Upper Bounds on Boolean-Width with Applications to Exact Algorithms

Yuri Rabinovich[1], Jan Arne Telle[2], and Martin Vatshelle[2]

[1] Department of Computer Science, University of Haifa, Israel
[2] Department of Informatics, University of Bergen, Norway

Abstract. Boolean-width is similar to clique-width, rank-width and NLC-width in that all these graph parameters are constantly bounded on the same classes of graphs. In many classes where these parameters are not constantly bounded, boolean-width is distinguished by its much lower value, such as in permutation graphs and interval graphs where boolean-width was shown to be $O(\log n)$ [1]. Together with FPT algorithms having runtime $O^*(c^{boolw})$ for a constant c this helped explain why a variety of problems could be solved in polynomial-time on these graph classes.

In this paper we continue this line of research and establish non-trivial upper-bounds on the boolean-width and linear boolean-width of *any* graph. Again we combine these bounds with FPT algorithms having runtime $O^*(c^{boolw})$, now to give a common framework of moderately-exponential exact algorithms that beat brute-force search for several independence and domination-type problems, on general graphs.

Boolean-width is closely related to the number of maximal independent sets in bipartite graphs. Our main result breaking the triviality bound of $n/3$ for boolean-width and $n/2$ for linear boolean-width is proved by new techniques for bounding the number of maximal independent sets in bipartite graphs.

1 Introduction

Boolean-width is a recently introduced graph parameter motivated by algorithms [2]. Having small boolean-width is witnessed by a decomposition of the graph into cuts with few different unions of neighborhoods - Boolean sums of neighborhoods - across the cut. This makes the decomposition natural to guide dynamic programming algorithms to solve problems where vertex sets having the same neighborhood across a cut can be treated as equivalent. Such dynamic programming on a given decomposition of boolean-width $boolw$ will for several problems related to independence and domination have runtime $O^*(c^{boolw})$ for a small constant c [2].

Boolean-width is similar to clique-width, rank-width and NLC-width in that all these graph parameters are constantly bounded on the same classes of graphs. However, in many classes where these parameters are not constantly bounded, boolean-width is distinguished by its much lower value. For example, permutation graphs, interval graphs, convex graphs and Dilworth k graphs all have boolean-width $O(\log n)$, and the decompositions are easy to find [1]. Since $O^*(c^{O(\log n)})$ is $n^{O(1)}$ this helps explain why several problems related to independence and domination are polynomial-time solvable on these graph classes.

G. Gutin and S. Szeider (Eds.): IPEC 2013, LNCS 8246, pp. 308–320, 2013.

In this paper we continue this line of research, combining $O^*(c^{boolw})$ dynamic programming for independence and domination problems with new bounds on boolean-width. Rather than giving a framework for polynomial-time algorithms on restricted graph classes, our goal in this paper is a framework for moderately-exponential exact algorithms on general graphs. Our main results are non-trivial upper-bounds of $(1-c)n/3$ on the boolean-width, and $(1-c)n/2$ on the linear boolean-width, of *any* graph, for some $c > 0$ and sufficiently large values of n. This is accompanied by a polynomial-time algorithm computing a decomposition witnessing the non-trivial bound on linear boolean-width.

We combine this with dynamic programming algorithms on decompositions of linear boolean-width k that solve INDEPENDENT SET in time $O^*(2^k)$ and DOMINATING SET, INDEPENDENT DOMINATING SET and TOTAL DOMINATING SET in time $O^*(2^{2k})$. The combination gives moderately-exponential exact algorithms on general graphs solving all these problems, also weighted versions and counting versions, by a runtime beating brute-force search. Note that faster algorithms do exist in the literature, our goal in this paper is mainly to show the viability of this line of research. This is the first time a non-trivial upper bound on the value of a graph width parameter has been shown to hold *for every graph*.

Boolean-width is defined based on branch decompositions of a graph, using as cut function what is called the boolean dimension $\mathrm{bd}(H)$ of a (bipartite) graph H. This corresponds to the logarithm (base 2) of the number of maximal independent sets $\mathrm{mis}(H)$ of H. Our upper bounds on (linear) boolean-width rely on new techniques for bounding the number of maximal independent sets. This number has received much attention both from the algorithmic and the structural perspectives. While it is known [3] that computing $\mathrm{mis}(G)$ is $\#P$-hard even for planar bipartite graphs, approximating it is a much more delicate problem. From the structural point of view, bounding the number of maximal independent sets in special as well as in general graphs leads to interesting hard problems. Let us just mention the entropy-based results about $\mathrm{bd}(G) = \log_2 \mathrm{mis}(G)$ of d-regular graphs [4], see the references therein for an updated picture of the state of research in this area.

We introduce three techniques for our bounds. The first (Theorem 5) is based on a vertex partition achieved from a packing of paths and goes via $\mathrm{im}(H)$, the size of a maximum induced matching in bipartite graph H. The second (Theorem 6) is based on a random partition and also goes via $\mathrm{im}(H)$. The third (Theorem 7) is based on Hoeffding's inequality and in contrast to the first two applies also to boolean-width rather than just linear boolean-width. As already mentioned, our goal is to show the viability of this line of research and these various techniques should be helpful in later attempts to improve the bounds.

Our paper is organized as follows. In Section 2 we give all definitions and some preliminary results, for example showing that a non-trivial upper bound on (linear) boolean-width of a graph G will follow from a non-trivial upper bound on the boolean dimension of some balanced partition of G. In Section 3 we aim at an understanding of the structure of bipartite graphs of high boolean dimension. It is well known and easy that $\mathrm{bd}(H)$ is at most $n/2$, and the maximum is attained by a size-$n/2$ matching. We tie $\mathrm{bd}(H)$ to $\mathrm{im}(H)$ and introduce and study the values of $\text{co-im}(H) = n/2 - \mathrm{im}(H)$

and $\text{co-bd}(H) = n/2 - \text{bd}(H)$, the high-end ranges of $\text{im}(H)$ and $\text{bd}(H)$. We show constant factor approximation algorithms for these values, as well as a stability result showing that the smaller the values are, the closer is the bipartite graph to the size-$n/2$ matching. In Section 4 we turn to general graphs, and show, constructively, by a polynomial-time algorithm, that every graph has a balanced partition where the boolean dimension of the associated bipartite graph beats the triviality bound. Combined with the result from Section 2 this implies a constructive result for linear boolean-width of general graphs, beating the triviality bound of $n/2$. In Section 5 we turn to the standard boolean-width parameter and show also in this case a non-trivial upper bound beating the triviality bound of $n/3$, also constructive, but now by a randomized and low-exponential-time algorithm.

2 Terminology and Preliminaries

We consider undirected unweighted simple graphs $G = (V, E)$ and bipartite graphs $H = (A, B, E)$. We also denote the vertex set V by $V(G)$. For $S \subseteq V$ we denote by $G|_S$ the subgraph induced by S. The *neighborhood* of a vertex $v \in V$ is denoted $N(v)$. The neighborhood of a set $S \subset V$ is $N(S) = \cup_{v \in S} N(v)$. Any $S \subseteq V$ defines a cut $(S, V–S)$, and a bipartite graph $G_{S,V–S} = (S, V–S, \{(u, v) \in E : u \in S \wedge v \in V–S\})$.

A *decomposition tree* of a graph $G = (V, E)$ is a pair (T, δ) where T is a ternary tree, i.e. all internal nodes are of degree three, and δ a bijection between the leaves of T and $V(G)$. Removing an edge (a, b) from T results in two subtrees T_a and T_b, and a bipartition of V into V_a and V_b corresponding, respectively, to the δ-labels of leaves of T_a and T_b, and a bipartite graph G_{V_a,V_b}.

Definition 1 (Boolean Dimension, Boolean Width and Linear Boolean Width). *For a bipartite graph $H = (A, B, E)$, let $\mathcal{N}_A = \{N(X) \subseteq B \mid X \subseteq A\}$ be the family of neighborhoods of all sets $X \subseteq A$. The* boolean dimension *of H is defined as* $\text{bd}(H) = \log_2 |\mathcal{N}_A|$.

The boolean-width *of a decomposition tree (T, δ) is the maximum value of* $\text{bd}(G_{V_a,V_b})$ *over all edges (a, b) of T. The* boolean-width *of G, denoted* $\text{bw}(G)$*, is the minimum boolean-width over all decomposition trees of G.*

The linear boolean-width *of G, denoted* lbw(G)*, is the minimum boolean-width over all decomposition trees (T, δ) of G where T is a path on $|V|$ inner nodes, each with an attached leaf, corresponding to a linear arrangement of V.*

Given a graph G there is a $O^*(2.52^n)$ algorithm computing its boolean-width exactly [5] and in FPT time parameterized by $bw(G)$ we can compute a decomposition of boolean-width $2^{2bw(G)}$ using the algorithm for decompositions of optimal rank-width [6]. The boolean-width parameter was originally introduced in [2] in the context of parameterized algorithms. In particular, using a natural dynamic programming approach it was shown there that

Theorem 1. *[2,5] Given a graph G and a decomposition tree of boolean-width k, one can solve weighted and counting versions of* INDEPENDENT SET *in time $O^*(2^{2k})$ and* DOMINATING SET, INDEPENDENT DOMINATING SET *and* TOTAL DOMINATING SET *in time $O^*(2^{3k})$.*

These are dynamic programming algorithms that choose a root of the decomposition tree and traverse it bottom-up, with each node of the tree representing the subgraph induced by vertices corresponding to the leaves of the subtree. In a decomposition tree for linear boolean-width we choose one end of the path of inner nodes as root so that one of the two children of any node will always represent a subgraph on a single vertex. The runtime on linear decompositions can for this reason be improved

Corollary 1. *[2,5] Given a linear arrangement of $V(G)$ of linear boolean-width k, one can solve weighted and counting versions of* INDEPENDENT SET *in time $O^*(2^k)$ and* DOMINATING SET, INDEPENDENT DOMINATING SET *and* TOTAL DOMINATING SET *in time $O^*(2^{2k})$.*

Note that for any bipartite graph $H = (A, B, E)$, we have $|\mathcal{N}_A| = |\mathcal{N}_B|$, see, e.g., [7]. A good combinatorial way to demonstrate this is by establishing a bijection between the elements of $\mathcal{N}_A$ (or $\mathcal{N}_B$) and the set of all maximal independent sets of $H = (A, B, E)$. Here is a sketch of the argument. Given a set $S \in \mathcal{N}_A$, let X be the maximal set in A such that $N(X) = S$. Then X is uniquely defined, and $X \cup B - S$ is maximal independent. In the other direction, given a maximal independent set I, $B - I \in \mathcal{N}_A$. Moreover, if I resulted from S, then S results from I. [1] Hence,

Proposition 1. *Let $\text{mis}(H)$ be the number of maximal independent sets in a bipartite graph H. Then, $\text{bd}(H) = \log_2 \text{mis}(H)$.*

The following simple property of $\text{bd}(H)$ will prove useful; it is an immediate consequence of the definition of bd.

Proposition 2. *$\text{bd}(H)$ is monotone decreasing with respect to vertex removal. Moreover, such removal may decrease $\text{bd}(H)$ by at most 1. Hence, for a bipartite graph $H = (A, B, E)$ we have $\text{bd}(H) \leq \min(|A|, |B|)$.*

More generally, given two bipartite graphs $G = (A, B, E)$ and $H = (A, B, E')$ on the same vertex set and the same two sides, it holds that $\text{bd}(G \cup H) \leq \text{bd}(G) + \text{bd}(H)$.

Proposition 2 implies that $\text{bd}(H) \leq n/2$, and this bound is met when H is a matching of size $n/2$. For a finer study of the structure of sub-extremal graphs H, we shall need the following notions.

Definition 2. *Define $\text{im}(G)$ as the size of a maximum induced matching in G, i.e. a maximum-size set of edges whose endpoints do not induce any other edges in G. Note that $\text{im}(G) \leq n/2$ and this bound is met only by a size-$n/2$ matching. To study the high-end range of $\text{bd}(H)$ and $\text{im}(G)$ we define $\text{co-im}(G) = \frac{n}{2} - \text{im}(G)$ and the boolean co-dimension $\text{co-bd}(H) = \frac{n}{2} - \text{bd}(H)$.*

The extremal values of $\text{bw}(G)$ and $\text{lbw}(\text{G})$ are not known. The following proposition provides a preliminary tool for the study of the former.

[1] Observed by Nathann Cohen in a course of discussion with the authors. Later we have learned that a similar observation was made in [8].

Proposition 3. *Let $A \subseteq V$ be a subset of vertices with $\frac{1}{3}n \leq |A| \leq \frac{2}{3}n$ and* $\mathrm{bd}(G_{A,V\text{–}A}) = (\frac{1}{3} - \epsilon)n$ *for some $\frac{1}{4} > \epsilon \geq 0$. Then one can construct a decomposition tree of boolean-width at most $(\frac{1}{3} - \frac{\epsilon}{3})n$.*

In particular, $\mathrm{bw}(G) \leq n/3$. *We call this the* triviality bound *for* $\mathrm{bw}(G)$.

Proof. Without loss of generality, $|A| \leq n/2$; otherwise we switch to $V\text{–}A$. Partition V into three sets $A \cup X, B_1, B_2$ where X is disjoint from A and $|X| \leq \frac{2}{3}\epsilon n$, $|A \cup X| \leq n/2$ and $|B_1|, |B_2| \leq (\frac{1}{3} - \frac{1}{3}\epsilon)n$. Refining this partition to a decomposition tree arbitrarily with the sole restriction that at the upper level $A \cup X$ is split into two approximately equal parts we argue that the proposition holds using Proposition 2. For the top cut $bd(G_{A\cup X,V\text{–}(A\cup X)}) \leq n/3 - \epsilon + |X| \leq (\frac{1}{3} - \frac{\epsilon}{3})n$. For any node representing a subset of B_1 or B_2 the proposition holds since the size of one side will be small. When splitting $A \cup X$ into two approximately equal parts, each part has size at most $n/4$, we have $n/4 \leq n/3 - \epsilon/3$ since $\epsilon < n/4$, and hence the proposition holds. The conclusion about $\mathrm{bw}(G) \leq n/3$ corresponds to choosing an A of size $\frac{1}{3}n$, and $\epsilon = 0$. $\square$

For $\mathrm{lbw}(\mathrm{G})$ one has an analogous statement:

Proposition 4. *Let $A \subseteq V$ be subset of vertices of size $n/2$ such that* $\mathrm{bd}(G_{A,V\text{–}A}) \leq (\frac{1}{2} - \epsilon) \cdot n$ *for some $\epsilon \geq 0$. Then one can construct a linear arrangement of the vertices of linear boolean-width at most $(\frac{1}{2} - \frac{1}{2}\epsilon) \cdot n$.*

In particular, $\mathrm{lbw}(\mathrm{G}) \leq \mathrm{n}/2$. *We call this the* triviality bound *for* $\mathrm{lbw}(\mathrm{G})$.

Proof. Take any linear arrangement whose first $n/2$ elements are precisely A. We claim that it has the desired property. Let A_i denote the set of the first (equivalently, the last) i elements in this arrangement. Let $G_i = G_{A_i,V\text{–}A_i}$. Proposition 2 implies that $\mathrm{bd}(G_i) \leq i$. It also implies that $|\mathrm{bd}(G_i) - \mathrm{bd}(G_{i+1})| \leq 1$ for every i, hence $\mathrm{bd}(G_i) \leq \mathrm{bd}(G_{n/2}) + (n/2 - i) \leq n - \epsilon n - i$. Combining the two bounds on $\mathrm{bd}(G_i)$, the statement follows. $\square$

The present paper, besides studying $\mathrm{bd}(H)$, is mostly dedicated to establishing upper bounds on $\mathrm{lbw}(\mathrm{G})$ and $\mathrm{bw}(G)$. Before starting with our toil, let us just mention that the above values can in general be as large as $\Omega(n)$, which is achieved e.g., when G is a constant-degree expander. Indeed, in this case any $G_{A,V\text{–}A}$ where both sides are $\geq n/3$ has at least $\Omega(n)$ edges, and hence, due to the constant degree, $\mathrm{im}(G_{A,V\text{–}A}) = \Omega(n)$. Since the size of an induced matching is a lower bound on the boolean dimension (see the next section for details), the conclusion follows. The constants obtained along this line of reasoning are, however, quite miserable. The extremal values of $\mathrm{lbw}(\mathrm{G})$, $\mathrm{bw}(\mathrm{G})$ and the structure of the corresponding extremal graphs remain a (highly inspiring) mystery.

3 On Boolean Co-Dimension

3.1 Boolean Dimension vs. the Size of Maximum Induced Matching

We start with a lemma relating the boolean dimension of a bipartite graph $G = (A, B, E)$, $|V(G)| = n$, to the size of the maximum induced matching in G:

Lemma 1. *When* $\mathrm{im}(G) \leq n/4$*, it holds that*

$$\mathrm{im}(G) \;\leq\; \mathrm{bd}(G) \;\leq\; \mathrm{im}(G) \cdot \log_2(n/\mathrm{im}(G)) \cdot \phi(2 \cdot \mathrm{im}(G)/n),$$

where ϕ *is a function which never exceeds* 1.088*, and tends to* 1 *as* $\mathrm{im}(G)$ *tends to* $n/4$*.*

Proof. The first inequality is obvious, as the boolean dimension is monotone with respect to taking induced subgraphs. For the second inequality, assume w.l.o.g., that $|A| \leq n/2$, and consider the family of neighborhoods $\mathcal{N}_A$ in B. For every $S \in \mathcal{N}_A$, there is a a minimal set $S^* \subseteq A$ such that $N(S^*) = S$. By minimality of S^*, each vertex v^* in it has a neighbour $v \in S$ not seen by the other vertices. Forming a set $S' \subseteq S \subseteq B$ by picking (one) such v for every $v^* \in S^*$, we conclude that the subgraph of G induced by (S^*, S') is an induced matching. In particular, it holds that $|S^*| \leq \mathrm{im}(G)$, and thus any $S \in \mathcal{N}_A$ is a neighbourhood of a subset of A of size $\leq \mathrm{im}(G)$. Consequently, using a standard estimation for the sum of binomial coefficients,

$$2^{\mathrm{bd}(G)} \;=\; |\mathcal{N}_A| \;\leq\; \sum_{i=0}^{\mathrm{im}(G)} \binom{|A|}{i} \;\leq\; \sum_{i=0}^{\mathrm{im}(G)} \binom{n/2}{i} \;\leq\; 2^{n/2 \cdot H(\mathrm{im}(G)/(n/2))}$$

where $H(p) = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$ is the entropy function. We introduce $p\log_2 \frac{1}{p} + p = p\log_2 \frac{2}{p}$ as an approximator of H, and set $\phi(p) = H(p)/(p\log_2 \frac{2}{p})$. Then,

$$\mathrm{bd}(G) \;\leq\; \frac{n}{2} \cdot H\left(\frac{\mathrm{im}(G)}{n/2}\right) \;=\; \mathrm{im}(G) \cdot \log_2\left(\frac{n}{\mathrm{im}(G)}\right) \cdot \phi\left(\frac{\mathrm{im}(G)}{n/2}\right)$$

Through numerical analysis we found that $\phi(0.157) \approx 1.08798$ is the global maximum of ϕ in the range $[0, 0.5]$. □

Thus, $\mathrm{im}(G)$ is a $\log n$-approximation of $\mathrm{bd}(G)$, and the quality of approximation improves as $\mathrm{im}(G)$ grows. However, when $\mathrm{im}(G) \geq n/4$, the approach of Lemma 1 fails to imply anything beyond the trivial upper bound $\mathrm{bd}(G) \leq n/2$. This makes Lemma 1 inapplicable to the study of co-$\mathrm{bd}(G)$ vs. co-$\mathrm{im}(G)$. The key result of this subsection is that when co-$\mathrm{bd}(G)$ is small, so is co-$\mathrm{im}(G)$, and, moreover, co-$\mathrm{bd}(G)$ and co-$\mathrm{im}(G)$ are linearly related.

We start with a special case:

Lemma 2. *Let* $G = (A, B, E)$ *be a bipartite graph of degree at most 2. Then,*

$$\text{co-im}(G) \;\geq\; \text{co-bd}(G) \;\geq\; 0.339 \cdot \text{co-im}(G).$$

Proof. The first inequality was already established in Lemma 1. For the second inequality, observe that both co-$\mathrm{bd}(G)$ and co-$\mathrm{im}(G)$ are additive with respect to disjoint union of graphs. Thus it suffices to consider connected G's, i.e. G is either C_n, the (even) n-cycle, or P_{n-1}, the path on $(n-1)$ edges. For such graphs both $\mathrm{im}(G)$ and $\mathrm{bd}(G)$ are tractable. For the maximum induced matching one easily gets $\mathrm{im}(C_n) = \lfloor \frac{n}{3} \rfloor$ and $\mathrm{im}(P_{n-1}) = \lfloor \frac{n+1}{3} \rfloor$. For boolean dimension, recall that by Proposition 1, $\mathrm{bd}(G) = \log_2 \mathrm{mis}(G)$, where $\mathrm{mis}(G)$ is the number of maximal independent sets in

G. Let $c(n) = \mathrm{mis}(C_n)$ and $p(n) = \mathrm{mis}(P_{n-1})$. The recurrence formulae for these values are well known (see e.g. [9]). Namely, $c(n) = c(n-2) + c(n-3)$, and $p(n) = p(n-2) + p(n-3)$. The initial conditions are $c(1) = 0$, $c(2) = 2$, $c(3) = 3$ and $p(1) = 1$, $p(2) = 2$, $p(3) = 2$ respectively.

Thus, to compare $\text{co-bd}(G)$ to $\text{co-im}(G)$ one needs to lower-bound the expressions

$$\frac{n/2 - \log_2(c(n))}{n/2 - \lfloor \frac{n}{3} \rfloor} \quad \text{and} \quad \frac{n/2 - \log_2(p(n))}{n/2 - \lfloor \frac{n+1}{3} \rfloor}\,.$$

Combining case analysis (according to $n \bmod 3$), numerical computations and an inductive argument, we conclude that the minimum is achieved on the 8-cycle C_8, and its value is $\frac{1}{2}(4 - \log_2 10) \approx 0.339036$. □

We continue with the general case.

Theorem 2. *Let $G = (A, B, E)$ be a bipartite graph. Then,*

$$\text{co-im}(G) \;\geq\; \text{co-bd}(G) \;\geq\; 0.0698 \cdot \text{co-im}(G) - 4\,.$$

Proof. As before, we shall be concerned only with the second inequality. Set $\Delta = \text{co-bd}(G)$. Keeping in mind that $\mathrm{mis}(G) = 2^{\mathrm{bd}(G)}$, observe that

$$\mathrm{mis}(G) \;\leq\; \mathrm{mis}(G|_{V-\{v\}}) + \mathrm{mis}(G|_{V-\{v\}-N(v)}), \tag{1}$$

where the first term (over-)counts the maximal independent sets not containing v, and the second term counts those containing v. In accordance with inequality, we define a splitting process, or a (weighted) rooted splitting tree T, as follows.

Each inner node x of T is labelled by (G_x, v), where G_x is an induced subgraph of G, and $v \in V(G_x)$ is a vertex of degree 3 or more in G_x. At the root $G_x = G$; the leaves correspond to induced subgraphs of degree at most 2. An inner node x has two children, one corresponding to the graph obtained from G_x by removing v, the other corresponding to the graph obtained from G_x by removing v and all its neighbours (at least 4 vertices removed). The weight of the respective edge is defined as the number of vertices (respectively) removed. The weight of the node x, $w(x)$, is the defined as the sum of weights on the path from the root to x.

In view of (1), it holds that

$$2^{\mathrm{bd}(G)} \;\leq\; \sum_{x:\ \text{leaf of } T} 2^{\mathrm{bd}(G_x)}\,. \tag{2}$$

The strategy of proof is as follows. The leaves L of T will be split into $L^+ = \{x \mid w(x) \geq z\}$ and $L^- = \{x \mid w(x) < z\}$ according to a suitably defined threshold value z. Then, it is shown that the leaves in L^+ contribute little to the above sum, while the graphs G_x corresponding to $x \in L^-$ have co-im comparable with Δ. The value of z will be set later, in the course of analysis.

Upper-bounding the contribution of L^+

By Proposition 2, $\mathrm{bd}(G_x) \leq (n - w(x))/2$ for any $x \in T$. Thus,

$$\sum_{x \in L^+} 2^{\mathrm{bd}(G_x)} \;\leq\; 2^{n/2} \cdot \sum_{x \in L^+} 2^{-w(x)/2}\,.$$

It is readily checked that the right-hand side is maximized when T is the complete 1-4 tree, where every inner node has an outgoing edge of weight 1 and an outgoing edge of weight 4. Moreover, for an inner node x and it two children x_1 and x_2 it holds that

$$2^{-w(x_1)/2} + 2^{-w(x_2)/2} \;=\; 2^{-w(x)/2} \cdot (2^{-1/2} + 2^{-4/2}) \;<\; 2^{-w(x)/2} .$$

Therefore, the right-hand side is maximized when the leaves are immediate descendants of the inner nodes of weight $< z$.

Let $s(i)$ denote the number of nodes of weight i in the complete 1-4 tree. Then, $|L^+| \;\leq\; s(z)+s(z+1)+s(z+2)+s(z+3)$. A closer look reveals that $s(0), s(1), s(2) = 1,\; s(3) = 2$, and that for $i \geq 4,\; s(i) = s(i-1)+s(i-4)$, where the first term counts the strings with leading "1", and the second term counts those with leading "4". Moreover, it holds that $s(z) + s(z+1) + s(z+2) + s(z+3) = s(z+6)$. Finding the maximal absolute-value root $\alpha = 1.38028$ of the equation $x^4 = x^3 + 1$, and using the estimation $s(z) \leq \alpha^z$, we conclude that

$$|L^+| \;\leq\; s(z+6) \;\leq\; 8 \cdot 1.38028^z \;\leq\; 8 \cdot 2^{0.4649589\, z} \;\; ,$$

and the total contribution of L^+ to the right-hand side of (2) is bounded from above by

$$2^{n/2} \cdot \sum_{x \in L^+} 2^{-w(x)/2} \;\leq\; 8 \cdot 2^{n/2} \cdot 2^{-z/2} \cdot 2^{0.4649589\, z} \;\leq\; 8 \cdot 2^{n/2} \cdot 2^{-0.035\, z} .$$

Setting $z = \lceil (\Delta + 4)/0.035 \rceil$, where $\Delta = \text{co-bd}(G) = n/2 - \text{bd}(G)$, ensures that the total contribution of L^+ is at most $0.5 \cdot 2^{n/2} \cdot 2^{-\Delta} = 0.5 \cdot 2^{\text{bd}(G)}$.

Consequently, the total contribution of L^- to the right-hand side of (2) is at least

$$\sum_{x \in L^-} 2^{\text{bd}(G_x)} \;\geq\; 0.5 \cdot 2^{\text{bd}(G)} \;=\; 2^{n/2} \cdot 2^{-\Delta - 1} \, . \tag{3}$$

Upper bounding the contribution of L^-

To get an estimation from above on $\sum_{x \in L^-} 2^{\text{bd}(G_x)}$, consider $\text{bd}(G_x)$ for a leaf x of T. Since G_x has degree ≤ 2, Lemma 2 implies that:

$$\text{bd}(G_x) \;=\; (n - w(x))/2 - \text{co-bd}(G_x) \;\leq\; (n - w(x))/2 \;-\; 0.339 \cdot \text{co-im}(G_x) \, .$$

Keeping in mind that $w(x) = n - |V(G_x)|$, it follows that $\text{co-im}(G_x) \;\geq\; \text{co-im}(G) - w(x)/2$. Substituting this in the previous line yields

$$\text{bd}(G_x) \;\leq\; n/2 - 0.33\, w(x) - 0.339 \cdot \text{co-im}(G) \, ,$$

and thus,

$$\sum_{x \in L^-} 2^{\text{bd}(G_x)} \leq \sum_{x \in L^-} 2^{\frac{n}{2} - 0.33 w(x) - 0.339 \text{co-im}(G)} = 2^{\frac{n}{2} - 0.339 \text{co-im}(G)} \sum_{x \in L^-} 2^{-0.33 w(x)}$$

As before, the complete 1-4 tree yields the most general (i.e., the weakest possible) upper bound on the sum $\sum_{x \in L^-} 2^{-0.33\, w(x)}$, as it maximizes the number of nodes of any weight i in T. Since this time the contribution of the father node is dominated by that of its sons, it suffices to analyse the case when the leaves of L^- have weights $z-4, z-3, z-2$ or $z-1$. Arguing as before, we conclude that $|L^-| \leq s(z+2) \leq 2^{0.4649589(z+2)} < 2^{0.4649589z+1}$. That is,

$$\sum_{x \in L^-} 2^{-0.33\,w(x)} \;\leq\; |L^-| \cdot 2^{-0.33(z-4)} \;\leq\; 2^{0.4649589z+1} \cdot 2^{-0.33(z-4)} \;<\; 2^{0.1349589z+2.32}\,.$$

Now, $z < (\Delta+4)/0.035+1$, implying $0.1349589z + 2.32 \;\leq\; 3.856\Delta + 18$. The bottom line is:

$$\sum_{x \in L^-} 2^{\mathrm{bd}(G_x)} \;\leq\; 2^{n/2} \cdot 2^{-0.339 \cdot \text{co-im}(G)} \cdot 2^{3.856\Delta+18}\,. \tag{4}$$

We are ready to conclude the proof of Theorem 2. Combining (3) and (4) yields

$$2^{n/2} \cdot 2^{-\Delta-1} \;\leq\; \sum_{x \in L^-} 2^{\mathrm{bd}(G_x)} \;\leq\; 2^{n/2} \cdot 2^{-0.339 \cdot \text{co-im}(G)} \cdot 2^{3.856\Delta+18}\,.$$

Combining the two sides, it follows that $0.339 \cdot \text{co-im}(G) \;\leq\; 4.856\Delta + 19$, and, finally, $\text{co-im}(G) \;\leq\; 14.33\,(\Delta+4) \;=\; 14.33\,(\text{co-bd}(G)+4)$. □

One curious structural implication following at once from Theorem 2 is the following result (for asymptotically tight results see [4]):

Corollary 2. *Let G be a d-regular bipartite graph,* $d > 1$. *Then* $\mathrm{mis}(G) \leq 2^{(\frac{1}{2}-\epsilon)n}$ *for some universal* $\epsilon > 0$.

The reason is that by a trivial computation, for such graphs one has $\mathrm{im}(G) \leq \frac{n}{2} \cdot \frac{d}{2d-1}$, and since $\text{co-bd}(G)$ is proportional to $\text{co-im}(G)$, the conclusion follows.

3.2 A Constant Factor Polynomial Approximation Algorithm for co-bd(G)

Let us first give a polynomial time constant-factor approximation algorithm for $\text{co-im}(G)$. As before, G is bipartite.

Approx-CoIm: *Construct (greedily or otherwise) a maximal vertex-disjoint packing* $\mathcal{P}$ *of* P_2*'s (paths on 2 edges) in G. Remove all the vertices in* $\mathcal{P}$. *Output* $\widetilde{M}$, *the set of the remaining edges.*

Theorem 3. *The above algorithm produces an induced matching* $\widetilde{M}$ *with* $n/2 - |\widetilde{M}| \;\leq\; 5 \cdot \text{co-im}(G)$. *In particular, it provides a 5-approximation for* $\text{co-im}(G)$.

Proof. Observe that after the removal of P_2's in $\mathcal{P}$, the remaining induced graph consists of singletons and isolated edges, and thus $\widetilde{M}$ is indeed an induced matching.

Let M^* denote the maximum induced matching of G. Since every P_2 in the packing must contain at least one vertex outside of M^*, the size of $\mathcal{P}$ is at most $n - 2|M^*|$. Now, since each P_2 in $\mathcal{P}$ may hit at most 2 edges of M^*, at least $|M^*| - 2|\mathcal{P}|$ edges of M^* will survive the removal of $\mathcal{P}$. Thus,

$$|\widetilde{M}| \;\geq\; |M^*| - 2|\mathcal{P}| \;\geq\; |M^*| - 2\cdot(n - 2|M^*|) \;=\; 5|M^*| - 2n\,;$$

$$n/2 - |\widetilde{M}| \;\leq\; 5/2\,n - 5|M^*| \;=\; 5\cdot(n/2 - |M^*|) \;=\; 5 \cdot \text{co-im}(G) \qquad \square.$$

Theorem 3 combined with Theorem 2 yields a constant factor approximation algorithm for $\text{co-bd}(G)$:

Theorem 4. *The co-size of* $\widetilde{M}$ *produced by* **Appox-CoIm** *on input G, i.e.,* $n/2 - |\widetilde{M}|$, *approximates* $\text{co-bd}(G)$ *within a multiplicative factor of* $5 \cdot 14.3 < 72$.

4 Linear Boolean Width: Beyond the Triviality Bound

We show (constructively and efficiently) that every size-n graph has a balanced bipartition of its vertex set such that the boolean dimension of the associated bipartite graph is $\leq (1/2 - c)n$ for some universal $c > 0$. Combined with Proposition 4, this implies that $\text{lbw(G)} \leq (1/2 - \text{c}/2)\text{n}$; the argument therein provides also the corresponding linear arrangement of the vertices.

Two constructions are provided, the first deterministic and somewhat elaborate, the other is just the random uniform bipartition. We start with the former.

GoodBipartition: *Construct (greedily or otherwise) a maximal vertex-disjoint packing* $\mathcal{P}$ *of* P_2*'s (paths on 2 edges) in G. For each* $P_2 \in \mathcal{P}$ *mark the middle vertex. Partition the vertices into two equal- (up to* ± 2*) size sets so that*

(i) *for every* $P_2 \in \mathcal{P}$*, the marked and the unmarked vertices lay on different sides;*
(ii) *no edge (parity permitting) remaining after the removal of* $V(\mathcal{P})$ *is split.*

Since $\mathcal{P}$ is maximal, removing $V(\mathcal{P})$ one obtains an (induced) graph that consists of isolated edges and vertices. The required partition is obtained by placing the middle vertices of P_2's, one by one, on alternating sides. The same is done for surviving isolated edges and isolated vertices.

Let $H = (A, B, E')$ be the graph defined by this bipartition.

Theorem 5. *It holds that* $\text{co-im}(H) \geq \frac{1}{10}\, n$, *and* $\text{bd}(H) \leq (\frac{1}{2} - \frac{1}{143})n + O(1)$.

Proof. Let IM be a set of edges corresponding to a maximum induced matching of H, on vertices $V(IM)$. For each $P_2 \in \mathcal{P}$ we have at least one vertex of this P_2 not belonging to $V(IM)$, let $h(P_2)$ be such a vertex. For each $(u, v) \in IM$ we have either u or v (or both) a vertex of some $P_{uv} \in \mathcal{P}$. Fix for each $(u, v) \in IM$ arbitrarily such a P_{uv} and define a function $f : IM \to V(H) - V(IM)$ by $f((u, v)) = h(P_{uv})$. Since IM is an induced matching this function assigns to each vertex $h(P_{uv})$ at most two edges of IM. Thus there are at least $|IM|/2$ vertices in $V(H) - V(IM)$ and we have $|V(H)| \geq 2|IM| + |IM|/2$ which gives $\text{im}(H) \leq \frac{2}{5}n$, and, equivalently, $\text{co-im}(H) \geq \frac{1}{2}n - \frac{2}{5}n = \frac{1}{10}n$. By Theorem 2, this implies that $\text{co-bd}(H) \geq \frac{1}{10 \cdot 14.3}\, n = \frac{1}{143}\, n$. □

Next, we show that a statement similar to that of Theorem 5 holds also for a random uniform bipartition of $V(G)$, when each vertex of V is assigned randomly and independently either to side A or to side B, resulting in $H(A, B, E'')$. While the bound will be weaker, this structural result is of independent interest. The proof is left out of this extended abstract.

Theorem 6. *For* H *as above, it holds almost surely that* $\text{co-im}(H) \geq \frac{1}{28}\, n - o(n)$. *Consequently,* $\text{co-bd}(H) \geq \frac{1}{801}\, n - o(n)$.

5 Boolean Width: Beyond the Triviality Bound

We turn to the boolean-width of general graphs. Since we currently have much less understanding of boolean dimension of unbalanced partitions than of the balanced ones, in this section we provide an existential argument, which can nevertheless be turned into an exponential time algorithm with a relatively small exponent.

We shall need the following standard estimation for the sum of binomial coefficients, to be called here the Entropy Bound:

$$\sum_{i=0}^{cm} \binom{m}{i} \leq 2^{H(c)\,m} \tag{5}$$

where $H(p) = p\log_2 \frac{1}{p} + (1-p)\log_2 \frac{1}{1-p}$ is the entropy function.

Lemma 3. *Every graph G has $A \subset V(G)$ with $|A| = \frac{n}{3} \pm o(n)$ such that* $\mathrm{bd}(G_{A,V\text{–}A}) \leq \frac{n}{3} - \frac{n}{226} + o(n)$.

Proof. If there exists $S \subseteq V(G)$ such that $|S| = \frac{n}{226}$ and $|N(S) \cup S| \leq \frac{n}{3}$ we just take any set A of size $n/3$ containing $N(S) \cup S$. Then, $\mathrm{bd}(G_{A,V\text{–}A}) \leq \frac{n}{3} - \frac{n}{226}$, as no vertex in S has neighbours in $V - A$.

Otherwise, every set S of size $\frac{n}{226}$ has $|N(S) - S| \geq \frac{n}{3} - \frac{n}{226}$. The set A will be constructed by a random procedure by choosing every vertex v with probability $\frac{1}{3}$, randomly and independently from the others. We claim that almost surely two following events take place. First, $|A| = n/3 \pm o(n)$, and second, all sets $S \subseteq V - A$ of size $\frac{n}{226}$ have $|N(S) \cap A| > \left(\frac{1}{3} - 0.202\right) \cdot \left(\frac{n}{3} - \frac{n}{226}\right)$, which we short-cut as αn for the suitable α. Such an A will be called good.

Since $1 - \Pr[X \cap Y] \leq (1 - \Pr[X]) + (1 - \Pr[Y])$, it suffices to show that each of the two events holds almost surely *separately*. To bound the probabilities of failure, we use a suitable Chernoff-Hoeffding Bound.

Let Σ be the number of successes in r i.i.d. 0/1 events, each happening with probability p. Then by [10], $\Pr[\,\Sigma \leq (p-t)r\,] \leq e^{-2t^2 r}$ and $\Pr[\,\Sigma \geq (p+t)r\,] \leq e^{-2t^2 r}$. The desired bound on the probability of the first event follows at once with $o(n)$ standing for any sublinear function majorizing $n^{0.5}$, e.g., $6(n^{0.5+\epsilon})$. For the second event, the analysis is more involved.

Let S be any subset of $V(G)$ of size $\frac{n}{226}$. The probability that S causes a failure is

$$\Pr\left[\{S \subseteq V - A\} \wedge \{|N(S) \cap A| \leq \alpha n\}\right] =$$

$$= \Pr\left[\{|N(S) \cap A| \leq \alpha n\} \mid \{S \subseteq V - A\}\right] \cdot \Pr[S \subset V - A].$$

We start with upper-bounding the first factor in the above product. Then, the set S is fixed and is in $V - A$. The choosing process on the unfixed vertices in $V - S$ remains, however, unaltered. In particular, the vertices in $N(S) - S$ are chosen randomly and independently as before. Recall that by our assumption there are $\geq n/3 - n/226$ vertices in this set. Choosing $t = 0.202$, we get from the above Hoeffding Bound: $\Pr\left[|N(S) \cap A| \leq \left(\frac{1}{3} - 0.202\right)\left(\frac{n}{3} - \frac{n}{226}\right) \mid \{S \subseteq V - A\}\right] < e^{-2\cdot 0.202^2\left(\frac{n}{3} - \frac{n}{226}\right)} < \; < e^{-0.02684\, n}$. Thus,

$\Pr[\,S \textit{ is bad}\,] < e^{-0.02684\,n} \cdot \Pr[S \subset V{-}A] = e^{-0.02684\,n} \cdot \left(\frac{2}{3}\right)^{n/226} < e^{-0.02863\,n}$. Next, we apply the union bound summing over all sets S of size $\frac{n}{226}$. As always, the binomial coefficients are upper-bounded using the Entropy Bound from Equation (5): $\Pr[\,\textit{there exists a bad } S\,] \leq e^{-0.02863\,n} \cdot \binom{n}{n/226} < e^{-0.02863\,n} \cdot 2^{H(1/226)\,n} <$ $< e^{-0.00023\,n} = o(1)$. Thus, a random A is good almost surely for a large enough n. We proceed with upper bounding $\mathrm{bd}(G_{A,V-A})$ for a good A by counting the sets in $\mathcal{N}_{V-A}$, the family of neighbourhoods of subsets of $V-A$ in A.

Recall that $|V-A| \approx 2n/3$. The sets $S \subset V-A$ of size $i < n/226$ may contribute only as many as $\sum_{i=0}^{n/226} \binom{2n/3+o(n)}{i}$ distinct neigbourhoods in A. The contribution of sets $S \subset V-A$ of size $\geq n/226$ may be bounded as follows. In each such S mark an arbitrary subset $X \subseteq S$ of size precisely $\lceil n/226 \rceil$. Call two large S's equivalent if the same X was marked in both of them. Then, since every X sees at least αn vertices n, the contribution of the entire equivalence class of large sets defined X is at most $2^{n/3-\alpha n}$. The number of X's is at most $\binom{2n/3+o(n)}{n/226}$. By plugging in the numerical value of α and using the Entropy Bound from Equation (5) for $\binom{2n/3}{n/226}$, the entire contribution can be bounded by:

$$|\mathcal{N}_{V-A}| \leq \sum_{i=0}^{n/226} \binom{2n/3+o(n)}{i} + \binom{2n/3+o(n)}{n/226} \cdot 2^{n/3-\alpha n} \leq 2^{0.3286n+o(1)} \leq$$

$$\leq 2^{\frac{n}{3}-\frac{n}{226}+o(1)} ,$$

and the upper bound on $\mathrm{bd}(G_{A,V-A}) = \log_2 |\mathcal{N}_{V-A}|$ follows.

As an immediate consequence of Lemma 3 and Proposition 3 we obtain the following result:

Theorem 7. *For any graph G, it holds that* $\mathrm{bw}(G) \leq \frac{n}{3} - \frac{n}{672} + o(n)$.

6 Conclusion

Our results are the first non-trivial upper bounds on the value of a graph width parameter that hold *for every graph*. In this paper we gave three techniques to show such bounds, respectively Theorems 5, 6 and 7. At the moment the first two work only for linear boolean-width and the third is here applied only to boolean-width but it should work also for the linear case. We believe these bounds can be substantially improved. Combining Corollary 1 with Proposition 4 and Theorem 5 we can solve MAXIMUM WEIGHT INDEPENDENT SET and COUNTING INDEPENDENT SETS OF SIZE K in time $O^*(1.4108^n)$, and solve MINIMUM WEIGHT DOMINATING SET, MINIMUM WEIGHT TOTAL DOMINATING SET, MAXIMUM/MINIMUM WEIGHT INDEPENDENT DOMINATING SET, and counting versions of these, in time $O^*(1.9904^n)$. These runtimes beat brute-force search but faster algorithms exist in the literature, see [11]. Our goal was mainly to prove the viability of this new line of research by establishing structural qualitative results. The natural directions for further work are to improve the bounds and hence the runtime, and to increase the class of problems handled by Corollary 1.

References

1. Belmonte, R., Vatshelle, M.: Graph classes with structured neighborhoods and algorithmic applications. In: TCS (2013), http://dx.doi.org/10.1016/j.tcs.2013.01.011
2. Bui-Xuan, B.-M., Telle, J.A., Vatshelle, M.: Boolean-width of graphs. Theoretical Computer Science 412(39), 5187–5204 (2011)
3. Rödl, V., Duffus, D., Frankl, P.: Maximal independent sets in bipartite graphs obtained from boolean lattices. Eur. J. Comb. 32(1), 1–9 (2011)
4. Fomin, F.V., Kratsch, D.: Exact Exponential Algorithms, 1st edn. Texts in Theoretical Computer Science (2010)
5. Füredi, Z.: The number of maximal independent sets in connected graphs. Journal of Graph Theory 11(4), 463–470 (1987)
6. Hlinený, P., Oum, S.I.: Finding branch-decompositions and rank-decompositions. SIAM J. Comput. 38(3), 1012–1032 (2008)
7. Hoeffding, W.: Probability inequalities for sums of bounded random variables. Journal of the American Statistical Association 58(301), 13–30 (1963)
8. Ilinca, L., Kahn, J.: Counting maximal antichains and independent sets. Order 30(2), 427–435 (2013)
9. Kim, K.H.: Boolean matrix theory and its applications. Monographs and textbooks in pure and applied mathematics. Marcel Dekker (1982)
10. Vadhan, S.P.: The complexity of counting in sparse, regular, and planar graphs. SIAM Journal on Computing 31, 398–427 (1997)
11. Vatshelle, M.: New Width Parameters of Graphs. PhD thesis, University of Bergen (2012) ISBN:978-82-308-2098-8