P_4-free Partition and Cover Numbers & Applications

Alexander R. Block
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Simina Brânzei
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Hemanta K. Maji
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Himanshi Mehta
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Tamalika Mukherjee
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Hai H. Nguyen
Department of Computer Science, Purdue University, West Lafayette, IN, USA

Abstract

P_4-free graphs– also known as cographs, complement-reducible graphs, or hereditary Dacey graphs– have been well studied in graph theory. Motivated by computer science and information theory applications, our work encodes (flat) joint probability distributions and Boolean functions as bipartite graphs and studies bipartite P_4-free graphs. For these applications, the graph properties of edge partitioning and covering a bipartite graph using the minimum number of these graphs are particularly relevant. Previously, such graph properties have appeared in leakage-resilient cryptography and (variants of) coloring problems.

Interestingly, our covering problem is closely related to the well-studied problem of product (a.k.a., Prague) dimension of loopless undirected graphs, which allows us to employ algebraic lower-bounding techniques for the product/Prague dimension. We prove that computing these numbers is NP-complete, even for bipartite graphs. We establish a connection to the (unsolved) Zarankiewicz problem to show that there are bipartite graphs with size-N partite sets such that these numbers are at least $\epsilon \cdot N^{1-2\epsilon}$, for $\epsilon \in \{1/3, 1/4, 1/5, \dots\}$. Finally, we accurately estimate these numbers for bipartite graphs encoding well-studied Boolean functions from circuit complexity, such as set intersection, set disjointness, and inequality.

For applications in information theory and communication & cryptographic complexity, we consider a system where a setup samples from a (flat) joint distribution and gives the participants, Alice and Bob, their portion from this joint sample. Alice and Bob's objective is to non-interactively establish a shared key and extract the left-over entropy from their portion of the samples as independent private randomness. A genie, who observes the joint sample, provides appropriate assistance to help Alice and Bob with their objective. Lower bounds to the minimum size of the genie's assistance translate into communication and cryptographic lower bounds. We show that (the $\log_2$ of) the P_4-free partition number of a graph encoding the joint distribution that the setup uses is equivalent to the size of the genie's assistance. Consequently, the joint distributions corresponding to the bipartite graphs constructed above with high P_4-free partition numbers correspond to joint distributions requiring more assistance from the genie.

As a representative application in non-deterministic communication complexity, we study the communication complexity of nondeterministic protocols augmented by access to the equality oracle at the output. We show that (the $\log_2$ of) the P_4-free cover number of the bipartite graph encoding a Boolean function f is equivalent to the minimum size of the nondeterministic input required by the parties (referred to as the communication complexity of f in this model). Consequently, the functions corresponding to the bipartite graphs with high P_4-free cover numbers have high communication complexity. Furthermore, there are functions with communication complexity close to the naïve protocol where the nondeterministic input reveals a party's input. Finally, the access to the equality

2nd Conference on Information-Theoretic Cryptography (ITC 2021).
Editor: Stefano Tessaro; Article No. 16; pp. 16:1–16:25

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

oracle reduces the communication complexity of computing set disjointness by a constant factor in contrast to the model where parties do not have access to the equality oracle. To compute the inequality function, we show an exponential reduction in the communication complexity, and this bound is optimal. On the other hand, access to the equality oracle is (nearly) useless for computing set intersection.

2012 ACM Subject Classification Security and privacy → Mathematical foundations of cryptography; Security and privacy → Information-theoretic techniques; Theory of computation → Communication complexity; Mathematics of computing → Graph theory

Keywords and phrases Secure keys, Secure private randomness, Gray-Wyner system, Cryptographic complexity, Nondeterministic communication complexity, Leakage-resilience, Combinatorial optimization, Product dimension, Zarankiewicz problem, Algebraic lower-bounding techniques, P_4-free partition number, P_4-free cover number

Digital Object Identifier 10.4230/LIPIcs.ITC.2021.16

Related Version *Full Version*: `https://eprint.iacr.org/2020/1605.pdf`

Funding Alexander R. Block is supported in part by NSF CCF #1910659. Alexander R. Block, Hemanta K. Maji, Tamalika Mukherjee, and Hai H. Nguyen are supported in part by an NSF CRII Award CNS–1566499, NSF SMALL Awards CNS–1618822 and CNS–2055605, the IARPA HECTOR project, MITRE Innovation Program Academic Cybersecurity Research Awards (2019–2020, 2020–2021), a Purdue Research Foundation (PRF) Award, and The Center for Science of Information, an NSF Science and Technology Center, Cooperative Agreement CCF–0939370.

1 Introduction

A graph is P_4-*free* if no four vertices induce a path of length three. Since the 1970s, P_4-free graphs – also known as cographs, complement-reducible graphs, or hereditary Dacey graphs from empirical logic [22] – have been widely studied in graph theory [45, 46, 36, 60, 62]. Motivated by computer science and information theory applications, our work encodes joint probability distributions and Boolean functions as bipartite graphs and studies *bipartite* P_4-free graphs.[1] For these applications, the graph properties of edge *partitioning* and *covering* a bipartite graph using the minimum number of these graphs are particularly relevant.[2]

The P_4-*free partition number* of a bipartite graph G is the minimum number of P_4-free subgraphs partitioning G's edges, denoted by $\mathsf{P_4\text{-}fp}(G)$. Similarly, the P_4-*free cover number* of a bipartite graph G is the minimum number of P_4-free subgraphs covering G's edges, denoted by $\mathsf{P_4\text{-}fc}(G)$. The definition extends to general graphs; however, our study focuses on bipartite graphs. We are given a bipartite graph as input, and the objective is to partition or cover its edges using P_4-free bipartite graphs. P_4-free partition and cover numbers are natural extensions of fundamental graph properties, such as product/Prague dimension, equivalence cover number, biclique partition, and cover numbers, arboricity, and star arboricity (refer to [63] for definitions). In turn, these graph properties have applications to theoretical computer science, information theory, and combinatorial optimization; for a discussion of these connections, see Appendix E in the full version.

[1] A bipartite P_4-free graph is a disjoint union of *bicliques*.
[2] In contrast, [31] introduced the *vertex* partitioning a graph into different color-classes so that the vertices of any color-class induces a P_4-free graph.

In addition to being motivated by intellectual curiosity, our work illustrates that the P_4-free partition and cover numbers appear in diverse computer science and information theory problems (refer to problems A and B in Section 1.1). Section 1.2 presents the equivalence between the P_4-free partition number and Problem A, and the consequences of the graph theory results for problem A. Next, Section 1.3 demonstrates the equivalence of Problem B and the P_4-free cover number, and the implications of the graph results for problem B. Interestingly, we prove that the P_4-free cover number of a bipartite graph is either identical to or one less than the well-studied product/Prague dimension [54, 55] of the complement graph (interpreted as a loopless undirected graph). Our work proves the following graph theory results (refer to Section 2 for formal statements).

1. Determining the P_4-free partition and cover numbers of general graphs, even bipartite ones, is NP-complete.
2. There are bipartite graphs with size-N partite sets whose P_4-free partition and cover numbers are at least $\epsilon \cdot N^{1-2\epsilon}$, for constant $\epsilon \in \{1/3, 1/4, 1/5, \dots\}$. Furthermore, Erdős-Rényi graphs (with constant parameter) have P_4-free partition and cover numbers $\geq N/\log N$, asymptotically almost surely.
3. Finally, we encode the Boolean set intersection and disjointness functions, and the inequality function as bipartite graphs. We present tight estimates of the P_4-free partition and cover numbers of these graphs.

1.1 Motivating Problems

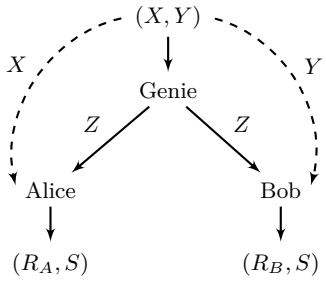
We encode joint probability distributions and Boolean functions as equivalent bipartite graphs and study the P_4-free partition and cover numbers of these graphs. Leveraging this connection, we present representative applications of these graph properties and their estimates to information theory and circuit complexity. In particular, consider the following illustrative representative problems from information theory and communication & cryptographic complexity motivating this study.

1.1.1 Problem A. Assistance for Correlation Distillation

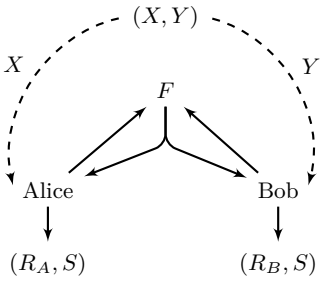
Extracting randomness [32, 56], establishing secret keys [49, 50, 51, 1, 2], and performing general secure computation [16, 17, 40, 41, 19, 42, 18, 64, 65, 39, 13] with maximum efficiency and resilience from noise sources is fundamental to theoretical computer science and information theory. Towards that objective, we study the communication and cryptographic complexity of parties to agree on a shared secret and extract private local randomness from a source.

A setup (see part (a) of Figure 1), the only source of randomness in the system, samples (x, y) according to the joint probability distribution p_{XY}, and (privately) sends x to Alice and y to Bob. Alice and Bob's objective is to agree on a shared secret key and private (independent) randomness without any additional public communication. A genie, who observes the sample (x, y), provides a public k-bit assistance z to Alice and Bob to facilitate their efforts. We emphasize that all agents Alice, Bob, and the genie are deterministic. After that, Alice and Bob locally compute the shared key s from their respective local views (x, z) and (y, z). Finally, Alice extracts the left-over entropy from x (conditioned on (s, z)) as her local private randomness r_A. Similarly, Bob extracts his local private randomness r_B from the left-over entropy of y.

For the security of Bob's local randomness, an honest but curious Alice cannot obtain any additional information on r_B beyond what is already revealed by z and s. Analogously, Bob's view should contain no additional information on Alice's view conditioned on z and s.



(a) Problem A: Assistance for Correlation Distillation



(b) Communication and Cryptographic lower bounds

Figure 1 Part (a). A pictorial summary of the system in our motivating problem A.
Part (b). The setup samples (x, y) according to the distribution p_{XY} and sends x to Alice and y to Bob. Alice and Bob use F adaptively multiple times to communicate with each other; F delivers its output to both Alice and Bob. The functionality F may be a communication protocol (i.e., a message forwarding functionality), or help Alice and Bob evaluate any (possibly, a stateful) functionality of their inputs. The objective of Alice and Bob is to generate a shared secret key s at the end of the protocol and extract the left-over entropy in their shares as independent local randomness.

Intuitively, conditioned on the genie's assistance Z, Alice-Bob samples' joint distribution splits into shared randomness and local independent randomness.

What is the *minimum* length k of the genie's assistance sufficient for Alice and Bob to agree on a shared key and obtain secure private randomness? In particular, which distributions p_{XY} need no assistance at all?

Mutual information and other common information variants (refer to Appendix D in the full version for discussion) cannot accurately measure this information-theoretic property; thus, motivating our study. This problem is equivalent to computing the P_4-free partition number of a bipartite graph encoding the (flat) joint probability distribution p_{XY}. In particular, lower bounds to k translates into lower bounds on (interactive) communication and cryptographic complexity (see part (b) of Figure 1).

1.1.2 Problem B. Nondeterministic Communication Complexity relative to the Equality Oracle

The nondeterministic communication complexity of the equality function is high [44]. However, what is the additional utility of an oracle call to the equality function in computing other functions?

Suppose Alice has input $x \in X$, Bob has input $y \in Y$, and are interested in computing the Boolean function $f\colon X \times Y \to \{0,1\}$ of their private inputs. They have access to an *equality oracle* $\mathsf{EQ}\colon \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$ defined by $\mathsf{EQ}(a,b) = 1$ if and only if $a = b$. They are interested in computing $f(x,y)$ using this equality oracle and a k-bit nondeterministic input *without any additional communication.*

The functions $A\colon X \times \{0,1\}^k \to \{0,1\}^*$ and $B\colon Y \times \{0,1\}^k \to \{0,1\}^*$ satisfying the following constraints define a *nondeterministic protocol* for f relative to the equality oracle.

1. For every input-pair $(x,y) \in X \times Y$ such that the output $f(x,y) = 1$, there exists a nondeterministic input $z \in \{0,1\}^k$ ensuring $\mathsf{EQ}(\ A(x,z)\ ,\ B(y,z)\ ) = 1$.
2. For every input-pair $(x,y) \in X \times Y$ such that the output $f(x,y) = 0$, for all nondeterministic inputs $z \in \{0,1\}^k$, we have $\mathsf{EQ}(\ A(x,z)\ ,\ B(y,z)\ ) = 0$.

The *communication complexity* of this protocol is k, i.e., the length of the nondeterministic input. What is the *minimum* communication complexity k of the function f?

Intuitively, we are augmenting the nondeterministic communication protocols with an equality oracle at the output. If the EQ oracle is useful to compute a function f, then its communication complexity in our model shall be significantly lower than where the parties cannot access the EQ oracle. We show that this problem is identical to the P_4-free cover number of a bipartite graph encoding the Boolean function f. Our results show that the access to the equality oracle reduces the communication complexity of computing set disjointness by a constant factor compared to the model where parties do not have access to the equality oracle. To compute the inequality function, perhaps surprisingly, we show an *exponential* reduction in the communication complexity. On the other hand, access to the equality oracle is virtually useless to computing the set intersection. Section 1.3 provides the details.

1.1.3 Additional Applications and History

In Appendix F of the full version, we present a representative scheduling problem that naturally reduces to computing P_4-free partition/cover numbers. Beyond the applications above, this example highlights the innate ability of P_4-free graphs to encode scheduling problems that are amenable to *parallelization*.

Edge-partitioning graphs using the minimum number of P_4-free graphs have found applications in *leakage-resilient cryptography* [9]. In particular, if k-bits of genie's assistance suffices for the setup in problem A, then k-bits of leakage also suffices for the adversary to destroy the possibility of performing general secure computation. Identifying a large P_4-free subgraph of a given graph is studied in clustering. For example, an *exclusive row and column bicluster* [48, 37] is identical to a P_4-free graph, with applications in analyzing biological data. [15] used P_4-free partition and cover numbers to approach a coloring conjecture (a variant of Ryser's conjecture) for bipartite graphs.

1.1.4 Related graph properties: Equivalence Cover Number and Product/Prague Dimension

The following discussion is specific to *loopless undirected graphs*. An *equivalence graph* is a (disjoint) union of cliques. The *equivalence cover number* of a graph G is the minimum number d of equivalence sub-graphs that cover the edges of G [54, 55]. Note that the P_4-free cover number is an extension of this concept to bipartite graphs. Furthermore, the equivalence cover number of G is identical to the *product/Prague dimension* of the complement of the graph G [63, 30], the minimum $d \in \mathbb{N}$ such that the complement of the graph G is an induced subgraph of $K_{\mathbb{N}}^d$ (the d-fold product of the infinite complete graph $K_{\mathbb{N}}$). Computing the equivalence cover number or the product dimension of a graph is NP-complete [54].

The P_4-free cover number (for bipartite graphs) has a close connection to the product (a.k.a., Prague) dimension.

▶ **Proposition 1.** *If a redundancy-free*[3] *bipartite graph $G = (L, R, E)$ has a size-d P_4-free edge-covering, then the complement bipartite graph $\overline{G} := (L, R, L \times R \setminus E)$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$.*

[3] A graph is redundancy-free if no two vertices have an identical neighborhood.

The converse of the proposition does not hold exactly (refer to Section 7). However, if $\overline{G}$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$, then G has a size-$(d+1)$ P_4-free cover. We prove that $\mathsf{P_4\text{-}fc}(G) \in \{\mathsf{pdim}(H), \mathsf{pdim}(H) - 1\}$, where $G = (L, R, E \subseteq L \times R)$ is a bipartite graph, $H = (L \cup R, L \times R \setminus E)$ is the loopless undirected graph representing the complement of the bipartite graph G, and $\mathsf{pdim}(H)$ is the product/Prague dimension of H (refer to Corollary 34 in Section 7). Figure 7 presents a graph showing the necessity of this slack in the characterization. However, for most applications, an additive slack of one should be acceptable. This proposition facilitates lower-bounding the $\mathsf{P_4\text{-}fc}(G)$ using the algebraic lower-bounding techniques for the product/Prague dimension [47, 4, 63, 5].

Despite this similarity, extremal properties of the equivalence cover number and product/Prague dimension need not translate into extremal properties of the P_4-free cover number. For example, an N-vertex star has an equivalence cover number $(N-1)$ [63]. On the other hand, the P_4-free cover number of any bipartite graph with size-N partite sets is at most its star arboricity (because star forests are P_4-free), which is at most (roughly) $N/2$ [3]. The bottleneck here is that the $\mathsf{P_4\text{-}fc}(G)$ is close to $\mathsf{pdim}(H)$, where H represents a bipartite graph, i.e., the graph H is structured (triangle-free in this particular case). The graphs realizing the extremal properties for equivalence cover number and product/Prague dimension need not have this structure. In particular, the construction of bipartite graphs with high P_4-free cover and partition numbers turns out to be non-trivial, and our work establishes a connection to the well-known (unsolved) Zarankiewicz problem [11] and relies on probabilistic techniques to demonstrate their existence.

Section 7 also presents a variant of the product/Prague dimension to estimate the P_4-free partition number (see Corollary 37). A lower bound for the P_4-free partition number is non-trivial if it is not already a lower bound to the P_4-free cover number. Unfortunately, no non-trivial lower-bounding techniques for general graphs are known for this new graph embedding property. When non-trivial lower bounds for this variant of the product/Prague dimension is proven, they shall transfer to the P_4-free partition number.

Among several notions of product dimension for graphs [30], most of which are unrelated to the property we wish to capture,[4] the graph property mentioned above is the closest and most relevant.

1.2 P_4-free Partition Number

We reduce problem A to computing the P_4-free partition number. We present the reduction's highlight. A bipartite graph G naturally represents a (flat) joint distribution p_{XY}, where the edge-set is the support of p_{XY} (see Figure 2 for examples). If G is already P_4-free, then Alice and Bob need no assistance from the genie; the connected component's identity is their shared key s, and (conditioned on the identity of the shared key) their samples $r_A = (x|s)$ and $r_B = (y|s)$ are independent private randomness. If G is not P_4-free, the genie decomposes G into $G_1, \ldots, G_d$ such that each G_i is P_4-free and the edge sets $E(G_1), \ldots .E(G_d)$ partition the edge set $E(G)$. For a joint sample $(u, v) \in E(G)$, the genie reveals the (unique) $z = i$ such that $(u, v) \in E(G_i)$. Conditioning on the genie's assistance $z = i$, Alice-Bob's samples come from the joint distribution G_i, which is P_4-free, so they agree on their shared key and secure private randomness as above. To minimize the genie's assistance, one needs to minimize $d \in \mathbb{N}$, identical to $\mathsf{P_4\text{-}fp}(G)$.

[4] Even the notions of dimension that are deceptively similar sounding, for example, the "product dimension of bipartite graphs" introduced by [59], are unrelated to the graph properties that this paper studies.

	00	11	01	10
00	1	1	0	0
11	1	1	0	0
01	0	0	1	1
10	0	0	1	1

(a) Forward or flip.

	00	01	10	11
00	1	1	0	0
01	0	1	1	0
10	0	0	1	1
11	1	0	0	1

(b) Noisy typewriter.

Figure 2 Pictorial representation of the probability distributions (a) forward or flip, and (b) noisy typewriter distributions, for $n = 2$. Rows correspond to Alice samples, and columns correspond to Bob samples. The (i, j)-th entry of a matrix being 1 represents that (i, j) is in the support of the distribution. The distribution is a uniform distribution over all the elements in the support. Let G_a be the bipartite graph whose adjacency matrix is defined by the matrix representation of the forward and flip distribution. The graph G_a is a disjoint union of 2^{n-1} copies of the $K_{2,2}$ biclique. Note that G_a is P_4-free, and, hence, $\mathsf{P_4\text{-}fp}(G_a) = 1$. Let G_b be the bipartite graph whose adjacency matrix is defined by the matrix representation of the noisy typewriter distribution. The graph G_b is a cycle of length 2^{n+1}. Note that G_b is *not* P_4-free, and $\mathsf{P_4\text{-}fp}(G_b) = 2$ (the graph decomposes into two matchings).

1.2.1 Discussion on Problem A

We begin by expanding how lower-bounding the information-theoretic measure in problem A translates into communication and cryptographic lower bounds (as in [8]). Suppose, in our model, one proves that the genie's assistance must be $k \geq k^*$ bits. Now consider the setting in part (b) of Figure 1 where there is no genie; however, the parties have access to a functionality F. The functionality F may be an arbitrary *communication protocol* or multiple calls to arbitrary *interactive stateful functionalities* that receive adaptive inputs from Alice and Bob. In particular, F may be multiple copies of the NAND-functionality, which is sufficient for general secure computation [68, 27, 42]. Observe that the genie can simulate the functionality F's entire output with access to (x, y). Consequently, we have the following result.

▶ **Proposition 2.** *If p_{XY} needs $k \geq k^*$ bits of assistance from the genie in our model, then Alice and Bob need to receive at least k^* bits from F in the Figure 1 part (b) model to establish a shared key s and extract the left-over entropy in their sample as independent private randomness.*

In information theory, Gray-Wyner systems/networks are well-studied [66]. However, existing measures like mutual information and various notions of common information are inadequate to capture the information-theoretic property in Problem A accurately. For example, there are two joint distributions with identical (Shannon's) mutual information [61]; however, one needs no assistance while the other needs one-bit assistance.[5] Refer to Figure 2 for the following discussion. Consider the first distribution (namely, the *forward or flip distribution*), where Alice gets i.i.d. uniformly random bits $x = (x_1, x_2, \ldots, x_n)$, and Bob either (with probability half) gets $y = x$ or $y = (\overline{x_1}, \ldots, \overline{x_n})$, i.e., every bit of x is flipped. In the second distribution (the *noisy typewriter distribution*), Alice gets a uniformly random sample $x \in \{0, 1, \ldots, 2^n - 1\}$, and Bob either gets $y = x$ or $y = (x + 1) \mod 2^n$ with probability half. The bipartite graph corresponding to the forward or flip distribution is, indeed, P_4-free, and the bipartite graph corresponding to the noisy typewriter distribution

[5] By tensorizing the distributions, one can increase the gap in the necessary assistance arbitrarily.

has P_4-free partition number 2 (i.e., one-bit assistance is necessary and sufficient). Both distributions have $(n-1)$ bits of mutual information; however, the first distribution needs no assistance, but the second distribution needs one-bit assistance[6] to agree on a secret key.

Wyner's common information [66] estimates the minimum assistance that removes any dependence between Alice-Bob samples. This quantity is a significant overestimation (for example, in the forward or flip distribution, it needs $(n-1)$-bits of assistance $z = (x_1, \ldots, x_{n-1})$), and Wyner's assistance eliminates the possibility of Alice and Bob agreeing on a secret key, which defeats the objective of this problem. Gács-Körner common information [25] estimates the length of the secret key that Alice and Bob can generate without any assistance from the genie, which results in pessimistic estimates. For example, starting with samples from the noisy typewriter distribution, Alice and Bob cannot even agree on a one-bit secret; however, appropriate one-bit assistance would help them generate an $(n-1)$-bit secret. Likewise, non-interactive correlation distillation [53, 52] enables parties to agree on a secret non-interactively *without any assistance*. However, even without the necessity to generate independent local randomness, strong hardness of computation results are known [53, 52, 67, 10, 14].

Refer to Appendix D in the full version for additional discussion on various forms of common information.

1.2.2 Our results for Problem A

Observe that the naïve assistance that reveals the XOR of the parties' inputs suffices; however, the minimum assistance may be exponentially smaller. Our work relies on suitably encoding (flat) joint distributions as bipartite graphs. We prove in Theorem 5 that ascertaining the minimum assistance is, in general, difficult. Furthermore, there are joint distributions where the minimum assistance that is needed is close to the naïve assistance mentioned above, yielding lower bounds in communication and cryptographic complexity. In other words, we obtain the following as a corollary to Theorem 6.

▶ **Corollary 3.** *Let $\Omega_X = \Omega_Y = \{0,1\}^n$. Fix $t \in \mathbb{N}$. There are joint distributions over the sample space $\Omega_X \times \Omega_Y$ that require Alice and Bob to (each) receive at least $\left(1 - \frac{2}{t+2}\right) n$ bits of communication in the model in Figure 1 part (b).*

Finally, we upper-bound the minimum assistance needed for a few well-studied probability distributions i.e. when p_{XY} is the INT_n[7] or the DISJ_n[8] joint distribution, then $\lceil n/2 \rceil$-bit assistance suffices (we explicitly provide the assistance that the genie provides and it is efficient to compute, see Theorem 8). For INEQ_N, where $N = 2^n$, the genie needs to provide $\lceil \log n \rceil$ bits of assistance. The assistance for INEQ_N is optimal because we prove a matching lower bound. In general, $\min\{\log_2 N, \frac{1}{2}\log_2 |\mathrm{Supp}(p_{XY})|\}$ bits of assistance suffices.[9]

---

[6] The genie notifies the parties whether $y = x$ or not.

[7] Alice receives random $X \subseteq \{1, 2, \ldots, n\}$, and Bob receives random $Y \subseteq \{1, 2, \ldots, n\}$ conditioned on $X \cap Y \neq \emptyset$.

[8] Alice receives random $X \subseteq \{1, 2, \ldots, n\}$, and Bob receives random $Y \subseteq \{1, 2, \ldots, n\}$ conditioned on $X \cap Y = \emptyset$.

[9] Because, $\mathsf{P_4\text{-}fp}(G) \leq \mathsf{sa}(G) \leq \mathcal{O}\left(\sqrt{|E(G)|}\right)$. The last bound on the star arboricity of G follows from an averaging argument and the bound of [3].

1.3 P_4-free Cover Number

We reduce Problem B to the P_4-free cover number. Boolean functions naturally encode a bipartite graph's adjacency matrix; an input-pair that evaluates to 1 denotes an edge in the graph. If the graph G (of a function f) is P_4-free, then parties need no nondeterministic input; they can evaluate f using the EQ oracle.[10] Otherwise, decompose G into P_4-free $G_1, \ldots, G_d$ such that the union of the edge-sets of $G_1, \ldots, G_d$ is the edge-set of G. For input (x, y) such that $f(x, y) = 1$, the nondeterministic input is $i \in \{1, \ldots, d\}$, where the edge-set of G_i contains the edge (x, y). Next, given this nondeterministic input, parties can evaluate f. For input (x, y) such that $f(x, y) = 0$, no nondeterministic input can make Alice and Bob output 1. One minimizes $d \in \mathbb{N}$ to minimize the nondeterministic communication complexity, which is identical to $\mathsf{P_4\text{-}fc}(G)$.

1.3.1 Discussion on Problem B

The equality function in the *standard* nondeterministic communication complexity model (where parties *do not* have access to the EQ oracle) has high nondeterministic communication complexity. Determining the minimum nondeterministic input is equivalent to covering the input-pairs where the output is 1 using a minimum number of *combinatorial rectangles*, a.k.a., the *biclique cover number* [35]. The motivating problem's objective is to characterize the utility of oracle access to the EQ function in computing other functions. If the EQ oracle is useful, then the nondeterministic communication complexity relative to the EQ oracle shall be lower than without accessing the EQ oracle. The particular notion of "reduction" considered above is similar to Karp-reduction [38], which permits only one call to the oracle and no post-processing of the oracle's output. Similarly, in circuit complexity, it is typical to augment a circuit class with a more expressive gate at the output that is not computable by circuits in that class. For example, one studies the effects of augmenting AC^0 circuits with a MAJ (majority) gate or a THR (threshold) gate at the output [7, 26, 33, 29], enabling a controlled exploration of the gap between the power of AC^0 and TC^0 circuits.

1.3.2 Our results for Problem B

Similar to the result for P_4-free partition number, we prove that computing the P_4-free cover number is difficult (see Theorem 5), and there are functions that need nondeterministic input (roughly) the size of the parties' inputs, in other words, we obtain the following as a corollary to Theorem 6.

▶ **Corollary 4.** *Fix $t \in \mathbb{N}$. There are Boolean functions $f\colon \{1, 2, \ldots, N\} \times \{1, 2, \ldots, N\} \to \{0, 1\}$ requiring at least $(1 - \frac{2}{t+2}) \log_2 N$ bits of nondeterministic input in the communication complexity model where parties have access to the EQ oracle.*

These functions are analogs of the "fooling sets" in our communication model. In the standard nondeterministic communication model, the EQ function is hard-to-compute and needs n-bits of nondeterministic input. The "fooling set" lower-bounding technique draws inspiration from this result. For a general f, this argument demonstrates pairs of Alice and Bob's input-sets where only the diagonal elements are 1; and the rest are 0. That is, the function f has an embedded EQ function. The size of this "embedded EQ" (a.k.a., the fooling set) in

[10] Parties compute the connected component where their private input belongs. Then, they use the EQ oracle to test if they belong to the same connected component.

f suffices to prove lower bounds on the nondeterministic input needed to compute f. In our setting, these functions that require $(1 - \frac{2}{t+2})n$-bit nondeterministic input serve as "fooling sets" in the nondeterministic communication complexity model where parties can access the EQ oracle.

Next, we provide estimates for some well-known functions in communication complexity (see Theorem 8). We prove that the P_4-free cover number of DISJ_n is (roughly) $\leq \sqrt{N}$. That is, only $n/2$ bits of nondeterministic input suffices to compute this function. Recall that, in the standard model, the function DISJ_n requires n-bit nondeterministic input because $\{(X, \{1, 2, \ldots, n\} \setminus X)\}_{X \subseteq \{1,2,\ldots,n\}}$ is a fooling set. Consequently, our result demonstrates a linear gap in the number of bits needed in our model, which indicates that the EQ oracle is non-trivially useful to compute DISJ_n. We prove a lower bound showing that $0.085n$-bit assistance is necessary.

Next, we prove that the P_4-free cover number of INT_n is between n and $n(1 - \frac{\log_2(n)}{n})$. Observe that the nondeterministic communication complexity of INT_n (without access to the EQ oracle) is already $\lceil \log_2 n \rceil$ bits. Consequently, EQ oracle's access is practically useless because the difference between the ceiling of the log of the lower and the upper bounds is at most 1 (asymptotically).

Finally, we show that INEQ_N needs only $\log_2 \log_2 N$ bit nondeterministic input using the EQ oracle. Intuitively, if $N = 2^{2^s}$ and all inputs are 2^s-bit binary strings, then the nondeterministic input is the s-bit index where the parties' input differ. Recall that in the standard model (without access to the EQ oracle), INEQ_N requires $\log_2 N$-bit nondeterministic input, which is exponentially higher. Furthermore, using the algebraic technique of [47, 63], we prove a matching lower bound to the P_4-free cover number of INEQ_N. Observe that we prove that $\mathsf{P_4\text{-}fp}(\mathsf{INEQ}_N)$, not just $\mathsf{P_4\text{-}fc}(\mathsf{INEQ}_N)$, matches the lower bound for the $\mathsf{P_4\text{-}fc}(\mathsf{INEQ}_N)$.

2 Our Contribution

We prove the NP-completeness of determining the P_4-free partition and cover numbers of a bipartite graph.

▶ **Theorem 5** (Hardness of P_4-free Partition and Cover). *The following languages are* NP*-complete.*

$$P_4\text{-}\mathit{FREE\text{-}PART} = \{\ \langle G \rangle\ \mid G \text{ is a bipartite graph and } \mathsf{P_4\text{-}fp}(G) \leq 2\},$$
$$P_4\text{-}\mathit{FREE\text{-}COV} = \{\ \langle G \rangle\ \mid G \text{ is a bipartite graph and } \mathsf{P_4\text{-}fc}(G) \leq 2\}.$$

Similar problems, for example, calculating the biclique partition number/cover [57] and star arboricity [34] (even for bipartite graphs) are NP-complete.

Next, we prove that there are graphs G with large P_4-free partition and cover numbers. Note that for a bipartite graph $G = (L, R, E)$, we have $\mathsf{P_4\text{-}fc}(G) \leq \mathsf{P_4\text{-}fp}(G) \leq \min\{|L|, |R|\}$ by decomposing the graph into stars rooted at vertices of the smaller partite set. Towards understanding the tightness of this naïve upper-bound, we show that, for any $N \in \mathbb{N}$ and constant $\epsilon \in \{1/3, 1/4, \ldots\}$, there are bipartite graphs with size-N partite sets and $\mathsf{P_4\text{-}fp}(G) \geq \mathsf{P_4\text{-}fc}(G) \geq \Omega(\epsilon \cdot N^{1-2\epsilon})$ (roughly).

▶ **Theorem 6** (High P_4- Free Partition and Cover Numbers). *Let C be an appropriate positive absolute constant and $t \in \mathbb{N}$ be a parameter. There exists $N_0 \in \mathbb{N}$ such that for all $N \in \mathbb{N}$ and $N \geq N_0$, there is a graph $G_{N,t} = (L, R, E)$ such that (1) $|L| = |R| = N$, and (2) $\mathsf{P_4\text{-}fp}(G_{N,t}) \geq \mathsf{P_4\text{-}fc}(G_{N,t}) \geq C \cdot \frac{1}{t} \cdot N^{1-\frac{2}{t+2}}$.*

Our constructions rely on extremal bipartite graphs that avoid $K_{t+1,t+1}$-subgraphs (the unsolved Zarankiewicz problem [11]), for which only probabilistic constructions are known (refer to the discussion in Section 4). Explicit constructions are known only for very specialized values of t. However, the P_4-free partition and cover numbers of $G_{N,t}$ cannot be too large. For any sparse bipartite graph G, using an averaging argument, its star-arboricity has the upper bound $\mathsf{sa}(G) \leq \mathcal{O}\left(\sqrt{|E(G)|}\right)$ [3]. Since star forests are P_4-free and $G_{N,t}$ has $\mathcal{O}\left(N^{2-\frac{2}{t+1}}\right)$ edges, it implies that $\mathsf{P_4\text{-}fp}(G_{N,t}) \leq \mathcal{O}\left(N^{1-\frac{1}{t+2}}\right)$.

In problem A, the joint distributions corresponding to these bipartite graphs require a lot of assistance from the genie. Consequently, these lower bounds translate into communication and cryptographic complexity lower bounds. The functions corresponding to these bipartite graphs are difficult to compute for parties with nondeterministic input and access to the EQ oracle. If these functions are embedded in another function, then that function must have high nondeterministic communication complexity as well.

As a corollary (of the proof technique presented above), we prove the following result for dense bipartite graphs drawn from the Erdős-Rényi distribution with (constant) parameter $p \in (0,1)$. Graphs drawn from $\mathsf{ER}(N,N,p)$ avoid bicliques with size-$(2\log_a N)$ partite sets. Therefore, we have the following result.

▶ **Corollary 7** (High P_4-Free Partition and Cover Number of Erdős-Rényi Graphs). *Let $p \in (0,1)$ be a constant parameter. Let $\mathsf{ER}(N,N,p)$ represent the distribution over the sample space of all bipartite graphs over size-N partite sets that includes every edge into the graph independently with probability p. Then, for $a = 1/p$, we have*

$$\Pr\left[\mathsf{P_4\text{-}fp}(G) \geq \mathsf{P_4\text{-}fc}(G) \geq \frac{pN}{4\log_a N} \cdot (1-o(1)) \colon G \xleftarrow{\$} \mathsf{ER}(N,N,p)\right] \geq 1 - o(1).$$

Upper bounds to the P_4-free cover and partition numbers for bipartite Erdős-Rényi graphs is potentially an extremely challenging problem. Upper-bounding the P_4-free partition number of Erdős-Rényi bipartite graphs remains open.

Finally, we estimate the P_4-free partition and cover numbers for the graphs $\mathsf{INT}_n, \mathsf{DISJ}_n$, and INEQ_N that are well-studied functions from communication theory and are defined below.

1. **The Intersection Graph.** For $n \in \mathbb{N}$, let $\mathsf{INT}_n = (\{0,1\}^n, \{0,1\}^n, E)$ be the bipartite graph defined as follows. For any $u, v \in \{0,1\}^n$, we have $(u,v) \in E$ if and only if the set $U \subseteq \{1,2,\ldots,n\}$ indicated by u, intersects the set $V \subseteq \{1,2,\ldots,n\}$ indicated by v.
2. **The Disjointness Graph.** For $n \in \mathbb{N}$, let $\mathsf{DISJ}_n = (\{0,1\}^n, \{0,1\}^n, E)$ be the bipartite graph defined as follows. For any $u, v \in \{0,1\}^n$, we have $(u,v) \in E$ if and only if the set $U \subseteq \{1,2,\ldots,n\}$ indicated by u, is disjoint from the set $V \subseteq \{1,2,\ldots,n\}$ indicated by v.
3. **The Inequality Graph.** For $N \in \mathbb{N}$, let $\mathsf{INEQ}_N = (\{1,2,\ldots,N\}, \{1,2,\ldots,N\}, E)$ be the bipartite graph defined as follows. For any $u, v \in \{1,2,\ldots,N\}$, we have $(u,v) \in E$ if and only if $u \neq v$.

▶ **Theorem 8** (Estimates for Particular Graphs). *For all $n, N \in \mathbb{N}$, the following statements hold.*

1. $n - \frac{1}{2}\lg(n) - \mathcal{O}(1) \leq \mathsf{P_4\text{-}fc}(\mathsf{INT}_n) \leq n$, *and* $\mathsf{P_4\text{-}fp}(\mathsf{INT}_n) \leq \begin{cases} 2\cdot 2^{n/2} - 2, & \text{even } n, \text{ and} \\ 3\cdot 2^{(n-1)/2} - 2, & \text{odd } n. \end{cases}$
2. $2^{0.085n} \leq \mathsf{P_4\text{-}fc}(\mathsf{DISJ}_n) \leq \mathsf{P_4\text{-}fp}(\mathsf{DISJ}_n) \leq 2^{\lceil n/2 \rceil}$.
3. $\mathsf{P_4\text{-}fc}(\mathsf{INEQ}_N) = \mathsf{P_4\text{-}fp}(\mathsf{INEQ}_N) = \lceil \log_2 N \rceil$.

Recall that for any Boolean function f, parties can calculate it with $\lceil \log_2 \mathsf{P_4\text{-}fc}\,(G(f))\rceil$-bit nondeterministic input and one call to the EQ oracle, where $G(f)$ is the bipartite graph representing the Boolean function f. Therefore, the bounds above translate into communication bounds.

Observe the exponential gap between the upper bounds on the P_4-free cover and partition numbers of INT_n. We conjecture that similar to the exponential gaps in the biclique cover and partition number of some graphs [58], INT_n is a candidate bipartite graph witnessing an exponential gap in its P_4-free cover and partition numbers. Currently, the authors are unaware of any general non-trivial lower bounding technique for the partition number that is not a lower bound to the cover number for this problem.

Lower-bounding the P_4-free cover numbers of INEQ_N and INT_n relies on Proposition 1 and the algebraic technique of [47, 63]. Furthermore, the P_4-free cover and partition numbers of INEQ_N are exact, previously unknown for the partition number. Finally, the lower bound on the P_4-free cover number of DISJ_n uses a new counting strategy.

3 Hardness of P_4-free Partition and Cover Numbers

In this section, we will prove Theorem 5. Our proof of hardness for both partition and cover number is based on a result from [28], which shows that computing the edge partition of a bipartite planar graph into two star forests is NP-complete.

▶ **Definition 9.** *A* star *is a tree with one internal node, in other words, a biclique in which either the left partite set or the right partite set has size one. A* star forest *is a forest whose connected components are stars. The* star arboricity *of a graph, represented by* $\mathsf{sa}\,(G)$*, is the minimum number of star forests that a graph can be partitioned into.*

▶ **Imported Theorem 10** (Gonçalves and Ochem [28])**.** *For any $g > 3$, deciding whether a bipartite planar graph G with girth*[11] *at least g and maximum degree 3 satisfies $\mathsf{sa}\,(G) \leq 2$ is NP-complete.*

Proof of Theorem 5. First we show the decision problem is in NP, that is, given a partition of the edge set of G into ≤ 2 components we can verify in polynomial time whether it is a P_4-free partition of size ≤ 2 of G or not. This can be done in polynomial time by checking if any set of four vertices (two in the left set and two in the right set) in each component is P_4-free.

Next we show that the decision problem from Theorem 10 is polynomial-time reducible to the P_4-free partition and cover number on bipartite graphs. The decision problem in Theorem 10 is NP-complete for any bipartite planar graph of girth at least $g > 3$; in particular, it holds for $g \geq 6$. Suppose we have a bipartite planar graph G with girth $g \geq 6$ and maximum degree 3. Since G has girth at least 6, there are no cycles of length less than 6 in G. It implies that $K_{2,2}$ is not a subgraph of G. Therefore, any disjoint union of bicliques in G is a star forest. This implies that $\mathsf{sa}\,(G) = \mathsf{P_4\text{-}fp}\,(G) = \mathsf{P_4\text{-}fc}\,(G)$, since $K_{2,2}$-free graphs have the property that the P_4-free partition and cover numbers are both identical to the star arboricity. Thus, the star arboricity of G is ≤ 2 if and only if the P_4-free partition number of G is ≤ 2. ◀

[11] The girth of an undirected graph is the length of the shortest cycle in the graph.

4 High P_4-free Partition and Cover Numbers

We shall prove Theorem 6 and Corollary 7 in this section. We begin with some terminologies in extremal graph theory. Fix a graph H. A classical problem in graph theory is to find the maximum number of edges in a graph on N vertices that does not contain a copy of H.

▶ **Definition 11** (Turán number). *Turán number denoted by $ex(N, H)$ is the maximum number of edges in an N-vertex graph that does not contain a copy of H.*

A sub-problem of special interest is when H is a complete bipartite graph, this problem is commonly referred to as the Zarankiewicz problem.

▶ **Definition 12** (Zarankiewicz function). *Zarankiewicz function, denoted by $z(M, N; s, t)$, is the maximum number of edges in a bipartite graph $G = (L, R, E)$, where $|L| = M$ and $|R| = N$, that does not contain a sub-graph of the form $K_{s,t}$.*

The Zarankiewicz function is well-studied [24]. The best general lower bound obtained by the probabilistic method [20] yields the following bound.

▶ **Imported Theorem 13** (Erdős and Spencer [20]). *For all $a, b \in \mathbb{N}$, we have $ex(N, K_{a,b}) \geq C \cdot N^{2-\frac{a+b-2}{ab-1}}$, where C is a positive absolute constant.*

An explicit construction for $K_{t+1,t+1}$-avoiding graphs for $t = 2$ is known [12], which has $\frac{1}{2}N^{\frac{5}{3}} + o(N^{\frac{5}{3}})$ edges.[12] Using *norm graphs*, constructions of $K_{t,s}$-avoiding graphs for fixed $t \geq 2$ and $s > (t-1)!$ are known as well [43, 6]. Note that the latter set of constructions do not apply to our setting for $t > 3$. Considering the adjacency matrix of a $K_{a,b}$-free graph on n vertices, we get $z(N, N, a, b) \geq 2ex(N, K_{a,b})$.

Let $G = (L, R, E)$ be a bipartite graph. A *combinatorial rectangle* is a set of the form $A \times B$, where $A \subseteq L$ and $B \subseteq R$. Observe that a combinatorial rectangle corresponds to a biclique if we restrict ourselves to rectangles of the form $\{A \times B : (u, v) \in A \times B \iff (u, v) \in E\}$. We shall use this fact in the sequel to show that the P_4-free partition number of a $K_{t+1,t+1}$-free bipartite graph is high.

▶ **Lemma 14.** *For a bipartite graph $G = (L, R, E)$ such that $|L| = |R| = N$, if G is $K_{t+1,t+1}$-free for some $t > 0$, then $\mathsf{P_4\text{-}fp}(G) \geq \frac{e(G)}{2Nt}$.*

Proof. Consider the adjacency matrix of the bipartite graph G. A biclique in G can be represented as a combinatorial rectangle in the adjacency matrix of G (as explained above). The *width* of this combinatorial rectangle is the smaller of its two dimensions, and the *length* of this combinatorial rectangle is the larger of the two dimensions. Observe that any P_4-free bipartite graph is the union of non-intersecting combinatorial rectangles.

Let G' be a P_4-free bipartite sub-graph of G. It is instructive to refer to Figure 3. For any combinatorial rectangle in G', *length* $\leq 2N$ and *width* $\leq t$, since if *width* $= t + 1 \leq$ *length*, then there exists a $K_{t+1,t+1}$-subgraph in G. This observation implies that $e(G') < 2Nt$, and consequently $\mathsf{P_4\text{-}fp}(G) \geq \frac{e(G)}{2Nt}$. ◀

The proof of Theorem 6 follows from the fact about Zarankiewicz function of $K_{t+1,t+1}$-free bipartite graphs and Lemma 14.

[12] For $t = 1$, Levi graph of a finite projective plane yields an explicit construction.

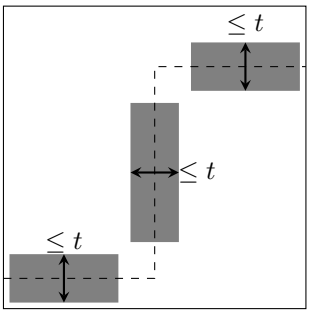


Figure 3 Let $t \in \mathbb{N}$ be a parameter. Proof intuition underlying the fact that a $K_{t+1,t+1}$-free bipartite graph cannot have a dense P_4-free subgraph.

Proof of Theorem 6. We construct a bipartite graph $G = (L, R, E)$ such that $|L| = |R| = N$ and it is $K_{t+1,t+1}$-free. By Imported Theorem 13,

$$e(G) = z(N, N; t+1, t+1) \geq 2ex(N, K_{t+1,t+1}) \geq 2CN^{2-\frac{2}{t+2}},$$

where C is a positive absolute constant. By Lemma 14, we get that

$$\mathsf{P_4\text{-}fp}\,(G) \geq \frac{e(G)}{2Nt} = \frac{2CN^{2-\frac{2}{t+2}}}{2Nt} = C \cdot \frac{1}{t} \cdot N^{1-\frac{2}{t+2}}\,. \qquad \blacktriangleleft$$

Similarly, to prove that $\mathsf{ER}(N, N, p)$ have high P_4-free partition and cover numbers (Corollary 7), we rely on the following two observations.

1. The number of edges in a bipartite graph $G \xleftarrow{\$} \mathsf{ER}(N, N, p)$ is at least $pN^2 \cdot (1 - o(1))$, with probability $1 - o(1)$.
2. Furthermore, $G \xleftarrow{\$} \mathsf{ER}(N, N, p)$ is $K_{t+1,t+1}$-avoiding with high probability, where $t + 1 = \lceil 2 \log_a N \rceil$.

The proof of the second observation follows from the standard outline for first moment techniques, see, for example, [23] Chapter 7.2. More concretely, let $t + 1 = \lceil 2 \log_a N \rceil$. Let $\mathbb{N}_{t+1}$ be the random variable counting the number of $K_{t+1,t+1}$ bicliques in G. Then, we have

$$\begin{aligned}\mathrm{E}[\mathbb{N}_{t+1}] = \binom{N}{t+1}^2 p^{(t+1)^2} &\leq \left(\frac{\mathrm{e}N}{t+1}\right)^{2(t+1)} p^{(t+1)^2} = \left(\frac{\mathrm{e}Np^{\frac{t+1}{2}}}{t+1}\right)^{2(t+1)} \\ &\leq \left(\frac{\mathrm{e}N \cdot \frac{1}{N}}{t+1}\right)^{2(t+1)} = o(1)\end{aligned}$$

Therefore, with probability $1 - o(1)$, there are no $K_{t+1,t+1}$ bicliques in G.

5 Upper Bounds for INT_n, DISJ_n, and INEQ_N

In this section, we establish the upper bounds for DISJ_n, INT_n, and INEQ_N as stated in Theorem 8. We also exhibit a non-trivial gap between the star arboricity, and the P_4-free partition number of DISJ_n (see Eq. 2 of Theorem 19).

5.1 P_4-free Partition/Cover Number and Graph Products

First, we introduce the notion of a graph product, and state some properties regarding the behavior of P_4-free partition/cover number on graph products. These concepts are used to solve recurrence relations for DISJ_n and INT_n in the sequel.

▶ **Definition 15** (Graph Product). *Let $G_1 = (L_1, R_1, E_1)$ and $G_2 = (L_2, R_2, E_2)$ be two bipartite graphs. Let G denote the* tensor product *of the two bipartite graphs G_1, and G_2, represented by $G_1 \times G_2$. The partite sets of G are $L_1 \times L_2$ and $R_1 \times R_2$, and the edge set is $E(G) \coloneqq \{(\,(u,a),(v,b)\,) : (u,v) \in E_1, (a,b) \in E_2\}$.*

▷ Claim 16 (Product of P_4-free bipartite graphs is P_4-free). Let G and H be two P_4-free bipartite graphs, then $G \times H$ is also P_4-free.

▷ Claim 17 (Sub-multiplicativity of the P_4-free Partition Number). Let G and H be two bipartite graphs, then the following holds for their graph product.

$$\mathsf{P_4\text{-}fp}\,(G \times H) \leq \mathsf{P_4\text{-}fp}\,(G) \cdot \mathsf{P_4\text{-}fp}\,(H)$$

Similarly, the P_4-free cover number is also sub-multiplicative.

▷ Claim 18 (Sub-multiplicativity of the P_4-free Cover Number). Let G and H be two bipartite graphs, then the following holds for their graph product.

$$\mathsf{P_4\text{-}fc}\,(G \times H) \leq \mathsf{P_4\text{-}fc}\,(G) \cdot \mathsf{P_4\text{-}fc}\,(H)$$

5.2 Bound on DISJ_n

We show an upper bound for $\mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_n)$ using the fact that DISJ_n is the tensor product $\mathsf{DISJ}_1^{\times n}$, and we show a lower bound for $\mathsf{sa}\,(\mathsf{DISJ}_n)$, thus exhibiting a gap between the two measures.

▶ **Theorem 19.** *For any $n \in \mathbb{N}$, the following bounds hold.*

1. $\mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_n) = \mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_1^n) \leq 2^{\lceil n/2 \rceil}$, *and*
2. $\mathsf{sa}\,(\mathsf{DISJ}_n) > \lceil (3/2)^n \rceil = \lceil 2.25^{n/2} \rceil$.

Proof. For the first bound, the proof proceeds by induction on n. For the base cases, observe that $\mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_1) = \mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_2) = 2$. Next, for any $2 < n \in \mathbb{N}$, we have

$$\begin{aligned}
\mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_n) &= \mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_{n-2} \times \mathsf{DISJ}_2) \\
&\leq \mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_{n-2}) \cdot \mathsf{P_4\text{-}fp}\,(\mathsf{DISJ}_2) && \text{(Claim 17)} \\
&\leq 2^{\lceil n-2/2 \rceil} \cdot 2 && \text{(Inductive Hypothesis)} \\
&= 2^{\lceil n/2 \rceil}
\end{aligned}$$

This observation completes the inductive proof.

For the second bound, note that a star forest over partite sets L and R has $< |L| + |R| = 2 \cdot 2^n$ edges in it. Note that $e(\mathsf{DISJ}_n) = 3^n$. Therefore, one needs $> \lceil (3/2)^n \rceil$ star forests to partition the edges of DISJ_n. ◀

5.3 Bound on INT_n

First, we show that $\mathsf{P_4\text{-}fc}\,(\mathsf{INT}_n) \leq n$. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. For each $1 \leq i \leq n$, construct a subgraph $G_i = (L_i, R_i, E_i)$ of INT_n that connect all sets that contain the element i in $[n]$. More formally, $L_i = R_i = \{S \subseteq [n] \colon S \ni i\}$, and $E_i = \{(S,T) \colon S \in L_i, T \in R_i\}$. Note that G_i is a biclique and it has 4^{n-1} edges. Note also that every edge in INT_n is covered by at least one graph G_i, for some $i \in [n]$ that witnesses the intersection of the two sets. It implies that $G_1, G_2, \ldots, G_n$ is a P_4-free cover of INT_n. Therefore, it holds that $\mathsf{P_4\text{-}fc}\,(\mathsf{INT}_n) \leq n = \lg N$.

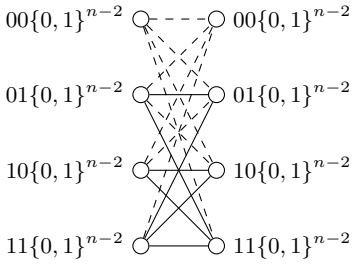


Figure 4 Partition of edges of INT_n into two sets.

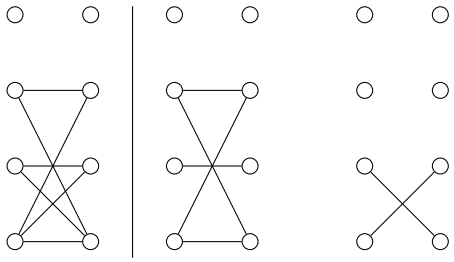


Figure 5 Partition of G_1 in Lemma 20 into two P_4-free graphs.

Next, we prove the upper bound for $\mathsf{P_4\text{-}fp}(\mathsf{INT}_n)$. Before we discuss our result, it is instructive to see that $\mathsf{P_4\text{-}fp}(\mathsf{INT}_n) \leq \mathsf{P_4\text{-}fp}(\mathsf{INT}_{n-1}) + \mathsf{P_4\text{-}fp}(\mathsf{DISJ}_{n-1})$, and by working out this recurrence relation we could have obtained a worse bound of $\mathsf{P_4\text{-}fp}(\mathsf{INT}_n) \leq 3 \cdot 2^{n/2} - 3$.

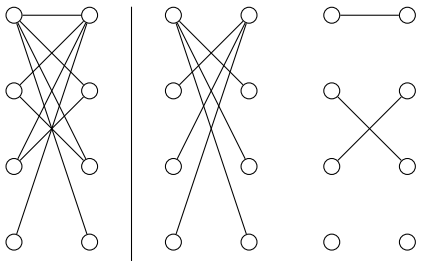


Figure 6 Partition of H_1 in Lemma 20 into two P_4-free graphs.

▶ **Lemma 20.** *For all $n \in \mathbb{N}$ and $n \geq 3$, $\mathsf{P_4\text{-}fp}(\mathsf{INT}_n) \leq 2\mathsf{P_4\text{-}fp}(\mathsf{INT}_{n-2}) + 2$*

Proof. Consider the graph INT_n. We partition the edges of INT_n into two sets. Consider an edge (u, v) where $u, v \in \{0,1\}^n$. Let $u' \in \{0,1\}^2$ represent the two most significant bits in u, define v' similarly. Let b_{uv} be an indicator variable that takes value 1 when u' and v' intersect, and 0 otherwise.

If for the edge (u, v), $b_{uv} = 1$, then we add the edge to the "bold" set. When $b_{uv} = 0$, we add the edge in the "dashed" set (refer to Figure 4). Let G be the subgraph induced by the bold edges, and let H be the subgraph induced by the dashed edges.

Next, we note that $G = K_{2^{n-2},2^{n-2}} \times G_1$ where G_1 is a graph with P_4-free partition number 2. See Figure 5 for an illustration. Similarly, $H = \mathsf{INT}_{n-2} \times H_1$ where H_1 has P_4-free partition number 2. See Figure 6 for an illustration. Combing the above observations, we get that

$$\begin{aligned}
\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) &\leq \mathsf{P_4\text{-}fp}\left(G\right) + \mathsf{P_4\text{-}fp}\left(H\right) \\
&\leq \mathsf{P_4\text{-}fp}\left(K_{2^{n-2},2^{n-2}} \times G_1\right) + \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2} \times H_1\right) \\
&\leq \mathsf{P_4\text{-}fp}\left(K_{2^{n-2},2^{n-2}}\right) \cdot \mathsf{P_4\text{-}fp}\left(G_1\right) + \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) \cdot \mathsf{P_4\text{-}fp}\left(H_1\right) \qquad \text{(Claim 17)} \\
&\leq 2 + 2\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) \qquad \blacktriangleleft
\end{aligned}$$

Applying Lemma 20 inductively, we have the following result as a consequence.

▶ **Theorem 21.** $\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) \leq \begin{cases} 2 \cdot 2^{n/2} - 2, & \text{for even } n, \\ 3 \cdot 2^{(n-1)/2} - 2, & \text{for odd } n. \end{cases}$

5.4 Bound on INEQ_N

In fact, we prove a more general result.

▷ Claim 22 (Complement of a P_4-free graph has a small P_4-free partition number). Let H be a P_4-free bipartite graph with $c \in \mathbb{N}$ connected components. Let G be the complement of H. Then, the following bound holds.

$$\mathsf{P_4\text{-}fc}\left(G\right) \leq \mathsf{P_4\text{-}fp}\left(G\right) \leq \begin{cases} \lceil \log_2 c \rceil, & \text{if } H \text{ has no isolated vertex,} \\ \lceil \log_2 c \rceil + 1, & \text{if } H \text{ has isolated vertices and } c > 1, \text{ and} \\ 2, & \text{if } H \text{ has isolated vertices and } c = 1. \end{cases}$$

Proposition 1 (along with a suitable embedding φ) implies the upper bound $\mathsf{P_4\text{-}fc}\left(G\right) \leq \lceil \log_2 c \rceil$. However, we prove the stronger result that $\mathsf{P_4\text{-}fp}\left(G\right) \leq \lceil \log_2 c \rceil$.

Our objective is to demonstrate a P_4-free partition for G of size $\lceil \log_2 c \rceil$. The proof starts by kernelizing the graph G using the rules in [21]. Essentially, without loss of generality, one can assume that H is a matching. For simplicity assume that H is a matching with c edges and assume that it has c vertices in each partite set (i.e., there are no isolated vertices).

Next, the idea is to break the problem into half the size while including only one P_4-free graph in the partition of G. Assume, without loss of generality, that the partite sets are $L = \{1, \ldots, c\}$ and $R = \{1, \ldots, c\}$, and the edges in H are (i, i), for $1 \leq i \leq c$.

Define $L_0 := \{1, \ldots, \lfloor c/2 \rfloor\}$ and $L_1 := L \setminus L_0$. Similarly, define $R_0 := \{1, \ldots, \lfloor c/2 \rfloor\}$ and $R_1 := R \setminus R_0$. Observe the following.

1. The edges induced by (L_0, R_1) and (L_1, R_0) in G are disjoint bicliques. Together, they shall form one P_4-free subgraph of G.
2. Next, the edges induced by (L_0, R_0) and (L_1, R_1) in G are disjoint and complements of matchings as well; albeit the matchings are of size $\lfloor c/2 \rfloor$ and $\lceil c/2 \rceil$, respectively. We recursively partition the disjoint union of these graphs.

Hence, Claim 22 is proved. Applying this claim for $G = \mathsf{INEQ}_N$ and $H = \mathsf{EQ}_N$, we have the following result.

▶ **Theorem 23.** *For any $N \in \mathbb{N}$, it holds that* $\mathsf{P_4\text{-}fp}\left(\mathsf{INEQ}_N\right) \leq \log_2 N$.

6 Lower Bounds for INT_n, DISJ_n, and INEQ_N

This section presents the proofs of the lower bounds in Theorem 8.

6.1 Bound for INEQ_N

We begin with a lower bound on $\mathsf{P_4\text{-}fc}(\mathsf{INEQ}_N)$ by outlining the proof of Proposition 1 below. Given a size-d P_4-free cover $\{G_1, \ldots, G_d\}$ of a bipartite graph $G = (L, R, E)$ consider the following function $\varphi\colon L \cup R \to \{1, 2\} \times \mathbb{N}^d$. For $i \in \{0, 1, \ldots, d\}$, $\varphi(u)_i$ refers to the i-th coordinate of the mapping $\varphi(u)$. Define $\varphi(u)_0 \coloneqq 1$ if $u \in L$; otherwise, if $u \in R$, define $\varphi(u)_0 \coloneqq 2$. If the edge $(u, v) \in E$ is covered in the G_i by the k-th connected component, then define $\varphi(u)_i = \varphi(v)_i \coloneqq k$. Since each connected component of G_i is a biclique, there are no inconsistencies introduced in defining the mapping φ. All remaining undefined coordinates of the mapping φ are completed with unique entries.

Observe that the mapping φ has the following property. For any $u \in L$ and $v \in R$, we have $(u, v) \in E$ if and only if $\varphi(u)_0 \neq \varphi(v)_0$, and there exists $i \in \{1, \ldots, d\}$ such that $\varphi(u)_i = \varphi(v)_i$. Equivalently, by taking the negation, one concludes that $(u, v) \in L \times R \setminus E$ if and only if, for all $i \in \{0, 1, \ldots, d\}$, we have $\varphi(u)_i \neq \varphi(v)_i$. Therefore, the complement of the bipartite graph G is a subgraph of $K_2 \times K_{\mathbb{N}}^d$, if φ is injective. Note that a redundancy-free graph cannot have $\varphi(u) = \varphi(v)$, for distinct vertices u and v. Consequently, we have Proposition 1. The other direction of the proposition does not hold because the first coordinate of the mapping φ need not be constant restricted over the vertices in L or R. However, given φ one can prepend a coordinate that is 1 for the vertices in L and 2 for the vertices in R. Therefore, if $\overline{G}$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$, then G has a size-$(d+1)$ P_4-free cover.

For deriving the lower bound, consider $G = \mathsf{INEQ}_N$, i.e., $\overline{G} = \mathsf{EQ}_N$. Using the algebraic lower-bounding technique of [47, 4, 63], one concludes $d \geq \lceil \log_2 N \rceil$. Therefore, we have the following result.

▶ **Theorem 24.** *For any $N \in \mathbb{N}$, it holds that $\mathsf{P_4\text{-}fc}(\mathsf{INEQ}_N) \geq \lceil \log_2 N \rceil$.*

6.2 Bound on DISJ_n

We rely on a counting technique to obtain this lower bound. Intuitively, existing algebraic technique are useful to obtain logarithmic lower bounds. However, in this problem, we seek to prove a polynomial lower bound.

▶ **Theorem 25.** *For all $n \in \mathbb{N}$, the following bound holds.*

$$\mathsf{P_4\text{-}fp}(\mathsf{DISJ}_n) \geq \mathsf{P_4\text{-}fc}(\mathsf{DISJ}_n) \geq N^{\log_2 3 - 3/2} \approx N^{0.085}$$

The following lemma is the key for the proof of Theorem 25.

▶ **Lemma 26.** *Any P_4-free subgraph of DISJ_n has at most $N\sqrt{N}$ edges.*

To prove Lemma 26, we shall use the following claims (see full version for their proofs).

▷ Claim 27. Any biclique subgraph of DISJ_n has at most N edges.

▷ Claim 28. Let $\{(a_i, b_i)\}_{i \in \mathbb{N}}$ be a sequence of non-negative numbers. Then,

$$\sum_{i \in \mathbb{N}} a_i b_i \leq \sqrt{\left(\max_{i \in \mathbb{N}} a_i b_i\right)\left(\sum_{i \in \mathbb{N}} a_i\right)\left(\sum_{i \in \mathbb{N}} b_i\right)}.$$

Furthermore, equality holds if and only if (a) for all $i \in \mathbb{N}$, one has $a_i > 0$ iff $b_i > 0$. (b) all positive a_i are constant, and (c) all positive b_i are constant.

Proof of Lemma 26. Suppose G is a P_4-free subgraph of DISJ_n. Let $K_{a_1,b_1}, K_{a_2,b_2}, \ldots, K_{a_m,b_m}$ be the (biclique) connected components of G, where $a_i \in \mathbb{N}, b_i \in \mathbb{N}$ for every $1 \leq i \leq m$ and $m \in \mathbb{N}$. The total number of edges in G is $\sum_{i=1}^{m} a_i b_i$. We shall show that $\sum_{i=1}^{m} a_i \cdot b_i \leq N\sqrt{N}$. By Claim 27, it holds that $a_i \cdot b_i \leq N$ for every $1 \leq i \leq m$. Since all the left partite sets of $K_{a_1,b_1}, K_{a_2,b_2}, \ldots, K_{a_m,b_m}$ are disjoint, it holds that $\sum_{i=1}^{m} a_i \leq N$. Similarly, $\sum_{i=1}^{m} b_i \leq N$. Therefore, applying Claim 28, the following inequality holds.

$$\begin{aligned}\sum_{i=1}^{m} a_i b_i &\leq \sqrt{\left(\max_i a_i b_i\right)\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{i=1}^{m} b_i\right)} \\ &\leq \sqrt{N \cdot N \cdot N} = N^{3/2}\end{aligned}$$

Thus, any P_4-free subgraph of DISJ_n has at most $N^{3/2}$ edges. ◀

Now, we are ready to prove Theorem 25.

Proof of Theorem 25. First, observe that there are 3^n edges in DISJ_n. By Lemma 26, any P_4-free subgraph of DISJ_n has at most $N\sqrt{N}$ edges. Therefore, we have

$$\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_n\right) \geq \mathsf{P_4\text{-}fc}\left(\mathsf{DISJ}_n\right) \geq \frac{3^n}{N\sqrt{N}} = N^{\log_2 3 - 3/2} \approx N^{0.085}$$

as desired. ◀

6.3 Bounds on P_4-free Cover Number of INT_n

We shall prove the following lower bound on the P_4-free cover number of INT_n.

▶ **Theorem 29.** *For all $n \in \mathbb{N}$, the following bounds hold.*

$$n - \frac{1}{2}\left(\lg \pi + \lg\left(\frac{n+1}{2} + \frac{1}{4} + \frac{1}{64(n+1)}\right)\right) \leq \mathsf{P_4\text{-}fc}\left(\mathsf{INT}_n\right).$$

First, we state claims needed for the proof of Theorem 29 (see full version for their proofs).

▷ Claim 30. For every $n \in \mathbb{N}$, the following bound holds.

$$\lg\binom{n}{\lfloor n/2 \rfloor} \geq n - \frac{1}{2}\left(\lg \pi + \lg\left(\frac{n+1}{2} + \frac{1}{4} + \frac{1}{64(n+1)}\right)\right)$$

▷ Claim 31. Let G be a bipartite graph. Then, for every induced subgraph H of G, the following inequality holds.

$$\mathsf{P_4\text{-}fc}\left(H\right) \leq \mathsf{P_4\text{-}fc}\left(G\right)$$

Proof of Theorem 29. Consider the induced subgraph $G = (L', R', E')$ of INT_n, where $L' = \{S \subseteq [n] \colon |S| = \lfloor \frac{n}{2} \rfloor\}$, $R' = \{T \subseteq [n] \colon |T| = \lceil \frac{n}{2} \rceil\}$. Observe that each vertex $S \in L'$ is connected to every $T \in R'$ except when $T = [n] \setminus S$. Thus, graph G is the complement of a matching of size M, where $M = \binom{n}{\lfloor n/2 \rfloor}$. Using the algebraic lower-bounding technique of [47] and Proposition 1, one concludes that

$$\mathsf{P_4\text{-}fc}\left(G\right) \geq \lceil \lg M \rceil \geq n - \frac{1}{2}\left(\lg \pi + \lg\left(\frac{n+1}{2} + \frac{1}{4} + \frac{1}{64(n+1)}\right)\right),$$

where the last inequality follows from Claim 30. Finally, by Claim 31, $\mathsf{P_4\text{-}fc}\left(G\right) \leq \mathsf{P_4\text{-}fc}\left(\mathsf{INT}_n\right)$. Therefore, we have

$$n - \frac{1}{2}\left(\lg \pi + \lg\left(\frac{n+1}{2} + \frac{1}{4} + \frac{1}{64(n+1)}\right)\right) \leq \mathsf{P_4\text{-}fc}\left(\mathsf{INT}_n\right),$$

as desired. ◀

7 Relation to Graph Embedding

This section presents the connection between P_4-free partition/cover number and product/-Prague dimension.

7.1 P_4-free Cover Number

▷ Claim 32. If a bipartite graph $G = (L, R, E)$ has a size-d P_4-free covering, then the complement bipartite graph $\overline{G} = (L, R, L \times R \setminus E)$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$.

Proof. Let $G_1, \ldots, G_d$ be a size-d P_4-free cover of G. Define a vertex mapping $\varphi\colon L \cup R \to K_2 \times K_{\mathbb{N}}^d$ as follows. Let $\varphi(u)_i$ denote the i-th coordinate of the mapping $\varphi(u)$. Define $\varphi(u)_0 = 0$, for all $u \in L$, and $\varphi(v)_0 = 1$, for all $v \in R$. For $i \in \{1, \ldots, d\}$, define $\varphi(u)_i = \varphi(v)_i = k$, for every edge (u, v) in the k-th connected component of G_i. All remaining entries of φ are filled with unique values. One can verify that $(u, v) \in L \times R \setminus E$ if and only if $\varphi(u)$ and $\varphi(v)$ differ in every coordinate, that is, $\varphi(u)_i \neq \varphi(v)_i$ for every $i \in \{0, 1, \ldots, d\}$. Therefore, the complement bipartite graph $\overline{G}$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$.

We emphasize that the vertex mapping φ has the additional property that $\varphi(u)$ and $\varphi(v)$ have t identical coordinates if and only if the edge (u, v) is covered in t P_4-free graphs among $G_1, \ldots, G_d$. This property shall be useful in the proof of Claim 35. ◁

▷ Claim 33. If a loopless undirected graph $H = (L \cup R, E)$ is an induced subgraph of $K_{\mathbb{N}}^d$ and $E \subseteq L \times R$, then the bipartite graph $H' = (L, R, L \times R \setminus E)$ has a size-d P_4-free covering.

Proof. Suppose a loopless undirected graph $H = (L \cup R, E)$ is an induced subgraph of $K_{\mathbb{N}}^d$ and $E \subseteq L \times R$. Then, there exists a vertex mapping $\varphi\colon L \cup R \to \mathbb{N}^d$ such that $(u, v) \in E$ if and only if there exists $i \in \{1, 2, \ldots, d\}$ such that $\varphi(u)_i = \varphi(v)_i$. Define a new vertex mapping $\varphi^+\colon L \cup R \to \{1, 2\} \times \mathbb{N}^d$ as follows.

$$\varphi^+(u) = \begin{cases} (1, \varphi(u)), & \text{if } u \in L \\ (2, \varphi(u)), & \text{otherwise.} \end{cases}$$

For $i \in \{1, 2, \ldots, d\}$, define $G_i = (L, R, E_i)$ such that E_i is the set of all $u \in L$ and $v \in R$ such that $\varphi^+(u)_i = \varphi^+(v)_i$. Observe that the set of vertices $u \in L$ such that $\varphi^+(u)_i = k$ and the set of vertices $v \in R$ such that $\varphi^+(u)_i = k$ for some $k \in \mathbb{N}$ form a biclique, and each E_i is a disjoint union of bicliques. Furthermore, an edge $(u, v) \in E$ if and only if there exists an $i \in \{1, 2, \ldots, d\}$ such that $\varphi(u)_i = \varphi(v)_i$ which is equivalent to $\varphi^+(u)_i = \varphi^+(v)_i$. This implies that E_i cover the edge (u, v). Therefore, $E_1, E_2, \ldots, E_d$ is a P_4-free cover of H.

The $G_1, \ldots, G_d$ have the property that if an edges (u, v) is covered t times by these P_4-free graphs, then $\varphi^+(u)$ intersects $\varphi^+(v)$ in exactly t coordinates. This property of the vertex mapping shall be useful in the proof of Claim 36. ◁

The following result is a consequence of Claim 32 and Claim 33.

▶ **Corollary 34.** *Let $G = (L, R, E)$ be a bipartite graph and $H = (L \cup R, E)$ be a loopless undirected graph. Then, the following identity holds.*

$$\mathsf{pdim}(H) \in \{\mathsf{P_4\text{-}fc}(G), \mathsf{P_4\text{-}fc}(G) + 1\},$$

or, equivalently,

$$\mathsf{P_4\text{-}fc}(G) \in \{\mathsf{pdim}(H) - 1, \mathsf{pdim}(H)\}.$$

Note that the additive slack of 1 in Corollary 34 is necessary. Figure 7 gives an example.

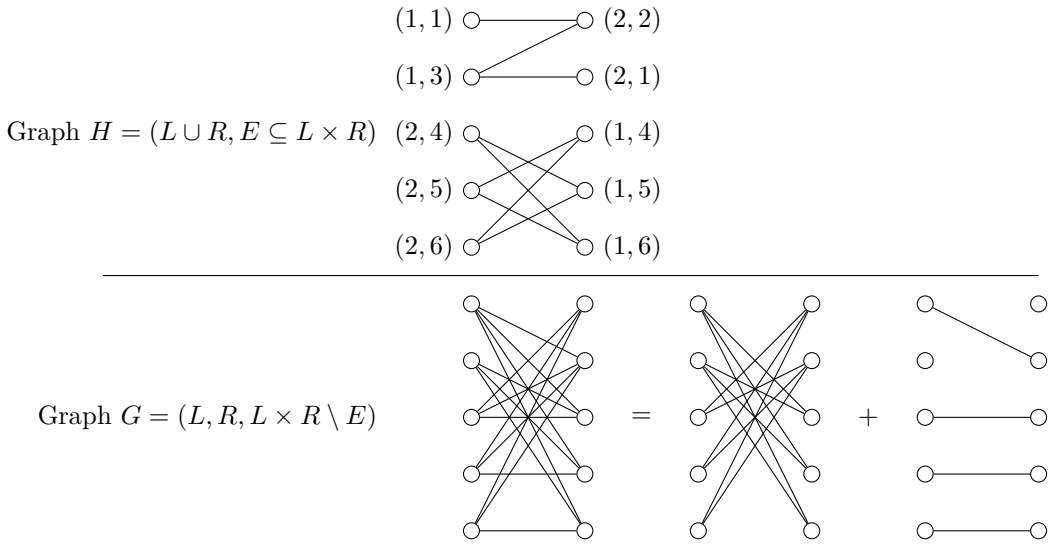


Figure 7 Example for the tightness of Corollary 34. Note that the loopless undirected graph $H = (L \cup R, E) = P_4 + C_6$, where $E \subseteq L \times R$, is an induced subgraph of $K_2 \times K_{\mathbb{N}}$. The (partition) vertex mapping of each vertex is explicitly mentioned next to it. However, the bipartite graph $G = (L, R, L \times R \setminus E)$ is not P_4-free and, hence, $\mathsf{P_4\text{-}fc}(G) \geq 2$; in fact, we have $\mathsf{P_4\text{-}fc}(G) = \mathsf{P_4\text{-}fp}(G) = 2$. The edges of G partition into $K_{2,3} + K_{3,2}$ and $4K_{1,1}$.

7.2 P_4-free Partition Number

Suppose a graph H is an induced subgraph of $K_{\mathbb{N}}^d$ via a vertex mapping $\varphi\colon V(H) \to \mathbb{N}^d$. The vertex mapping φ is a *partition* if the following conditions are satisfied.

1. If $(u, v) \in E(H)$, then $\varphi(u)_i \neq \varphi(v)_i$, for all $i \in \{1, 2, \ldots, d\}$.
2. If $(u, v) \notin E(H)$, then there exists a *unique* $i \in \{1, 2, \ldots, d\}$ such that $\varphi(u)_i = \varphi(v)_i$.

We emphasize that in an unrestricted vertex mapping, instead of (2) above, we insist that there exists an $i \in \{1, 2, \ldots, d\}$ (not necessarily a *unique* i). Let $\mathsf{pdim}^*(H)$ represent the minimum $d \in \mathbb{N}$ such that H is an induced subgraph of $K_{\mathbb{N}}^d$ via a partition vertex mapping.

▷ Claim 35. If a bipartite graph $G = (L, R, E)$ has a size-d P_4-free partitioning, then the complement bipartite graph $\overline{G} = (L, R, L \times R \setminus E)$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$ via a partition vertex mapping.

▷ Claim 36. If a loopless undirected graph $H = (L \cup R, E)$ is an induced subgraph of $K_{\mathbb{N}}^d$ via a partition vertex mapping and $E \subseteq L \times R$, then the bipartite graph $H' = (L, R, L \times R \setminus E)$ has a size-d P_4-free partitioning.

The proofs of Claim 35 and Claim 36 are identical to the proofs of Claim 32 and Claim 33, respectively, utilizing the fact that the vertex mapping is a partition. As a consequence of Claim 35 and Claim 36, we have the following result.

▶ **Corollary 37.** *Let $G = (L, R, E)$ be a bipartite graph and $H = (L \cup R, E)$ be a loopless undirected graph. Then, the following identity holds.*

$$\mathsf{pdim}^*(H) \in \{\mathsf{P_4\text{-}fp}(G), \mathsf{P_4\text{-}fp}(G) + 1\},$$

or equivalently

$$\mathsf{P_4\text{-}fp}(G) \in \{\mathsf{pdim}^*(H) - 1, \mathsf{pdim}^*(H)\}.$$

References

1 Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography - I: secret sharing. *IEEE Trans. Inf. Theory*, 39(4):1121–1132, 1993. doi:10.1109/18.243431.

2 Rudolf Ahlswede and Imre Csiszár. Common randomness in information theory and cryptography - part II: CR capacity. *IEEE Trans. Inf. Theory*, 44(1):225–240, 1998. doi:10.1109/18.651026.

3 I. Algor and Noga Alon. The star arboricity of graphs. *Discret. Math.*, 75(1-3):11–22, 1989. doi:10.1016/0012-365X(89)90073-3.

4 Noga Alon. Covering graphs by the minimum number of equivalence relations. *Combinatorica*, 6(3):201–206, 1986.

5 Noga Alon and Ryan Alweiss. On the product dimension of clique factors. *European Journal of Combinatorics*, 86:103097, 2020.

6 Noga Alon, Lajos Rónyai, and Tibor Szabó. Norm-graphs: Variations and applications. *J. Comb. Theory, Ser. B*, 76(2):280–290, 1999. doi:10.1006/jctb.1999.1906.

7 James Aspnes, Richard Beigel, Merrick L. Furst, and Steven Rudich. The expressive power of voting polynomials. In *23rd Annual ACM Symposium on Theory of Computing*, pages 402–409, New Orleans, LA, USA, May 6–8 1991. ACM Press. doi:10.1145/103418.103461.

8 Amos Beimel, Yuval Ishai, Ranjit Kumaresan, and Eyal Kushilevitz. On the cryptographic complexity of the worst functions. In Yehuda Lindell, editor, *TCC 2014: 11th Theory of Cryptography Conference*, volume 8349 of *Lecture Notes in Computer Science*, pages 317–342, San Diego, CA, USA, February 24–26 2014. Springer, Heidelberg, Germany. doi:10.1007/978-3-642-54242-8_14.

9 Alexander R. Block, Hemanta K. Maji, and Hai H. Nguyen. Secure computation based on leaky correlations: High resilience setting. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology – CRYPTO 2017, Part II*, volume 10402 of *Lecture Notes in Computer Science*, pages 3–32, Santa Barbara, CA, USA, August 20–24 2017. Springer, Heidelberg, Germany. doi:10.1007/978-3-319-63715-0_1.

10 Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *IEEE Trans. Inf. Theory*, 57(10):6351–6355, 2011. doi:10.1109/TIT.2011.2134067.

11 Béla Bollobás. *Extremal graph theory*. Courier Corporation, 2004.

12 W. G. Brown. On graphs that do not contain a thomsen graph. *Canadian Mathematical Bulletin*, 9(3):281–285, 1966. doi:10.4153/CMB-1966-036-2.

13 Ignacio Cascudo, Ivan Damgård, Felipe Lacerda, and Samuel Ranellucci. Oblivious transfer from any non-trivial elastic noisy channel via secret key agreement. In Martin Hirt and Adam D. Smith, editors, *TCC 2016-B: 14th Theory of Cryptography Conference, Part I*, volume 9985 of *Lecture Notes in Computer Science*, pages 204–234, Beijing, China, October 31 – November 3 2016. Springer, Heidelberg, Germany. doi:10.1007/978-3-662-53641-4_9.

14 Siu On Chan, Elchanan Mossel, and Joe Neeman. On extracting common random bits from correlated sources on large alphabets. *IEEE Trans. Inf. Theory*, 60(3):1630–1637, 2014. doi:10.1109/TIT.2014.2301155.

15 G. Chen, S. Fujita, A. Gyarfas, J. Lehel, and A. Toth. Around a biclique cover conjecture, 2012. arXiv:1212.6861.

16 Claude Crépeau and Joe Kilian. Achieving oblivious transfer using weakened security assumptions (extended abstract). In *29th Annual Symposium on Foundations of Computer Science*, pages 42–52, White Plains, NY, USA, October 24–26 1988. IEEE Computer Society Press. doi:10.1109/SFCS.1988.21920.

17 Claude Crépeau and Joe Kilian. Weakening security assumptions and oblivious transfer (abstract). In Shafi Goldwasser, editor, *Advances in Cryptology – CRYPTO'88*, volume 403 of *Lecture Notes in Computer Science*, pages 2–7, Santa Barbara, CA, USA, August 21–25 1990. Springer, Heidelberg, Germany. doi:10.1007/0-387-34799-2_1.

18 Claude Crépeau, Kirill Morozov, and Stefan Wolf. Efficient unconditional oblivious transfer from almost any noisy channel. In Carlo Blundo and Stelvio Cimato, editors, *SCN 04: 4th International Conference on Security in Communication Networks*, volume 3352 of *Lecture Notes in Computer Science*, pages 47–59, Amalfi, Italy, September 8–10 2005. Springer, Heidelberg, Germany. doi:10.1007/978-3-540-30598-9_4.

19 Ivan Damgård, Joe Kilian, and Louis Salvail. On the (im)possibility of basing oblivious transfer and bit commitment on weakened security assumptions. In Jacques Stern, editor, *Advances in Cryptology – EUROCRYPT'99*, volume 1592 of *Lecture Notes in Computer Science*, pages 56–73, Prague, Czech Republic, May 2–6 1999. Springer, Heidelberg, Germany. doi:10.1007/3-540-48910-X_5.

20 Paul Erdös and Joel Spencer. *Probabilistic methods in combinatorics*, volume 17. Academic Press New York, 1974.

21 Herbert Fleischner, Egbert Mujuni, Daniël Paulusma, and Stefan Szeider. Covering graphs with few complete bipartite subgraphs. *Theoretical Computer Science*, 410(21):2045–2053, 2009. doi:10.1016/j.tcs.2008.12.059.

22 DJ Foulis. Empirical logic, xeroxed course notes. *University of Massachusetts, Amherst, Massachusetts (1969-1970)*, 1969.

23 Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

24 Zoltán Füredi and Miklós Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

25 Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973.

26 Mikael Goldmann. On the power of a threshold gate at the top. *Inf. Process. Lett.*, 63(6):287–293, 1997. doi:10.1016/S0020-0190(97)00141-5.

27 Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game or A completeness theorem for protocols with honest majority. In Alfred Aho, editor, *19th Annual ACM Symposium on Theory of Computing*, pages 218–229, New York City, NY, USA, May 25–27 1987. ACM Press. doi:10.1145/28395.28420.

28 Daniel Gonçalves and Pascal Ochem. On star and caterpillar arboricity. *Discret. Math.*, 309(11):3694–3702, 2009. doi:10.1016/j.disc.2008.01.041.

29 Parikshit Gopalan and Rocco A. Servedio. Learning and lower bounds for ac^0 with threshold gates. In Maria J. Serna, Ronen Shaltiel, Klaus Jansen, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, volume 6302 of *Lecture Notes in Computer Science*, pages 588–601. Springer, 2010. doi:10.1007/978-3-642-15369-3_44.

30 Richard Hammack, Wilfried Imrich, and Sandi Klavžar. *Handbook of product graphs*. CRC press, 2011.

31 Chinh T. Hoàng and Van Bang Le. P_4-Colorings and P_4-Bipartite Graphs. *Discrete Mathematics and Theoretical Computer Science*, 4(2):109–122, 2001. URL: https://hal.inria.fr/hal-00958951.

32 Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions (extended abstracts). In *21st Annual ACM Symposium on Theory of Computing*, pages 12–24, Seattle, WA, USA, May 15–17 1989. ACM Press. doi:10.1145/73007.73009.

33 Jeffrey C. Jackson, Adam Klivans, and Rocco A. Servedio. Learnability beyond AC0. In *34th Annual ACM Symposium on Theory of Computing*, pages 776–784, Montréal, Québec, Canada, May 19–21 2002. ACM Press. doi:10.1145/509907.510018.

34 Minghui Jiang. Trees, paths, stars, caterpillars and spiders. *Algorithmica*, 80(6):1964–1982, 2018.

35 Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*. Springer Publishing Company, Incorporated, 2012.

36 Heinz A Jung. On a class of posets and the corresponding comparability graphs. *Journal of Combinatorial Theory, Series B*, 24(2):125–133, 1978.

37 Sebastian Kaiser. *Biclustering: methods, software and application.* PhD thesis, lmu, 2011.

38 Richard M. Karp. Reducibility among combinatorial problems. In Raymond E. Miller and James W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York, USA*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972. doi:10.1007/978-1-4684-2001-2_9.

39 Dakshita Khurana, Hemanta K. Maji, and Amit Sahai. Secure computation from elastic noisy channels. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology – EUROCRYPT 2016, Part II*, volume 9666 of *Lecture Notes in Computer Science*, pages 184–212, Vienna, Austria, May 8–12 2016. Springer, Heidelberg, Germany. doi:10.1007/978-3-662-49896-5_7.

40 Joe Kilian. Founding cryptography on oblivious transfer. In *20th Annual ACM Symposium on Theory of Computing*, pages 20–31, Chicago, IL, USA, May 2–4 1988. ACM Press. doi:10.1145/62212.62215.

41 Joe Kilian. A general completeness theorem for two-party games. In *23rd Annual ACM Symposium on Theory of Computing*, pages 553–560, New Orleans, LA, USA, May 6–8 1991. ACM Press. doi:10.1145/103418.103475.

42 Joe Kilian. More general completeness theorems for secure two-party computation. In *32nd Annual ACM Symposium on Theory of Computing*, pages 316–324, Portland, OR, USA, May 21–23 2000. ACM Press. doi:10.1145/335305.335342.

43 János Kollár, Lajos Rónyai, and Tibor Szabó. Norm-graphs and bipartite turán numbers. *Combinatorica*, 16(3):399–406, 1996. doi:10.1007/bf01261323.

44 Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, 1997.

45 H Lerchs. On cliques and kernels. *Department of Computer Science, University of Toronto*, 1971.

46 H Lerchs. On the clique-kernel structure of graphs. *Dept. of Computer Science, University of Toronto*, 1972.

47 László Lovász, J Nešetšil, and Ales Pultr. On a product dimension of graphs. *Journal of Combinatorial Theory, Series B*, 29(1):47–67, 1980.

48 Sara C Madeira and Arlindo L Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM transactions on computational biology and bioinformatics*, 1(1):24–45, 2004.

49 Ueli M. Maurer. Perfect cryptographic security from partially independent channels. In *23rd Annual ACM Symposium on Theory of Computing*, pages 561–571, New Orleans, LA, USA, May 6–8 1991. ACM Press. doi:10.1145/103418.103476.

50 Ueli M. Maurer. A universal statistical test for random bit generators. *Journal of Cryptology*, 5(2):89–105, January 1992. doi:10.1007/BF00193563.

51 Ueli M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. Inf. Theory*, 39(3):733–742, 1993. doi:10.1109/18.256484.

52 Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Structures & Algorithms*, 26(4):418–436, 2005. doi:10.1002/rsa.20062.

53 Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006.

54 J Nešetřil and Ales Pultr. A dushnik-miller type dimension of graphs and its complexity. In *International Conference on Fundamentals of Computation Theory*, pages 482–493. Springer, 1977.

55 Jaroslav Nešetřil and Vojtěch Rödl. A simple proof of the galvin-ramsey property of the class of all finite graphs and a dimension of a graph. *Discrete Mathematics*, 23(1):49–55, 1978.

56 Noam Nisan and David Zuckerman. More deterministic simulation in logspace. In *25th Annual ACM Symposium on Theory of Computing*, pages 235–244, San Diego, CA, USA, May 16–18 1993. ACM Press. doi:10.1145/167088.167162.

57 James Orlin. Contentment in graph theory: Covering graphs with cliques. *Indagationes Mathematicae (Proceedings)*, 80(5):406–424, 1977. doi:10.1016/1385-7258(77)90055-5.

58 Trevor Pinto. Biclique covers and partitions. *arXiv preprint arXiv:1307.6363*, 2013.

59 Svatopluk Poljak, D Rödl, and Ales Pultr. On a product dimension of bipartite graphs. *Journal of graph theory*, 7(4):475–486, 1983.

60 Dieter Seinsche. On a property of the class of n-colorable graphs. *Journal of Combinatorial Theory, Series B*, 16(2):191–193, 1974.

61 Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

62 David P Sumner. Dacey graphs. *Journal of the Australian Mathematical Society*, 18(4):492–502, 1974.

63 Douglas Brent West et al. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River, NJ, 1996.

64 Jürg Wullschleger. Oblivious-transfer amplification. In Moni Naor, editor, *Advances in Cryptology – EUROCRYPT 2007*, volume 4515 of *Lecture Notes in Computer Science*, pages 555–572, Barcelona, Spain, May 20–24 2007. Springer, Heidelberg, Germany. doi:10.1007/978-3-540-72540-4_32.

65 Jürg Wullschleger. Oblivious transfer from weak noisy channels. In Omer Reingold, editor, *TCC 2009: 6th Theory of Cryptography Conference*, volume 5444 of *Lecture Notes in Computer Science*, pages 332–349. Springer, Heidelberg, Germany, March 15–17 2009. doi:10.1007/978-3-642-00457-5_20.

66 Aaron Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, 1975. doi:10.1109/TIT.1975.1055346.

67 Ke Yang. On the (im)possibility of non-interactive correlation distillation. In Martin Farach-Colton, editor, *LATIN 2004: Theoretical Informatics, 6th Latin American Symposium*, volume 2976 of *Lecture Notes in Computer Science*, pages 222–231, Buenos Aires, Argentina, April 5–8 2004. Springer, Heidelberg, Germany.

68 Andrew Chi-Chih Yao. Protocols for secure computations (extended abstract). In *23rd Annual Symposium on Foundations of Computer Science*, pages 160–164, Chicago, Illinois, November 3–5 1982. IEEE Computer Society Press. doi:10.1109/SFCS.1982.38.