

$$\begin{aligned}\operatorname{dist}(\mathbb{A}(G),\mathbb{A}(H)) &= 2\cdot|E(G)\bigtriangleup E(H)|+|L(G)\bigtriangleup L(H)|\\ &\le 2\cdot(|E(G)|-1)+1\\ &= 2\cdot|E(G)|-1,\end{aligned}$$

thus $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) < 2\cdot|E(G)|$, as claimed. □

Proposition 3.3. *For any finite undirected graph G, we have*

$$\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \le 2\cdot|E(G)|,$$

and equality holds here if and only if G is triangle-free and loops appear only on isolated vertices.

Proof. We have already proved the inequality $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \le 2\cdot|E(G)|$ in Lemma 3.1, so we only need to describe the graphs for which we have equality. Isolated vertices (with or without loops) do not influence the value of $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$, and they do not count into $|E(G)|$ either. Therefore, we can assume without loss of generality that G has no isolated vertices. Moreover, by Lemma 3.2, we may also suppose that G is irreflexive.

Assume first that G contains a triangle xyz. Let us construct the graph H such that $V(H)=V(G)$, $E(H)=\{xy,xz,yz\}$ and $L(H)=\{x,y,z\}$, i.e., H is obtained from G by deleting all edges except the three edges of the triangle xyz and adding a loop to x, y and z (if there was no loop there). By Corollary 2.13, H has an associative graph algebra, since its only nontrivial component is isomorphic to $K_3^{\circlearrowleft}$. According to (1), we have

$$\begin{aligned}\operatorname{dist}(\mathbb{A}(G),\mathbb{A}(H)) &= 2\cdot|E(G)\bigtriangleup E(H)|+|L(G)\bigtriangleup L(H)|\\ &\le 2\cdot(|E(G)|-3)+3\\ &= 2\cdot|E(G)|-3,\end{aligned}$$

thus $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) < 2\cdot|E(G)|$, as claimed.

It remains to prove that if G is an irreflexive triangle-free graph without isolated vertices, then $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \ge 2\cdot|E(G)|$. Let $H\in\operatorname{AssGr}(V)$, where $V=V(G)$, and let us verify that $\operatorname{dist}(\mathbb{A}(G),\mathbb{A}(H)) \ge 2\cdot|E(G)|$. By Corollary 2.13, all nontrivial connected components of H are reflexive complete graphs. Denote the vertex sets of the nontrivial components of H by V_i $(i=1,\dots,k)$ and let V_0 denote the set of isolated vertices of H (we can assume without loss of generality that H has no loops on isolated vertices). Of course, it may happen that H has no nontrivial connected components; in that case we have $k=0$ and $V_0=V$. Let e_i $(i=1,\dots,k)$ denote the number of edges in $G|_{V_i}$ and let e_0 denote the number of the remaining edges in G. Note that e_0 counts the edges of G within V_0 as well as the edges across the sets V_i $(i=0,1,\dots,k)$.

This time it will be easier to compute $\operatorname{dist}(\mathbb{A}(G),\mathbb{A}(H))$ directly by considering the operation tables of $\mathbb{A}(G)$ and $\mathbb{A}(H)$, instead of using (1). The two operation tables differ at $|V_i|^2-2e_i$ many places in $V_i\times V_i$ for $i=1,\dots,k$ and at $2e_0$ many places outside the set $\bigcup_{i=1}^k V_i\times V_i$. Therefore, we have $\operatorname{dist}(\mathbb{A}(G),\mathbb{A}(H)) = 2e_0+\sum_{i=1}^k(|V_i|^2-2e_i)$. Since $G|_{V_i}$ $(i=1,\dots,k)$ is an irreflexive triangle-free graph, $e_i\le|V_i|^2/4$ holds by Theorem 2.10. This implies $|V_i|^2-2e_i\ge 2e_i$ for

$i = 1, \ldots, k$, hence

$$\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H)) = 2e_0 + \sum_{i=1}^{k} (|V_i|^2 - 2e_i) \geq 2e_0 + \sum_{i=1}^{k} 2e_i = 2|E(G)|,$$

and this completes the proof. □

Before stating and proving the main results of this section that describe possible values of $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ (Theorem 3.5) and the graphs attaining the maximal value (Theorem 3.6), we need to look at what happens to $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ if we add a loop to a vertex.

Lemma 3.4. *Let G be a finite undirected graph with $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = m$, and assume that there is no loop on the vertex $v \in V(G)$. Let G^* be the graph obtained from G by adding a loop to v. Then $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G^*)) \in \{m, m-1\}$.*

Proof. Clearly $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(G^*)) = 1$, thus the triangle inequality implies that $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G^*)) \geq m - 1$. In order to prove $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \leq m$, let us consider $H \in \operatorname{AssGr}(V)$, with $V = V(G)$ and $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H)) = m$. If $v \in L(H)$ then $\operatorname{dist}(\mathbb{A}(G^*), \mathbb{A}(H)) = m - 1 \leq m$. If $v \notin L(H)$ then, by Corollary 2.13, v is an isolated vertex in H and the graph H^* obtained from H by adding a loop to the vertex v also belongs to $\operatorname{AssGr}(V)$; furthermore, $\operatorname{dist}(\mathbb{A}(G^*), \mathbb{A}(H^*)) = m$ in this case. This proves that in both cases there is an associative graph algebra of distance at most m from $\mathbb{A}(G)$, thus $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \leq m$, as claimed. □

Let $\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}})$ denote the range of $\operatorname{sdist}_{\mathrm{gr}}$ on graph algebras of n-vertex undirected graphs:

$$\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}}) = \big\{\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) : G \text{ is an undirected graph on } n \text{ vertices}\big\}.$$

Theorem 3.5. *For any positive integer n, we have*

$$\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}}) = \Big\{0, 1, \ldots, \lfloor n^2/2 \rfloor \Big\}.$$

Proof. Let us consider bipartite graphs G with color classes A and B, where $|A| = \lfloor n/2 \rfloor$, $|B| = \lceil n/2 \rceil$. The number of edges of such a graph can be any number between 0 and $\lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$. Since bipartite graphs are triangle-free, we have $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 2 \cdot |E(G)|$ by Proposition 3.3, thus $\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}})$ contains all even numbers between 0 and $2 \cdot \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil = \lfloor n^2/2 \rfloor$.

To obtain the odd numbers in this interval, let G be any one of the graphs considered above that has a non-isolated vertex x. (The latter assumption exludes only the empty graph.) If G^* is the graph obtained from G by adding a loop to the vertex x, then we have $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G^*)) = 2 \cdot |E(G)| - 1$ by lemmas 3.2 and 3.4. This proves that $\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}})$ contains all odd numbers between 1 and $\lfloor n^2/2 \rfloor - 1$.

It remains to prove that $\operatorname{range}_n(\operatorname{sdist}_{\mathrm{gr}})$ does not contain any number greater than $\lfloor n^2/2 \rfloor$. Let $V = V(G)$ have n elements, let H_0 be the empty graph on V (with no edges and no loops), and let H_1 be the reflexive complete graph on V (thus $H_1 \cong K_n^{\circlearrowleft}$). Clearly, $H_0, H_1 \in \operatorname{AssGr}(V)$ and $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H_0)) + \operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H_1)) = n^2$. This implies that the smaller one of $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H_0))$ and $\operatorname{dist}(\mathbb{A}(G), \mathbb{A}(H_1))$ is at most $\lfloor n^2/2 \rfloor$, hence $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) \leq \lfloor n^2/2 \rfloor$. □

Theorem 3.6. *For an arbitrary finite undirected graph G on n vertices, we have $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = \lfloor n^2/2 \rfloor$ if and only if $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$.*

Proof. It is clear from Proposition 3.3 that $\mathrm{sdist_{gr}}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil}) = 2 \cdot \lfloor n/2\rfloor \cdot \lceil n/2\rceil = \lfloor n^2/2\rfloor$. Conversely, assume that $V = V(G)$ has n elements and $\mathrm{sdist_{gr}}(\mathbb{A}(G)) = \lfloor n^2/2\rfloor$, and let us prove that $G \cong K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$. We separate two cases on whether G contains loops or not.

Case 1: There are no loops in G (i.e., G is irreflexive). Let H_0 be the empty graph on V (with no edges and no loops), and let H_1 be the reflexive complete graph on V (thus $H_1 \cong K_n^{\circlearrowleft}$). Then we have

$$\mathrm{dist}(\mathbb{A}(G), \mathbb{A}(H_0)) = 2 \cdot |E(G)| \geq \mathrm{sdist_{gr}}(\mathbb{A}(G)) = \lfloor n^2/2\rfloor$$
$$\mathrm{dist}(\mathbb{A}(G), \mathbb{A}(H_1)) = n^2 - 2 \cdot |E(G)| \geq \mathrm{sdist_{gr}}(\mathbb{A}(G)) = \lfloor n^2/2\rfloor .$$

From these two inequalities we get $\lfloor n^2/2\rfloor \leq 2 \cdot |E(G)| \leq \lceil n^2/2\rceil$. If n is odd, then $\lceil n^2/2\rceil$ is an odd number, hence $2 \cdot |E(G)| = \lceil n^2/2\rceil$ is impossible. So we have $2 \cdot |E(G)| = \lfloor n^2/2\rfloor$, and this is certainly true also if n is even. Therefore, $2 \cdot |E(G)| = \mathrm{sdist_{gr}}(\mathbb{A}(G))$, and then G is triangle-free, according to Proposition 3.3. Furthermore, G is irreflexive and $|E(G)| = 1/2 \cdot \lfloor n^2/2\rfloor = \lfloor n^2/4\rfloor$, thus we can use Theorem 2.10 to conclude that $G \cong K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$.

Case 2: There is at least one loop in G. By Lemma 3.4, $\mathrm{sdist_{gr}}(\mathbb{A}(G))$ could only increase if we remove loops, but by Theorem 3.5, $\mathrm{sdist_{gr}}(\mathbb{A}(G)) = \lfloor n^2/2\rfloor$ is the maximal element of $\mathrm{range}_n(\mathrm{sdist_{gr}})$. This means that $\mathrm{sdist_{gr}}(\mathbb{A}(\hat{G})) = \lfloor n^2/2\rfloor$ holds for the graph $\hat{G}$ that we obtain from G by removing all loops. Now $\hat{G} \cong K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ follows by Case 1, hence $E(\hat{G}) = \lfloor n/2\rfloor \cdot \lceil n/2\rceil$. Since G contains a loop but has no isolated vertices, we can apply Lemma 3.2:

$$\mathrm{sdist_{gr}}(\mathbb{A}(G)) < 2 \cdot |E(G)| = 2 \cdot |E(\hat{G})| = 2 \cdot \lfloor n/2\rfloor \cdot \lceil n/2\rceil = \lfloor n^2/2\rfloor .$$

However, this contradicts our assumption $\mathrm{sdist_{gr}}(\mathbb{A}(G)) = \lfloor n^2/2\rfloor$, so Case 2 is actually not possible. □

4. Index of nonassociativity of graph algebras

As the next proposition shows, the index of nonassociativity is much easier to compute than $\mathrm{sdist_{gr}}(\mathbb{A}(G))$: we just need to count cherries and edges. We also see that $\mathrm{ns}(\mathbb{A}(G))$ is always an even number for undirected graphs.

Proposition 4.1. *For any finite undirected graph G, we have*

$$\mathrm{ns}(\mathbb{A}(G)) = 4 \cdot |Ch(G)| + 4 \cdot |E_0(G)| + 2 \cdot |E_1(G)|.$$

Proof. Let uvw be a cherry, say, with $uv, uw \in E(G)$ and $vw \notin E(G)$. It is straightforward to verify that 4 of the 6 permutations of u, v, w give a nonassociative triple, namely (u, v, w), (u, w, v), (v, u, w) and (w, u, v). We claim that all nonassociative triples with pairwise distinct entries do arise this way from a cherry. Indeed, if (x, y, z) is a nonassociative triple, then either $(xy)z = 0 \neq x = x(yz)$ or $(xy)z = x \neq 0 = x(yz)$. The first case holds if and only if $xy \in E(G)$, $yz \in E(G)$, $xz \notin E(G)$, while the second case is equivalent to $xy \in E(G)$, $yz \notin E(G)$, $xz \in E(G)$. We see that in both cases xyz is a cherry, provided that x, y, z are pairwise distinct. Thus the number of nonassociative triples consisting of three different entries is $4 \cdot |Ch(G)|$.

Now if u and v are two different vertices and $uv \notin E(G)$ or $uv \in E_2(G)$, then all triples formed from u and v are associative. If $uv \in E_0(G)$, then we get the 4 nonassociative triples (u, v, v), (u, v, u), (v, u, u) and (v, u, v). If $uv \in E_1(G)$, say, with $u \in L(G)$ and $v \notin L(G)$, then we get only 2 nonassociative triples, namely (u, v, v) and (v, u, v). Thus the number of nonassociative triples consisting of two different entries is $4 \cdot |E_0(G)| + 2 \cdot |E_1(G)|$. To finish the proof, we only need to note that triples of the form (u, u, u) are associative (as well as those that contain the zero element). □

Proposition 4.2. *For any finite undirected graph G, we have*

$$\mathrm{ns}(\mathbb{A}(G)) \le 2 \cdot \sum_{v \in V(G)} d(v)^2,$$

and equality holds here if and only if G is triangle-free and loops appear only on isolated vertices.

Proof. Isolated vertices (with or without loops) do not influence the index of nonassociativity, and they do not count into vertex degrees either. Therefore, we can assume without loss of generality that G has no isolated vertices. By Proposition 4.1, $\mathrm{ns}(\mathbb{A}(G)) \le 4 \cdot |Ch(G)| + 4 \cdot |E(G)|$, with equality if and only if G is irreflexive. Each vertex v is the "top" vertex of at most $\binom{d(v)}{2}$ cherries, and if v is not contained in any triangle, then (and only then), v is the "top" vertex of exactly $\binom{d(v)}{2}$ cherries. Thus we can estimate $\mathrm{ns}(\mathbb{A}(G))$ as follows:

$$\begin{aligned}\mathrm{ns}(\mathbb{A}(G)) &\le 4 \cdot |Ch(G)| + 4 \cdot |E(G)| \\ &\le 4 \cdot \sum_{v \in V(G)} \binom{d(v)}{2} + 4 \cdot |E(G)| \\ &= 2 \cdot \sum_{v \in V(G)} d(v)^2 - 2 \cdot \sum_{v \in V(G)} d(v) + 4 \cdot |E(G)| \\ &= 2 \cdot \sum_{v \in V(G)} d(v)^2\end{aligned}$$

(in the last step we used the well-known fact that the sum of the degrees is twice the number of edges). It is clear from the arguments above that this estimate is sharp if and only if G contains neither loops nor triangles. □

Example 4.3. It follows immediately from the proposition above that for the complete bipartite graph $K_{a,b}$, we have $\mathrm{ns}(\mathbb{A}(K_{a,b})) = 2ab^2 + 2a^2b = 2(a+b)ab$.

In the following we take a route similar to Section 3: we estimate $\mathrm{ns}(\mathbb{A}(G))$ by the number of edges (Proposition 4.4), then we see what happens to $\mathrm{ns}(\mathbb{A}(G))$ when we add a loop (Lemma 4.5), and we use these to prove our main results about the range of $\mathrm{ns}(\mathbb{A}(G))$ (Theorem 4.6, Corollary 4.8 and Theorem 4.10).

Proposition 4.4. *For any finite undirected graph G with n vertices, we have*

$$\mathrm{ns}(\mathbb{A}(G)) \le 2n \cdot |E(G)|,$$

and equality holds here if and only if G is either a complete bipartite graph or G has no edges.

Proof. Example 4.3 shows that $\mathrm{ns}(\mathbb{A}(G)) = 2n \cdot |E(G)|$ for complete bipartite graphs, and this is trivially true for "edgeless" graphs, too.

Now assume that G has at least one edge, and for any $e \in E(G)$, let $ch(e)$ denote the number of cherries that contain the edge e. Clearly, we have $ch(e) \le (n-2)$ for each $e \in E(G)$; furthermore, $\sum_{e \in E(G)} ch(e) = 2 \cdot |Ch(G)|$, as each cherry contains exactly two edges. Using these observations together with Proposition 4.1, we can estimate $\mathrm{ns}(\mathbb{A}(G))$ as follows:

$$\begin{aligned}\mathrm{ns}(\mathbb{A}(G)) &\leq 4 \cdot |Ch(G)| + 4 \cdot |E(G)| \\ &= 2 \cdot \sum_{e \in E(G)} ch(e) + 4 \cdot |E(G)| \\ &\leq 2(n-2) \cdot |E(G)| + 4 \cdot |E(G)| \\ &= 2n \cdot |E(G)|.\end{aligned}$$

The inequalities above turn into equalities if and only if the endpoints of each edge are loopless, and every edge is contained in $n-2$ cherries, i.e.,

(2) $\forall xy \in E(G)\ \forall v \in V(G) \setminus \{x, y\} \colon xv \in E(G)$ or $yv \in E(G)$, but not both.

Let us assume that G is such a graph, and suppose for contradiction that G contains a cycle C of odd length. Any *chord* of C (i.e., an edge connecting two non-consecutive vertices of C) cuts C into two shorter cycles, one of which has odd length. Therefore, if we choose C to be of minimal odd length ℓ, then C is a chordless cycle. However, the existence of a chordless cycle of length ℓ contradicts (2), unless $\ell = 4$. This proves that G does not contain any cycles of odd length, hence G is bipartite. Let A and B be the two color classes, and let $xy \in E(G)$ with $x \in A$ and $y \in B$. If v is any vertex from A, then (2) gives that $yv \in E(G)$, whereas if v belongs to B, then we have $xv \in E(G)$. Since this is true for every edge xy of G, we can conclude that G is a complete bipartite graph. □

Lemma 4.5. *Let G be a finite undirected graph, let v be a loopless vertex of G, and let G^* denote the graph obtained from G by adding a loop to v. Then we have*

$$\mathrm{ns}(\mathbb{A}(G^*)) = \mathrm{ns}(\mathbb{A}(G)) - 2 \cdot d(v).$$

Proof. Let p denote the number of loopless neighbors of v, and let q denote the number of neighbors of v that have a loop. Then we have

$$|Ch(G^*)| = |Ch(G)|, \quad |E_0(G^*)| = |E_0(G)| - p, \quad |E_1(G^*)| = |E_1(G)| + p - q.$$

Now we can compute $\mathrm{ns}(\mathbb{A}(G^*))$ with the help of Proposition 4.1:

$$\begin{aligned}\mathrm{ns}(\mathbb{A}(G^*)) &= 4 \cdot |Ch(G)| + 4 \cdot (|E_0(G)| - p) + 2 \cdot (|E_1(G)| + p - q) \\ &= \mathrm{ns}(\mathbb{A}(G)) - 2(p+q) \\ &= \mathrm{ns}(\mathbb{A}(G)) - 2d(v).\end{aligned}$$

□

Theorem 4.6. *For any finite irreflexive undirected graph G on n vertices, the following hold:*

(i) *if $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, then $\mathrm{ns}(\mathbb{A}(G)) = 2n \lfloor n/2 \rfloor \lceil n/2 \rceil = n \lfloor n^2/2 \rfloor$;*

(ii) *if $G \ncong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, then $\mathrm{ns}(\mathbb{A}(G)) \leq 2n \lfloor n/2 \rfloor \lceil n/2 \rceil - 4 \lfloor n/2 \rfloor + 4$.*

Proof. The proof is quite long and technical, so we present it separately in Appendix A. □

Remark 4.7. Let G be the graph obtained from $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ by connecting two vertices in the color class of size $\lfloor n/2 \rfloor$ by an edge. Then we have $\mathrm{ns}(\mathbb{A}(G)) = 2n \lfloor n/2 \rfloor \lceil n/2 \rceil - 4 \lfloor n/2 \rfloor + 4$, and this shows that the estimate in item ((ii)) of Theorem 4.6 cannot be improved.

Let $\mathrm{range}_n(\mathrm{ns})$ denote the range of ns on graph algebras of n-vertex undirected graphs:

$$\mathrm{range}_n(\mathrm{ns}) = \bigl\{\mathrm{ns}(\mathbb{A}(G)) : G \text{ is an undirected graph on } n \text{ vertices}\bigr\}.$$

Corollary 4.8. *Let n be a positive integer, and assume that $n \geq 8$.*

(i) *If n is even, then the three largest elements of* $\mathrm{range}_n(\mathrm{ns})$ *are*

$$\frac{n^3}{2} - (2n-4), \quad \frac{n^3}{2} - n, \quad \frac{n^3}{2}.$$

(ii) *If n is odd, then the four largest elements of* $\mathrm{range}_n(\mathrm{ns})$ *are*

$$\frac{n^3-n}{2} - (2n-6), \quad \frac{n^3-n}{2} - (n+1), \quad \frac{n^3-n}{2} - (n-1), \quad \frac{n^3-n}{2}.$$

Proof. The corollary can be derived from Theorem 4.6 and Lemma 4.5 as follows. Let G be a graph on n vertices, and let $\hat{G}$ denote the graph that we obtain from G by deleting all loops. If $\hat{G} \not\cong K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$, then

$$\mathrm{ns}(\mathbb{A}(G)) = \mathrm{ns}(\mathbb{A}(\hat{G})) - \sum_{v\in L(G)} 2d(v) \leq 2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil - 4\left\lfloor\frac{n}{2}\right\rfloor + 4,$$

and we can have equality here, for example, if G is the graph described in Remark 4.7. If $\hat{G} \cong K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$, then

$$\mathrm{ns}(\mathbb{A}(G)) = \mathrm{ns}(\mathbb{A}(K_{\lfloor n/2\rfloor,\lceil n/2\rceil})) - \sum_{v\in L(G)} 2d(v) = 2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil - \sum_{v\in L(G)} 2d(v).$$

This number can be larger than $2n\lfloor n/2\rfloor\lceil n/2\rceil - 4\lfloor n/2\rfloor + 4$ only if G has at most one loop, and then the subtrahend $\sum_{v\in L(G)} 2d(v)$ is either zero, $2\lfloor n/2\rfloor$ or $2\lceil n/2\rceil$. □

Remark 4.9. Corollary 4.8 is valid also for $n < 8$, in the sense that the largest elements of $\mathrm{range}_n(\mathrm{ns})$ are

$$2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil-4\left\lfloor\frac{n}{2}\right\rfloor+4,\ 2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil-2\left\lceil\frac{n}{2}\right\rceil,\ 2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil-2\left\lfloor\frac{n}{2}\right\rfloor,\ 2n\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil,$$

but some of these numbers might coincide (even for odd n), and their order might be different.

We have seen that there are some gaps close to the top of $\mathrm{range}_n(\mathrm{ns})$; however, the bottom of $\mathrm{range}_n(\mathrm{ns})$ contains a long sequence of consecutive even numbers (asymptotically as long as the whole range).

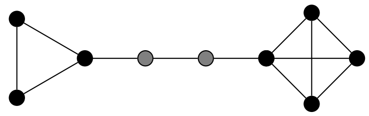
Theorem 4.10. *Let r_n be the greatest even integer such that all even numbers up to r_n belong to* $\mathrm{range}_n(\mathrm{ns})$. *Then we have*

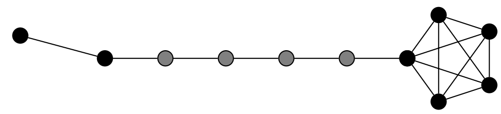
$$\lim_{n\to\infty}\frac{r_n}{n^3} = \frac{1}{2}.$$

Proof. The proof is quite long and technical, so we present it separately in Appendix B. □

It is easy to determine graphs with the least possible nonzero index of nonassociativity (we will do this in Section 5, but we invite the reader to do this on their own now), and most of these graphs are not connected. Connected graphs are much more interesting in this respect, so we conclude this section by a characterization of the "most associative" connected graphs.

Let $K_a \overset{\ell}{\text{——}} K_b$ denote the graph that is constructed by connecting a vertex of K_a and a vertex of K_b by a path of length ℓ. Here, by the length of a path we mean the number of edges in the path, i.e., $K_a \overset{\ell}{\text{——}} K_b$ has $a + b + \ell - 1$ vertices. As an illustration, let us draw $K_3 \overset{3}{\text{——}} K_4$ and $K_2 \overset{5}{\text{——}} K_5$ (vertices of the connecting paths are colored gray):





Clearly, $K_a \overset{\ell}{\text{——}} K_b \cong K_b \overset{\ell}{\text{——}} K_a$; furthermore, the graphs $K_2 \overset{\ell}{\text{——}} K_2$, $K_2 \overset{\ell+1}{\text{——}} K_1$ and $K_1 \overset{\ell+2}{\text{——}} K_1$ are isomorphic, as all of them are paths of length $\ell+2$. Therefore, in the following we will always assume that $a \geq 2$ and $b \geq 2$ when we consider the graphs $K_a \overset{\ell}{\text{——}} K_b$. This way the path of length 2 cannot be written in the form $K_a \overset{\ell}{\text{——}} K_b$, but it is included in the second item of the proposition below as $K_3 - \rotatebox{90}{$\bullet\!\!-\!\!\bullet$}$.

Proposition 4.11. *For any finite connected irreflexive undirected graph G on $n \geq 3$ vertices, the following hold:*

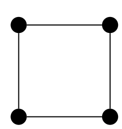
(i) *if $G \cong K_n$, then $|Ch(G)| = 0$;*

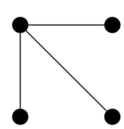
(ii) *if $G \cong K_n - \rotatebox{90}{$\bullet\!\!-\!\!\bullet$}$, then $|Ch(G)| = n - 2$;*

(iii) *if $G \cong K_r \overset{\ell}{\text{——}} K_s$ for some $r, s \geq 2$, $\ell \geq 1$ with $r + s + \ell - 1 = n$, then $|Ch(G)| = n - 2$;*

(iv) *if G is not isomorphic to any of the above mentioned graphs, then $|Ch(G)| > n - 2$.*

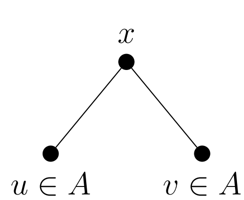
Proof. The first three statements of the proposition are straightforward to verify. We prove (iv) by induction on n. The case $n = 3$ is void, and for $n = 4$ there are only two graphs (up to isomorphism) that are relevant to (iv), and both of them indeed have more than $n - 2 = 2$ cherries:



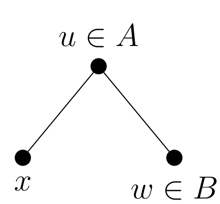


From now on we consider the case $n \geq 5$. If every vertex of G has degree 1 or $n - 1$, then G is isomorphic either to the complete graph K_n or to the star $K_{1,n-1}$. In the latter case $|Ch(G)| = \binom{n-1}{2} > n - 2$, since $n \geq 5$.

Therefore, we may assume that there is a vertex x with $2 \leq d(x) \leq n - 2$. Setting $A = N(x)$ and $B = V(G) \setminus (A \cup \{x\})$, we have $|A| = d(x) \geq 2$ and $|B| = n - 1 - |A| \geq 1$. There are two types of cherries in G that contain the vertex x:



(a)



(b)

For type (a) we need that $uv \notin E(G)$, thus the number of cherries of this type is the number of non-edges in the induced subgraph $G|_A$; we will denote this number by $n_A = \binom{|A|}{2} - |E(G|_A)|$. Since $xw \notin E(G)$ for all $w \in B$ (by the very definition of the set B), the number of cherries of type (b) equals the number of edges uw with $u \in A$ and $w \in B$, which we will denote by e_{AB}. Thus we have the following relationship between the number of cherries in G and in $G \setminus \{x\}$:

$$|Ch(G)| = |Ch(G \setminus \{x\})| + n_A + e_{AB}. \tag{3}$$

It is possible that $n_A = 0$, but the connectedness of G guarantees that $e_{AB} \geq 1$. We discuss four cases for $G \setminus \{x\}$ corresponding to ((i)), ((ii)), ((iii)) and ((iv)).

If $G \setminus \{x\} \cong K_{n-1}$, then $|Ch(G \setminus \{x\})| = 0$, $n_A = 0$ and $e_{AB} = |A| \cdot |B|$. The case $|A| = n-2$ is not possible, since then G would be isomorphic to $K_n - \updownarrow$. Therefore, $2 \leq |A| \leq n-3$, which together with $|A| + |B| = n-1$ implies $|A| \cdot |B| \geq 2 \cdot (n-3) > n-2$. Now (3) gives $|Ch(G)| = 0 + 0 + |A| \cdot |B| > n-2$, and this is what we had to prove.

If $G \setminus \{x\} \cong K_{n-1} - \updownarrow$, then $|Ch(G \setminus \{x\})| = n-3$, $n_A \in \{0,1\}$ and $e_{AB} \geq 2$. (The last inequality follows from the fact that for $n \geq 5$ it is not possible to divide the vertices of $K_{n-1} - \updownarrow$ into two parts in such a way that only one edge goes across the two parts.) By (3), we have $|Ch(G)| \geq (n-3) + 0 + 2 > n-2$, as required.

If $G \setminus \{x\} \cong K_r \overset{\ell}{\text{---}} K_s$, then $|Ch(G \setminus \{x\})| = n-3$, $n_A \geq 0$ and $e_{AB} \geq 1$, hence (3) yields $|Ch(G)| \geq (n-3) + 0 + 1 \geq n-2$. We have equality here if and only if $n_A = 0$ and $e_{AB} = 1$. This means that A induces a complete subgraph of $G \setminus \{x\}$ that can be separated from the rest of $G \setminus \{x\}$ by the removal of a single edge, which is possible only if A corresponds to K_r or K_s at the isomorphism $G \setminus \{x\} \cong K_r \overset{\ell}{\text{---}} K_s$. Since G is obtained from $G \setminus \{x\}$ by connecting the new vertex x to each element of A, either $G \cong K_{r+1} \overset{\ell}{\text{---}} K_s$ or $G \cong K_r \overset{\ell}{\text{---}} K_{s+1}$, contradicting the assumption of ((iv)). This proves that $|Ch(G)| > n-2$.

In the remaining cases the induction hypothesis gives that $|Ch(G \setminus \{x\})| > n-3$, hence $|Ch(G)| > (n-3) + 0 + 1 = n-2$ follows by (3), as $n_A \geq 0$ and $e_{AB} \geq 1$. □

Theorem 4.12. *For any finite connected undirected graph G on $n \geq 3$ vertices, the following hold:*

- (i/a) *if $G \cong K_n^{\circlearrowleft}$, then* $\mathrm{ns}(\mathbb{A}(G)) = 0$*;*
- (i/b) *if $G \cong K_n^{\circlearrowleft} - \circlearrowleft$, then* $\mathrm{ns}(\mathbb{A}(G)) = 2(n-1)$*;*
- (ii) *if $G \cong K_n^{\circlearrowleft} - \updownarrow$, then* $\mathrm{ns}(\mathbb{A}(G)) = 4(n-2)$*;*
- (iii) *if $G \cong (K_r \overset{\ell}{\text{---}} K_s)^{\circlearrowleft}$ for some $r, s \geq 2$, $\ell \geq 1$ with $r + s + \ell - 1 = n$, then* $\mathrm{ns}(\mathbb{A}(G)) = 4(n-2)$*;*
- (iv) *if G is not isomorphic to any of the above mentioned graphs, then* $\mathrm{ns}(\mathbb{A}(G)) > 4(n-2)$.

Proof. The first four statements follow from propositions 4.1 and 4.11. In order to prove ((iv)), consider the graph $\hat{G}$ that is obtained from G by removing all loops.

If $\hat{G} \cong K_n$, then Proposition 4.1 shows that $\mathrm{ns}(\mathbb{A}(G)) = 2(n-1) \cdot (n - |L(G)|)$. Here we must have $|L(G)| \leq n-2$ (otherwise we are in case (i/a) or (i/b)), hence $\mathrm{ns}(\mathbb{A}(G)) \geq 2(n-1) \cdot 2 > 4(n-2)$.

If $\hat{G} \cong K_n - \overset{\bullet}{\underset{\bullet}{|}}$ or $\hat{G} \cong K_r \overset{\ell}{\text{——}} K_s$, then G must have at least one loopless vertex (otherwise we are in case ((ii)) or ((iii))); therefore, $\mathrm{ns}(\mathbb{A}(G)) > 4 \cdot |Ch(G)|$ according to Proposition 4.1, and the latter equals $n-2$ by Proposition 4.11. This proves that $\mathrm{ns}(\mathbb{A}(G)) > 4(n-2)$.

If G is not isomorphic to any of the above mentioned graphs, then propositions 4.1 and 4.11 give $\mathrm{ns}(\mathbb{A}(G)) \geq 4 \cdot |Ch(G)| > 4(n-2)$. □

5. Conclusion

In this section we compare the extremal cases for the three measures of associativity ($\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$, $\mathrm{ns}(\mathbb{A}(G))$ and $\mathrm{spec}(\mathbb{A}(G))$) for undirected graphs. Before doing so, let us complement the results of the previous two sections by the following description of associative spectra of graph algebras of undirected graphs (the first item in the theorem is just a repetition of Corollary 2.13).

Theorem 5.1 ([16]). *Let G be an undirected graph.*

(i) *If every nontrivial connected component of G is a reflexive complete graph, then $s_n(\mathbb{A}(G)) = 1$ for all $n \in \mathbb{N}$.*

(ii) *If every nontrivial connected component of G is either a reflexive complete graph or a complete bipartite graph, and the last case occurs at least once, then $s_n(\mathbb{A}(G)) = 2^{n-2}$ for all $n \geq 2$.*

(iii) *Otherwise $s_n(\mathbb{A}(G)) = C_{n-1}$ for all $n \in \mathbb{N}$.*

By theorems 3.6 and 4.6, the only antiassociative graph algebra on n vertices is $\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil})$, if we measure associativity by $\mathrm{sdist}_{\mathrm{gr}}$ or by ns. However, $\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil})$ is not antiassociative with respect to the associative spectrum (in fact, it is almost associative!), according to the theorem above. On the other hand, most graph algebras are antiassociative in the "spectral" sense by Theorem 5.1, and the other two measures of associativity can be very small for these graphs (see below).

The least nonzero value of $\mathrm{sdist}_{\mathrm{gr}}$ is 1, and we have $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$ if and only if G is obtained from a graph $H \in \mathrm{AssGr}(V(G))$ by removing one loop, but G itself does not belong to $\mathrm{AssGr}(V(G))$. By Corollary 2.13, this means that one component of G is isomorphic to $K_r^{\circlearrowleft} - \circlearrowleft$ for some $r \geq 2$, and all other nontrivial components are reflexive complete graphs. For such a graph, we have $\mathrm{ns}(\mathbb{A}(G)) = 2(r-1)$ by Proposition 4.1. Therefore, $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$ implies $\mathrm{ns}(\mathbb{A}(G)) \leq 2(n-1)$ for graphs G with n vertices. (This explains the $(n-1)$ dots in the second column in tables 1, 2 and 3.) Theorem 5.1 shows that these graphs have a Catalan spectrum.

Since the index of nonassociativity of a graph algebra is always an even number, the least possible nonzero value is 2, and, by Proposition 4.1, we have $\mathrm{ns}(\mathbb{A}(G)) = 2$ if and only if $Ch(G) = E_0(G) = \emptyset$ and $|E_1(G)| = 1$. It is straightforward to verify that this happens if and only if one component of G is isomorphic to $K_2^{\circlearrowleft} - \circlearrowleft$ (an edge with a loop at one endpoint) and all other nontrivial components are reflexive complete graphs. If G is such a graph, then $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$, and G has a Catalan spectrum. (This explains why we only have one dot in the second row in tables 1, 2 and 3.)

We see that if $\mathbb{A}(G)$ is almost associative with respect to ns (i.e., $\mathrm{ns}(\mathbb{A}(G)) = 2$), then it is almost associative also with respect to $\mathrm{sdist}_{\mathrm{gr}}$ (i.e., $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$), but the converse implication is not true. However, if we restrict our attention to connected graphs with at least four vertices, then these two notions of almost associativity actually coincide. Indeed, for connected graphs with n vertices, the least nonzero value of $\mathrm{ns}(\mathbb{A}(G))$ is $2(n-1)$, and this value is attained only for $K_n^{\circlearrowleft} - \circlearrowleft$ if $n \geq 4$ (see Theorem 4.12). On the other hand, it follows from the discussion above that if G is connected, then $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$ holds also only if $G \cong K_n^{\circlearrowleft} - \circlearrowleft$.

We can conclude that for undirected graphs, $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ and $\mathrm{ns}(\mathbb{A}(G))$ match nicely (at least as far as antiassociativity and almost associativity are concerned), but the associative spectrum is quite unrelated to them. The following theorem, which is an analogue of Theorem 2.9 and Conjecture 5.3 below also show connections between $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ and $\mathrm{ns}(\mathbb{A}(G))$.

Theorem 5.2. *For every undirected graph G with n vertices, we have* $\mathrm{ns}(\mathbb{A}(G)) \leq 2(n-1) \cdot \mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$.

Proof. Let $H \in \mathrm{AssGr}(V(G))$ such that $\mathrm{dist}(\mathbb{A}(G), \mathbb{A}(H)) = \mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$. By (1), we have $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 2k + \ell$, where $k = |E(G) \triangle E(H)|$ and $\ell = |L(G) \triangle L(H)|$. We introduce some more notation to facilitate the proof: let $e_i = |E_i(G)|$ and $e_i' = |E_i(G) \cap E(H)|$ for $i = 0, 1$. We claim that

$$2e_0 + e_1 \leq 2k + 2e_0' + e_1'. \tag{4}$$

Indeed, by the definition of k, we have

$$\begin{aligned} 2k \geq 2 \cdot |E(G) \setminus E(H)| &\geq 2 \cdot |E_0(G) \setminus E(H)| + 2 \cdot |E_1(G) \setminus E(H)| \\ &= 2(e_0 - e_0') + 2(e_1 - e_1') \\ &\geq 2(e_0 - e_0') + (e_1 - e_1'), \end{aligned}$$

and this is equivalent to (4).

Let us consider the following sets for $i = 0, 1$:

$$\Theta_i = \big\{(xy, x) : xy \in E_i(G) \cap E(H) \text{ and } x \in L(G) \triangle L(H)\big\}.$$

If $xy \in E_0(G) \cap E(H)$, then xy is an edge of H, hence, by Corollary 2.13 both x and y must have a loop in H. Since these two vertices do not have a loop in G, both (xy, x) and (xy, y) belong to Θ_0; consequently, $|\Theta_0| = 2e_0'$. Similarly, if $xy \in E_1(G) \cap E(H)$, then exactly one of (xy, x) and (xy, y) belongs to Θ_1, hence $|\Theta_1| = e_1'$. On the other hand, for each $x \in L(G) \triangle L(H)$, there are at most $n-1$ vertices y such that $(xy, x) \in \Theta_0 \cup \Theta_1$, thus $|\Theta_0 \cup \Theta_1| \leq (n-1) \cdot \ell$. This implies that

$$2e_0' + e_1' = |\Theta_0| + |\Theta_1| = |\Theta_0 \cup \Theta_1| \leq (n-1) \cdot \ell$$

(note that Θ_0 and Θ_1 are disjoint). Comparing this with (4), we obtain that

$$2e_0 + e_1 \leq 2k + (n-1) \cdot \ell. \tag{5}$$

Next we consider the following set:

$$\Xi = \big\{(xyz, xy) : xyz \in Ch(G) \text{ and } xy \in E(G) \triangle E(H)\big\}.$$

Since H contains no cherries by Corollary 2.13, all cherries of G must be "destroyed", i.e., each cherry of G appears at least once as the first component of an element of Ξ, thus $|\Xi| \geq |Ch(G)|$. On the other hand, for each $xy \in E(G) \triangle E(H)$, there are at most $n-2$ vertices z such that $(xyz, xy) \in \Xi$, hence $|\Xi| \leq (n-2) \cdot k$. This implies that

$$|Ch(G)| \leq (n-2) \cdot k. \tag{6}$$

Now we can prove the desired inequality with the help of (5), (6) and Proposition 4.1:

$$\begin{aligned}\operatorname{ns}(\mathbb{A}(G)) &= 4 \cdot |Ch(G)| + 4e_0 + 2e_1 \\ &\leq 4(n-2) \cdot k + 4k + 2(n-1) \cdot \ell \\ &= 2(n-1) \cdot (2k + \ell) \\ &= 2(n-1) \cdot \operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)). \qquad \square\end{aligned}$$

Conjecture 5.3. For every undirected graph G with n vertices, we have $\operatorname{ns}(\mathbb{A}(G)) \geq 2 \cdot \operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$.

Remark 5.4. As we have already mentioned, we have $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = 1$ and $\operatorname{ns}(\mathbb{A}(G)) = 2(n-1)$ for $G = K_n^{\circlearrowleft} - \circlearrowleft$. This shows that the estimate in Theorem 5.2 cannot be improved. Similarly, if G is a disjoint union of r copies of $K_2^{\circlearrowleft} - \circlearrowleft$, then $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(G)) = r$ and $\operatorname{ns}(\mathbb{A}(G)) = 2r$, hence Conjecture 5.3 is also sharp (if it is true at all). This conjecture can clearly seen in Tables 1, 2 and 3: there are no dots above the main diagonal.

Finally, let us list some topics for further research:

1. Prove or disprove Conjecture 5.3.
2. Complete the description of the set $\operatorname{range}_n(\operatorname{ns})$ by determining all gaps in the range.
3. In Section 3 we considered $\operatorname{sdist}_{\mathrm{gr}}$ instead of the "real" semigroup distance. Determine the range of $\operatorname{sdist}(\mathbb{A}(G))$ for graph algebras of n-vertex undirected graphs and characterize graphs corresponding to the extremal cases. Let us mention here that $\operatorname{sdist}(\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil})) < \operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}))$ if $n \geq 3$. Indeed, let H be the directed graph that we obtain from $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ by making all edges one-way, pointing to the color class of size $\lfloor n/2 \rfloor$, and adding a loop to each vertex of this color class. By Theorem 2.12, $\mathbb{A}(H)$ is associative, and we have $\operatorname{dist}(\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}), \mathbb{A}(H)) = \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil + \lfloor n/2 \rfloor$, which is less than $\operatorname{sdist}_{\mathrm{gr}}(\mathbb{A}(K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil})) = 2 \cdot \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil$ for $n \geq 3$.
4. Investigate measures of associativity for graph algebras of directed graphs (see [17] for a study of associative spectra of these graph algebras).

Appendix A. Proof of Theorem 4.6

Graphs with the maximal number of cherries were detetmined in [21] (cherries are called open triangles in that paper). We will use the same line of reasoning to find the maxmal value of $\operatorname{ns}(\mathbb{A}(G))$. We need two technical lemmas: in Lemma A.1 we determine the maximum of a function that will appear in the proof of Theorem 4.6, and in Lemma A.2 we deal with a special case that would cause some trouble during the proof.

Lemma A.1. *Let us define a four-variable function f on real numbers by*

$$f(a,b,c,d) = (a+c+1)(a/2+b+d) + (b+c+1)(b/2+a+d) + 2c - 2d.$$

For a given $t \in \mathbb{N}$, let $\Sigma_t = \{(a,b,c,d) \in \mathbb{N}_0^4 : a+b+c+d = t\}$. The maximum of f on Σ_t is

$$M_t := \frac{1}{2}t^2 + \frac{3}{2}t + \left\lfloor \frac{t}{2} \right\rfloor \left\lceil \frac{t}{2} \right\rceil.$$

Depending on the parity of t, the maximal value is attained at one or three tuples.

(i) *If t is even, then $f(a,b,c,d) = M_t$ if and only if*

$$(a,b,c,d) = \left(\frac{t}{2}, \frac{t}{2}, 0, 0\right).$$

(ii) *If t is odd, then $f(a,b,c,d) = M_t$ if and only if (a,b,c,d) is one of the following three tuples:*

$$\left(\frac{t-1}{2}, \frac{t+1}{2}, 0, 0\right), \quad \left(\frac{t+1}{2}, \frac{t-1}{2}, 0, 0\right), \quad \left(\frac{t-1}{2}, \frac{t-1}{2}, 1, 0\right).$$

Proof. Let $(a,b,c,d) \in \Sigma_t$ with $c \geq 1$ and $d \geq 1$. Then $(a+1, b+1, c-1, d-1) \in \Sigma_t$ and it is straightforward to verify that

$$f(a+1, b+1, c-1, d-1) = f(a,b,c,d) + (a+b)/2 + c + 1 > f(a,b,c,d).$$

Therefore, we can increase the value of f by simultaneously increasing a and b by 1 while decreasing c and d by 1. Hence the maximum value of f can be attained only at tuples with $c = 0$ or $d = 0$ or both. We investigate the three cases $c > 0, d = 0$ and $c = 0, d > 0$ and $c = 0, d = 0$ separately.

1. If $c \geq 1$ and $d = 0$, then $(a+1, b, c-1, 0), (a, b+1, c-1, 0) \in \Sigma_t$ and

$$\begin{aligned}
f(a,b,c,0) &= (a+c+1)(a/2+b) + (b+c+1)(b/2+a) + 2c, \\
f(a+1,b,c-1,0) &= f(a,b,c,0) - a/2 + b/2 + 3c/2 - 3/2, \\
f(a,b+1,c-1,0) &= f(a,b,c,0) + a/2 - b/2 + 3c/2 - 3/2, \\
f(a+1,b,c-1,0) + f(a,b+1,c-1,0) &= 2f(a,b,c,0) + 3c - 3.
\end{aligned}$$

   If $c \geq 2$ then $3c-3 > 0$, thus the above calculations show that at least one of the inequalities $f(a+1, b, c-1, 0) > f(a,b,c,0)$ and $f(a, b+1, c-1, 0) > f(a,b,c,0)$ hold. This means that $f(a,b,c,0)$ cannot be maximal if $c \geq 2$. If $c = 1$, then we have

$$\begin{aligned}
f(a+1,b,0,0) &= f(a,b,1,0) - (a/2 + b/2), \\
f(a,b+1,0,0) &= f(a,b,1,0) + (a/2 - b/2).
\end{aligned}$$

   This implies that if $a \neq b$, then one of $f(a+1, b, 0, 0)$ and $f(a, b+1, 0, 0)$ is greater than $f(a,b,1,0)$, hence the latter cannot be the maximal value of f. We can conclude that in the case $c \geq 1, d = 0$, the maximum of f can be attained only at $((t-1)/2, (t-1)/2, 1, 0)$, which is possible only if t is odd.

2. If $c = 0$ and $d \geq 1$, then $(a+1, b, 0, d-1) \in \Sigma_t$ and

$$\begin{aligned}
f(a,b,0,d) &= (a+1)(a/2+b+d) + (b+1)(b/2+a+d) - 2d, \\
f(a+1,b,0,d-1) &= f(a,b,0,d) + b + d + 1 > f(a,b,c,d).
\end{aligned}$$

   Therefore, we can increase the value of f by simultaneously increasing a by 1 and decreasing d by 1, hence cannot get the maximal value of f in this case.

3. Finally we assume that $c = 0$ and $d = 0$. Then $(a, b, 0, 0) \in \Sigma_t$ implies $a + b = t$, hence
$$f(a, b, 0, 0) = (a + 1)(a/2 + b) + (b + 1)(b/2 + a) = \frac{1}{2}t^2 + \frac{3}{2}t + ab.$$
Since t is constant on Σ_t, in this case the maximum of f can be attained only if $a = \lfloor (t-1)/2 \rfloor$, $b = \lceil (t-1)/2 \rceil$ or $a = \lceil (t-1)/2 \rceil$, $b = \lfloor (t-1)/2 \rfloor$.

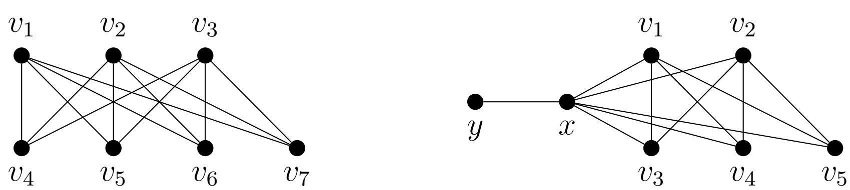
We have proved that f can take its maximal value only at the tuples listed in items ((i)) and ((ii)) in the statement of the lemma. To finish the proof, one only needs to verify that f takes the same value at these tuples, and this common value is
$$M_t = \frac{1}{2}t^2 + \frac{3}{2}t + \left\lfloor \frac{t}{2} \right\rfloor \left\lceil \frac{t}{2} \right\rceil .$$
□

Lemma A.2. *Let G be a finite irreflexive undirected graph on $n \geq 7$ vertices, and for any edge $e \in E(G)$, let $\Delta(e)$ denote the number of triangles containing e. Let $m = \min\{\Delta(e) : e \in E(G)\}$, and assume that for every edge $e = xy \in E(G)$ with $\Delta(e) = m$, we have $G \setminus \{x, y\} \cong K_{\lfloor (n-2)/2 \rfloor, \lceil (n-2)/2 \rceil}$. Then one of the following holds:*

(i) *$m = 0$ and G is isomorphic to $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, or*

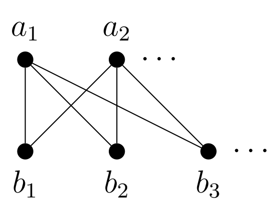
(ii) *$m = 1$ and G has the following structure:*
- *$V(G) = \{v_1, \ldots, v_{n-2}, x, y\}$;*
- *the subgraph induced on $\{v_1, \ldots, v_{n-2}\}$ is isomorphic to $K_{\lfloor (n-2)/2 \rfloor, \lceil (n-2)/2 \rceil}$;*
- *x is connected by an edge to all other vertices, whereas y is connected only to x.*

Example A.3. Here are the two graphs considered in the lemma above for the case $n = 7$:



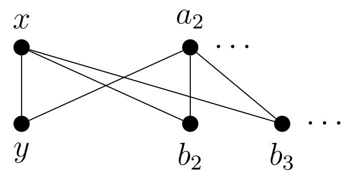
Proof. Let $xy \in E(G)$ be an arbitrary edge such that $\Delta(xy) = m$. Then $G \setminus \{x, y\}$ is a balanced complete bipartite graph with at least five vertices:

$G \setminus \{x, y\}$:

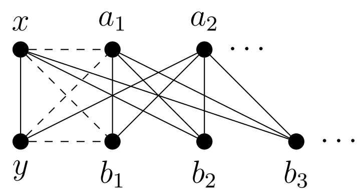


Since $G \setminus \{x, y\}$ has no triangles, a triangle containing an edge of the form $a_i b_j$ must have either x or y as the third vertex, hence $\Delta(a_i b_j) \leq 2$, thus $m \leq 2$. We separate two cases on whether there is an edge $a_i b_j$ with $\Delta(a_i b_j) = m$.

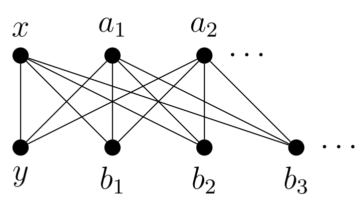
Case 1: We have $\Delta(a_i b_j) = m$ for some i and j, say, $\Delta(a_1 b_1) = m$. Then $G \setminus \{a_1, b_1\}$ is a balanced complete bipartite graph (we may also assume without loss of generality that x is a "upper" vertex and y is an "lower" vertex in this bipartite graph):



Let us summarize what we know about our graph:

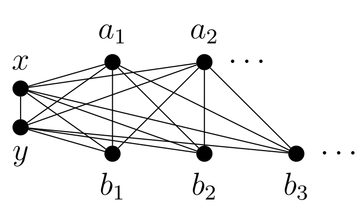


The dashed lines indicate that so far we do not know if xa_1, xb_1, ya_1, yb_1 are edges in G. However, despite the lacking information, we can conclude that the edge a_2b_2 is not included in any triangle, hence $\Delta(a_2b_2) = 0 = m$. This implies that $G \setminus \{a_2, b_2\}$ is a balanced complete bipartite graph, thus $xb_1, ya_1 \in E(G)$ and $xa_1, yb_1 \notin E(G)$. Now we have completely determined the structure of G, and we see that $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$:



Case 2: We have $\Delta(a_ib_j) > m$ for all i and j. This implies $\Delta(a_ib_j) \geq 1$, thus each edge a_ib_j is included in at least one triangle. Recall also that a_ib_j can be included in at most two triangles (namely a_ib_jx and a_ib_jy), hence $m < \Delta(a_ib_j) \leq 2$ and consequently $m \leq 1$. We split Case 2 into two subcases depending on the edges between the sets $\{x, y\}$ and $\{a_1, a_2, \ldots, b_1, b_2, b_3 \ldots\}$.

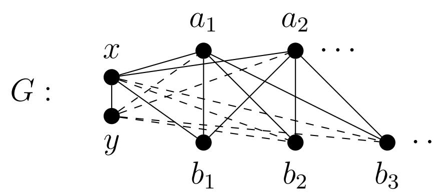
Case 2a: For all i and j, we have $xa_i, xb_j, ya_i, yb_j \in E(G)$. In this case our graph looks like this:



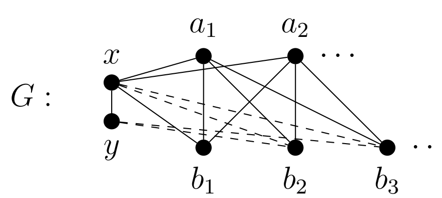
There exists a triangle xyv for all $v \in \{a_1, a_2, \ldots, b_1, b_2, b_3, \ldots\}$, thus $\Delta(xy) = n - 2$. Now $n - 2 = m \leq 1$ follows, contradicting our assumption $n \geq 7$.

Case 2b: There exist $v \in \{x, y\}$ and $w \in \{a_1, a_2, \ldots, b_1, b_2, b_3, \ldots\}$ such that $vw \notin E(G)$. We may assume without loss of generality that $yb_1 \notin E(G)$. Then the only triangle containing the edge a_ib_1 can be a_ib_1x, so we must have

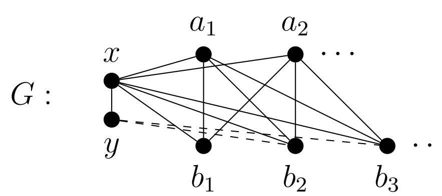
$xb_1 \in E(G)$ and $xa_i \in E(G)$ for all i:



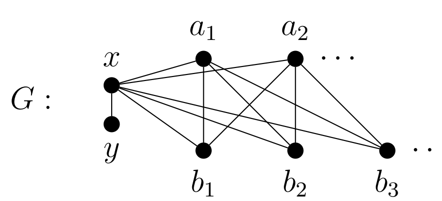
It remains to find out the status of the dashed (non)edges. We see that $\Delta(a_i b_1) = 1 > m$, hence $m = 0$, i.e., xy is contained in no triangle. This implies that y cannot be connected to any a_i:



Now the only triangle containing the edge $a_i b_j$ can be $a_i b_j x$, so we must have $xb_j \in E(G)$ for all j (previously we saw this only for $j = 1$):

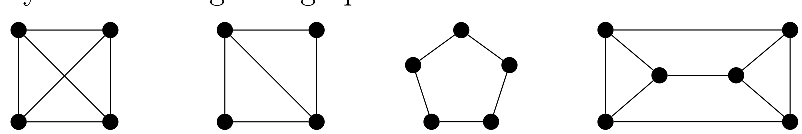


The final step is to observe that $\Delta(xy) = 0$ implies $yb_j \notin E(G)$ for all $j \geq 2$, and now we see that our graph has the structure described in item ((ii)) in the statement of the lemma:



□

Remark A.4. The assumption $n \geq 7$ cannot be dropped in Lemma A.2, as witnessed by the following four graphs:



Now we are ready to prove Theorem 4.6.

Proof of Theorem 4.6. The first statement of the theorem follows from Example 4.3. We prove the second statement by induction on the number of vertices. For $n \leq 6$ we have verifed ((ii)) by computer. Let us now assume that $G \ncong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$ with $n = |V(G)| \geq 7$ and that ((ii)) holds for graphs with less than n vertices. Let $m = \min\{\Delta(e) : e \in E(G)\}$.

If $G \setminus \{x,y\} \cong K_{\lfloor (n-2)/2 \rfloor, \lceil (n-2)/2 \rceil}$ holds for every edge $e = xy \in E(G)$ with $\Delta(e) = m$, then G is the second graph described in Lemma A.2. It is straightforward to compute by counting cherries and edges that

$$\mathrm{ns}(\mathbb{A}(G)) = 2(n-4)\left\lfloor \frac{n-2}{2} \right\rfloor \left\lceil \frac{n-2}{2} \right\rceil + 2n^2 - 2n.$$

After simple but tedious calculations we get that the inequality of ((ii)) is equivalent to $8\left\lfloor \frac{n-2}{2} \right\rfloor \left\lceil \frac{n-2}{2} \right\rceil - 4\left\lfloor \frac{n-2}{2} \right\rfloor \geq 0$, which is indeed true (for all $n \geq 2$).

Now suppose that there is an edge $e = xy \in E(G)$ with $\Delta(e) = m$ such that $G \setminus \{x,y\} \ncong K_{\lfloor (n-2)/2 \rfloor, \lceil (n-2)/2 \rceil}$. We introduce some notation to facilitate cherry-counting:

$$\begin{aligned}
ch(x) &= \text{ the number of cherries in } G \text{ that contain the vertex } x;\\
ch(y) &= \text{ the number of cherries in } G \text{ that contain the vertex } y;\\
ch(e) &= \text{ the number of cherries in } G \text{ that contain the edge } e;\\
A &= \{v \in V(G) \setminus \{x,y\} : vx \in E(G) \text{ and } vy \notin E(G)\}, \quad a = |A|;\\
B &= \{v \in V(G) \setminus \{x,y\} : vx \notin E(G) \text{ and } vy \in E(G)\}, \quad b = |B|;\\
C &= \{v \in V(G) \setminus \{x,y\} : vx \in E(G) \text{ and } vy \in E(G)\}, \quad c = |C|;\\
D &= \{v \in V(G) \setminus \{x,y\} : vx \notin E(G) \text{ and } vy \notin E(G)\}, \quad d = |D|.
\end{aligned}$$

The triangles containing the edge e are of the form xyv, where $v \in C$, thus $\Delta(e) = m = c$. Comparing the graphs G and $G \setminus \{x,y\}$, we see that

$$\begin{aligned}
|Ch(G)| &= |Ch(G \setminus \{x,y\})| + ch(x) + ch(y) - ch(e);\\
|E(G)| &= |E(G \setminus \{x,y\})| + a + b + 2c + 1.
\end{aligned}$$

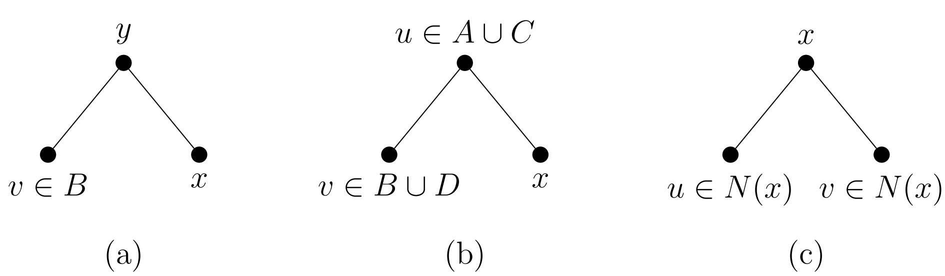
Therefore, the index of nonassociativity of $\mathbb{A}(G)$ can be computed as follows (we divide by 4, for convenience):

$$\frac{1}{4}\mathrm{ns}(\mathbb{A}(G)) = \frac{1}{4}\mathrm{ns}(G \setminus \{x,y\}) + ch(x) + ch(y) - ch(e) + a + b + 2c + 1.$$

Applying the induction hypothesis to the graph $G \setminus \{x,y\}$ and observing that $ch(e) = a + b$ (since xyv is a cherry if and only if $v \in A \cup B$), we obtain

$$\frac{1}{4}\mathrm{ns}(\mathbb{A}(G)) \leq \frac{n-2}{2}\left\lfloor \frac{n-2}{2} \right\rfloor \left\lceil \frac{n-2}{2} \right\rceil - \left\lfloor \frac{n-2}{2} \right\rfloor + ch(x) + ch(y) + 2c + 2. \tag{7}$$

In order to estimate $ch(x)$, we note that there are three types of cherries that contain the vertex x:



The number of cherries of type (a) is b, as $yv \in E(G)$ and $xv \notin E(G)$ for all $v \in B$. For cherries of type (b), we observe that $xu \in E(G)$ for all $u \in A \cup C$ and $xv \notin E(G)$ for all $v \in B \cup D$, thus the number of such cherries is equal to the number of edges between $A \cup C$ and $B \cup D$, which is at most $(a+c)(b+d)$. To get a cherry of type (c), we can choose the pair $\{u, v\} \subseteq N(x) = A \cup C \cup \{y\}$ in $\binom{a+c+1}{2}$

many ways, but we must make sure that $uv \notin E(G)$ so that we indeed obtain a cherry. Therefore, the number of cherries of type (c) is $\binom{a+c+1}{2} - E(G|_{N(x)})$, where $G|_{N(x)}$ denotes the induced subgraph of G on the vertex set $N(x)$. For any $u \in N(x)$, we have $\Delta(xu) = d_{G|_{N(x)}}(u)$, since xuv is a triangle if and only if $v \in N(x)$ and $uv \in E(G)$. We have seen that $\min\{\Delta(e) : e \in E(G)\} = c$, so we can conclude that $d_{G|_{N(x)}}(u) \geq c$. Thus every vertex of the graph $G|_{N(x)}$ has degree at least c, hence $E(G|_{N(x)}) \geq \frac{1}{2} \cdot |N(x)| \cdot c = \frac{1}{2}(a+c+1)c$. Consequently, the number of cherries of type (c) is at most $\binom{a+c+1}{2} - \frac{1}{2}(a+c+1)c$. Adding our estimates for cherries of type (a), (b) and (c), we get

$$ch(x) \leq b + (a+c)(b+d) + \binom{a+c+1}{2} - \frac{1}{2}(a+c+1)c. \tag{8}$$

Let us record for later reference that equality holds in this estimate if and only if

$$\forall u \in A \cup C \ \forall v \in B \cup D\colon uv \in E(G) \text{ and } \forall u \in N(x)\colon d_{G|_{N(x)}}(u) = c. \tag{9}$$

We can treat $ch(y)$ in a similar manner, we only need to interchange A and B:

$$ch(y) \leq a + (b+c)(a+d) + \binom{b+c+1}{2} - \frac{1}{2}(b+c+1)c, \tag{10}$$

and equality holds here if and only if

$$\forall u \in A \cup D \ \forall v \in B \cup C\colon uv \in E(G) \text{ and } \forall v \in N(y)\colon d_{G|_{N(y)}}(v) = c. \tag{11}$$

Substituting (8) and (10) into (7), we get

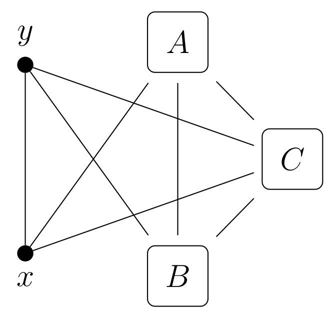
$$\frac{1}{4}\,\mathrm{ns}(\mathbb{A}(G)) \leq \frac{n-2}{2}\left\lfloor\frac{n-2}{2}\right\rfloor\left\lceil\frac{n-2}{2}\right\rceil - \left\lfloor\frac{n-2}{2}\right\rfloor + 2 + f(a,b,c,d) \tag{12}$$

where f is the function considered in Lemma A.1. Since $a+b+c+d = n-2$, we have $f(a,b,c,d) \leq M_{n-2}$ by Lemma A.1, and this gives

$$\begin{aligned} \frac{1}{4}\,\mathrm{ns}(\mathbb{A}(G)) &\leq \frac{n-2}{2}\left\lfloor\frac{n-2}{2}\right\rfloor\left\lceil\frac{n-2}{2}\right\rceil - \left\lfloor\frac{n-2}{2}\right\rfloor + 2 + M_{n-2} \\ &= \frac{n}{2}\left\lfloor\frac{n}{2}\right\rfloor\left\lceil\frac{n}{2}\right\rceil - \left\lfloor\frac{n}{2}\right\rfloor + 2, \end{aligned} \tag{13}$$

which is unfortunately 1 more than what we would need for ((ii)). As a remedy, we shall prove that it is not possible that $f(a,b,c,d) = M_{n-2}$, and, at the same time, equality holds in (8) as well as in (10).

So let us assume for contradiction that $f(a,b,c,d) = M_{n-2}$ and that (9) and (11) both hold. Then, according to Lemma A.1, we have $d = 0$, $c \leq 1$ and the difference of a and b is at most one. By the first parts of (9) and (11), all edges across the sets A, B, C are drawn, thus our graph has the following structure:



Since each vertex of A is connected to each vertex of C by an edge, the second part of (9) implies that there are no edges inside the set A (recall that $N(x) = A \cup C \cup \{y\}$); similarly, there are no edges within B either, by the second part of (11). This shows that if $c = 0$, then $G \cong K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, contradicting the assumption of ((ii)). If $c = 1$ and u is the unique element of C, then $d_{G|_{N(x)}}(u) = a + 1$, hence the second part of (9) gives $c = a + 1$, i.e., $a = 0$. Similarly, we have $b = 0$, and then $n = a + b + c + d + 2 = 3$, which is again a contradiction.

We have proved that the inequality in (13) cannot be equality, and this implies $\frac{1}{4} \operatorname{ns}(\mathbb{A}(G)) \leq n/2 \cdot \lfloor n/2 \rfloor \cdot \lceil n/2 \rceil - \lfloor n/2 \rfloor + 1$, which is equivalent to ((ii)). □

Appendix B. Proof of Theorem 4.10

We will construct a family of graphs that cover a large part of $\operatorname{range}_n(\operatorname{ns})$. The graphs are parameterized by tuples of natural numbers, and in the first three lemmas we establish the required properties of these tuples.

For positive integers h and k, let

$$\Omega_{h,k} = \{(\omega_1, \ldots, \omega_h) : 0 \leq \omega_1 \leq \cdots \leq \omega_h \leq k\} \subseteq \{0, 1, \ldots, k\}^h.$$

We consider the following property for tuples $\omega = (\omega_1, \ldots, \omega_h) \in \Omega_{h,k}$:

$$\forall i \in \{1, \ldots, h\} \colon \omega_i \leq \omega_1 + \cdots + \omega_{i-1} + 1. \tag{14}$$

(Note that this implies that the leftmost nonzero entry of the tuple ω is 1.) For $(\omega_1, \ldots, \omega_h) \in \Omega_{h,k}$, this property guarantees that each integer between 0 and $\omega_1 + \cdots + \omega_h$ can be written as the sum of some of the numbers $\omega_1, \ldots, \omega_h$. This fact is almost trivial, but we provide the proof in the next lemma (however, it might be easier for the reader to prove it for themselves than reading the proof).

Lemma B.1. *If $\omega = (\omega_1, \ldots, \omega_h) \in \Omega_{h,k}$ has property* (14)*, then for each nonnegative integer s with $s \leq \omega_1 + \cdots + \omega_h$, there exist $\varepsilon_1, \ldots, \varepsilon_h \in \{0, 1\}$ such that $s = \varepsilon_1 \omega_1 + \cdots + \varepsilon_h \omega_h$.*

Proof. We may assume without loss of generality that $\omega_1 = 1$ (hence $\omega_i > 0$ for each i), since the zero entries in ω (if there are any) are irrelevant for the lemma. We prove the lemma by induction on h. For $h = 1$ the claim is obvious, as $\omega_1 = 1$, hence s is either 0 or ω_1. Now let $h \geq 2$ and let $s \leq \omega_1 + \cdots + \omega_h$. If $s \leq \omega_1 + \cdots + \omega_{h-1}$, then we can apply the induction hypothesis to the tuple $(\omega_1, \ldots, \omega_{h-1}) \in \Omega_{h-1,k}$ with the number s. If $s > \omega_1 + \cdots + \omega_{h-1}$, then $s \geq \omega_h$, since $\omega_h \leq \omega_1 + \cdots + \omega_{h-1} + 1$ holds by (14). Thus $s - \omega_h$ is a nonnegative integer, and $s - \omega_h \leq \omega_1 + \cdots + \omega_{h-1}$. Therefore, we can apply the induction hypothesis to the tuple $(\omega_1, \ldots, \omega_{h-1}) \in \Omega_{h-1,k}$ with the number $s - \omega_h$: there are coefficients $\varepsilon_1, \ldots, \varepsilon_{h-1} \in \{0, 1\}$ such that $s - \omega_h = \varepsilon_1 \omega_1 + \cdots + \varepsilon_{h-1} \omega_{h-1}$. Setting $\varepsilon_h = 1$, we can write s as $s = (s - \omega_h) + \omega_h = \varepsilon_1 \omega_1 + \cdots + \varepsilon_{h-1} \omega_{h-1} + \varepsilon_h \omega_h$. □

For $\omega = (\omega_1, \ldots, \omega_h) \in \Omega_{h,k} \setminus \{(0, \ldots, 0)\}$, let $m = m(\omega)$ denote the multiplicity of the maximal (i.e., rightmost) entry: $\omega_{h-m} < \omega_{h-m+1} = \cdots = \omega_h$.

Lemma B.2. *If $\omega = (\omega_1, \ldots, \omega_h) \in \Omega_{h,k} \setminus \{(0, \ldots, 0)\}$ has property* (14)*, then for each nonnegative integer s with $s \leq m(\omega) + \omega_1 + \cdots + \omega_h$, there exist $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_h \in \{0, 1\}$ such that $s = \varepsilon_0 m(\omega) + \varepsilon_1 \omega_1 + \cdots + \varepsilon_h \omega_h$.*

Proof. Let $m = m(\omega)$ and $\sigma = \omega_1 + \cdots + \omega_h$, for brevity. Lemma B.1 shows that each integer in the interval $[0, \sigma]$ can be written as the sum of some of the numbers $\omega_1, \ldots, \omega_h$. Adding m to each of these numbers, we find that each integer in the interval $[m, m+\sigma]$ can be written as the sum of some of the numbers $m, \omega_1, \dots, \omega_h$. To finish the proof, we just need to verify that the intervals $[0, \sigma]$ and $[m, m+\sigma]$ cover $[0, m+\sigma]$. Indeed, $m \leq \sigma$, as the sum $\sigma = \omega_1 + \dots + \omega_h$ contains m copies of the positive integer ω_h. $\square$

For $\omega = (\omega_1, \dots, \omega_h) \in \Omega_{h,k} \setminus \{(0, \dots, 0)\}$, we define ω^- as the tuple that is obtained from ω by decreasing the leftmost occurrence of ω_h by 1:

$$\omega^- = (\omega_1, \dots, \omega_{h-m}, \, \omega_{h-m+1} - 1, \, \omega_{h-m+2}, \dots, \omega_h),$$

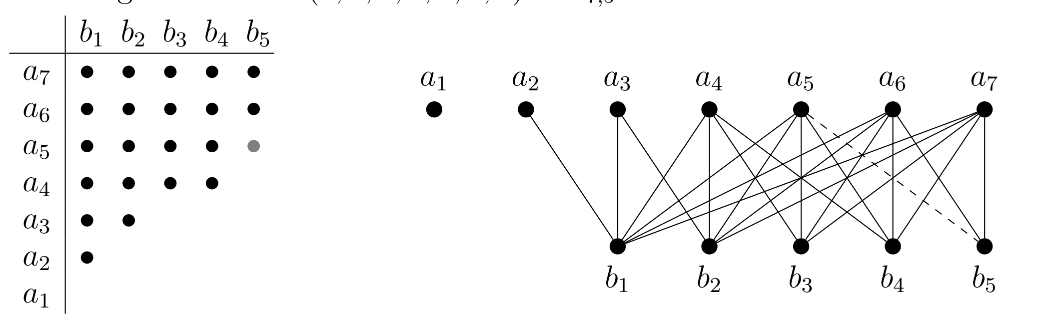
where $m = m(\omega)$. For example, if $\omega = (0, 1, 2, 4, 5, 5, 5)$, then $\omega^- = (0, 1, 2, 4, 4, 5, 5)$, and if $\omega = (0, 1, 2, 3, 4, 7, 9)$, then $\omega^- = (0, 1, 2, 3, 4, 7, 8)$. It is clear that if $\omega \in \Omega_{h,k} \setminus \{(0, \dots, 0)\}$, then $\omega^- \in \Omega_{h,k}$; moreover, property (14) is preserved under the map $\omega \mapsto \omega^-$ (this is straightforward to verify, so we omit the proof).

Lemma B.3. *If $\omega \in \Omega_{h,k} \setminus \{(0, \dots, 0)\}$ has property* (14)*, then ω^- also satisfies* (14)*.*

Now we describe the promised graph construction: we assign a bipartite graph G_ω to each tuple $\omega \in \Omega_{h,k}$. Let $V(G_\omega) = \{a_1, \dots, a_h, b_1, \dots, b_k\}$ and

$$E(G_\omega) = \{a_i b_j : 1 \leq i \leq h \text{ and } 1 \leq j \leq \omega_i\}.$$

We can regard ω as a partition of the number $\omega_1 + \dots + \omega_h$, and we can visualize this by a Ferrers diagram. The left hand side of the figure below shows this Ferrers diagram for $\omega = (0, 1, 2, 4, 5, 5, 5) \in \Omega_{7,5}$.



This Ferrers diagram can be considered as a kind of adjacency matrix for the bipartite graph G_ω, as indicated by the row and column indices of the diagram. The graph G_ω can be seen on the right hand side the figure above. (Removing the gray dot we obtain the diagram of $\omega^- = (0, 1, 2, 4, 4, 5, 5)$, and removing the dashed edge from the graph we get G_{ω^-}.) The degrees of the vertices a_i are determined by the rows of the Ferrers diagram: $d(a_i) = \omega_i$ for $i = 1, \dots, h$. The vertices b_j correspond to the columns of the diagram. One could expess the degrees $d(b_j)$ in terms of the conjugate partition, but we will not formalize this; we will only need that $d(b_{\omega_h}) = m(\omega)$.

The next lemma is crucial to obtain graphs whose indices of nonassociativity cover all even numbers in a "large" interval.

Lemma B.4. *Let $\omega \in \Omega_{h,k} \setminus \{(0, \dots, 0)\}$, let $G = G_\omega$ and $G^- = G_{\omega^-}$. If t is an even number and $\operatorname{ns}(\mathbb{A}(G^-)) \leq t \leq \operatorname{ns}(\mathbb{A}(G))$, then there is a graph G^* with $n = h + k$ vertices, such that $\operatorname{ns}(\mathbb{A}(G^*)) = t$.*

Proof. By Proposition 4.1, the index of nonassociativity of any graph algebra is even, thus we can write $\operatorname{ns}(\mathbb{A}(G)) - t = 2s$ for some nonnegative integer s with $s \leq \frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - \frac{1}{2}\operatorname{ns}(\mathbb{A}(G^-))$. We have to construct a graph G^* such that $\frac{1}{2}\operatorname{ns}(\mathbb{A}(G^*)) = \frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - s$. We will do this by adding loops to appropriate vertices of G.

Let $V = \{a_1, \dots, a_h, b_1, \dots, b_k\}$ be the common vertex set of G and G^-. We can obtain G^- from G by deleting one edge: $E(G^-) = E(G) \setminus \{a_{h-m+1}b_{\omega_h}\}$. Therefore, $d_G(a_{h-m+1}) = \omega_h$, $d_{G^-}(a_{h-m+1}) = \omega_h - 1$ and $d_G(b_{\omega_h}) = m$, $d_{G^-}(b_{\omega_h}) = m-1$, where $m = m(\omega)$, but apart from these two exceptions, the degrees of the vertices are the same in G and in G^-. This implies, with the help of Proposition 4.2 that

$$\begin{aligned}\frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - \frac{1}{2}\operatorname{ns}(\mathbb{A}(G^-)) &= \sum_{v\in V} d_G(v)^2 - \sum_{v\in V} d_{G^-}(v)^2 \\ &= d_G(a_{h-m+1})^2 - d_{G^-}(a_{h-m+1})^2 \ + \ d_G(b_{\omega_h})^2 - d_{G^-}(b_{\omega_h})^2 \\ &= \omega_h^2 - (\omega_h - 1)^2 \ + \ m^2 - (m-1)^2 \\ &= 2\omega_h + 2m - 2.\end{aligned}$$

This proves that $s \le 2\omega_h + 2m - 2$.

We would like to use Lemma B.2, so we need to verify that $2\omega_h + 2m - 2 \le m + \omega_1 + \cdots + \omega_h$. By property (14) and by the definition of $m = m(\omega)$, we can estimate $\omega_1 + \cdots + \omega_h$ as follows:

$$\omega_1 + \cdots + \omega_h = \underbrace{\omega_1 + \cdots + \omega_{h-m}}_{\ge \omega_h - 1} + \underbrace{\omega_h + \cdots + \omega_h}_{= m\cdot\omega_h} \ge (m+1)\cdot\omega_h - 1.$$

We can derive the desired inequality as follows:

$$\begin{aligned}m + \omega_1 + \cdots + \omega_h &\ge m + (m+1)\cdot\omega_h - 1 \\ &= (m-1)(\omega_h - 1) + 2\omega_h + 2m - 2 \\ &\ge 2\omega_h + 2m - 2 \ge s.\end{aligned}$$

Now we can apply Lemma B.2: $s = \varepsilon_0 m(\omega) + \varepsilon_1\omega_1 + \cdots + \varepsilon_h\omega_h$ for suitable $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_h \in \{0,1\}$. Let us modify the graph G by adding a loop to b_{ω_h} if $\varepsilon_0 = 1$, and let us add a loop to a_i if $\varepsilon_i = 1$. By Lemma 4.5, the index of nonassociativity of the resulting graph G^* can be computed as follows:

$$\begin{aligned}\frac{1}{2}\operatorname{ns}(\mathbb{A}(G^*)) &= \frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - \varepsilon_0 d_G(b_{\omega_h}) - \sum_{i=1}^{h} \varepsilon_i d_G(a_i) \\ &= \frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - \bigl(\varepsilon_0 m(\omega) + \varepsilon_1\omega_1 + \cdots + \varepsilon_h\omega_h\bigr) \\ &= \frac{1}{2}\operatorname{ns}(\mathbb{A}(G)) - s.\end{aligned}$$

As we have already mentioned, this is equivalent to $\operatorname{ns}(\mathbb{A}(G^*)) = t$. □

Now we are ready to prove Theorem 4.10.

Proof of Theorem 4.10. Theorem 4.6 shows that $r_n \le n^3/2$, so our task is to give a lower estimate that is asymptotically $n^3/2$. To this end, let us fix positive integers h and k such that $h + k = n$, and let us consider the graphs G_ω for $\omega \in \Omega_{h,k}$.

We claim that if $\omega \in \Omega_{h,k}$ satisfies (14), then $r_n \ge \operatorname{ns}(\mathbb{A}(G_\omega))$. By Lemma B.4, $\operatorname{range}_n(\operatorname{ns})$ contains all even numbers from $\operatorname{ns}(\mathbb{A}(G_{\omega^-}))$ to $\operatorname{ns}(\mathbb{A}(G_\omega))$. Since, by Lemma B.3, ω^- also has property (14), we can applying Lemma B.4 again, and we get that $\operatorname{range}_n(\operatorname{ns})$ contains all even numbers from $\operatorname{ns}(\mathbb{A}(G_{(\omega^-)^-}))$ to $\operatorname{ns}(\mathbb{A}(G_{\omega^-}))$. We can continue this process, until, after applying the transformation $\omega \mapsto \omega^-$ sufficiently many times (more precisely: $\omega_1 + \cdots + \omega_h$ times), we reach the tuple $(0, \dots, 0)$ and the corresponding "edgeless" graph, whose graph algebra is obviously associative. This proves that each even number from 0 to $\mathrm{ns}(\mathbb{A}(G_\omega))$ occurs as index of nonassociativity of a graph algebra of a graph with n vertices, hence $r_n \geq \mathrm{ns}(\mathbb{A}(G_\omega))$, as claimed.

Now we only need to find a tuple $\omega \in \Omega_{h,k}$ satisfying (14), such that $\mathrm{ns}(\mathbb{A}(G_\omega))$ is large enough. Let $h = \lfloor n/2 \rfloor$, $k = \lceil n/2 \rceil$, and let $\omega = (1, 2, 4, \dots, 2^\ell, k, \dots, k)$, where $\ell = \lfloor \log_2 k \rfloor$ and the number of copies of k is $m = m(\omega) = h - \ell - 1$. The identity $1+2+4+\cdots+2^{i-1} = 2^i - 1$ guarantees that $\omega \in \Omega_{h,k}$ has property (14). The graph G_ω has m vertices of degree k (namely, $a_{\ell+2}, \dots, a_h$), and each of the vertices $b_1, \dots, b_k$ have degree at least m. Therefore, since G_ω is irreflexive and triangle-free, Proposition 4.2 implies that

$$\mathrm{ns}(\mathbb{A}(G_\omega)) = 2 \cdot \sum_{v \in V(G)} d(v)^2 \geq 2 \cdot (mk^2 + km^2).$$

Clearly, $\lim_{n\to\infty} k/n = \lim_{n\to\infty} m/n = 1/2$, hence $\lim_{n\to\infty} 2(mk^2 + km^2)/n^3 = 1/2$, and this gives the desired lower estimate for r_n. □

Appendix C. Tables

	0	1	2	3	4	$\mathrm{sdist}_{\mathrm{gr}}$
0	•					
2		•				
4		•	•			
6			•			
8			•	•		
10				•		
12				•	•	
ns						

Table 1. Possible values of $\mathrm{ns}(\mathbb{A}(G))$ and $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ for undirected graphs on 3 vertices

	0	1	2	3	4	5	6	7	8	$\mathrm{sdist}_{\mathrm{gr}}$
0	•									
2		•								
4		•	•							
6		•	•	•						
8			•	•	•					
10			•	•						
12			•	•	•					
14				•	•					
16				•	•	•				
18				•	•	•				
20					•	•	•			
22						•				
24					•	•	•			
26										
28							•	•		
30										
32									•	
ns										

Table 2. Possible values of $\mathrm{ns}(\mathbb{A}(G))$ and $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ for undirected graphs on 4 vertices

ns \ $\mathrm{sdist}_{\mathrm{gr}}$	0	1	2	3	4	5	6	7	8	9	10	11	12
0	•												
2		•											
4		•	•										
6		•	•	•									
8		•	•	•	•								
10			•	•	•								
12			•	•	•	•							
14			•	•	•	•							
16			•	•	•	•	•						
18				•	•	•							
20				•	•	•	•						
22				•	•	•	•						
24				•	•	•	•	•					
26					•	•	•	•					
28					•	•	•	•	•				
30						•	•	•					
32					•	•	•	•	•				
34						•	•	•	•				
36						•	•	•	•	•			
38							•	•	•	•			
40						•	•	•	•	•	•		
42							•	•	•	•			
44								•	•	•	•		
46										•			
48								•	•	•	•		
50									•	•	•		
52											•		
54												•	
56										•	•	•	
58													
60													•

TABLE 3. Possible values of $\mathrm{ns}(\mathbb{A}(G))$ and $\mathrm{sdist}_{\mathrm{gr}}(\mathbb{A}(G))$ for undirected graphs on 5 vertices

REFERENCES

[1] B. Batíková, Š, Dvořáková, M. Trch, *Semigroup distances of finite groupoids*, Ital. J. Pure Appl. Math. **32** (2014), 547–560.

[2] Al. C. Climescu, *Études sur la théorie des systèmes multiplicatifs uniformes I. L'indice de non-associativité*, Bull. École Polytech. Jassy **2** (1947), 347–371. (French)

[3] B. Csákány, T. Waldhauser, *Associative spectra of binary operations*, Mult.-Valued Log. **5** (2000), 175–200.

[4] R. Diestel, *Graph theory*, Graduate Texts in Mathematics, Volume 173, 5th electronic edition, Springer-Verlag, Heidelberg, 2016.

[5] P. Hájek, *Die Szászschen Gruppoide*, Mat.-Fyz. Časopis. Sloven. Akad. Vied. **15** (1965), 15–42.

[6] T. Kepka, M. Trch, *Groupoids and the associative law I. (Associative triples)*, Acta Univ. Carol. Math. Phys. **33** (1992), 69–86.

[7] T. Kepka, M. Trch, *Groupoids and the associative law II. (Groupoids with small semigroup distance)*, Acta Univ. Carol. Math. Phys. **34** (1993), 67–83.

[8] T. Kepka, M. Trch, *Groupoids and the associative law III. (Szász-Hájek groupoids)*, Acta Univ. Carol. Math. Phys. **36** (1995), 17–30.

[9] M. Trch, *Groupoids and the associative law IIIA. (Primitive extensions of SH-groupoids and their semigroup distances)*, Acta Univ. Carol. Math. Phys. **50** (2009), 73–82.

[10] T. Kepka, M. Trch, *Groupoids and the associative law IV. (Szász-Hájek groupoids of type (A,B,A))*, Acta Univ. Carol. Math. Phys. **35** (1994), 31–42.
[11] T. Kepka, M. Trch, *Groupoids and the associative law V. (Szász-Hájek groupoids of type (A,A,B))*, Acta Univ. Carol. Math. Phys. **36** (1995), 31–44.
[12] T. Kepka, M. Trch, *Groupoids and the associative law VI. (Szász-Hájek groupoids of type (a,b,c))*, Acta Univ. Carol. Math. Phys. **38** (1997), 13–22.
[13] M. Trch, *Groupoids and the associative law VII. (Semigroup distance of SH-groupoids)*, Acta Univ. Carol. Math. Phys. **47** (2006), 57–63.
[14] M. Trch, *Groupoids and the associative law VIIA. (SH-groupoids of type (A,B,A) and their semigroup distances)*, Acta Univ. Carol. Math. Phys. **48** (2007), 43–54.
[15] M. Trch, *Groupoids and the associative law VIIB. (SH-groupoids and simply generated congruences)*, Acta Univ. Carol. Math. Phys. **52** (2011), 77–87.
[16] E. Lehtonen, T. Waldhauser, *Associative spectra of graph algebras I. Foundations, undirected graphs, antiassociative graphs*, J. Algebraic Combin. **53** (2021), 615–638.
[17] E. Lehtonen, T. Waldhauser, *Associative spectra of graph algebras II. Satisfaction of bracketing identities, spectrum dichotomy*, J. Algebraic Combin. **55** (2022), 533–557.
[18] S. Liebscher, T. Waldhauser, *On associative spectra of operations*, Acta Sci. Math. (Szeged) **75** (2009), 433–456.
[19] T. Poomsa-ard, *Hyperidentities in associative graph algebras*, Discuss. Math. Gen. Algebra Appl. **20** (2000), 169–182.
[20] R. Pöschel, W. Wessel, *Classes of graphs definable by graph algebra identites or quasi-identities*, Comment. Math. Univ. Carolin. **28** (1987), 581–592.
[21] A. Pyatkin, E. Lykhovyd, S. Butenko, *The maximum number of induced open triangles in graphs of a given order*, Optim. Lett. **13** (2019), 1927–1935.
[22] C. R. Shallon, *Non-finitely based binary algebras derived from lattices*, PhD thesis, UCLA, 1979.
[23] G. Szász, *Über die Unabhängigkeit der Assoziativitätsbedingungen*, Acta Sci. Math. (Szeged) **15** (1953), 20–28.

(K. Kátai-Urbán) Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H6720 Szeged, Hungary
Email address, K. Kátai-Urbán: `katai@math.u-szeged.hu`

(T. Waldhauser) Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H6720 Szeged, Hungary
Email address, T. Waldhauser: `twaldha@math.u-szeged.hu`