On Set Intersection Representations of Graphs *

S. Jukna †‡

Abstract

The intersection dimension of a bipartite graph with respect to a type L is the smallest number t for which it is possible to assign sets $A_x \subseteq \{1,\ldots,t\}$ of labels to vertices x so that any two vertices x and y from different parts are adjacent if and only if $|A_x \cap A_y| \in L$. The weight of such a representation is the sum $\sum_x |A_x|$ over all vertices x. We exhibit explicit bipartite $n \times n$ graphs whose intersection dimension is: (i) at least $n^{1/|L|}$ with respect to any type L, (ii) at least $\sqrt{n}$ with respect to any type of the form $L = \{k, k+1, \ldots\}$, and (iii) at least $n^{1/|R|}$ with respect to any type of the form $L = \{k \mid k \bmod p \in R\}$, where p is a prime number. We also show that any intersection representation of a Hadamard graph must have weight about $n \ln n / \ln\ln n$, independent on the used type L. Finally, we formulate several problems about intersection dimensions of graphs related to some basic open problems in the complexity of boolean functions.

1 Introduction

We consider representations of graphs as intersection graphs of families of sets. The size of the underlying set serves as a measure of complexity. Various conditions on when we draw an edge between the sets give various measures. Our motivation is that, for *bipartite* graphs, these measures capture the computational complexity of boolean functions (see Section 4).

We are interested in representing a graph $G = (V,E)$ by assigning to each vertex $x \in V$ a finite set A_x of *labels* so that we can then distinguish those pairs of vertices x, y that are edges from those that are not simply by looking at the number $|A_x \cap A_y|$ of their common labels. The parameter we are interested in is the total number $|\bigcup_{x \in V} A_x|$ of used labels. A *type* of such a representation is a set L of nonnegative integers such that

$$|A_x \cap A_y| \in L \quad \text{iff} \quad xy \in E. \tag{1}$$

Given a subset $L \subseteq \{0,1,\ldots\}$, the *intersection dimension* $\Theta_L(G)$ of a graph G with respect to L is the smallest number of labels used in an intersection representation of G under the intersection rule (1). The *absolute dimension* is the minimum $\Theta(G) = \min_L \Theta_L(G)$ over all types L.

*To appear in: *J. Graph Theory*

†Research supported by a DFG grant SCHN 503/4-1.

‡Institute of Mathematics, Akademijos 4, LT-80663 Vilnius, Lithuania

As customary, the complement of a graph $G=(V,E)$ is the graph $\overline{G}=(V,F)$, where $xy\in F$ iff $xy\notin E$. Since every intersection representation of G with respect to a type L is also an intersection representation of the complement graph $\overline{G}$ with respect to the complementary type $\overline{L}=\{0,1,\ldots\}\setminus L$, we have that $\Theta_L(\overline{G})=\Theta_{\overline{L}}(G)$. In particular, $\Theta(\overline{G})=\Theta(G)$. Also, if H is an *induced* subgraph of G, then $\Theta_L(H)\leq\Theta_L(G)$ for all types L.

Different types L lead to different measures $\Theta_L(G)$. The following types have recently drawn considerable attention.

- The *threshold-k dimension* $\Theta_k(G)$ is the intersection dimension with respect to the type $L=\{k,k+1,\ldots\}$. The *threshold dimension* is the minimum $\Theta_{\rm thr}(G)=\min_k\Theta_k(G)$ over all threshold values k.

- The *modular dimension* of G is the minimum $\Theta_{\rm mod}(G)=\min_L\Theta_L(G)$ over all modular types of the form $L=\{\ell\mid\ell \bmod p\in R\}$, where $p\geq 2$ is an arbitrary integer and $R\subseteq\{0,1,\ldots,p-1\}$ is an arbitrary set of residues. The simplest is the *parity dimension* $\Theta_{\rm odd}(G)$ corresponding to the case when $p=2$ and $R=\{1\}$.

We will now briefly summarize what was known about these measures.

1.1 Known results

Among all intersection dimensions the most intensively studied was the threshold-1 dimension $\Theta_1(G)$. In this case the intersection rule is: $xy\in E$ iff $A_x\cap A_y\neq\emptyset$. The threshold-1 dimension is also known as the *edge clique covering number*, that is, the smallest number t of complete subgraphs $G_1,\ldots,G_t$ of a given graph G covering all its edges. To see the connection, just associate with each vertex $x\in V(G)$ the set of labels $A_x=\{i\mid x\in V(G_i)\}$.

For the threshold-1 dimension the following bounds are known.

- A classical result of Erdős, Goodman and Pósa [16] states that $\Theta_1(G)\leq\Theta_1(K_{n,n})=n^2$ for every graph G on $2n$ vertices, where $K_{n,n}$ is a complete bipartite $n\times n$ graph. Moreover, a desired covering of G can be achieved by only using simplest complete subgraphs: edges and triangles.

- Chung and West [9], and Eaton, Gould and Rödl [12] extended this to $\Theta_k(K_{n,n})\sim n^2/k$.

- That random graphs have smaller threshold-1 dimension $\Theta_1(G)$ about $n^2/(\ln n)^2$ was proved by Frieze and Reed [17], by improving a similar estimate of Bollobás, Erdős, Spencer and West in [7].

- Eaton and Grable [11] extended this result to threshold-k dimensions $\Theta_k(G)$, k being any fixed number.

- Using a probabilistic construction, Alon [1] proved a general upper bound $\Theta_1(\overline{G})=O(\Delta^2\ln n)$ for all n-vertex graphs G of maximum degree Δ.

- Eaton and Rödl [10] proved that this upper bound in not very far from the truth: n-vertex graphs G of maximum degree Δ with $\Theta_1(\overline{G}) = \Omega(\Delta^2 \ln(n/\Delta)/\ln\Delta)$ exist.

If an n-vertex graph G has maximal degree Δ, then we already know that $\Theta_1(\overline{G}) = O(\Delta^2 \ln n)$ [1]. Note, however, that we cannot expect similar upper bound for the graph G itself, just because $\Theta_1(G)$ is equal to the number of edges if the graph is triangle-free. Still, such upper bounds can be achieved if one allows larger threshold values k.

- $\Theta_k(G) = O(\Delta^2(n/\Delta)^{1/k})$ for every fixed k (Eaton, Gould and Rödl [12]).
- $\Theta_{\text{thr}}(G) = O(\Delta^2 \ln n)$ (Eaton and Rödl [10]).
- $\Theta_{\text{thr}}(\overline{G}) = O(\Delta(\ln\Delta)^2 \ln n)$ if the graph G is bipartite (Eaton and Rödl [10]).

A related result of Alon [2] states that about $kn^{1/k}$ complete bipartite subgraphs are necessary and sufficient in order to cover the edges of a complete n-vertex graph K_n so that each edge belongs to at most k of the subgraphs. This generalizes (from $k = 1$ to an arbitrary k) the classical result of Graham and Pollak [18] that $n-1$ edge-disjoint complete bipartite subgraphs are necessary and sufficient in order to cover K_n.

For the maximum $\Theta_{\text{odd}}(n)$ of the parity dimension $\Theta_{\text{odd}}(G)$ over all n-vertex graphs, the following bounds are known:

- $n - \sqrt{2n} - \lceil \log_2 n \rceil \leq \Theta_{\text{odd}}(n) \leq n-1$ (Eaton and Grable [11]).

Some structural properties of modular dimensions with respect to the types of the form $L = \{\ell \mid (\ell \bmod p) \geq k\}$ were considered by McMorris and Wang in [25].

1.2 Intersection dimensions and complexity of boolean functions

In this paper we are interested in various intersection dimensions of *bipartite* graphs $G = (V,E)$ with a fixed partition $V = V_1 \cup V_2$ of their vertices; the sets V_1 and V_2 are sometimes referred to as *color classes* of the graph. In this case we relax the intersection condition, and only require that the intersection rule (1) must be satisfied by pairs of vertices from different color classes—the sets of labels of pairs of vertices from the same color class may intersect arbitrarily! That is, the relaxed intersection rule for bipartite graphs $G = (V_1 \cup V_2, E)$ is:

$$|A_x \cap A_y| \in L \text{ iff } xy \in E, \text{ as long as } x \in V_1 \text{ and } y \in V_2.$$

Hence, in the bipartite case, we do not care about what the value of $|A_x \cap A_y|$ is, if both vertices x and y belong to the same color class. The relaxed measures will be denoted by lower case theta: $\theta_L(G), \theta_{\text{thr}}(G), \theta_{\text{mod}}(G)$, $\theta(G)$, etc. Most interesting of these measures is the absolute dimension $\theta(G) = \min_L \theta_L(G)$.

Note that the relaxed measures $\theta_L(G)$ are only defined for bipartite graphs $G = (V_1 \cup V_2, E)$, and for such graphs we always have that $\theta_L(G) \leq \Theta_L(G)$. But for some bipartite graphs the gap between these two measures may be very large: for example, $\Theta_1(K_{n,n}) = n^2$ but $\theta_1(K_{n,n}) = 1$.

As in the case of general graphs, the threshold-1 dimension $\theta_1(G)$ of a bipartite graph $G = (V_1 \cup V_2, E)$ is equal to its *edge biclique covering number*, that is, the smallest number t of complete bipartite subgraphs $B_1, \ldots, B_t$ of G covering all its edges. To see this connection, just associate with each vertex $x \in V(G)$ the set of labels $A_x = \{i \mid x \in V(B_i)\}$.

Remark 1. [Motivation] The relaxation of intersection rules for bipartite graphs is motivated by an intimate relation between the resulting intersection dimensions of graphs and the computational complexity of boolean functions. Given a bipartite graph $G = (V_1 \cup V_2, E)$ with $|V_1| = |V_2| = 2^m$, we can encode the vertices x in each color class by vectors $\vec{x} \in \{0,1\}^m$. After the encoding is fixed, the graph G defines a boolean function $f_G : \{0,1\}^{2m} \to \{0,1\}$ in $2m$ variables by $f_G(\vec{x},\vec{y}) = 1$ iff $xy \in E$. Let $SYM(f_G)$ be the smallest number t such that the function $f_G(\vec{x},\vec{y})$ can be computed by a depth-2 formula of the form

$$f_G(\vec{x},\vec{y}) = \varphi(g_1(\vec{x},\vec{y}), \ldots, g_t(\vec{x},\vec{y})),$$

where each g_i is an AND of some variables and/or their negations, and $\varphi : \{0,1\}^t \to \{0,1\}$ is an arbitrary symmetric boolean function, that is, a boolean function whose output only depends on the number of 1's in the input vector. Since each set $B_i = \{xy \mid g_i(\vec{x},\vec{y}) = 1\}$ forms a complete bipartite graph, this implies that $\theta_L(G) \le t$, where L is the subset of $\{0,1,\ldots,t\}$ such that $\varphi(a_1,\ldots,a_t) = 1$ iff $|\{i \mid a_i = 1\}| \in L$. Hence, $SYM(f_G) \ge \theta(G)$.

To find an explicit boolean function f in m variables with $SYM(f) \ge 2^{(\log m)^\alpha}$ for some $\alpha \to \infty$ is an old problem whose solution would have important consequences in computational complexity (see Section 4). By what was said, this problem would be solved by exhibiting an explicit bipartite $n \times n$ graph G of absolute dimension $\theta(G) \ge 2^{(\ln\ln n)^\alpha}$.

Another bridge between bipartite graphs and boolean functions is given by the fact that $\log_2 \theta_1(G)$ is precisely the nondeterministic communication complexity of f_G [26].

Motivated by the connection with the computational complexity of boolean functions, in this paper we are interested in finding *explicit* bipartite $n \times n$ graphs whose intersection dimension is large. An ultimate goal is to construct graphs whose absolute dimension $\theta(G) = \min_L \theta_L(G)$ or at least the modular dimension $\theta_{\text{mod}}(G)$ is $\Omega(n^\varepsilon)$ for a constant $\varepsilon > 0$. As mentioned above, this would have important consequences in complexity theory of boolean functions (see Section 4 for further discussion). Easy counting shows (see Proposition 2) that

bipartite $n \times n$ graphs of absolute dimension $\theta(G) = \Omega(n)$ *exist*.

The problem, however, is to prove a similar *explicit* lower bound, that is, a lower bound $\theta(G) = \Omega(n^\varepsilon)$ for *explicitly given* graphs G. Unfortunately, no explicit graphs with $\theta(G) \gg \log_2 n$ or even $\Theta(G) \gg \log_2 n$ are known so far.

Note that $\theta(G) \ge \log_2 n$ is a trivial lower bound achieved by any bipartite *twin-free* $n \times n$ graph, that is, by any bipartite graph where no two vertices in one color class have the same set of neighbors: different vertices then require different sets of labels.

Larger explicit lower bounds were only known for several simplest intersection dimensions. Let M_n be a bipartite $n \times n$ graph consisting of n vertex disjoint edges (a perfect matching). Since

$\theta_{\text{odd}}(G)$ is just the rank over $GF(2)$ of the adjacency matrix of G (see Proposition 3), we obtain $\theta_1(M_n) = \theta_{\text{odd}}(M_n) = n$. In fact, it is shown in [21] that if a bipartite graph G contains M_n, then $\theta_1(G) \geq n^2/|E(G)|$. Based on the observation of Eaton and Rödl [10] that $\theta_k(G) \geq \theta_1(G)^{1/k}$ (see Section 2.3), hight lower bounds can be also obtained on the threshold-k dimension for any constant threshold value k. Except of these, however, no explicit lower bounds $\theta_L(G) = \Omega(n^{\varepsilon})$ were known for other types L.

In this paper we prove lower bounds $\Omega(n^{\varepsilon})$ for some modular types as well as for *arbitrary* threshold types (Theorems 2 and 3). We also prove such a lower bound for the absolute dimension $\theta(G)$ but only under a restriction on sets of labels used in the representation of graphs (Theorem 4). An unconditional super-linear lower bound $\Omega(n \ln / \ln \ln n)$ with no restrictions on the used type L or on the form of used sets A_x of labels, is proved for the "weight" of representations $\{A_x \mid x \in V(G)\}$, that is, for the total sum $\sum_{x \in V(G)} |A_x|$ of numbers of labels (Theorem 5). Finally, we give one result (Theorem 6) related to the Log-Rank Conjecture in communication complexity. In the last section we list several open problems whose solution would have great consequences in the computational complexity of boolean functions.

2 Our results

In this section we present our main results; their proof are given in the next section. But before we begin, let us first show that the measures $\theta_L(G)$ are well-defined for all graphs G and all types L, two trivial types $L = \emptyset$ and $L = \{0, 1, 2, \ldots\}$ being the only exceptions.

Proposition 1. *For every nontrivial type L and for every bipartite $n \times n$ graph G, we have that $\theta_L(G) \leq n + k$, where k is the smallest integer with $|\{k, k+1\} \cap L| = 1$.*

Proof. Let $G = (V_1 \cup V_2, E)$ be a bipartite $n \times n$ graph, and L be any nontrivial type. Suppose that $k \notin L$ and $k+1 \in L$ (the case when $k \in L$ and $k+1 \notin L$ is dual). Take some set K with $|K| = k$ elements, and let $N(x) \subseteq V_2$ be the set of all neighbors of $x \in V_1$. Assign the set $A_x = K \cup N(x)$ to each vertex $x \in V_1$, and the set $A_y = K \cup \{y\}$ to each vertex $y \in V_2$. We then have that $|A_x \cap A_y|$ is equal to $k+1 \in L$ if $xy \in E$, and is equal to $k \notin L$ if $xy \notin L$. □

2.1 General bounds

We start with some general bounds on the absolute dimension $\theta(G) = \min_L \theta_L(G)$ of bipartite $n \times n$ graphs G. Since every such graph can be covered by at most n stars, we have a trivial upper bound $\theta(G) \leq \theta_1(G) \leq n$. Moreover, this trivial upper bound cannot be substantially improved.

Proposition 2. *For every n, bipartite $n \times n$ graphs G with $\theta(G) \geq (n-1)/2$ exist.*

Proof. We have at most 2^{2tn} possible encodings of $2n$ vertices by subsets of $\{1, \ldots, t\}$, and at most 2^{t+1} possibilities to choose the type $L \subseteq \{0, 1, \ldots, t\}$. Hence, at most $2^{2tn+t+1}$ bipartite $n \times n$ graphs can have absolute dimension at most t. On the other hand, we have 2^{n^2} such graphs, implying that some of them need $t \geq (n^2-1)/(2n+1) \geq (n-1)/2$ labels. □

For n-vertex graphs of bounded maximal degree Δ, we have tighter upper bounds: since $\Theta(\overline{G}) = \Theta(G)$, the above mentioned upper bound of Alon [1] implies that $\Theta(G) = O(\Delta^2 \ln n)$.

Our first result gives an upper bound on $\theta(G)$ for *bipartite* graphs which is Δ times smaller than that for arbitrary graphs. By a *bipartite complement* of a bipartite graph $G = (V_1 \cup V_2, E)$ we will mean the bipartite graph $G^c = (V_1 \cup V_2, E^c)$, where $E^c = (V_1 \times V_2) \setminus E$.

Theorem 1. *For every bipartite $n \times n$ graph G of maximal degree Δ, $\theta_1(G^c) = O(\Delta \ln n)$. Moreover, graphs of maximal degree Δ with $\theta(G) = \Omega(\Delta \ln(n/\Delta))$ exist.*

2.2 Modular dimensions

The simplest among all modular dimensions is the parity dimension $\theta_{odd}(G)$ of G with respect to the type $L = \{\ell \mid \ell \bmod 2 = 1\}$. The adjacency matrix of a (labeled) bipartite graph $G = (V_1 \cup V_2, E)$ is a $0/1$ matrix M whose rows correspond to vertices $x \in V_1$, columns to vertices $y \in V_2$, and $M[x, y] = 1$ iff $xy \in E$.

Proposition 3. *For every bipartite graph G, $\theta_{odd}(G)$ is equal to the rank of the adjacency matrix of G over $GF(2)$.*

Proof. From linear algebra we know that the rank of any $0/1$ matrix M over $GF(2)$ is the smallest number r such that it is possible to assign vectors $\vec{x} = (x_1, \ldots, x_r)$ in $\{0,1\}^r$ to rows/columns $x \in V_1 \cup V_2$ so that the entry $M[x, y]$ in the x-th row and y-th column is the scalar product $\langle \vec{x}, \vec{y} \rangle$ of $\vec{x}$ and $\vec{y}$ over $GF(2)$. To obtain an intersection representation of G with respect to the parity type $L = \{\ell \mid \ell \bmod 2 = 1\}$, it is enough to take $A_x = \{i \mid x_i = 1\}$ and observe that $\langle \vec{x}, \vec{y} \rangle = |A_x \cap A_y| \bmod 2$. □

To capture more complicated modular measures, in Section 3.2 we introduce a less direct algebraic argument, and use it to prove the following

Theorem 2. *Let $L = \{k \mid k \bmod p \in R\}$ for some prime number p and some subset R of $r = |R|$ residues. Then, for every bipartite $n \times n$ graph G of maximum degree Δ, we have*

$$\theta_L(G) \geq (n/r\Delta)^{1/(p-1)} \quad \textit{and} \quad \theta_L(G^c) \geq (n/\Delta)^{1/r}.$$

Since for every natural number s there is a prime p with $s < p \leq 2s$, Theorem 2 implies

Corollary 1. *For any type L with $|L| = s > 0$, and for any bipartite $n \times n$ graph G of maximum degree Δ, we have*

$$\theta_L(G) \geq (n/s\Delta)^{1/(2s-1)} \quad \textit{and} \quad \theta_L(G^c) \geq (n/\Delta)^{1/s}.$$

2.3 Threshold dimensions

Next, we consider threshold dimensions of graphs. For every positive integer k and every bipartite graph G, we have

$$\theta_1(G)^{1/k} \le \theta_k(G) \le \theta_{k-1}(G)+1 \le \theta_1(G)+k-1.$$

Nontrivial here is only the first inequality which was observed by Eaton and Rödl in [10]: having a threshold-k representation $\{A_x\}$ of G, we can look at k-element subsets of labels as new labels, and assign to each vertex x the set A'_x of all k-element subsets of A_x; this gives a threshold-1 representation of G using $\binom{\theta_k(G)}{k} \le \theta_k(G)^k$ labels.

If M is an *n-matching*, that is, a bipartite $n\times n$ graph consisting of n vertex-disjoint edges, then clearly $\Theta_1(M)=\theta_1(M)=n$, because no two such edges can be contained in a complete (even complete bipartite) graph. Hence, $\theta_k(M)\ge n^{1/k}$ for every threshold value k. However, the general upper bound $\Theta_{\rm thr}(G)=O(\Delta^2\ln n)$ proved in [10] for graphs of maximal degree Δ implies that

$$\theta_{\rm thr}(M)\le\Theta_{\rm thr}(M)=O(\ln n). \qquad (2)$$

This can also be shown directly: if $\binom{t}{k}\ge n$, then we can assign to both endpoints of each edge of M its *own* k-element subset of $\{1,\dots,t\}$.

Hence, at least for some graphs, using larger threshold values k may drastically decrease their threshold dimension. So, a natural question is: What graphs have large threshold dimension *independent* of the used threshold value k?

We show that Hadamard graphs are such graphs. Recall that an *Hadamard matrix* of order n is an $n\times n$ matrix with entries ±1 and with row vectors mutually orthogonal (over the reals). A graph associated with an Hadamard matrix (or just an Hadamard graph) of order n is a bipartite $n\times n$ graph H where two vertices are adjacent if and only if the corresponding entry of the Hadamard matrix is equal $+1$.

Theorem 3. *For every bipartite $n\times n$ Hadamard graph H, both $\theta_{\rm thr}(H)$ and $\theta_{\rm thr}(H^{\rm c})$ are at least $\Omega(\sqrt{n})$.*

The bounds above imply that the parity and threshold dimensions are incomparable, and the gaps may be even *exponential* in both directions. For this it is enough to compare the corresponding intersection dimensions of an $n\times n$ matching M and of a bipartite Sylvester graph S: this last graph is a bipartite $n\times n$ graph with $n=2^r$, where vertices in each color class are identified with subsets of $\{1,\dots,r\}$, and two vertices x and y are adjacent iff $|x\cap y|$ is odd. Hence, $\theta_{\rm odd}(S)\le r=\log_2 n$. On the other hand, Proposition 3 implies that $\theta_{\rm odd}(M)=n$. Since each Sylvester graph is also a Hadamard graph, Theorem 3 together with the upper bound (2) yields the following trade-offs between the parity and threshold dimensions.

Corollary 2. $\theta_{\rm odd}(M)/\theta_{\rm thr}(M)=\Omega(n/\ln n)$ *and* $\theta_{\rm thr}(S)/\theta_{\rm odd}(S)=\Omega(\sqrt{n}/\ln n)$.

The example of Sylvester graphs shows another interesting fact: some Ramsey graphs have very small parity dimension. Namely, Pudlák and Rödl show in [30] that S contains an induced $\sqrt{n}\times\sqrt{n}$ subgraph which is Ramsey, meaning that neither the graph nor its bipartite complement contains a copy of $K_{s,s}$, for s much larger than $\ln n$. Since $\theta_{\text{odd}}(S) \leq \log_2 n$ and since the intersection dimension of *induced* subgraphs does not exceed that of the original graph, this implies that some bipartite Ramsey graphs have logarithmic parity dimension.

2.4 Arbitrary types

When trying to prove large lower bounds (larger than $\log_2 n$) on the absolute dimension $\theta(G) = \min_L \theta_L(G)$ of explicit bipartite graphs, we are faced with two problems: the "adversary" (trying to represent the graph with as few labels as possible) is allowed

(i) to choose an *arbitrary* type L, and

(ii) to assign vertices x *arbitrary* sets A_x.

If both are allowed then, as mentioned in the Introduction, we are unable to say anything more than that some explicit graphs—namely, all twin-free graphs—require at least $\log_2 n$ labels. The lower bounds above correspond to the case when we allow (ii) but restrict (i). Now we look at what happens if we allow (i) but restrict (ii).

2.4.1 Balanced representations

A natural intersection representation of any (non necessarily bipartite) graph is to assign each vertex x the set A_x of its incident edges. This gives a threshold-1 representation of the graph. The representation itself has, however, an additional property that $|A_u \cap A_v \cap A_x| = 0$ for any triple u, v, x of distinct vertices.

Motivated by this example, we say that a representation $\{A_x \mid x \in V_1\} \cup \{B_u \mid u \in V_2\}$ of a bipartite graph $G = (V_1 \cup V_2, E)$ is *balanced* if there are two vertices $x \neq y \in V_1$ such that

$$|A_x \cap B_u \cap B_v| = |A_y \cap B_u \cap B_v| \text{ for all vertices } u \neq v \in V_2.$$

It is easy to see that every bipartite $n \times n$ graph has a balanced threshold-1 representation using at most n labels: assign to each vertex $x \in V_1$ the set $A_x \subseteq V_2$ of its neighbors, and assign to each vertex $u \in V_2$ a single-element set $B_u = \{u\}$. This is a balanced intersection representation with respect to the type $L = \{1\}$. A natural question is: Can the number of labels be substantially reduced by using another types L? Our next result says that, at least for Hadamard graphs, this is not possible.

Theorem 4. *Every balanced representation of a bipartite $n \times n$ Hadamard graph with respect to any type $L \subseteq \{0, 1, \ldots\}$ must use at least $n/4$ labels.*

2.4.2 The weight of representations

So far we were interested in the size $|\bigcup_{x \in V} A_x|$ of representations $\{A_x \mid x \in V\}$, that is, in the total number of used labels. Another important measure of representations is their *weight* $\sum_{x \in V} |A_x|$.

For a bipartite graph G, let $w_L(G)$ denote the weight analog of $\theta_L(G)$, that is, the smallest weight of an intersection representation of G with respect to the type L.

These measures were mainly considered with respect to the threshold-1 type $L = \{1, 2, \ldots\}$, and to the parity type consisting of all odd natural numbers. For the minimal weight $w_1(G)$ of intersection representations with respect to the threshold-1 type the following bounds were proved by Chung, Erdős and Spencer [8], and independently by Tuza [34]:

- $w_1(G) = O(n^2/\ln n)$ for every n-vertex graph $G = (V, E)$, and
- $w_1(G) = \Omega(|E(G)|/r)$ if G contains no complete bipartite $r \times r$ subgraph;
- hence, graphs G with $w_1(G) = \Omega(n^2/\ln n)$ exist.

Erdős and Pyber [15] improved the upper bound $w_1(G) = O(n^2/\ln n)$ by showing that every n-vertex graph has a threshold-1 intersection representation such that each label is used by at most $O(n/\ln n)$ vertices.

The interest in the minimal weight $w_{\text{odd}}(G)$ with respect to the parity type is motivated by the fact that this is precisely the smallest number of wires in a depth-2 circuit with unbounded fanin parity gates computing the linear transformation $M\vec{x} = \vec{y}$ over $GF(2)$, where M is the adjacency matrix of G (see [4]). A super-linear lower bound on this measure was proved by Alon, Karchmer and Wigderson [4]:

- For any bipartite $n \times n$ Hadamard graph H we have $w_{\text{odd}}(H) = \Omega(n \ln n)$.

For Hadamard graphs, this lower bound is optimal, because $w_{\text{odd}}(S) \leq n \log_2 n$ for a bipartite $n \times n$ Sylvester graph S.

We show that the argument, used in [4] for the parity type, can be extended to *arbitrary* types.

Theorem 5. *For every bipartite $n \times n$ Hadamard graph H and for every type $L \subseteq \{0, 1, \ldots\}$, we have that $w_L(H) = \Omega(n \ln n / \ln \ln n)$.*

In fact, this lower bound, as well as a lower bound $\Omega(n)$ on the number of labels in a balanced representation (Theorem 4), holds for any bipartite $n \times n$ graph $G = (V_1 \cup V_2, E)$ with the following property: For every two vertices $x \neq y \in V_1$, there is a set $S \subseteq V_2$ of $|S| = \Omega(n)$ vertices such that every vertex $u \in S$ is adjacent to x and non-adjacent to y.

2.5 Tight representations and the Log-Rank Conjecture

Finally, we prove one result related to the so-called "Log-Rank Conjecture" in communication complexity stating that the deterministic communication complexity of any $0/1$ matrix is at most poly-logarithmic in its real rank (see, for example, [26]). Lovász and Saks [24] noted that this conjecture is equivalent to the following Rank-Coloring Conjecture for graphs stating that $\chi(G) \leq 2^{(\ln r)^{O(1)}}$ for any graph G, where $\chi(G)$ is the chromatic number of G and $r = \text{rk}(G)$ is the real rank of the adjacency matrix of G.

At some time it was thought that $\chi(G) \leq \text{rk}(G)$. This was conjectured in 1976 by C. van Nuffelen [27]. The first counterexample to van Nuffelen's conjecture was obtained by Alon and

Seymour [3]. They constructed a graph with chromatic number 32 and with an adjacency matrix of rank 29. Razborov [32] then showed that the gap between the chromatic number and the rank of the adjacency matrix can be super-linear, and Raz and Spieker [31] showed that the gap can even be super-polynomial. The best result known so far is due to Nisan and Wigderson [26]. It gives an infinite family of graphs with rank r and with chromatic number $\chi(G) = 2^{\Omega(\log_2 r)^{\alpha}}$, where $\alpha = \log_3 6 > 1$.

Nisan and Wigderson [26] have also found a yet another equivalent formulation of the Log-Rank Conjecture in terms of the maximal number $\mathrm{cliq}(G)$ of edges of a complete bipartite graph lying in G or in G^{c}. The conjecture then translates to: There is a constant $c > 0$ such that $\mathrm{cliq}(G) \geq nm/2^{(\ln r)^c}$ for every bipartite $n \times m$ graph G with $\mathrm{rk}(G) = r$.

To approach this conjecture, Sgal [33] suggested to first solve it with the rank r of G replaced by the intersection dimension of G under "tight" representations. A *k-tight* representation of a bipartite graph $G = (V_1 \cup V_2, E)$ is an intersection representation of type $L = \{k\}$ with an additional condition that $|A_x \cap A_y| \in \{k-1, k\}$ for all $x \in V_1$ and $y \in V_2$. Let $\theta_{\mathrm{tight}}(G)$ be the minimum, over all integers $k \geq 1$, of the smallest number of labels in a k-tight representation of G. The intersection matrix of each k-tight representation of G has entries $k-1$ and k only, and if we subtract $k-1$ from each entry, we obtain a $0/1$ adjacency matrix of G whose rank is at most 1 plus the number of used labels. Hence, $\mathrm{rk}(G) \leq \theta_{\mathrm{tight}}(G)$.

A weakened version of Log-Rank Conjecture, suggested by Sgal [33], states that there is a constant $c > 0$ such that $\mathrm{cliq}(G) \geq nm/2^{(\ln r)^c}$ for every bipartite $n \times m$ graph G with $\theta_{\mathrm{tight}}(G) = r$.

We show that Sgal's conjecture is true for all k-tight representations, as long as k is at most poly-logarithmic in the total number of used labels.

Theorem 6. *If a bipartite $n \times m$ graph G has a k-tight representation using r labels, then*

$$\mathrm{cliq}(G) \geq nm/4r^{2k}.$$

Now we turn to the proofs.

3 Proofs

3.1 General bounds: Proof of Theorem 1

We have at most 2^{2tn} possible encodings of $2n$ vertices by subsets of $\{1, \ldots, t\}$, and at most 2^{t+1} possibilities to choose the type $L \subseteq \{0, 1, \ldots, t\}$. Hence, at most $2^{2tn+t+1}$ bipartite $n \times n$ graphs can have intersection dimension at most t. Since we have at least $(n/\Delta)^{n\Delta/2}$ bipartite $n \times n$ graphs of maximum degree at most Δ (see, for example, Proposition 2.1 in [10]), this implies that some of degree-Δ graphs require $t = \Omega(\Delta \ln(n/\Delta))$ labels, independent on what type L we use. This proves the second claim of Theorem 1.

To prove the first claim, let $G = (V_1 \cup V_2, E)$ be a bipartite $n \times n$ graph of maximal degree Δ, and let $G^{\mathrm{c}} = (V_1 \cup V_2, E^{\mathrm{c}})$ with $E^{\mathrm{c}} = (V_1 \times V_2) \setminus E$ be its bipartite complement. Our goal is to cover the set E^{c} of edges of G^{c} by $O(\Delta \ln n)$ complete bipartite subgraphs $S \times T \subseteq F$ of G^{c}.

To do this, we construct $S \times T$ via the following probabilistic procedure: pick every vertex $x \in V_1$ independently, with probability $p = 1/\Delta$ to get a random subset $S \subseteq V_1$, and let

$$T = \{y \in V_2 \mid xy \in E^{\mathrm{c}} \text{ for all } x \in S\}$$

be the set of all those vertices $y \in V_2$ that are adjacent in G^{c} to *all* vertices in S. It is clear that each so constructed complete bipartite graph $S \times T$ is a subgraph of G^{c}. An edge $xy \in E^{\mathrm{c}}$ of G^{c} is covered by such a subgraph if x was chosen in S and none of (at most Δ) neighbors of y in G was chosen in S. Hence, this happens with probability at least $p(1-p)^{\Delta} \geq p\mathrm{e}^{-p\Delta} = p/\mathrm{e}$.

If we apply this procedure t times to get t complete bipartite subgraphs, then the probability that xy is covered by *none* of these subgraphs does not exceed $(1-p/\mathrm{e})^t \leq \mathrm{e}^{-tp/\mathrm{e}}$. Hence, the probability that some edge of G^{c} remains uncovered is smaller than

$$n^2 \mathrm{e}^{-tp/\mathrm{e}} = \exp(2\ln n - t/(\mathrm{e}\Delta)),$$

which is smaller than 1 for $t = 2\mathrm{e}\Delta \ln n$. □

3.2 Modular dimension: Proof of Theorem 2

In this section we first describe an algebraic approach to proving lower bounds on modular dimensions, and then prove Theorem 2 itself.

3.2.1 An approach

Let $G = (V_1 \cup V_2, E)$ be a labeled bipartite $n \times n$ graph with color classes $V_1 = \{x_1, \ldots, x_n\}$ and $V_2 = \{y_1, \ldots, y_n\}$. When trying to estimate the (relaxed) intersection dimension $\theta_L(G)$ of G with respect to some type L, we are faced with the following problem. We have two systems $\mathcal{A} = \{A_1, \ldots, A_n\}$ and $\mathcal{B} = \{B_1, \ldots, B_n\}$ of (not necessarily distinct) subsets of $\{1, \ldots, t\}$ and the only knowledge about these systems is that the intersection sizes $|A_i \cap B_j|$ must be consistent with the given graph G: $|A_i \cap B_j| \in L$ iff $x_i y_j \in E$. Hence, the whole information about the pair $\mathcal{A}, \mathcal{B}$ we are interested in is given by its intersection matrix

$$I(\mathcal{A}, \mathcal{B}) = \left\{ |A_i \cap B_j| \;:\; 1 \leq i, j \leq n \right\}.$$

Since $I(\mathcal{A}, \mathcal{B})$ is a matrix of scalar products of the corresponding characteristic vectors of length t, the total number t of labels used must be at least the rank of $I(\mathcal{A}, \mathcal{B})$ over the reals.

In general, however, it is difficult to estimate the rank of the intersection matrix because our knowledge about its entries $|A_i \cap B_j|$ is rather poor: we only know that some of them lie within the set L and the other lie outside this set. In such a situation, one can try to transform the original matrix into a matrix whose rank r is easier to estimate and is still not much smaller than $\theta_L(G)$.

To be more precise, let F be some fixed field, and let $f : \{0,1\}^t \to F$ be a function. Define the f-intersection matrix of $\mathcal{A}$ and $\mathcal{B}$ as

$$I_f(\mathcal{A}, \mathcal{B}) = \left\{ f(A_i \cap B_j) \;:\; 1 \leq i, j \leq n \right\},$$

where here and in what follows, the value $f(C)$ of f on a subset $C \subseteq \{1,\dots,t\}$ stands for the value $f(\vec{c})$ of the incidence $0/1$ vector $\vec{c} = (c_1,\dots,c_t)$ of C given by $c_i = 1$ iff $i \in C$. Note that the intersection matrix $I(\mathcal{A},\mathcal{B})$ corresponds to the case when $f(x_1,\dots,x_t) = x_1 + x_2 + \cdots + x_t$.

A multilinear polynomial over F of *degree* d and *weight* w is a sum of w monomials $a_I X_I$, where $a_I \in F$ and $X_I = \prod_{j\in I} x_j$ with $|I| \le d$. Given an arbitrary (multivariate) polynomial $f(x_1,\dots,x_t)$, we define its *weight* $w(f)$ as the smallest number w such that, when restricted to $\{0,1\}^t \subseteq F^t$, f can be written as a multilinear polynomial of weight w. Note that any polynomial $f(x_1,\dots,x_t)$ of degree d has weight

$$w(f) \le \binom{t}{0} + \binom{t}{1} + \cdots + \binom{t}{d},$$

which is at most t^d for growing t.

Having found a polynomial f of weight $w(f) \le t^d$ such that the corresponding f-intersection matrix has rank at least r, the following lemma implies a lower bound $t \ge r^{1/d}$ on the number t of used labels.

Lemma 1. *For every polynomial f over a field F, every f-intersection matrix has rank at most $w(f)$ over F.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be systems of subsets of $\{1,\dots,t\}$, and let f be a polynomial of weight $w = w(f)$. Then the restriction of f to the binary cube $\{0,1\}^t$ can be written as a linear combination $f = a_1X_1 + \cdots + a_wX_w$ of monomials. Since for every monomial $X_i = \prod_{j\in I_i} x_j$, we have

$$X_i(A\cap B) = 1 \text{ iff } I_i \subseteq A\cap B \text{ iff } I_i \subseteq A \text{ and } I_i \subseteq B \text{ iff } X_i(A)\cdot X_i(B) = 1,$$

the value $f(A\cap B)$ is just the scalar product of two vectors

$$(a_1X_1(A),\dots,a_wX_w(A)) \text{ and } (X_1(B),\dots,X_w(B))$$

of length w over F, implying that the rank of $I_f(\mathcal{A},\mathcal{B})$ cannot exceed $w = w(f)$. □

Lemma 1 is particularly appealing when dealing with *modular* dimensions of graphs since in this case the choice of the appropriate polynomial f is quite natural.

3.2.2 Proof of Theorem 2

A bipartite $n\times n$ graph $G = (V_1 \cup V_2, E)$ is *increasing* if it is possible to enumerate its vertices $V_1 = \{x_1,\dots,x_n\}$ and $V_2 = \{y_1,\dots,y_n\}$ so that $x_iy_i \in E$ and $x_iy_j \notin E$ for all $i > j$. In this case, the adjacency matrix of G is a lower triangular matrix with nonzero elements on the diagonal. In particular, every matching is an increasing graph.

The following simple fact allows us to concentrate on increasing graphs.

Proposition 4. *If G is a bipartite $n\times n$ graph of maximum degree Δ and with no isolated vertices, then G contains an induced bipartite $(n/\Delta)\times(n/\Delta)$ increasing subgraph.*

Proof. We can construct an induced increasing subgraph of $G=(V_1\cup V_2,E)$ inductively as follows. Suppose we have already constructed an induced increasing $m\times m$ subgraph $G'=(V_1'\cup V_2',E')$ of G with $m<n/\Delta$. Then we can enlarge G' to an induced increasing $(m+1)\times(m+1)$ subgraph as follows. Remove from V_1 all vertices having at least one neighbor in V_2'. Since $m\Delta<n$, at least one vertex $x_{m+1}\in V_1\setminus V_1'$ must survive. Since this vertex is not isolated and has no neighbors in V_2', some vertex $y_{m+1}\in V_2\setminus V_2'$ must be adjacent to x_{m+1}. Since the subgraph G' induced by V_1' and V_2' was increasing, the subgraph induced by $V_1'\cup\{x_{m+1}\}$ and $V_2'\cup\{y_{m+1}\}$ must be increasing as well. □

Since *induced* subgraphs can only have smaller dimension than the original graph, Theorem 2 is a direct consequence of Proposition 4 and the following lemma.

Lemma 2. *Let G be an increasing bipartite $n\times n$ graph, p a prime number and $1\le r<p$ an integer. Then for every type of the form $L=\{k\mid k \bmod p\in R\}$ with $|R|=r$ we have $\theta_L(G)\ge (n/r)^{1/(p-1)}$ and $\theta_L(G^c)\ge n^{1/r}$.*

Proof. Let G be an arbitrary increasing bipartite $n\times n$ graph, and let $\mathcal{A}=\{A_1,\dots,A_n\}$ and $\mathcal{B}=\{B_1,\dots,B_n\}$ be systems of subsets of $\{1,\dots,t\}$ associated with its vertices. Suppose that these two systems form an intersection representation of G with respect to a type $L=\{k\mid k \bmod p\in R\}$ with $|R|=r$. When taken modulo p, the diagonal entries of the intersection matrix $I(\mathcal{A},\mathcal{B})$ must belong to R, and none of the entries below the diagonal can belong to R. We can therefore find a number $a\in R$ and a subset $I\subseteq\{1,\dots,n\}$ of size $|I|\ge n/|R|=n/r$ such that $|A_i\cap B_i| \bmod p=a$ for all $i\in I$, and $|A_i\cap B_j| \bmod p\ne a$ for $i>j$. Let $\mathcal{A}'=\{A_i\mid i\in I\}$ and $\mathcal{B}'=\{B_i\mid i\in I\}$, and consider a polynomial

$$g_a(z_1,\dots,z_t)=z_1+\cdots+z_t-a.$$

By what was said, the corresponding g_a-intersection submatrix $\{g_a(A_i\cap B_j)\mid i,j\in I\}$ modulo p has zeroes on the diagonal, and has nonzero entries below the diagonal. By Fermat's Little Theorem, the corresponding f-intersection matrix $I_f(\mathcal{A}',\mathcal{B}')$, with $f(z_1,\dots,z_t)$ defined by

$$f(z_1,\dots,z_t)=1-g_a(z_1,\dots,z_t)^{p-1},$$

is a lower triangular matrix with nonzero diagonal entries, and must therefore have full rank $|I|\ge n/r$ over $GF(p)$. Since f is a polynomial of degree $p-1$, we have that $w(f)\le t^{p-1}$. Lemma 1 implies $n/r\le|I|\le w(f)\le t^{p-1}$, and the lower bound $t\ge(n/r)^{1/(p-1)}$ follows.

If $\mathcal{A},\mathcal{B}$ is an intersection representation of the complement graph G^c with respect to the type $L=\{k\mid k \bmod p\in R\}$, then this is an intersection representation of the graph G itself with respect to the type $\{k\mid k \bmod p\notin R\}$. Since the graph G is increasing, this means that $|A_i\cap B_i| \bmod p\notin R$ for all i, and $|A_i\cap B_j| \bmod p\in R$ for $i>j$. Hence, if we take

$$h(z_1,\dots,z_t)=\prod_{a\in R} g_a(z_1,\dots,z_t)$$

then, modulo p, the h-intersection matrix $I_h(\mathcal{A},\mathcal{B})$ itself is a lower triangular matrix with nonzero diagonal entries, and must therefore have full rank n over $GF(p)$. Since h is a polynomial of degree $|R|\le r$, we have that $w(h)\le t^r$. Lemma 1 implies in this case $n\le w(h)\le t^r$, and the desired lower bound $t\ge n^{1/r}$ follows. □

3.3 Threshold dimensions: Proof of Theorem 3

A standard tool when dealing with threshold-type coverings is to introduce a particular notion of discrepancy between the fractions of "correctly" and "wrongly" covered elements.

Let $\mathcal{B}$ be a family of subsets of a finite set X. For a subset $A \subseteq X$, let $\mathrm{thr}_{\mathcal{B}}(A)$ denote the minimum number t for which there exist t members $B_1, \ldots, B_t$ of $\mathcal{B}$ and a number $0 \leq k \leq t$ such that, for every $x \in X$, $x \in A$ if and only if x belongs to at least k of B_i's. Define the *relative discrepancy* of the set A with respect to a family $\mathcal{B}$ by

$$\mathrm{disc}_{\mathcal{B}}(A) = \max_{B \in \mathcal{B}} \left| \frac{|A \cap B|}{|A|} - \frac{|\overline{A} \cap B|}{|\overline{A}|} \right|.$$

Proposition 5. $\mathrm{thr}_{\mathcal{B}}(A) \geq 1/\mathrm{disc}_{\mathcal{B}}(A)$.

Proof. Let $B_1, \ldots, B_t \in \mathcal{B}$ be a threshold-k covering of A, for some number $k \geq 1$. Then $x \in A$ iff x belongs to at least k of B_i's. Since every element of A belongs to at least k of the sets $A \cap B_i$, the average size of these sets must be at least k. Since no element of $\overline{A}$ belongs to more than $k-1$ of the sets $\overline{A} \cap B_i$, the average size of these sets must be at most $k-1$. Hence,

$$1 \quad \leq \quad \frac{1}{|A|} \sum_{i=1}^{t} |A \cap B_i| - \frac{1}{|\overline{A}|} \sum_{i=1}^{t} |\overline{A} \cap B_i| \leq t \cdot \max_{1 \leq i \leq t} \left| \frac{|A \cap B_i|}{|A|} - \frac{|\overline{A} \cap B_i|}{|\overline{A}|} \right| \leq t \cdot \mathrm{disc}_{\mathcal{B}}(A).$$

□

We can now prove Theorem 3 as follows. Let $H = (V_1 \cup V_2, E)$ be a bipartite Hadamard $n \times n$ graph, and let $t = \theta_{\mathrm{thr}}(H)$. Then there exists an integer $k \geq 1$, a set I of $|I| = t$ labels and a system of sets $\{A_x \subseteq I \mid x \in V_1 \cup V_2\}$ such that, for every pair xy of vertices $x \in V_1$ and $y \in V_2$, $xy \in E$ iff $|A_x \cap A_y| \geq k$. Associate with each label $i \in I$ a complete bipartite graph $B_i = S_i \times T_i$, where $S_i = \{x \in V_1 \mid i \in A_x\}$ and $T_i = \{y \in V_2 \mid i \in A_y\}$. Then $xy \in E$ iff $|A_x \cap A_y| \geq k$ iff xy belongs to at least k of B_i's. Hence $\theta_{\mathrm{thr}}(H) = t \geq \mathrm{thr}_{\mathcal{B}}(E)$ with $\mathcal{B} = \{B_1, \ldots, B_t\}$. By Proposition 5, it remains to show that $\mathrm{disc}_{\mathcal{B}}(E) = O(n^{-1/2})$.

For this, take a complete bipartite graph $B = S \times T$ in $\mathcal{B}$ achieving the maximum in the definition of $\mathrm{disc}_{\mathcal{B}}(E)$. The well-known Lindsey's Lemma (see, for example, [14], p. 88) says that the sum of all entries in any $s \times t$ submatrix of an $n \times n$ Hadamard ± 1 matrix lies between $-\sqrt{stn}$ and $\sqrt{stn}$. In particular, this sum lies between $-n^{3/2}$ and $n^{3/2}$ for any submatrix. In terms of graphs, we obtain that the absolute value of the difference $|E \cap B| - |E^{\mathrm{c}} \cap B|$ does not exceed $n^{3/2}$; here, as before, $E^{\mathrm{c}} = (V_1 \times V_2) \setminus E$ is the set of non-edges of H. Since the number of edges as well as of non-edges of H is at least cn^2 for a constant $c > 0$, this implies that $\mathrm{disc}_{\mathcal{B}}(E) = O(n^{3/2}/n^2) = O(n^{-1/2})$. □

3.4 Balanced representations: Proof of Theorem 4

We will give a lower bound on the number of labels in balanced representations in terms of the following characteristic of graphs, which we will also use in the next section.

A bipartite graph $G = (V_1 \cup V_2, E)$ is *k-isolated* if, for every two vertices $x \neq y \in V_1$, there is a set $S \subseteq V_2$ of $|S| = k$ vertices such that every vertex $u \in S$ is adjacent to x and non-adjacent to y. For example, a bipartite Hadamard $n \times n$ graph is k-isolated with $k \geq n/4$. Hence, Theorem 4 is a special case of the following

Theorem 7. *Every balanced representation of a k-isolated bipartite $n \times n$ graph must use at least k labels.*

Proof. Let $\mathcal{A} = \{A_x \mid x \in V\}$ be a balanced intersection representation of G using t labels. Let $V = V_1 \cup V_2$ be the bipartition of G. Since the representation is balanced, there must exist two vertices $x \neq y \in V_1$ such that their sets of labels $X = A_x$ and $Y = A_y$ satisfy

$$|A_u \cap A_v \cap X| = |A_u \cap A_v \cap Y| \qquad \text{for all } u \neq v \in V_2. \tag{3}$$

On the other hand, since the graph is k-isolated, there must be a subset $S \subseteq V_2$ of $|S| = k$ vertices such that every vertex $u \in S$ is adjacent to x and non-adjacent to y. Hence, independent on what type L was used for the representation, we must have that

$$|A_u \cap X| \neq |A_u \cap Y| \qquad \text{for all } u \in S. \tag{4}$$

For every subset $C \subseteq \{1, \ldots, t\}$, the value $f(C)$ of a real polynomial

$$f(z_1, \ldots, z_t) = \sum_{i \in X} z_i - \sum_{i \in Y} z_i$$

is the difference between $|C \cap X|$ and $|C \cap Y|$. Hence, by taking $C = A_u \cap A_v$, (3) implies that $f(A_u \cap A_v) = 0$ for all $u \neq v \in S$, and (4) implies that $f(A_u \cap A_u) \neq 0$ for all $u \in S$. That is, the f-intersection matrix $I_f(\mathcal{A}', \mathcal{A}')$ of $\mathcal{A}' = \{A_u \mid u \in S\}$ is a real diagonal matrix with nonzero diagonal entries. Lemma 1 implies $t \geq w(f) \geq |S| = k$. □

3.5 Weight: Proof of Theorem 5

Theorem 5 is a direct consequence of the following

Theorem 8. *If a bipartite $n \times n$ graph G is k-isolated, then $w_L(G) = \Omega(k \ln n / \ln \ln n)$ for every type L.*

Proof. Let $G = (V_1 \cup V_2, E)$ be a bipartite k-isolated $n \times n$ graph. Fix an arbitrary intersection representation $A_1, \ldots, A_n, B_1, \ldots, B_n$ of G with respect to some type L. We may assume that $k > 0$ (since for $k = 0$ there is nothing to prove). Hence, all sets $A_1, \ldots, A_n$ must be distinct. Let $m = c \ln n / \ln \ln n$ for a sufficiently small absolute constant $c > 0$. If $\sum_{i=1}^n |A_i| > mn$, then we are done. So, assume that $\sum_{i=1}^n |A_i| \leq mn$. Our goal is to show that then $\sum_{j=1}^n |B_j| \geq mk$.

A classical result of Erdős and Rado [10] says that every family of $r!s^r$ sets, each of which has cardinality less than r, contains a sunflower with s petals, that is, a family $F_1, \ldots, F_s$ of sets of the form $F_i = P_i \cup C$, where the P_i's are pairwise disjoint; the set C is the *core* of the sunflower, and the P_i's are called the *petals*.

Now, if $\sum_{i=1}^{n} |A_i| \leq mn$, then at least $n/2$ of the sets A_i must be of size at most $r = 2m$. By the sunflower theorem, these sets must contain a sunflower with $s = 2m$ petals. Having such a sunflower with a core C, we can pair its members arbitrarily, $(A_{u_1}, A_{v_1}), \ldots, (A_{u_m}, A_{v_m})$. Important for us is that all m symmetric differences $D_i = A_{u_i} \oplus A_{v_i} = (A_{u_i} \cup A_{v_i}) \setminus C$ are mutually disjoint.

Since the graph is k-isolated, each two vertices $u_i \neq v_i$ have a set $S_i \subseteq V_2$ of $|S_i| = k$ vertices, all of which are adjacent to u_i and none of which is adjacent to v_i. Hence, independent on the type L, we have that $|A_{u_i} \cap B_j| \neq |A_{v_i} \cap B_j|$ must hold for all $j \in S_i$. This implies that each set B_j with $j \in S_i$ must have at least one element in the symmetric difference $D_i = A_{u_i} \oplus A_{v_i}$. Hence,

$$\sum_{j=1}^{n} |D_i \cap B_j| \geq \sum_{j \in S_i} |D_i \cap B_j| \geq |S_i| = k \qquad \text{for each } i = 1, \ldots, m.$$

Since the sets $D_1, \ldots, D_m$ are disjoint, this implies

$$\sum_{j=1}^{n} |B_j| \geq \sum_{j=1}^{n} \sum_{i=1}^{m} |D_i \cap B_j| = \sum_{i=1}^{m} \sum_{j=1}^{n} |D_i \cap B_j| \geq \sum_{i=1}^{m} k = mk.$$

□

3.6 Tight representations: Proof of Theorem 6

Since every k-tight representation is also a threshold-k representation, Theorem 6 is a direct consequence of the following

Theorem 9. *Let G be a bipartite $n \times m$ graph of threshold-k dimension r. Then either the graph G contains a complete bipartite subgraph with $nm/4\binom{r}{k}^2$ edges, or the bipartite complement G^{c} of G contains a complete bipartite subgraph with $nm/4$ edges.*

Proof. Let $\{A_x \mid x \in V_1 \cup V_2\}$ be a threshold-k representation of $G = (V_1 \cup V_2, E)$ using r labels. Hence, $xy \in E$ iff $|A_x \cap A_y| \geq k$. We say that a set S of labels *appears* in a vertex x if $S \subseteq A_x$. Set $\alpha = 1/2\binom{r}{k}$, and call a set S of labels *left popular* (resp., *right popular*) if it appears in at least α-fraction of vertices of V_1 (resp., of V_2).

Case 1 At least one k-element set S of labels is left popular as well as right popular. In this case, the set S appears in at least αn vertices in V_1 as well as in at least αm vertices in V_2. The corresponding vertices all share the same k-element set S, and hence, form a complete bipartite subgraph of G with at least $(\alpha n)(\alpha m) = \alpha^2 nm$ edges.

Case 2 No k-element set S of labels is both left and right popular. If a set S is not left popular, then it can appear in at most αn of vertices in V_1. Hence, the number of vertices in V_1 containing at least one k-element subset, which is not left popular, cannot exceed $\binom{r}{k} \cdot \alpha n = n/2$. We can therefore find a subset $L \subseteq V_1$ of $|L| \geq n/2$ vertices $x \in V_1$, *all* whose k-element subsets $S \subseteq A_x$ are left popular. By symmetry, we can find a subset $R \subseteq V_2$ of $|R| \geq m/2$ vertices $y \in V_2$, *all* whose k-element subsets $S \subseteq A_y$ are right popular. Since, by our assumption, no k-element subset of labels can be both left and right popular, no k-element set can be contained in any intersection $A_x \cap A_y$ with $x \in L$ and $y \in R$. Hence, $|A_x \cap A_y| < k$ for all $x \in L$ and $y \in R$. That is, in this case the complement graph G^{c} contains a complete bipartite $(n/2) \times (m/2)$ subgraph $L \times R$. □

4 Conclusion and open problems

As mentioned in the Introduction, high lower bounds on the intersection dimension of explicit bipartite graphs would resolve some old problems in the computational complexity of boolean functions. In this paper we obtained such lower bounds when either the form of the type L or the form of used sets of labels is restricted. Our results, as well as previous ones, are still too weak to have new consequences for boolean functions. Below we shortly describe what we need to have such consequences. In all these problems we are looking for an *explicit* sequence of bipartite $n \times n$ graphs of large intersection dimension.

Problem 10. *Prove* $\theta_{\mathrm{mod}}(G) \geq 2^{(\ln \ln n)^{\alpha}}$ *for some* $\alpha \to \infty$.

By results of Yao [37] and Beigel and Tarui [6], this would yield the first super-polynomial lower bound for constant depth circuits with arbitrary modular gates (so-called *ACC*-circuits), thus resolving a long-standing open problem in computational complexity. Actually, as shown by Green *et al.* [19], it would be enough to prove such a lower bound on $\theta_L(G)$ for a special kind of modular types L consisting of all natural numbers whose binary representations have a 1 in the middle. Such types (called *middle-bit* types) consist of disjoint intervals of consecutive numbers. According to this result, it would be enough to prove an explicit lower bound $2^{(\ln \ln n)^{\alpha}}$ on the minimum of $\theta_L(G)$ over all modular types of the form $L = \{\ell \mid (\ell \bmod p) \geq p/2\}$. Theorem 2 yields such lower bounds, as long as $p \leq (\ln n)/(\ln \ln n)^{\alpha}$. The problem is to extend this for longer intervals.

Each interval type $L = \{a, a+1, \ldots, b\}$ is an intersection of two threshold types $\{a, a+1, \ldots\}$ and $\{0, 1, \ldots, b\}$. We already know (see Theorem 3) that, with respect to both these types, Hadamard $n \times n$ graphs H must have dimension about $\sqrt{n}$. Can the dimension of H with respect to $\{a, a+1, \ldots, b\}$ be much smaller than $\sqrt{n}$?

Problem 11. *What is the intersection dimension of Hadamard graphs with respect to interval types?*

In the context of boolean functions, the next important measure is the following generalization of the threshold-1 dimension $\theta_1(G)$ of graphs, that is, of the edge clique covering number of graphs. Namely, let $\rho_s(G)$ be the smallest number r such that G can be written as an intersection of at most s graphs $G_1, \ldots, G_s$ with $\theta_1(G_i) \leq r$ for all $i = 1, \ldots, s$. That is, in oder to reduce the number of complete subgraphs in the covering, we now allow to replace up to a $1 - 1/s$ fraction of non-edges of G by new edges.

Define the *resistance* $\rho(G)$ of a graph G by $\rho(G) = \min\{s \mid \rho_s(G) \leq s\}$. It can be shown (see [20]) that $\rho(G)$ is the smallest number r for which it is possible to associate with each vertex x an $r \times r$ matrix M_x with entries in $\{0, 1\}$ so that $xy \in E$ iff the diagonal of the product matrix $M_x \cdot M_y^{\top}$ over the reals has no zero on the diagonal.

Problem 12. *Prove* $\rho(G) = \Omega(n^{\varepsilon})$ *for a constant* $\varepsilon > 0$.

Together with the well-known reduction of log-depth circuits to depth-3 circuits, due to Valiant [35], this would give the first super-linear lower bound for boolean circuits of logarithmic depth (see [20] for more details). To prove such a lower bound is one of the central open problems in the computational complexity of boolean functions.

Easy counting shows that bipartite $n \times n$ graphs of absolute dimension $\rho(G) = \Omega(\sqrt{n})$ exist: we have 2^{n^2} bipartite $n \times n$ graphs, but only 2^{2nr^2} possibilities to assign $r \times r$ $0/1$ matrices M_x to the vertices x. The problem, however, is to prove a comparable lower bound for an *explicit* sequence of graphs. The best we can do so far is a lower bound $\rho(G) = \Omega(\ln^{3/2} n)$ proved by Lokam [23] for Hadamard graphs. It is also known that $\rho_s(H) = \Omega(n^{1/2-\varepsilon})$, as long as $s \leq \varepsilon \log_2 n$ [20]. Hence, it is important to understand what happens, when s approaches the border of $\log_2 n$. That this may be indeed a critical border can be seen on an example of an $n \times n$ matching M: then $\rho_1(M) = \theta_1(M) = n$ but $\rho_s(M) \leq 2$ for $s = \log_2 n$. To see this, encode each vertex x in one color class by its *own* vector $\vec{x} \in \{0,1\}^s$, and assign the matched vertex in the other color class the same vector $\vec{x}$. We can then write M as an intersection of s graphs $G_1, \ldots, G_s$, where G_i, consists of all edges whose endpoints have the same bit in the i-th coordinate. Since each G_i is a union of just two complete bipartite graphs, $\rho_s(M) \leq 2$ follows.

Problem 13. *What is $\rho_s(H)$ for an $n \times n$ Hadamard graph H when $s = \log_2 n$?*

Even a mere *existence* of bipartite graphs, whose resistance is much larger than the resistance of their bipartite complements, is not known.

Problem 14. *Does there exists a sequence G_n of bipartite $n \times n$ graphs such that $\ln \rho(G_n) \leq (\ln \ln n)^c$ for a constant c, but $\ln \rho(\overline{G_n}) \geq (\ln \ln n)^{\alpha}$ for some α tending to infinity as $n \to \infty$?*

If not, then this would separate the second level of the communication complexity hierarchy and resolve a more than 20 years old problem raised by Babai, Frankl and Simon [5].

The last problem concerns the *weight* $w_{\mathrm{odd}}(G)$ of intersection representations with respect to the type consisting of all odd integers.

Problem 15. *Prove $w_{\mathrm{odd}}(G) = \Omega(n^{1+\varepsilon})$ for a constant $\varepsilon > 0$.*

The highest known lower bound $w_{\mathrm{odd}}(G) = \Omega(n \ln^{3/2} n)$ is due to Pudlák and Rödl [29]. When dealing with this measure, the following equivalent reformulation could be useful: $w_{\mathrm{odd}}(G)$ is the smallest number w for which the adjacency matrix of G can be written as a product AB of two 0-1 matrices A and B over $GF(2)$ such that the total number of nonzero entries in A and B does not exceed w. An indication that Hadamard graphs may be not good for this purpose is given in [29]: Sylvester matrix *can* be decomposed into the product of *three* matrices with only linear number of nonzero elements.

References

[1] N. Alon, Covering graphs by the minimum number of equivalence relations, Combinatorica 6 (1986) 201–206.

[2] N. Alon, Neighborly families of boxes and bipartite coverings, in: The Mathematics of Paul Erdős, R. L. Graham and J. Nesetril, eds., Springer Verlag, Vol II, Berlin (1997) 27–31.

[3] N. Alon, P. Seymour, A counterexample to the rank-coloring conjecture, J. Graph Theory 13 (1989) 523–525.

[4] N. Alon, M. Karchmer, A. Wigderson, Linear circuits over GF(2), SIAM. J. Comput. 19(6) (1990) 1064–1067.

[5] L. Babai, P. Frankl, J. Simon, Complexity classes in communication complexity. In Proc. of 26th IEEE Symp. on Foundations of Comput. Sci. (1986) 337–347.

[6] R. Beigel, J. Tarui, On ACC, Computat. Complexity 4 (1994) 350–366

[7] B. Bollobás, P. Erdős, J. Spencer, D. B. West, Clique coverings of the edges of a random graph, Combinatorica 13 (1993) 1–5.

[8] F. R. H. Chung, P. Erdős, J. Spencer, On the decomposition of graphs into complete bipartite subgraphs, in: Studies in pure mathematics (Birkhuser, Basel, 1983) 95–101.

[9] M. S. Chung, D. B. West, The p-intersection number of a complete bipartite graph and orthogonal double coverings of a clique, Combinatorica 14 (1994) 453–461.

[10] N. Eaton, V. Rödl, Graphs of small dimensions, *Combinatorica* 16(1) (1996) 59–85.

[11] N. Eaton, D. Grable, Set representations for almost all graphs, J. Graph Theory 23(3) (1996) 1–12.

[12] N. Eaton, R. J. Gould, V. Rödl, On p-intersection representations, J. Graph Theory 21(4) (1996) 377–392.

[13] P. Erdős, R. Rado, Intersection theorems for systems of sets, J. London Math. Soc. 35 (1960) 85–90.

[14] P. Erdős, J. Spencer, Probabilistic methods in combinatorics, Akademiai Kiado, Budapest (1974).

[15] P. Erdős, L. Pyber, Covering a graph by complete bipartite graphs, Discrete Math. 170 (1997) 249–251.

[16] P. Erdős, A. W. Goodman, L. Pósa, The representation of graphs by set intersections, Canadian J. of Math. 18 (1966) 106–112.

[17] A. Frieze, B. Reed, Covering the edges of a random graph by cliques, Combinatorica 15 (1995) 1–9.

[18] R. L. Graham, H. O. Pollak, On the addressing problem for loop switching, Bell System Techn. Jour. 50 (1971) 2495–2519.

[19] F. Green, J. Köbler, K. W. Regan, T. Schwentick, J. Torn, The power of the middle bit of a # P function, J.Comput. Syst. Sci. nummer50(3) (1995) 456–467.

[20] S. Jukna, On graph complexity, Comb., Prob. and Comput. 15(6) (2006) 855–876.

[21] S. Jukna, A. Kulikov, On covering graphs by complete bipartite subgraphs, Discrete Math., to appear.

[22] E. Kushilevitz, N. Nisan, Communication Complexity, Cambridge University Press (2006).

[23] S. V. Lokam, Graph complexity and slice functions, Theory of Comput. Syst. 36(1) (2003) 1–88.

[24] L. Lovász, M. Saks, Lattices, Möbius functions and communication complexity, in: Proc. of 29th IEEE Symp. on Foundations of Comput. Sci. (1988) 81–90.

[25] F. R. McMorris, Chi Wang, Modular intersection graphs, Graphs and Combinatorics 12 (1996) 267–281.

[26] N. Nisan, A. Wigderson, On rank vs. communication, complexity, Combinatorica 15(4) (1995) 557–566.

[27] C. van Nuffelen, A bound for the chromatic number of a graph, Amer. Math. Monthly 83 (1976) 265–266.

[28] P. Pudlák, V. Rödl, P. Savický, Graph complexity, Acta Inform. 25 (1988) 515–535.

[29] P. Pudlák, V. Rödl, Some combinatorial-algebraic problems from complexity theory, Discrete Math. 136 (1994) 253-279.

[30] P. Pudlák, V. Rödl, Pseudorandom sets and explicit constructions of Ramsey graphs, in: J. Krajicek (ed.) Quad. Mat., 13 (2004) 327-346.

[31] R. Raz, B. Spiker, On the Log-Rank conjecture in communication complexity, Combinatorica 15(4) (1995) 567–588.

[32] A. Razborov, The gap between the chromatic number of a graph and the rank of its adjacency matrix is superlinear, *Discrete Math.* 108 (1992) 393–396.

[33] J. Sgal, Bounds on pairs of families with restricted intersections, Combinatorica 19 (1999) 555–566.

[34] Zs. Tuza, Coverings of graphs by complete bipartite subgraphs, complexity of 0-1 matrices, Combinatorica 4(1) (1984) 111–116.

[35] L. Valiant, Graph-theoretic methods in low-level complexity, in: Proc. of 6th Conf. on Math. Foundations of Comput. Sci., Springer Lect. Notes in Comput. Sci. 53 (1977) 162–176.

[36] L. Valiant, The Log-Rank Conjecture and low degree polynomials, Inf. Process. Letters 89 (2004) 99–103.

[37] A. C. Yao, On ACC and threshold circuits, in: Proc. of 31th IEEE Symp. on Foundations of Comput. Sci. (1990) 619–627.