Upper Bounds on the Boolean Width of Graphs with an Application to Exact Algorithms

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

1 **Department of Computer Science, University of Haifa**
Haifa, Israel
`yuri@cs.haifa.ac.il`

2 **Department of Informatics, University of Bergen**
Bergen, Norway
`[telle,vatshelle]@ii.uib.no`

Abstract

The *boolean dimension* bd of a bipartite graph and the *boolean width* bw of a graph were introduced in the recent [1] in the context of Fixed Parameter Tractable algorithms. While the algorithms based on these parameters compare well to those based e.g., on treewidth, it was also noticed that there are also some fundamental differences between these parameters and other standard parameters. Motivated in part by a perspective of designing a uniform framework for non-trivial exact algorithms for some NP- and $\#P$-Complete problems, and in part by a surprising structural observation that the boolean dimension of a bipartite graph H turns out to be precisely $\log_2(\mathrm{mis(H)})$ the logarithm of the number of the maximal independent sets in H (a very interesting and little understood combinatorial parameter), we study the extremal values of this parameter, as well as the structure of the corresponding extremal graphs. We establish a constant factor polynomial time approximation algorithm for the *boolean co-dimension* $|V(H)|/2 - \mathrm{bd(H)}$, and show that the boolean width of any graph G is at most $(1/3 - c_1)n$ for some constant $c_1 > 0$. We also show that the boolean dimension of a random balanced bipartition of any G is almost surely at most $(1/2 - c_2)n$ for another constant $c_2 > 0$, with application to *linear boolean width*.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases Exact Algorithm, Maximal Independent sets, Boolean-width

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

The boolean width bw(G) and the linear boolean width lbw(G) of a graph G were introduced in [1, 7] in the context of introducing a uniform framework beating brute-force search for several NP-hard graph problems. The framework is based on a branch decompositions of graphs, employing boolean dimension bd(G) as the cut function. In particular, it was shown in [1, 7] that weighted and counting versions of INDEPENDENT SET and DOMINATING SET can be solved in time $O^*(2^{2\mathrm{bw(G)}})$ or $O^*(2^{\mathrm{lbw(G)}})$, and time $O^*(2^{3\mathrm{bw(G)}})$ or $O^*(2^{2\mathrm{bw(G)}})$ respectively. At first glance these bounds are hardly exciting, as they are quite similar to bounds resulting from standard FPT algorithms based on treewdith, cliquewidth etc. However, it soon turned out that bw(G) and lbw(G) behave quite diffently from these other standard graph theoretic parameters. In particular (see [7] and references therein)

* Part of this work done while in Bergen supported by the Research Council of Norway.
† Supported by the Research Council of Norway.

Conference title on which this volume is based on.
Editors: Billy Editor, Bill Editors; pp. 1–12
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

it was shown that for interval graphs $\mathrm{bw(G)} \leq \mathrm{O(\log n)}$, and that for random graphs $\mathrm{bw(G)} \leq \mathrm{O(\log^2 n)}$. Moreover, the correspoding decompostions can be efficiently found. This contrasts strongly with other standard parameters, which on these families of graphs attain value $\Omega(n)$. It is natural to ask how far this approached can be pushed, i.e., what is the maximum possible value of bw(G) and lbw(G) on *any* graph G on n vertices. This is precisely the problem studied in the present paper.

Right from the outset it is clear that $\mathrm{bw(G)}, \mathrm{lbw(G)}$ can be a large as $\Omega(n)$, as shown by constant-degree expanders. However, the hidden constant is tiny, and appears to be less than 0.01. (The lower bound is not a part of this paper; it will be elborated and hopefully stregthened elsewhere.) Observe that in order to improve the running time of the best know (exact mildly exponential) algorithm for Maximum Independent Set it would suffice to show that $\mathrm{lbw(G)} \leq \mathrm{n}/4$, and provide means for finding suitable decomposition.

The trivial upper bounds for bw(G) and lbw(G) are $n/3$ and $n/2$ respectively. The main technical contribution of this paper is to show the existance of universal constants $c_1 > 0, c_2 > 0$ such that $\mathrm{bw(G)} < \mathrm{n}/3 - \mathrm{c}_1$, and $\mathrm{bw(G)} < \mathrm{n}/2 - \mathrm{c}_2$. While the constants we establish here are quite tiny and do not lead to algorithms improving those currently known, the fact that such constants exist at all, as opposed to all other standard parameters we are aware of, is fascinating. And there is a hope that in the future one will be able to get better upper bounds that will lead to significant better algorithms.

The key parameter, the boolean dimension bd(H) of a bipartite graph $H = (A, B, E)$ can be defined in several equivalent ways. The original definition is the logarithm (base 2) of the number of distinct neighborhoods (subsets of B) of all subsets of A. Equivalently, it is the *boolean rank* of the incidence (A vs. B) matrix M of H. [1]

But there is also an altogether different reason for our interest in bd(G), hitherto unnoticed and unexploited. There exists a surprising link between the boolean dimension of a (bipartite) graph H, and the number of maximal independent sets in H. More precisely, $\mathrm{bd(H)} = \log_2 \mathrm{mis(H)}$. Thus, the study of boolean dimension, and in turn of boolean width and linear boolean width, is intimately related to the study of mis(H), where mis(H) is the number of the maximal independent sets of H. The latter parameter, as well as its logarithm, has gained a considerable attention from both theoretical CS and combinatorial perspective. In particular, computing mis(H) is NP-Hard even for bipartite planar graphs [6], and approximating $\log_2 \mathrm{mis(H)}$ is a well know hard open problem even for graphs of minimal degree 4. In Combinatorics a considerable effort was devoted to finding good bounds on mis(H) for various classes of graphs. See, e.g., the recent [4] and the references therein. Thelink between bd(H) and mis(H) provide an unexpected and exiting (at least for us) perspective to the results of the current paper.

The technical contribution of this paper.
We focus here mostly on the basic structural properties of boolean width and boolean dimension, and develop suitable mathematical and algorithmical tools. Our main goal is first to understand the structure of bipartite graphs of high boolean width, and then to apply this understanding to producing non-trivial upper bound on bw(G) and lbw(G) of size-n graphs.

Unlike the other standard width parameters where the trivial (in a somewhat vaguely defined but quite concrete sense) upper bound is met e.g., by random graphs, the extremal

[1] Boolean rank of M is the logarithm (base 2) of the number of distinct vectors obtainable by taking boolean sums ($0{+}0 = 0$, $1{+}0 = 0{+}1 = 1{+}1 = 1$) of its column vectors. The first thing about it is that M and M^T have the same boolean rank. See [5] for details.

values of bw(G) and of lbw(G), as well as the structure of corresponding extremal graphs, is quite evasive. The trivial upper bounds here are $n/3$ and $n/2$ respectively, and it it also easy to show a lower bound of cn with a very small constant c using constant-degree expanders. One of our main results is to get below the trivial upper bound and show that $\mathrm{bw(G)} < \mathrm{n}(1/3 - \mathrm{c}_1)$ for another tiny constant $c_1 > 0$. This is mostly a qualitative result, as c_1 can clearly be improve, and the gap between c and $1/3 - c_1$ is huge. Similarly, we show that $\mathrm{lbw(G)} < \mathrm{n}(1/2 - \mathrm{c}_2)$ for some $c_2 > 0$. Yet, even this weak results lead to improved (i.e., non-trivial, but also not best known) exponential algorithms for hard standard problems. We hope that a future improvement in upper bounds on bw(G) and lbw(G) may lead to a uniform framework and significantly improved algorithms for this and many other hard problems.

Turning the maximal independent sets in a graph, it is well known and easy that $\log_2 \mathrm{mis(G)} = \mathrm{bd(G)}$ for a bipartite graph is at most $n/2$, and the maximum is attained (only) by a size-$n/2$ matching. We introduce and study the value of co-bd$(G) = n/2 - \mathrm{bd(G)}$, i.e., the high-end range of bd(G). We show a constant factor approximation algorithm for this value, as well as a stability result showing that the smaller this value is, the closer is the graph to the size-$n/2$ matching.

We also study the boolean dimension of balanced bi-partitions of G, and show, among other things, that a random bipartition has $\mathrm{bd} < \mathrm{n}(1/2 - \mathrm{c}^*)$, beating the triviality bound.

One of the contributions of this paper lies in the new appealing (at least in our opinion) open problems it raises. In addition to the central open problem of finding tight bounds on $\mathrm{bw(G)}, \mathrm{lbw(G)}$, other open problems are about the structure of graphs with $\mathrm{bw(G)} = \Omega(\mathrm{n})$ (i.e., does this class contain essentially anything beyond constant degree expanders?), the stability of mis(G) for general, not only bipartite, graphs, etc. We believe that our results and open problems provide a new perspective and a new motivation for classical hard algorithmic and combinatorial problems.

The 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 $(S, V{-}S)$ 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 $\log_2 \mathrm{mis(H)} = \mathrm{bd(H)}$ is at most $n/2$, and the maximum is attained (only) by a size-$n/2$ matching. We tie bd(H) to the size of the maximum induced matching of H. We introduce and study the value of co-bd$(H) = n/2 - \mathrm{bd(H)}$, i.e., the high-end range of bd(H). We show a constant factor approximation algorithm for this value, as well as a stability result showing that the smaller this value is, 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 simple graphs $G = (V, E)$. If G is bipartite, we use the standard notation $G(A, B, E)$. A neighborhood of a vertex $v \in V$ is denoted by $N(v)$, and $N(S)$ will denote a neighborhood of a set $S \subset V$, being the union of neighborhoods of $v \in S$, i.e., $N(S) = \cup_{v \in S} N(v)$. The *outer* neighbothood of S is $N(S) - S$. Any $S \subseteq V$ defines a cut $(S, V - S)$ and the corresponding bipartite graph $G_{S,V-S} = (S, V - S, E(S, V - S))$.

▶ **Definition 1** (Boolean dimension, boolean width and linear boolean width).
For a bipartite graph $H = (A, B, E)$, let $\mathcal{N}_A = \{N(X) \subseteq B : X \subseteq A\}$ be the family of neighborhoods of defined by sets $X \subseteq A$. The **boolean dimension** of H is defined as $\mathrm{bd(H)} = \log_2 |\mathcal{N}_\mathrm{A}|$.

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, a bipartition of V into A and B corresponding, respectively, to the δ-labels of leaves of T_a and T_b, and a bipartite graph $G_{A,B}$. The *boolean width* of (T, δ) is the maximum value of $\mathrm{bd(G_{A,B})}$ over all edges $\{a, b\}$ of T. The **boolean width** of G, denoted $\mathrm{bw(G)}$, is the minimum boolean width over all decomposition trees of G.

The **linear boolean width** of G, denoted $\mathrm{lbw(G)}$, is the minimum boolean width over all decomposition trees (T, δ) of G, where T is a caterpillar. Since the caterpillars correspond to linear arrangements of $V(G)$, $\mathrm{lbw(G)}$ is the maximal boolean dimension among the bipartite graphs defined by (prefix-suffix) bipartitions according to the best linear arrangement of $V(G)$.

One may wonder whether the $\mathrm{bd(H)}$ is well defined, since it is not obvious that $|\mathcal{N}_A| = |\mathcal{N}_B|$. A good answer to this is that there exists a bijection (actually, an involution) between the sets in $\mathcal{N}_A$ and set of all maximal independent sets of H. Indeed, given a set $S \in \mathcal{N}_A$, let X be the maximal set in A such that $N(X) = S$. Then X is unique, and $X \cup B - S$ is maximal independent. In the other direction, given a maximal independent set I, $B–I \subseteq \mathcal{N}_A$. Moreover, if I resulted from S, then S results from I.[2] Hence, $\mathrm{bd(H)} = \log_2 \mathrm{mis(H)}$

The boolean width parameter was originally introduced in [1] in the context of parameterized algorithms. In particular, using a natural dynamic programming approach it was shown there that

▶ **Theorem 2.** *[1] Given a graph G and a tree decomposition of boolean width k, one can solve weighted and counting versions of* Independent set *and* Dominating Set *in time $O^*(2^{2k})$ and $O^*(2^{3k})$, respectively.*

Using a similar approach, one can show that

▶ **Corollary 3.** *[1, 7] Given a given a linear arrangement of $V(G)$ of linear boolean width k, one can solve weighted and counting versions of* Independent set *and* Dominating Set *in time $O^*(2^k)$ and $O^*(2^{2k})$, respectively.*

Thus, the results of the current paper have implications on complexity of the exact algorithms for these problems. We shall discuss these implications in Section 6.

For $\mathrm{bd(G)}$, the following simple properties hold.

---

[2] This was observed by Nathann Cohen in a course of discussion with the authors. Later we have learned that a similar observation was made in [2].

▶ Proposition 1. $\mathrm{bd(G)}$ is monotone decreasing with respect to vertex removal. Moreover, such removal may decrease $\mathrm{bd(G)}$ by at most 1. Hence, $\mathrm{bd(G)} \leq |\mathrm{A}|, |\mathrm{B}|, |\mathrm{V}|/2$. This inequality is tight and met by a matching of size $|V|/2$.

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

Both properties easily follow from the original definition of bd, and we omit the details from this preliminary version.

For $\mathrm{bw(G)}$ we have the following. Let $n = |V|$.

▶ **Lemma 4.** *Let $A \subseteq V$ be a subset of vertices with $\frac{1}{3}n \leq |A| \leq \frac{2}{3}n$ and $G_{A,V-A}$, such that $\mathrm{bd(G_{A,V-A})} = (\frac{1}{3} - \epsilon)\mathrm{n}$ for some small $\epsilon \geq 0$. Then one can construct a decomposition tree of width $\leq (\frac{1}{3} - \frac{\epsilon}{3})n$. In particular, $\mathrm{bw(G)} \leq (\frac{1}{3} - \frac{\epsilon}{3})\mathrm{n}$.*

Proof. Without loss of generality, $|A| \leq n/2$; otherwise we switch to $V - A$. Partition V into three sets $A \cup X, B_1, B_2$ where X is disjoint from A and $|X| = \frac{2}{3}\epsilon n$, and also $|B_1|, |B_2| \leq (\frac{1}{3} - \frac{1}{3}\epsilon)n$. Taking an arbitrary refinement of this partition, and using Proposition 1, we conclude that it has the desired properties. ◀

Similarly, for $\mathrm{lbw(G)}$ one gets the following:

▶ **Lemma 5.** *Let $A \subseteq V$ be subset of vertices of size $n/2$ such that $\mathrm{bd(G_{A,V-A})} \leq (\frac{1}{2} - \epsilon) \cdot \mathrm{n}$. Then one can construct a linear arrangement of the vertices of the linear boolean width at most $(\frac{1}{2} - \frac{1}{2}\epsilon) \cdot n$. In particular, $\mathrm{lbw(G)} \leq (\frac{1}{2} - \frac{1}{2}\epsilon) \cdot \mathrm{n}$.*

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 i elements in this arrangement. In view of Proposition 1, it suffices to examine $\mathrm{bd(G_i)}$ where $G_i = G_{A_i,V-A_i}$ and $n/2 - \epsilon/2 \leq in/2 + \epsilon/2$. However, Proposition 1 implies that $|\mathrm{bd(G_i)} - \mathrm{bd(G_{i+1})}| \leq 1$. This, together with the assumption on $\mathrm{bd(A_{n/2})}$ yields the required bound. ◀

The main technical goal of this paper is the study of extremal values of $\mathrm{bw(G)}, \mathrm{lbw(G)}$. The following notions will prove useful in this study:

▶ **Definition 6.** Let G be a graph on n vertices. Defined $\mathrm{im(G)}$ 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 $\mathrm{im(G)} \leq \mathrm{n}/2$ and this bound is met only by a size-$n/2$ matching. To study the high-end range of $\mathrm{bd(G)}$ and $\mathrm{im(G)}$ we define co-im$(G) = \frac{n}{2} - \mathrm{im(G)}$ and the boolean co-dimension co-bd$(G) = \frac{n}{2} - \mathrm{bd(G)}$.

3 The Boolean Dimension and Boolean co-Dimension of a Bipartite graph

We start with the following proposition 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:

▶ Proposition 2. It holds that

$$\mathrm{im(G)} \;\leq\; \mathrm{bd(G)} \;\leq\; \mathrm{im(G)} \cdot \log_2(\mathrm{n}/\mathrm{im(G)}) \cdot (1 + \phi(\mathrm{n}/\mathrm{im(G)})),$$

where $\phi(x)$ never exceeds 0.088, and tends to 0 as x tends to 2.

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 $|A| \le n/2$, 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) v for every $v^* \in S^*$, we conclude that the subgraph of G induced by (S^*, S') is an induced matching. Thus, $|\mathrm{S}^*| \le \mathrm{im(G)}$, and consequently

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

where $H(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$ is the entropy function. Approximating $H(p)$ by $F(p) = p \log_2 \frac{1}{p} + p$ and computing the value of $n/2 \cdot H(p) = n/2 \cdot F(p) \cdot [H(p)/F(p)]$ at $p = \mathrm{im(G)/(n/2)}$, we arrive at the desired conclusion. ◀

Thus, im(G) is a $\log n$-approximation of bd(G), and the quality of approximation improves as bd(G) grows. However, when bd(G) is close to $n/2$, i.e., when co-bd(G) is small, Proposition 2 yields only that im(G) cannot be much less than $n/4$. We will strengthen this, and show that when co-bd(G) is small, so is co-im(G). Moreover, co-bd(G) and co-im(G) are linearly related. We start with a special case:

▶ Proposition 3. Let $G = (A, B, E)$ be a bipartite graph of degree at most 2. Then,

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

Proof. The first inequality was already established in Proposition 2. For the second inequality, observe that both co-bd(G) and co-im(G) are additive with respect to disjoint union of graphs, and therefore one may w.l.o.g., assume that G is connected. I.e., it is either C_n, the (even) n-cycle, or P_n, the n-path. Since $\mathrm{bd(G)} = \log_2 \mathrm{im(G)}$, the reccurence relation from [3] for computing $c(k) = \mathrm{im(C_k)}$ and $p(k) = \mathrm{im(P_k)}$ may be employed. Namely, $c(k) = c(k-2) + c(k-3)$, and also $p(k) = p(k-2) + p(k-3)$. The initial conditions are $c(1) = 0$, $c(2) = 2$, $c(3) = 3$ and $p(1) = 1$, $p(2) = 2$, $p(3) = 2$. The size of the maximum induced matching is $\lfloor \frac{k}{3} \rfloor$ and $\lfloor \frac{k+1}{3} \rfloor$ respectively. To sum up, one needs to lower-bound the expressions

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

Combining numerical computations and inductive arguments, we conclude that the minimum is achieved on the 8-cycle C_8, and its value is about 0.339036. ◀

We continue with the general case.

▶ **Theorem 7.** *Let $G = (A, B, E)$ be a bipartite graph. Then, there exists a universal constant $c > 0.01$ such that*

$$\text{co-im}(G) \;\ge\; \text{co-bd}(G) \;\ge\; c \cdot \text{co-im}(G).$$

As before, we shall be concerned only with the second inequality. Since we believe this is an interesting result of independent interest, and our numerical constants are likely to be far from the truth, we present two different proofs. Both reduce the problem to the degree 2 case. The first is simpler, the second more technical, but yielding a better constant.

Proof. A: Assume that co-bd$(G) = \Delta$. The basic process is as follows. Start with G, and, as long as there exists a vertex v of degree 3 or more, remove v and (any) three of its neighbours. What remains is a graph G' of degree at most 2, and Proposition 3 applies. The key point is to bound the number of the removed vertices and neighbours in terms of Δ.

Let K be the set of the removed vertices, let K^* be the set of the removed neighbours, and let $k = |K|$. Clearly, $|K^*| = 3k$. How many maximum independent sets I such that $|I \cap K| = i$ can there be? At most $\binom{k}{i}$ times 2 to the power of half the number of the remaining free vertices, i.e., $V(G) - K - N(I)$. Since $|V(G) - K - N(I)| \le n - k - 3i$, one gets

$$\mathrm{im(G)} \; = \; 2^{\mathrm{n/2-\Delta}} \; \le \; \sum_{\mathrm{i=0}}^{\mathrm{k}} \binom{\mathrm{k}}{\mathrm{i}} 2^{\mathrm{(n-k-3i)/2}} \; = \; 2^{\mathrm{n/2}} \cdot 2^{\mathrm{-k/2}} \cdot \sum_{\mathrm{i=0}}^{\mathrm{k}} \binom{\mathrm{k}}{\mathrm{i}} 2^{-1.5\mathrm{i}} \; =$$

$$2^{n/2} \cdot 2^{-k/2} \cdot (1 + 2^{-1.5})^k \; \le \; 2^{n/2 - k/2 + 0.437k} \; = \; 2^{n/2 - 0.063k} \; .$$

Thus, $0.063\,k \le \Delta$, and therefore $k < 15.88\,\Delta$.

Now, observe that the size of the graph G' is $n - 4k$, and that $\mathrm{bd(G')} \; \ge \; \mathrm{n/2} - \Delta - 4\mathrm{k} \; = \; |\mathrm{G'}|/2 - \Delta - 2\mathrm{k}$. Thus, co-bd$(G') \; \le \; 32.76\,\Delta$. Observe also that $\mathrm{im(G)} - 4\mathrm{k} \; \le \; \mathrm{im(G')} \; \le \; \mathrm{im(G)}$, and therefore co-im$(G') \; = \; n/2 - 2k - \mathrm{im(G')}$, is between co-im$(G) - 2k$ and co-im$(G) + 2k$. A finer anaysis shows that one can gain a bit: removing a vertex and its three neighbours may destroy only 3 (and not 4) edges from the maximum induced matching of G. Thus, $\mathrm{im(G)} - 3\mathrm{k} \; \le \; \mathrm{im(G')}$, implying co-im$(G') \; \le \;$ co-im$(G) - k$. By Proposition 3 applied to G',

$$\text{co-im}(G) \; \le \; k + \text{co-im}(G') \; \le \; k + (1/0.339) \cdot \text{co-bd}(G') \; \le \; (15.88 + 1/0.339 \cdot 32.76)\Delta \; \le \; 112.52\Delta.$$

Thus, co-im$(G) \; \le \; 112.52 \cdot$ co-bd(G), or co-bd$(G) \ge 0.0088 \cdot$ co-im(G). ◀

Proof. B: As before, let co-bd$(G) = \Delta$. Observe that

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

where the first term counts the maximal independent sets containing v, and the second term counts those not containing v. We define a splitting process, or a (weighted) rooted splitting tree T, where each inner node is labelled by (G^*, v), where G^* is an induced subgraph of G, and v is a vertex of G^* of degree 3 or more. Such a node will have two children, one corresponding to the graph obtained from G^* by removing v, the other corresponding to the graph obtained from G^* by removing v and all its neighbours. The weight of the respective edge is the number of vertices just removed. At the root there is G. The leaves correspond to induced subgraphs G^* of degree at most 2.

To sum up, we have constructed a rooted binary tree T such that each inner node has one outgoing edge of weight 1, and another of weight at least 4. Each leaf x of T corresponds to an induced subgraph G_x of degree at most 2. The weight of x, $w(x)$, is the sum of weights on the path from the root to x. In view of (1), it holds that

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

The strategy will be as follows. We introduce a threshold value z (to be explicitly presented later) such that the nodes of weight $w(x) \ge z$ contribute little to the above sum, while the graphs G_x corresponding to leaves of weight $w(x) < z$ typically have small co-im.

We start with upper-bounding the contribution of the above-threshold (i.e., heavy) leaves of T to this sum. Call this set of leaves L^+.

Since $\mathrm{bd}(G_x) \leq\ n/2 - w(x)/2$ for any $x \in T$, it holds that $\sum_{x\in L^+} 2^{\mathrm{bd}(\mathrm{G_x})} \ \leq\ 2^{n/2} \cdot \sum_{x\in L^+} 2^{-w(x)/2}$. Observe that for a node x in T with two sons x_1, x_2, it holds that $2^{-w(x_1)/2} + 2^{-w(x_1)/2} < 2^{-w(x)/2}$, as $2^{-1/2} + 2^{-4/2} < 1$. Thus, we may assume w.l.o.g., that all $x \in L^+$ are immediate descendants of nodes of weight $< z$ in T.

It is easy to get convinced that the worst tree T for us is the complete 1-4 tree, where every inner node has an outgoing edge of weight 1 and outgoing edges of weight 4. Let $s(i)$ denote the number of nodes of weight i in such a tree. Clearly, $s(i)$ is the number of 1-4 strings that sum up to i. It holds that $s(0), s(1), s(2) = 1,\ s(3) = 2$, and for $i \geq 4,\ \ s(i) = s(i-1) + s(i-4)$, where the first term counts the string with leading "1", and the second term counts those with leading "4". Moreover, for such T it holds that $|L^+| \ \leq\ s(z) + s(z+1) + s(z+2) + s(z+3) = s(z+6)$. Finding the maximal absolute value root of the equation $x^4 = x^3 - 1$, we conclude that

$$|L^+| \ \leq\ s(z+6) \ \leq\ 5.5 \cdot 1.38028^z \ \leq\ 5.5 \cdot 2^{0.4649589\, z} \quad .$$

Thus, the total contribution of L^+ is at most

$$2^{n/2} \cdot \sum_{x\in L^+} 2^{-w(x)/2} \ \leq\ 5.5 \cdot 2^{n/2} \cdot 2^{-z/2} \cdot 2^{0.4649589\, z} \ \leq\ 5.5 \cdot 2^{n/2} \cdot 2^{-0.035\, z}.$$

Choosing $z = \lceil(\Delta{+}5)/0.035\rceil$ we ensure that the contribution of L^+ is at most $0.173{\cdot}2^{n/2}{\cdot}2^{-\Delta}$, and therefore the contribution of the below-threshold leaves must be at least $0.827 \cdot 2^{n/2} \cdot 2^{-\Delta}$. Call the set of these leaves L^-.

To analyse the contribution of L^-, we combine the last conclusion with (2) and with Proposition 3, and get

$$0.827{\cdot}2^{n/2}{\cdot}2^{-\Delta} \ \leq \sum_{x\in L^-} 2^{\mathrm{bd}(\mathrm{G_x})} \ = \sum_{x\in L^-} 2^{(n/2-\ w(x)/2)\ -\ \text{co-bd}(G_x)} \ \leq \sum_{x\in L^-} 2^{n/2-\ w(x)/2\ -\ 0.339\cdot\text{co-im}(G_x)}\ .$$

Since $\text{co-im}(G_x) \ \geq\ \text{co-im}(G) - w(x)/2$, it follows that

$$0.827 \cdot 2^{-\Delta} \ \leq\ 2^{-0.339\cdot\text{co-im}(G)} \cdot \sum_{x\in L^-} 2^{-(1/2-0.339/2)\,\cdot\, w(x)}$$

and

$$0.827 \cdot 2^{-\Delta} \cdot 2^{0.339\cdot\text{co-im}(G)} \ \leq\ \sum_{x\in L^-} 2^{-0.33w(x)}\ .$$

As before, the 1-4 tree is extremal, as it maximizes the number of nodes of any weight i in T. Since this time the contribution of the father is dominated by that of its sons, it suffices to analyse the case when the leaves of L^- have weight $z-4, z-3, z-2$ or $z-1$, and thus $|L^-| \leq s(z+2) \leq 0.827 \cdot 2^{0.4649589z}$. Thus, the right-hand-side is upper-bounded by

$$|L^-| \cdot 2^{-0.33(z-4)} \ \leq\ 0.827 \cdot 2^{0.4649589z} \cdot 2^{-0.33(z-4)} \ <\ 0.827 \cdot 2^{0.1349589z+1.4}\ ,$$

and, keeping in mind that $z < (\Delta+5)/0.035 + 1$, we conclude that

$$\text{co-im}(G) \ \leq\ (0.1349589\, z{+}\Delta{+}1.4)/0.339 \ \leq\ 14.33{\cdot}(\Delta{+}4.3) \ =\ 14.33{\cdot}(\text{co-bd}(G){+}4.3)\ .$$

◀

One curious structural implication following at once from Theorem 7 is we following (known) result:

▶ **Corollary 8.** *If G is a regular bipartite graph, then* $\mathrm{bd(G)} \leq (1/2 - \epsilon)\mathrm{n}$ *for some* ϵ*.*

Having established the linear relation between co-bd(G) and co-im(G), we are now in position to design a constant factor approximation algorithm for co-bd(G). We start with presenting a polynomial time constant-factor approximation algorithm for co-im(G). As before, G is bipartite.

Appox-CoIm: *Construct (greedily or otherwise) a maximal vertex-disjoint packing $\mathcal{P}$ of P_3's (paths of length 2) in G. Remove all the vertices in $\mathcal{P}$. Output the set of the remaining edges.*

▶ **Theorem 9.** *The above algoritm produces an induced matching $\widetilde{M}$ with* $n/2 - |\widetilde{M}| \leq 5 \cdot \text{co-im}(G)$*.*

Proof. Observe that after the removal of P_3's in $\mathcal{P}$, the remaining graph consists of singletons and isolated edges, and thus the output is indeed an induced matching.

Let IM denote the maximum induced matching of G. Observe that every P_3 in the packing must contain at least one vertex outside of IM. Thus, the size of $\mathcal{P}$ is at most $2\,\text{co-im}(G)$. Moreover, since each P_3 in $\mathcal{P}$ may meet at most 2 edges of IM, removing all of them leaves at least $|IM| - 4\,\text{co-im}(G) = n/2 - 5\,\text{co-im}(G)$ edges. ◀

Theorem 9 together with Theorem 7 yields a constant factor approximation algorithm for co-bd(G):

▶ **Theorem 10.** *There exists a polynomial-time algoritm approximating* co-bd(G) *within a factor of* $5 \cdot 14.3 < 72$*.*

4 Linear Boolean Width: Breaking the Triviality Bound

We turn to general graphs, and show, constructively, by a polynomial-time algorithm, that every graph G has a balanced bipartition of its vertex set where the boolean dimension of the associated bipartite graph beats the triviality bound. Combined with Lemma 5 this implies a constructive result for linear boolean-width of general graphs, beating the triviality bound.

Let us first note that it is not hard to show a lower bound of $\Omega(n)$. E.g., as shown by Kostochka and Melnikov [], the *bisection width* of a random 3-regular graph H is almost surely $> 0.1n$. That is, any G_H corresponding to a balanced bipartition of H has at least that many edges. Thus, it suffices to show that bd(G) of a bipartite G, of degree at most 3 with $|E(G)| > 0.1n$, is $\Omega(n)$. It follows, e.g., from the trivial observations that for such G, $\mathrm{im(G)} = \Omega(\mathrm{G})$. The argument is quite coarse, and it yields quite a miserable constant. We postpone its elaboration (or replacement) to the future work.

For simplicity, assume $G = (V, E)$ has an even number of vertices and recall that $A \subseteq V$ is balanced if $|A| = \frac{1}{2}|V|$. Our goal will be to find a balanced subset $A \subseteq V$ such that $\mathrm{bd(G_{A,V-A})} \leq (\frac{1}{2} - \mathrm{c})|\mathrm{V}|$ for some $c > 0$. As with the constant in the above lower bound, the constant c that we are presently able to establish seems to be too small, and far off the mark. On the other hand, the mere fact that such c exists for Boolean dimension is new and somewhat surprising.

We start with an approximation algorithm.

Approx-Bipartition: *Construct (greedily or otherwise) a maximal vertex-disjoint packing $\mathcal{P}$ of P_3's (paths of length 2) and C_3's (cycles of length 3) in G. For each $P_3 \in \mathcal{P}$ mark the middle vertex, and for each $C_3 \in \mathcal{P}$ mark any (single) vertex. Partition the vertices in a balanced manner so that*

(i) *for every $P_3, C_3 \in \mathcal{P}$, 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.*

Let H be the graph defined by this bipartition.

▶ **Theorem 11.** *It holds that* $\text{co-im}(H) \geq n/10$. *Consequently,* $\text{bd(H)} \leq \text{n}/2 - 1/144\,\text{n}$.

Proof. The packing $\mathcal{P}$ of G, being maximal, induces a maximal packing $\mathcal{P}'$ of H. I.e., $\mathcal{P}'$ consists of disjoint P_3's in H, and the removal of $V(\mathcal{P}')$ yields a graph of degree 1. We start with lower-bounding co-im(H).

Given the vertex partition (A, B), $\mathcal{P}$ and its maximality, and forgetting about the actual structure of H, we claim that the maximum possible value of im(H) is attained e.g., by an induced matching IM with the following properties. The union of IM and $\mathcal{P}'$ should consists of disjoint union of *as many as possible* P_5's, and a subset of original P_3's. Each P_5 has the two outer edges in IM and the two inner edges in $\mathcal{P}'$. The remaining edges of IM are a subset of the edges of $\mathcal{P}'$. The easy proof of extremality follows from case analysis and simple variational arguments.

Having characterized the structure of IM, it remains to compute its size. Let D be the number of P_3's in $\mathcal{P}'$. Let x be maximum possible number of P_5's as above. Then, $|IM| = 2x + (D - x) = x + D$. On the other hand, $x \leq \min\{(n - 3D)/2,\ D\}$. Thus, if $D \leq n/5$, then $|IM| = 2D$, and if $D \geq n/5$, then $|IM| = (n - D)/2$. In any case, $|IM| \leq n/2 - n/10$.

We have shown that for the graph H obtained from G via bipartition of **Approx-Bipartition**, it holds that co-im$(H) \geq n/10$. Consequently, by Th. 7, co-bd$(H) \geq n/(10 \cdot 14.3)$. ◀

Interestingly, a statement similar to that of Th. 11 holds for a random balanced bipartition of G as well.

▶ **Theorem 12.** *For H obtained by random balanced partition of $V(G)$, almost surely it holds that*
co-im$(H) \geq n/36$. *Consequently,* $\text{bd(H)} \leq \text{n}/2 - 1/515\,\text{n}$.

Proof. (Sketch) Consider the same packing $\mathcal{P}$ of G; let $D = |\mathcal{P}|$. The random bipartition of the vertices will, typically, split a quarter of P_3's, and three quarters of C_3's in it. Call this splitted set $\mathcal{S} \subseteq \mathcal{P}$, and let $\mathcal{S}'$ be the corresponding set in H.

Consider two cases: $D \geq n/9$ and $D \leq n/9$.

In the first case, employing the argument already used in the proof of Th. 9, we conclude that
$2\,\text{co-im}(H) \geq |\mathcal{S}'| \approx D/4$, and thus co-im$(H) \geq n/36$.

In the second case, let $K = V(\mathcal{S})$, $R = V(\mathcal{P}) - K$, and $W = V - K - R$. A simple variational argument shows that the maximum possible size of IM is attained when it consists only of edges of the type (K, W), (R, W) and (W, W). The number of edges of the first type is at most $2/3\,|K|$; those are the P_5's as in the proof of Th. 11. The number of edges of the second type is at most $|R|$. Finally, the number of the edges of the type (W, W) is at most (about) $W/4$, since $G|_W$ is a degree 1 graph, and we use a random partition. Therefore,

$$\text{im(G)} \leq \text{D}/4 \cdot 2 + 3\text{D}/4 \cdot 3 + (\text{n} - 3\text{D})/4 = \text{n}/4 + 2\text{D}\,,$$

and co-im$(G) \geq n/4 - 2D \geq n/4 - 2/9 \cdot n = n/36$.

In both cases co-im$(H) \geq n/36$, and by Th. 7, we conclude that co-bd$(H) \geq n/(36 \cdot 14.3)$. ◀

5 Boolean Width: Breaking 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 neverthless be turned into an exponential time algorithm with a relatively small exponent.

▶ **Lemma 13.** *Every graph G has $A \subset V(G)$ with $\frac{n}{3} \leq |A| \leq \frac{n}{2}$ such that* $\mathrm{bd}(\mathrm{G}_{\mathrm{A,V\text{-}A}}) \leq \frac{\mathrm{n}}{3} - \frac{\mathrm{n}}{226}$.

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 a set A of size $n/3$ containing $N(S) \cup S$. Then, $\mathrm{bd}(\mathrm{G}_{\mathrm{A,V\text{-}A}}) \leq \frac{\mathrm{n}}{3} - \frac{\mathrm{n}}{226})$ since no vertex in S has a neighbour 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 $1/3$, randomly and independently from the others. We claim that almost surely the two events hold: first, $|A| = (1 - o(1)) \cdot n/3$, 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 shortcut as αn for a 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 *serapately.* To bound the probabilities of failure, we use the Hoeffding Bound. Let X be the number of successes in r i.i.d. 0/1 events, each happening with probability p. Then,

$$\Pr(X \leq (p - t)r) \leq e^{-2t^2 r}.$$

The desired bound on the probability of the first event follows at once. For the second event, the analysis is more technical.

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

$$\Pr\left(\{S \subset V – A\} \wedge \{|N(S) \cap A| \leq \alpha n\}\right) = \Pr\left(\{|N(S) \cap A| \leq \alpha n\} \mid \{S \subset V – A\}\right) \cdot \Pr(S \subset V – A).$$

The probability of the first factor in the above product can be bounded as follows. The set S is fixed and is in $V - A$. The choosing process on the remaining $V - S$ vertices remains, however, unaltered. Since the vertices in $N(S) \setminus S$ are chosen randomly and independently as before, and there are $n/3 - n/226$ vertices in this set, taking $t = 0.202$, we get from Hoeffdings bound:

$$\Pr\left(|N(S) \cap A| \leq \left(\frac{1}{3} - 0.202\right)\left(\frac{n}{3} - \frac{n}{226}\right) \mid \{S \subset V – A\}\right) < e^{-2 \cdot 0.202^2 (n/3 - n/226)} < e^{-0.0268n}.$$

We now use the union bound summing over all sets S of size $\frac{n}{226}$.

$$\Pr(\text{There exists a bad } S) \leq e^{-0.0268n} \cdot \Pr(S \subset V – A) \cdot \binom{n}{n/226} = e^{-0.0268n} \cdot \binom{2n/3}{n/226} \approx$$

$$e^{-0.0268n} \cdot 2^{2/3 \cdot H(3/2 \cdot 1/226)} = o(1) .$$

Thus, a random A is good almost surely for a large enough n. We proceed with upper bounding $\mathrm{bd}(\mathrm{G}_{\mathrm{A,V\text{-}A}})$ 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 $\sum_{i=0}^{n/226} \binom{2n/3}{i}$ 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 $\binom{2n/3}{n/226}$. Thus, pluging in the numerical value of α and using the entropy bound for $\binom{2n/3}{n/226}$, the entire contribution can be bounded by:

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

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

As an immediate consequence of Lemma 13 and Lemma 4 we get

▶ **Theorem 14.** *For any graph G, it holds that* $\mathrm{bw(G)} \leq \frac{\mathrm{n}}{3} - \frac{\mathrm{n}}{672}$.

6 Discussion

None of the other parameters comparable to boolean-width seems to have the same potential to give exact algorithms directly, since either their value can be as large as n such as for tree-width and clique-width or there are no known FPT algorithms with good enough runtime to beat the brute force algorithms.

Bounding the boolean-width in it self poses interesting challenges with lots of connections to well studied concepts in mathematics.

References

1 Binh-Minh Bui-Xuan, Jan Arne Telle, and Martin Vatshelle. Boolean-width of graphs. *Theoretical Computer Science*, 412(39):5187–5204, 2011.

2 Vojtech Rödl Dwight Duffus, Peter Frankl. Maximal independent sets in bipartite graphs obtained from boolean lattices. *Eur. J. Comb.*, 32(1):1–9, 2011.

3 Zoltán Füredi. The number of maximal independent sets in connected graphs. *Journal of Graph Theory*, 11(4):463–470, 1987.

4 Liviu Ilinca and Jeff Kahn. Counting maximal antichains and independent sets. *ArXiv e-prints*, February 2012.

5 Ki Hang Kim. *Boolean matrix theory and its applications.* Monographs and textbooks in pure and applied mathematics. Marcel Dekker, 1982.

6 Salil P. Vadhan. The complexity of counting in sparse, regular, and planar graphs. *SIAM Journal on Computing*, 31:398–427, 1997.

7 Martin Vatshelle. *New Width Parameters of Graphs.* PhD thesis, University of Bergen, 2012. ISBN:978-82-308-2098-8.