P_4-free Partition and Cover Numbers

Alexander R. Block Simina Brânzei Hemanta K. Maji Himanshi Mehta
Tamalika Mukherjee Hai H. Nguyen

Abstract

P_4-free graphs– also known as cographs, complement-reducible graphs, or hereditary Dacey graphs–have been well studied in graph theory. We introduce the graph properties of partitioning and covering the edges of a graph with the minimum number of P_4-free graphs, namely, the P_4-free partition and P_4-free cover numbers. We prove that computing these numbers is NP-complete, even for bipartite graphs. We present bipartite graph constructions where these numbers are at least $\epsilon N^{1-2\epsilon}$, for $\epsilon \in \{1/3, 1/4, 1/5, \dots\}$, where N is the number of vertices in each partite set. Finally, we upper bound these numbers for bipartite graphs encoding well-studied Boolean functions from circuit complexity, such as set intersection, set disjointness, and inequality.

Our work encodes joint probability distributions and Boolean functions as equivalent bipartite graphs and studies 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.

For applications in information theory, we consider a system where a setup samples from a joint distribution and gives the participants, Alice and Bob, their portion from this joint sample. The objective of Alice and Bob 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 an appropriate assistance to help Alice and Bob with their objective. Lower bounds to the minimum size of the genie's assistance translates 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 number correspond to joint distributions requiring more assistance from the genie.

As a representative application in communication complexity, we study communication complexity of non-deterministic 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 non-deterministic 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 number have high communication complexity. Furthermore, there are functions with communication complexity close to the naïve protocol where the non-deterministic input reveals the input of a party. Finally, the access to the equality oracle reduces the communication complexity of computing set intersection and disjointness by a constant factor in contrast to the model where parties do not have access to the equality oracle. In the case of computing the inequality function, we show an exponential reduction in the communication complexity.

This page is intentionally left blank.

1 Introduction

A bipartite graph is P_4*-free* if none of its 2×2 subgraphs induce a path of length three.[1] Since the 1970s, P_4-free graphs—also known as cographs, complement-reducible graphs, or hereditary Dacey graphs from empirical logic [Fou69]—have been widely studied in graph theory [Ler71, Ler72, Jun78, Sei74, Sum74]. Our work studies the graph properties of *partitioning* and *covering* the edges of a given bipartite graph using the minimum number of P_4-free bipartite graphs. [2]

The P_4-free partition number of a bipartite graph G is the minimum number of P_4-free bipartite graphs needed to partition the edges of G, 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 bipartite graphs needed to cover the edges of G, denoted by $\mathsf{P_4\text{-}fc}\,(G)$. The definition extends to general graphs, however, our study focuses on bipartite graphs. That is, we are given a bipartite graph as input and the goal is to partition or cover its edges using bipartite graphs. P_4-free partition and cover numbers are a natural generalization of fundamental graph properties, such as biclique partition and cover number, arboricity, and star-arboricity, which, in turn, have applications to theoretical computer science, information theory, and combinatorial optimization; for a discussion of these connections, Appendix E. In addition to being motivated by intellectual curiosity, the P_4-free partition and cover numbers (and their close variants) appear in diverse computer science and information theory problems as well, which is unsurprising given the well-established central role that graph properties such as biclique partition and cover numbers, arboricity, and star-arboricity enjoy.

Our work proves the following results (refer to Section 2 for formal statements).

1. It is NP-complete to determine the P_4-free partition and cover numbers of a general graph, even bipartite graphs.

2. We construct 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, we show that Erdős-Rényi graphs (with constant parameter) have P_4-free partition and cover numbers $\geqslant N/\log N$ asymptotically almost surely.

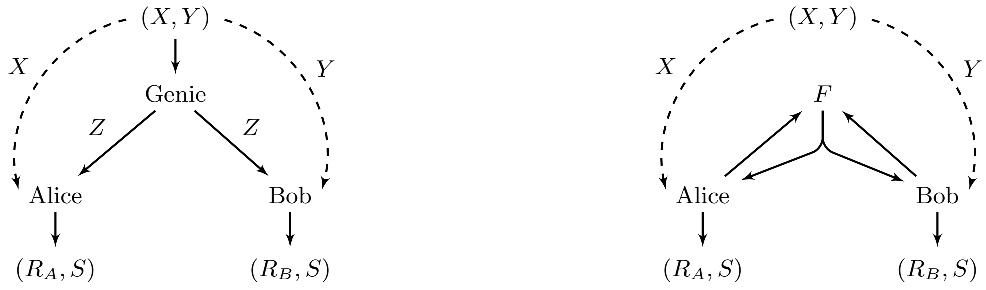
3. Finally, we encode the Boolean set intersection and disjointness functions, and the inequality function as bipartite graphs. We upper bound the P_4-free partition and cover numbers of these graphs. Furthermore, we prove that our estimate for the inequality function is tight using algebraic techniques.

Roadmap of the paper. Next we first discuss two motivating problems (called problems A and B) for the graph properties introduced. Section 1.1 presents the equivalence between P_4-free partition number and Problem A. Section 1.2 demonstrates the equivalence of Problem B and the P_4-free cover number. Section 2 states our contributions. Section 3 provides a technical overview of our proof techniques. Appendix contains all the formal definitions and the omitted proofs.

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 and cryptographic complexity motivating this study.

[1] Any 2×2 subgraph of a P_4-free bipartite graph either has 4 edges, or has at most 2 edges.

[2] [HL01] introduced the vertex partitioning a graph into different color-classes so that the subgraph induced by the vertices of any color-class is P_4-free.



(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.

Problem A. Assistance for Correlation Distillation. 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 with non-zero entropy 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. Intuitively, conditioned on the genie's suggestion Z, Alice-Bob samples' joint distribution decompose into shared randomness and local independent randomness.

What is the *minimum* length k of the genie's suggestion 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 variants of common information cannot accurately measure this information-theoretic measure; thus, motivating our study. This problem is equivalent to computing the P_4-free partition number of a bipartite graph encoding the joint probability distribution. In particular, lower bounds to k shall translate into lower bounds on (interactive) communication and cryptographic complexity (see part (b) of Figure 1).

Proposition 1. *If p_{XY} needs $k \geqslant k^*$ 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.*

This result holds because the genie can simulate the entire output of the functionality F with access to (x, y). We emphasize that F may be a communication protocol, or (multiple) calls to the

NAND-gate functionality, which suffices for general two-party secure computation.

Refer to Section 1.1 for the details.

Problem B. Non-deterministic Communication Complexity relative to Equality Oracle. Suppose Alice has input $x \in X$, Bob has input $y \in Y$, and the players are interested in computing the Boolean function $f\colon X \times Y \to \{0,1\}$ of their private inputs. Both parties 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 non-deterministic suggestion *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 *non-deterministic protocol* successfully computing f relative to the equality oracle without additional communication.

1. For every input-pair $(x,y) \in X \times Y$ such that the output $f(x,y) = 1$, there exists a non-deterministic suggestion $z \in \{0,1\}^k$ 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 non-deterministic suggestions $z \in \{0,1\}^k$, the functions satisfy $\mathsf{EQ}(\ A(x,z)\ ,\ B(y,z)\ ) = 0$.

The *communication complexity* of this protocol is k, i.e., the length of the non-deterministic suggestion. What is the *minimum* communication complexity k for which such a protocol exists?

Intuitively, we are augmenting the non-deterministic 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 intersection and disjointness by a constant factor when compared to the model where parties do not have access to the equality oracle. In the case of computing the inequality function, perhaps surprisingly, we show an *exponential* reduction in the communication complexity. Section 1.2 provides the details.

Concluding remarks (other applications and extensions). Edge-Partitioning of graphs using the minimum number of P_4-free graphs has found applications in cryptography [BMN17], where a malicious party performs leakage on the setup 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* [MO04, Kai11] is identical to a P_4-free bipartite graph, indicating applications in analyzing biological data. In Appendix I, we also present a representative scheduling problem that naturally reduces to P_4-free partition/cover numbers. This example highlights the innate ability of P_4-free graphs to encode scheduling problems that are amenable to *parallelization.*

Related graph properties: Equivalence cover number and Product dimension. The following discussion is particular 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 needed to cover the edges of G [NP77, NR78]. Note that the P_4-free cover number, which is introduced by our work, is an extension of this concept to bipartite graphs. Furthermore, the equivalence cover number of G is identical to the *product dimension* of the complement of the graph G, i.e., the minimum number $d \in \mathbb{N}$ 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 and the product dimension of a graph are NP-complete [NP77].

However, the equivalence cover number behaves significantly differently from P_4-free cover number. For example, an N-vertex star has equivalence partition number $(N-1)$, i.e., the equivalence cover number of graphs can be very high. 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, which is at most (roughly) $N/2$ [AA89]. The construction of bipartite graphs with high P_4-free cover and partition numbers turns out to be non-trivial, and our work relies on probabilistic techniques to demonstrate their existence.

1.1 P_4-free Partition Number

We reduce motivating problem A to computing the P_4-free partition number in Appendix B. In this section, we establish connections to well-studied problems in information theory and computer science.

1.1.1 Discussion of Problem A

We begin by expanding on how lower-bounding the information-theoretic measure in problem A translates into communication and cryptographic lower bounds (as in [BIKK14]). Suppose, in our model, one proves that $k \geqslant k^*$ bits of the genie's assistance is necessary. 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 interactive protocol or multiple calls to arbitrary interactive functionalities of choice 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 [Yao82, GMW87, Kil00]. The lower bound in our model implies that the *total communication received* by Alice and Bob from the functionality F must be $\geqslant k^*$ to agree on the shared key and extract independent private randomness.

In information-theory, Gray-Wyner systems/networks are well-studied [Wyn75]. However, existing measures like mutual information and various notions of common information are inadequate to accurately capture the information-theoretic property mentioned above. For example, there are two joint distributions with identical (Shannon's) mutual information [Sha48]; however, one needs no assistance while the other needs one-bit assistance.[3] 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 $x = (x_1, x_2, \dots, x_n)$, and Bob either (with probability half) gets $y = x$ or $y = (\overline{x_1}, \dots, \overline{x_n})$, i.e., every bit of x is flipped. In the second distribution (the *noisy typewriter distribution*), Alice gets a uniformly random sample $x \in \{0, 1, \dots, 2^n - 1\}$, and Bob either gets $y = x$ or $y = (x+1) \mod 2^n$ with probability half. Both distributions have $(n-1)$ bits of mutual information; however, the first distribution needs no assistance, but the second distribution needs one-bit assistance[4] to agree on a secret key.

Wyner's common information [Wyn75] 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 $z = (x_1, \dots, x_{n-1})$), 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 [GK73] estimates the length of the secret key that Alice and Bob can generate without any assistance from the genie,

[3] By tensorizing the distributions, one can increase the necessary assistance arbitrarily.

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

$$\begin{array}{c|cccc} & 00 & 11 & 01 & 10 \\ \hline 00 & 1 & 1 & 0 & 0 \\ 11 & 1 & 1 & 0 & 0 \\ 01 & 0 & 0 & 1 & 1 \\ 10 & 0 & 0 & 1 & 1 \end{array}$$

(a) Forward or flip.

$$\begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 00 & 1 & 1 & 0 & 0 \\ 01 & 0 & 1 & 1 & 0 \\ 10 & 0 & 0 & 1 & 1 \\ 11 & 1 & 0 & 0 & 1 \end{array}$$

(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).

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 [MOR$^+$06, MO05] enables parties to agree on a secret non-interactively *without any assistance*. Even without the necessity to generate independent local randomness, strong hardness of computation results are known [MOR$^+$06, MO05, Yan04, BM11, CMN14].

1.1.2 Our results for Problem A

Observe that the naïve assistance that reveals one party's sample to the other party suffices; however, the minimum assistance may be exponentially smaller. Our work relies on suitably encoding joint distributions as bipartite graphs. Appendix B argues that the set of all P_4-free graphs *precisely* encodes the joint distributions that, without *any assistance*, enables the parties to agree on a secret key s and extract the left-over entropy of their respective samples as private independent randomness. Next, the motivating problem above corresponds to partitioning the bipartite graph that encodes the joint probability distribution p_{XY} using a minimum number of P_4-free graphs (refer to Appendix A for formal definitions). Looking back, observe that 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 has P_4-free partition number 2 (i.e., one-bit assistance is necessary and sufficient).

We prove in Theorem 1 that ascertaining the minimum assistance is, in general, difficult. Furthermore, there are joint distributions where the minimum assistance 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 2.

Corollary 1. *Let $\Omega_X = \Omega_Y = \{0,1\}^n$. Fix $t \in \mathbb{N}$. There exist 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[5] or the DISJ_n[6] joint distribution then $\lceil n/2 \rceil$-bit assistance suffices (we explicitly provide the suggestion that the genie provides and it is efficient to compute). Our bounds are exponentially smaller than those entailed by existing techniques (see Theorem 3).

Refer to Appendix D for additional discussion on various forms of common information.

1.2 P_4-free Cover Number

We reduce motivating Problem B to the P_4-free cover number in Appendix C. In this section, we discuss our results with respect to Problem B.

1.2.1 Discussion on Problem B

The equality function in the *plain* non-deterministic communication complexity model (where parties do not have access to the EQ oracle) has high communication complexity. Determining the minimum non-deterministic input is equivalent to covering the input-pairs where the output is 1 using a minimum number of *combinatorial rectangles*, a.k.a., *biclique cover number* [Juk12]. 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 non-deterministic 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 [Kar72], 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 [ABFR91, Gol97, JKS02, GS10], enabling a controlled exploration of the gap between the power of AC^0 and TC^0 circuits.

1.2.2 Summary of our results for Problem B

Appendix C demonstrates that the set of all P_4-free graphs corresponds to the set of functions f that need no non-deterministic input to compute it using one call to the EQ oracle. Analogous to the biclique cover number, the P_4-free cover number of the bipartite graph encoding a function determines the size of the non-deterministic input needed to compute it in our model. Similar to the result for P_4-free partition number, we prove that computing the P_4-free cover number is difficult (see Theorem 1), and there are functions that need non-deterministic input (roughly) the size of the parties' inputs, in other words we obtain the following as a corollary to Theorem 2.

Corollary 2. *Fix $t \in \mathbb{N}$. There exist Boolean functions $f\colon \{1,2,\ldots,N\} \times \{1,2,\ldots,N\} \to \{0,1\}$ that require at least $(1-\frac{2}{t+2})\log_2 N$ bits of non-deterministic 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. Consider the plain model for non-deterministic communication complexity model, that is, one where parties do not have access to the EQ oracle. In this model the EQ function is hard-to-compute and needs n-bits of non-deterministic input. The "fooling set" lower-bounding technique draws inspiration from this result. For a general f it demonstrates pairs of inputs-sets of Alice and Bob where only the diagonal elements are 1 and rest are 0. That is, the function f has an embedded EQ function. The

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

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

$$\begin{matrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{matrix} = \begin{matrix}0&1&1&1\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{matrix} + \begin{matrix}0&0&0&0\\1&0&1&1\\0&0&0&0\\0&0&0&0\end{matrix} + \begin{matrix}0&0&0&0\\0&0&0&0\\1&1&0&1\\0&0&0&0\end{matrix} + \begin{matrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\1&1&1&0\end{matrix}$$

Figure 3: Figure illustrating $\mathsf{bp}(\mathsf{INEQ}_N) = N$, for $N = 4$.

$$\begin{matrix}0&1&1&1\\1&0&1&1\\1&1&0&1\\1&1&1&0\end{matrix} = \begin{matrix}0&0&1&1\\0&0&1&1\\1&1&0&0\\1&1&0&0\end{matrix} + \begin{matrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{matrix}$$

Figure 4: Figure illustrating that $\mathsf{P_4\text{-}fp}(\mathsf{INEQ}_4) = \lceil \log_2 N \rceil$, for $N = 4$.

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

Next, we provide estimates for some well-known functions in communication complexity (see Theorem 3). We prove that the P_4-free cover number of DISJ_n is (roughly) $\leqslant \sqrt{N}$. That is, only $n/2$ bits of non-deterministic input suffices to compute this function. Recall that, in the plain model, the function DISJ_n requires n-bit non-deterministic 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.

Finally, perhaps surprisingly, we show that INEQ_N needs only $\log_2 \log_2 N$-bit non-deterministic input using the EQ oracle (see Figure 4). Recall that in the plain model (without access to the EQ oracle), INEQ_N requires $\log_2 N$-bit non-deterministic input, which is exponentially higher (see Figure 3). Furthermore, using the algebraic technique of [LNP80], we prove a matching lower bound as well.

2 Our Contribution

We prove that it is NP-complete to determine the P_4-free partition and cover number of a bipartite graph.

Theorem 1 (Hardness of P_4-free Partition and Cover)**.** *The following languages are* NP*-complete:*

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

Similar problems, for example, calculating the biclique partition number/cover [Orl77] and star arboricity [Jia18] (even for bipartite graphs) are also known to be NP-complete.

Next, we prove there exist graphs G with large P_4-free partition and cover number. Note that for a bipartite graph $G = (L, R, E)$, we have $\mathsf{P_4\text{-}fc}(G) \leqslant \mathsf{P_4\text{-}fp}(G) \leqslant \min\{|L|, |R|\}$ 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 exists an N-vertex bipartite graph with $\mathsf{P_4\text{-}fp}(G) \geqslant \mathsf{P_4\text{-}fc}(G) \geqslant \Omega(\epsilon \cdot N^{1-2\epsilon})$ (roughly).

Theorem 2 (High P_4- Free Partition and Cover Numbers)**.** *Let C be a 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 $N \geqslant N_0$, there is a graph $G_{N,t} = (L, R, E)$ such that*

1. $|L| = |R| = N$, *and*

2. $\mathsf{P}_4\text{-}\mathsf{fp}\left(G_{N,t}\right) \geqslant \mathsf{P}_4\text{-}\mathsf{fc}\left(G_{N,t}\right) \geqslant C \cdot \frac{N^{1-\frac{2}{t+2}}}{t}$,

Our constructions rely on extremal bipartite graphs that avoid $K_{t+1,t+1}$-subgraphs, for which probabilistic and explicit constructions are known (refer to the discussion in Section 3.2).

In motivating 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 non-deterministic input and access to the EQ oracle. If these functions are embedded in another function, then that function must have high complexity as well.

As a corollary (of the construction technique of the previous result), we prove the following result for dense bipartite graphs drawn from the Erdős-Rényi distribution with (constant) parameter $p \in (0, 1)$.

Corollary 3 (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. This distribution includes every edge into the graph independently with probability p. Then,*

$$\Pr\left[\mathsf{P}_4\text{-}\mathsf{fp}\left(G\right) \geqslant \mathsf{P}_4\text{-}\mathsf{fc}\left(G\right) \geqslant \frac{pN}{4\log_a N} \cdot (1 - o(1)) \colon G \xleftarrow{\$} \mathsf{ER}(N, N, p)\right] \geqslant 1 - o(1),$$

where $a = 1/p$.

Again, we rely on the fact that graphs drawn from $\mathsf{ER}(N, N, p)$ avoid bicliques with size-$(2\log_a N)$ partite sets.

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

1. $\mathsf{P}_4\text{-}\mathsf{fc}\left(\mathsf{INT}_n\right) \leqslant \mathsf{P}_4\text{-}\mathsf{fp}\left(\mathsf{INT}_n\right) \leqslant \begin{cases} 2 \cdot 2^{n/2} - 2, & \textit{even } n, \textit{ and} \\ 3 \cdot 2^{(n-1)/2} - 2, & \textit{odd } n. \end{cases}$

2. $\mathsf{P}_4\text{-}\mathsf{fc}\left(\mathsf{DISJ}_n\right) \leqslant \mathsf{P}_4\text{-}\mathsf{fp}\left(\mathsf{DISJ}_n\right) \leqslant 2^{\lceil n/2 \rceil}$*. In particular,* $\mathsf{P}_4\text{-}\mathsf{fc}\left(\mathsf{DISJ}_1\right) = \mathsf{P}_4\text{-}\mathsf{fp}\left(\mathsf{DISJ}_1\right) = 2$.

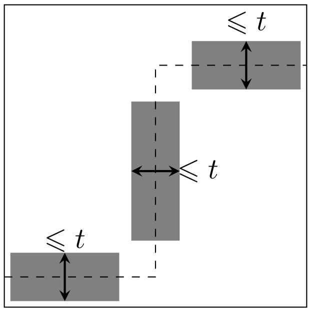


Figure 5: Let $t \in \mathbb{N}$ be a parameter. Proof intuition underlying the fact that a $K_{t+1,t+1}$-free bipartite graph cannot have a dense P_4-free subgraph.

3. $\mathsf{P_4\text{-}fc}\left(\mathsf{INEQ}_N\right) = \mathsf{P_4\text{-}fp}\left(\mathsf{INEQ}_N\right) = \lceil \log_2 N \rceil$.

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

Note that DISJ_n has 3^n edges. Using the arboricity bound of [AA89] and an averaging argument, one can show that the star arboricity of graph G is $\mathcal{O}\left(\sqrt{e(G)}\right)$, where $e(G)$ is the total number of edges in G. For DISJ_n, consequently, we have that the star arboricity is bounded by $\left\lfloor 3^{n/2} \right\rfloor$. Our upper bound on DISJ_n is an exponential improvement from the upper bound implied by bound on start arboricity. Furthermore, any star forest of DISJ_n can have atmost $2^n - 1$ edges. This implies that the star arboricity of DISJ_n must be greater than $\lceil 2.25^{n/2} \rceil$, which introduces a large gap compared to $\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_n\right)$.

3 Technical Overview

This section provides a high-level overview of the proof strategies for our results.

3.1 Proof of Theorem 1

The crucial idea underlying the hardness of computation result for computing the P_4-free partition and cover numbers is the following. If a bipartite graph is $K_{2,2}$-free then any P_4-free subgraph of this graph is a star forest. Furthermore, the minimum number of star forests needed to cover and partition a graph are equal. Consequently, computing the P_4-free partition or cover number of any $K_{2,2}$-free graph is equivalent to computing the star arboricity of that graph. Appendix F presents the full proof.

3.2 Proof of Theorem 2 and Corollary 3

Our objective is to consider bipartite graphs G that do not have P_4-free subgraphs with a large number of edges. Intuitively, the adjacency matrix M of the bipartite graph G should avoid "large combinatorial rectangles" of all 1s. So, consider a bipartite graph $G = (L, R, E)$ that is $K_{t+1,t+1}$-avoiding. For any combinatorial rectangle that is a subgraph of G, define its *width* to be the smaller of its two dimensions. Note that the width of any combinatorial rectangle that is a subgraph of G has to be $\leqslant t$; otherwise, a $K_{t+1,t+1}$-subgraph of G shall exist.

Let H be a P_4-free subgraph of G. It is instructive to refer to Figure 5. The width of the combinatorial rectangle corresponding to any of its connected components is $\leqslant t$. The sum of the lengths (the longer dimension of a combinatorial rectangle) of the combinatorial rectangles corresponding to each connected component is $\leqslant |L| + |R|$. Because, the length can either belong to the left partite set or to the right partite set. So, the total number of edges in H is $\leqslant t\,(|L| + |R|)$. Consequently, any partition or cover of G requires at least $\frac{|E(G)|}{t(|L|+|R|)}$ P_4-free subgraphs.

So, an appropriate choice for G is a $K_{t+1,t+1}$-avoiding graph with as many edges as possible. These extremal properties are well-studied [FS13]. The best general lower bound obtained by the probabilistic method [ES74] yields

$$|E(G)| \geqslant C' N^{2-\frac{2}{t+2}}$$

where C' is a positive absolute constant.

An explicit construction for $K_{t+1,t+1}$-avoiding graphs for $t = 2$ is known [Bro66], which has $\frac{1}{2}N^{\frac{5}{3}} + o(N^{\frac{5}{3}})$ edges. Using *norm graphs*, constructions of $K_{t,s}$-avoiding graphs for fixed $t \geqslant 2$ and $s > (t-1)!$ are known as well [KRS96, ARS99]. Note that the latter set of constructions do not apply to our setting for $t > 3$.

Similarly, to prove that $\mathsf{ER}(N, N, p)$ have high P_4-free partition and cover numbers (that is Corollary 3, we rely on the following two observations.

1. The number of edges in a bipartite graph $G \overset{\$}{\leftarrow} \mathsf{ER}(N, N, p)$ is at least $pN \cdot (1 - o(1))$, with probability $1 - o(1)$.

2. Furthermore, $G \overset{\$}{\leftarrow} \mathsf{ER}(N, N, p)$ is $K_{t+1,t+1}$-avoiding, where $t+1 = \lceil 2\log_a N \rceil$. For completeness, we follow the exposition of [FK16] to prove this result in Appendix G using the first moment technique.

Appendix G presents the complete proofs.

3.3 Estimates for $\mathsf{INT}_n, \mathsf{DISJ}_n$, and INEQ_N

Bound for INT_n. First, we prove the following general result.

Claim 4 (Submultiplicity of P_4-free partition number)**.** *Let G and G' are two bipartite graphs. Then, the following bound holds.*

$$\mathsf{P_4\text{-}fp}\left(G \times G'\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right) \cdot \mathsf{P_4\text{-}fp}\left(G'\right).$$

Using this claim, we shall inductively upper-bound $\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right)$. Our base cases are $\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_1\right) = 1$ and $\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_2\right) = 2$.

Recall that INT_n indicates the intersection between a subset $X \subseteq \{1, 2, \ldots, n\}$ and $Y \subseteq \{1, 2, \ldots, n\}$. Consider the edges in INT_n where the witness of the intersection is 1, or 2. Let A be the subgraph of INT_n formed by these edges. We argue that $\mathsf{P_4\text{-}fp}\left(A\right) \leqslant 2$.

On the remainder of the edges we recurse. The remainder of the edges form a graph B that is $\mathsf{INT}_{n-2} \times H$, where H is a graph satisfying $\mathsf{P_4\text{-}fp}\left(H\right) = 2$.

So, we get the recursion

$$\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) \leqslant \mathsf{P_4\text{-}fp}\left(B\right) + 2 \leqslant \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) \cdot \mathsf{P_4\text{-}fp}\left(H\right) + 2 = 2 \cdot \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) + 2.$$

Consequently, we get that

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

Bound for DISJ_n. It is well-known that DISJ_n is the tensor product $\mathsf{DISJ}_1^{\times n}$. We prove that $\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_2\right) = 2$. Consequently, we get that $\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_n\right) \leqslant 2^{\lceil n/2 \rceil}$, by the submultiplicity of P_4-free partition number.

Bound for INEQ_N. In fact, we prove a more general result.

Let G be the complement of a P_4-free bipartite graph H. We prove the following result.

Claim 5 (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 connected components. Let G be the complement of H. Then, the following bound holds.*

$$\mathsf{P_4\text{-}fc}\left(G\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right) \leqslant \begin{cases} \lceil \log_2 c \rceil\,, & \textit{if } H \textit{ has no isolated vertex,} \\ \lceil \log_2 c \rceil + 1, & \textit{otherwise.} \end{cases}$$

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 [FMPS09]. 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 $L = \{1, \dots, c\}$ and $R = \{1, \dots, c\}$, and the edges in H are (i, i), for $1 \leqslant i \leqslant c$.

Define $L_0 \coloneqq \{1, \dots, \lfloor c/2 \rfloor\}$ and $L_1 \coloneqq L \setminus L_0$. Similarly, define $R_0 \coloneqq \{1, \dots, \lfloor c/2 \rfloor\}$ and $R_1 \coloneqq R \setminus R_0$. 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, we get our result. Appendix H presents the full proofs of all the three results.

For a tight lower bound on $\mathsf{P_4\text{-}fc}\left(\mathsf{INEQ}_N\right)$ consider the following reduction for a general bipartite graph. Given a size-d P_4-free cover $\{G_1, \dots, G_d\}$ 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 $i \in \{0, 1, \dots, d\}$, $\varphi(u)_i$ 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

1. $\varphi(u)_0 \neq \varphi(v)_0$, and
2. There exists $i \in \{1, \dots, d\}$ such that $\varphi(u)_i = \varphi(v)_i$.

Equivalently, one concludes that $(u, v) \in L \times R \setminus E$ if and only if, for all $i \in \{0, 1, \dots, d\}$, 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.

A graph G is redundancy-free if no two vertices have an identical neighborhood. Note that a redundancy-free graph cannot have $\varphi(u) = \varphi(v)$, for distinct vertices u and v. Consequently, we have the following proposition.

Proposition 2. *If a redundancy-free bipartite graph $G = (L, R, E)$ has a size-d P_4-free edge-covering, then the bipartite graph $\overline{G} \coloneqq (L, R, L \times R \setminus E)$ is an induced subgraph of $K_2 \times K_{\mathbb{N}}^d$.*

This result helps apply the algebraic techniques to lower bound the $\mathsf{P_4\text{-}fc}\,(G)$.

For example, consider $G = \mathsf{INEQ}_N$. Then, we have $\overline{G} = \mathsf{EQ}_N$. Using the algebraic lower-bounding technique of [LNP80], one concludes that $d \geqslant \lceil \log_2 N \rceil$. Therefore, we have $\mathsf{P_4\text{-}fc}\,(\mathsf{INEQ}_N) \geqslant \lceil \log_2 N \rceil$.

References

[AA89] 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, 9, 27

[ABFR91] 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. 6

[AG76] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability*, pages 925–939, 1976. 25

[AGKN13] Venkat Anantharam, Amin Gohari, Sudeep Kamath, and Chandra Nair. On maximal correlation, hypercontractivity, and the data processing inequality studied by erkip and cover. *arXiv preprint arXiv:1304.6133*, 2013. 25

[ARS99] 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. 10

[BF88] László Babai and Péter Frankl. *Linear algebra methods in combinatorics.* University of Chicago, 1988. 27

[BG15] Salman Beigi and Amin Gohari. On the duality of additivity and tensorization. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2381–2385. IEEE, 2015. doi:10.1109/ISIT.2015.7282882. 25

[BIKK14] 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. 4

[BM11] 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. 5, 26

[BMN17] 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. 3, 26

[Bor82] Christer Borell. Positivity improving operators and hypercontractivity. *Mathematische Zeitschrift*, 180(3):225–234, 1982. doi:10.1007/BF01318906. 25

[Bro66] 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. 10

[CHHK14] Parinya Chalermsook, Sandy Heydrich, Eugenia Holm, and Andreas Karrenbauer. Nearly tight approximability results for minimum biclique cover and partition. In

Andreas S. Schulz and Dorothea Wagner, editors, *Algorithms - ESA 2014 - 22th Annual European Symposium, Wroclaw, Poland, September 8-10, 2014. Proceedings*, volume 8737 of *Lecture Notes in Computer Science*, pages 235–246. Springer, 2014. doi:10.1007/978-3-662-44777-2_20. 27, 28, 35

[CMN14] 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. 5, 26

[DMN18] Anindya De, Elchanan Mossel, and Joe Neeman. Non interactive simulation of correlated distributions is decidable. In Artur Czumaj, editor, *29th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2728–2746, New Orleans, LA, USA, January 7–10, 2018. ACM-SIAM. doi:10.1137/1.9781611975031.174. 25

[EHM+08] Alina Ene, William G. Horne, Nikola Milosavljevic, Prasad Rao, Robert Schreiber, and Robert Endre Tarjan. Fast exact and heuristic methods for role minimization problems. In Indrakshi Ray and Ninghui Li, editors, *13th ACM Symposium on Access Control Models and Technologies, SACMAT 2008, Estes Park, CO, USA, June 11-13, 2008, Proceedings*, pages 1–10. ACM, 2008. doi:10.1145/1377836.1377838. 35

[ES74] Paul Erdös and Joel Spencer. *Probabilistic methods in combinatorics*, volume 17. Academic Press New York, 1974. 10, 29

[EU18] Alessandro Epasto and Eli Upfal. Efficient approximation for restricted biclique cover problems. *Algorithms*, 11(6):84, 2018. doi:10.3390/a11060084. 28, 35

[FK16] Alan Frieze and Michał Karoński. *Introduction to random graphs.* Cambridge University Press, 2016. 10, 30

[FMPS07] Herbert Fleischner, Egbert Mujuni, Daniël Paulusma, and Stefan Szeider. Covering graphs with few complete bipartite subgraphs. In Vikraman Arvind and Sanjiva Prasad, editors, *FSTTCS 2007: Foundations of Software Technology and Theoretical Computer Science, 27th International Conference, New Delhi, India, December 12-14, 2007, Proceedings*, volume 4855 of *Lecture Notes in Computer Science*, pages 340–351. Springer, 2007. doi:10.1007/978-3-540-77050-3_28. 28, 35

[FMPS09] Herbert Fleischner, Egbert Mujuni, Daniël Paulusma, and Stefan Szeider. Covering graphs with few complete bipartite subgraphs. *Theoretical Computer Science*, 410(21):2045 – 2053, 2009. URL: http://www.sciencedirect.com/science/article/pii/S0304397508009407, doi:https://doi.org/10.1016/j.tcs.2008.12.059. 11, 33

[Fou69] DJ Foulis. Empirical logic, xeroxed course notes. *University of Massachusetts, Amherst, Massachusetts (1969-1970)*, 1969. 1

[FS13] Zoltán Füredi and Miklós Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013. 10

[GH07] Hermann Gruber and Markus Holzer. Inapproximability of nondeterministic state and transition complexity assuming p=!np. In Tero Harju, Juhani Karhumäki, and Arto Lepistö, editors, *Developments in Language Theory, 11th International Conference, DLT 2007, Turku, Finland, July 3-6, 2007, Proceedings*, volume 4588 of *Lecture Notes in Computer Science*, pages 205–216. Springer, 2007. doi:10.1007/978-3-540-73208-2_21. 28, 35

[GK73] Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 4, 25, 26

[GKS16] Badih Ghazi, Pritish Kamath, and Madhu Sudan. Decidability of non-interactive simulation of joint distributions. In Irit Dinur, editor, *57th Annual Symposium on Foundations of Computer Science*, pages 545–554, New Brunswick, NJ, USA, October 9–11, 2016. IEEE Computer Society Press. doi:10.1109/FOCS.2016.65. 25

[GMW87] 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. 4

[GO09] 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. 28

[Gol97] 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. 6

[GP71] Ronald L Graham and Henry O Pollak. On the addressing problem for loop switching. *Bell System Technical Journal*, 50(8):2495–2519, 1971. 27

[GP72] Ronald L Graham and Henry O Pollak. On embedding graphs in squashed cubes. In *Graph theory and applications*, pages 99–110. Springer, 1972. 27, 28

[GS10] 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. 6

[GW88] Harold N. Gabow and Herbert H. Westermann. Forests, frames and games: Algorithms for matroid sums and applications. In *20th Annual ACM Symposium on Theory of Computing*, pages 407–421, Chicago, IL, USA, May 2–4, 1988. ACM Press. doi:10.1145/62212.62252. 26

[GW92] Harold N. Gabow and Herbert H. Westermann. Forests, frames, and games: Algorithms for matroid sums and applications. *Algorithmica*, 7(5&6):465–497, 1992. doi:10.1007/BF01758774. 26

[Hir35] Hermann O Hirschfeld. A connection between correlation and contingency. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31, pages 520–524. Cambridge University Press, 1935. doi:10.1017/S0305004100013517. 25

[HL01] 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. 1

[HMR06] Michael Hirsch, Henk Meijer, and David Rappaport. Biclique edge cover graphs and confluent drawings. In Michael Kaufmann and Dorothea Wagner, editors, *Graph Drawing, 14th International Symposium, GD 2006, Karlsruhe, Germany, September 18-20, 2006. Revised Papers*, volume 4372 of *Lecture Notes in Computer Science*, pages 405–416. Springer, 2006. doi:10.1007/978-3-540-70904-6_39. 28, 35

[Hof72] AJ Hoffman. Eigenvalues and partitionings of the edges of a graph. *Linear Algebra and Its Applications*, 5(2):137–146, 1972. 27

[Jia18] Minghui Jiang. Trees, paths, stars, caterpillars and spiders. *Algorithmica*, 80(6):1964–1982, 2018. 7

[JK09] Stasys Jukna and Alexander S Kulikov. On covering graphs by complete bipartite subgraphs. *Discrete Mathematics*, 309(10):3399–3403, 2009. 28, 35

[JKS02] 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. 6

[Juk12] Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*. Springer Publishing Company, Incorporated, 2012. 6, 23

[Jun78] Heinz A Jung. On a class of posets and the corresponding comparability graphs. *Journal of Combinatorial Theory, Series B*, 24(2):125–133, 1978. 1

[KA16] Sudeep Kamath and Venkat Anantharam. On non-interactive simulation of joint distributions. *IEEE Transactions on Information Theory*, 62(6):3419–3435, 2016. 25

[Kai11] Sebastian Kaiser. *Biclustering: methods, software and application*. PhD thesis, lmu, 2011. 3

[Kar72] 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. 6

[Kil00] 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. 4, 26

[KMN20] Hamidreza Amini Khorasgani, Hemanta K. Maji, and Hai H. Nguyen. Secure non-interactive simulation: Hardness & feasibility. *IACR Cryptol. ePrint Arch.*, 2020:252, 2020. URL: https://eprint.iacr.org/2020/252. 26

[KPRW19] Ravi Kumar, Rina Panigrahy, Ali Rahimi, and David P. Woodruff. Faster algorithms for binary matrix factorization. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3551–3559. PMLR, 2019. URL: http://proceedings.mlr.press/v97/kumar19a.html. 27

[KRS96] János Kollár, Lajos Rónyai, and Tibor Szabó. Norm-graphs and bipartite turán numbers. *Combinatorica*, 16(3):399–406, 1996. doi:10.1007/bf01261323. 10

[KRW88] Thomas Kratzke, Bruce Reznick, and Douglas West. Eigensharp graphs: Decomposition into complete bipartite subgraphs. *Transactions of the American Mathematical Society*, 308(2):637–653, 1988. 27

[Ler71] H Lerchs. On cliques and kernels. *Department of Computer Science, University of Toronto*, 1971. 1

[Ler72] H Lerchs. On the clique-kernel structure of graphs. *Dept. of Computer Science, University of Toronto*, 1972. 1

[LNP80] László Lovász, J Nešetšil, and Ales Pultr. On a product dimension of graphs. *Journal of Combinatorial Theory, Series B*, 29(1):47–67, 1980. 7, 12

[MMG+08] Pauli Miettinen, Taneli Mielikäinen, Aristides Gionis, Gautam Das, and Heikki Mannila. The discrete basis problem. *IEEE transactions on knowledge and data engineering*, 20(10):1348–1362, 2008. 28, 35

[MO04] 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. 3

[MO05] 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. 5, 25, 26

[MOR+06] 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. 5, 25, 26

[MOS13] Elchanan Mossel, Krzysztof Oleszkiewicz, and Arnab Sen. On reverse hypercontractivity. *Geometric and Functional Analysis*, 23(3):1062–1097, 2013. 25

[Mül96] Haiko Müller. On edge perfectness and classes of bipartite graphs. *Discrete Mathematics*, 149(1-3):159–187, 1996. 28

[NED+13] Igor Nor, Jan Engelstädter, Olivier Duron, Max Reuter, Marie-France Sagot, and Sylvain Charlat. On the genetic architecture of cytoplasmic incompatibility: inference from phenotypic data. *The American Naturalist*, 182(1):E15–E24, 2013. 28, 35

[NHC+12] Igor Nor, Danny Hermelin, Sylvain Charlat, Jan Engelstadter, Max Reuter, Olivier Duron, and Marie-France Sagot. Mod/resc parsimony inference: Theory and application. *Information and Computation*, 213:23 – 32, 2012. Special Issue: Combinatorial Pattern Matching (CPM 2010). URL: http://www.sciencedirect.com/science/article/pii/S0890540112000247, doi:https://doi.org/10.1016/j.ic.2011.03.008. 28, 35

[NMWA78] Dana S Nau, George Markowsky, Max A Woodbury, and D Bernard Amos. A mathematical analysis of human leukocyte antigen serology. *Mathematical Biosciences*, 40(3-4):243–270, 1978. 28, 35

[NP77] 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. 3, 4

[NR78] Jaroslav Nešetřil and Vojtěch Rōdl. 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. 3

[Orl77] James Orlin. Contentment in graph theory: Covering graphs with cliques. *Indagationes Mathematicae (Proceedings)*, 80(5):406 – 424, 1977. URL: http://www.sciencedirect.com/science/article/pii/1385725877900555, doi:https://doi.org/10.1016/1385-7258(77)90055-5. 7, 27, 28, 35

[Pec84] GW Peck. A new proof of a theorem of graham and pollak. *Discrete mathematics*, 49(3):327–328, 1984. 27, 28

[Pin13] Trevor Pinto. Biclique covers and partitions. *arXiv preprint arXiv:1307.6363*, 2013. 28

[PP10] Vinod M. Prabhakaran and Manoj Prabhakaran. Assisted common information. In *IEEE International Symposium on Information Theory, ISIT 2010, June 13-18, 2010, Austin, Texas, USA, Proceedings*, pages 2602–2606. IEEE, 2010. doi:10.1109/ISIT.2010.5513743. 26

[PP11] Vinod M. Prabhakaran and Manoj Prabhakaran. Assisted common information: Further results. In Alexander Kuleshov, Vladimir M. Blinovsky, and Anthony Ephremides, editors, *2011 IEEE International Symposium on Information Theory Proceedings, ISIT 2011, St. Petersburg, Russia, July 31 - August 5, 2011*, pages 2861–2865. IEEE, 2011. doi:10.1109/ISIT.2011.6034098. 26

[PP14] Vinod M. Prabhakaran and Manoj Prabhakaran. Assisted common information with an application to secure two-party sampling. *IEEE Trans. Inf. Theory*, 60(6):3413–3434, 2014. doi:10.1109/TIT.2014.2316011. 26

[Rén59] Alfréd Rényi. On measures of dependence. *Acta mathematica hungarica*, 10(3-4):441–451, 1959. doi:10.1007/BF02024507. 25

[RP14] K. Sankeerth Rao and Vinod M. Prabhakaran. A new upperbound for the oblivious transfer capacity of discrete memoryless channels. In *2014 IEEE Information Theory Workshop, ITW 2014, Hobart, Tasmania, Australia, November 2-5, 2014*, pages 35–39. IEEE, 2014. doi:10.1109/ITW.2014.6970787. 26

[Sei74] Dieter Seinsche. On a property of the class of n-colorable graphs. *Journal of Combinatorial Theory, Series B*, 16(2):191–193, 1974. 1

[Sha48] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948. 4

[Sim90] Hans Ulrich Simon. On approximate solutions for combinatorial optimization problems. *SIAM J. Discret. Math.*, 3(2):294–310, 1990. doi:10.1137/0403025. 28, 35

[SLY06] Guoqiang Shu, David Lee, and Mihalis Yannakakis. A note on broadcast encryption key management with applications to large scale emergency alert systems. In *20th International Parallel and Distributed Processing Symposium (IPDPS 2006), Proceedings, 25-29 April 2006, Rhodes Island, Greece*. IEEE, 2006. doi:10.1109/IPDPS.2006.1639680. 35

[STW20] Madhu Sudan, Himanshu Tyagi, and Shun Watanabe. Communication for generating correlation: A unifying survey. *IEEE Trans. Inf. Theory*, 66(1):5–37, 2020. doi:10.1109/TIT.2019.2946364. 25

[Sum74] David P Sumner. Dacey graphs. *Journal of the Australian Mathematical Society*, 18(4):492–502, 1974. 1

[SW03] B Schölkopf and MK Warmuth. Learning theory and kernel machines: 16th annual conference on learning theory and 7th kernel workshop (colt/kernel 2003). In *16th Annual Conference on Learning Theory and 7th Kernel Workshop (COLT/Kernel 2003)*. Springer, 2003. 28, 35

[Tve82] Helge Tverberg. On the decomposition of kn into complete bipartite graphs. *Journal of Graph Theory*, 6(4):493–494, 1982. 27, 28

[Vis08] Sundar Vishwanathan. A polynomial space proof of the graham–pollak theorem. *Journal of Combinatorial Theory, Series A*, 115(4):674–676, 2008. 27, 28

[Vis13] Sundar Vishwanathan. A counting proof of the graham–pollak theorem. *Discrete Mathematics*, 313(6):765–766, 2013. 27, 28

[VL85] JH Van Lint. {0, 1,*} distance problems in combinatorics. 1985. 27

[vLW01] Jacobus Hendricus van Lint and Richard Michael Wilson. *A course in combinatorics*. Cambridge university press, 2001. 27

[Wit75] Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. doi:doi.org/10.1137/0128010. 25

[WW05] Stefan Wolf and Jürg Wullschleger. New monotones and lower bounds in unconditional two-party computation. In Victor Shoup, editor, *Advances in Cryptology – CRYPTO 2005*, volume 3621 of *Lecture Notes in Computer Science*, pages 467–477, Santa Barbara, CA, USA, August 14–18, 2005. Springer, Heidelberg, Germany. doi:10.1007/11535218_28. 26

[Wyn75] 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. 4, 25, 26, 27

[Yan91] Mihalis Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. Syst. Sci.*, 43(3):441–466, 1991. doi:10.1016/0022-0000(91)90024-Y. 28

[Yan04] 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. 5, 26

[Yao82] 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`. 4

[YY06] Weigen Yan and Yeong-Nan Yeh. A simple proof of graham and pollak's theorem. *Journal of Combinatorial Theory, Series A*, 113(5):892–893, 2006. 27

A Relevant Background and Terminology

In this section, we formally introduce relevant concepts from graph theory, information theory, and communication complexity.

Introductory Graph-theory. Let $G = (L, R, E)$ represent an undirected bipartite graph with partite sets L and R, and edge set $E \subseteq L \times R$. The *complement* of G is the bipartite graph $G^c := (L, R, (L \times R) \setminus E)$. A *biclique* is a bipartite graph $G = (L, R, E)$ such that there exist subsets $L' \subseteq L$, $R' \subseteq R$, and $E = L' \times R'$, that is, all vertices in L' are connected to all vertices in R'.

Proposition 3. *A P_4-free bipartite graph G is a graph where each of its connected components is a biclique.*

That is, there exists $c \in \mathbb{N}$ (the number of components of the bipartite graph), disjoint subsets $L_1, L_2, \ldots, L_c \subseteq L$, and disjoint subsets $R_1, R_2, \ldots, R_c \subseteq R$, such that the edge-set satisfies $E = \bigcup_{i=1}^{c} L_i \times R_i$. Alternatively, P_4-free bipartite graphs are a (disjoint) *union of bicliques*. Figure 6 provides a pictorial representation of bicliques and P_4-free graphs.

Remark 1. *Cluster graphs are a similar notion in graph theory. However, they are* not *bipartite and are a union of* cliques*; (non-bipartite) graphs where every vertex connects to every other vertex. In contrast, P_4-free bipartite graphs are a union of bicliques, a.k.a.,* biclusters.

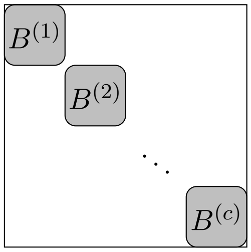


Figure 6: Pictorial representation of a P_4-free bipartite graph with c connected components after rearranging the rows and columns appropriately. Each block $B^{(i)}$ represents a connected component in the graph, which is a biclique. It is possible that the graph has isolated vertices.

The *P_4-free partition number* of a bipartite graph $G = (L, R, E)$, represented by $\mathsf{P_4\text{-}fp}(G)$, is the minimum number m such that there exist P_4-free graphs $G_i = (L, R, E_i)$, for $1 \leqslant i \leqslant m$, and the edge sets $E_1, E_2, \ldots, E_m$ partition E, the edge set of G. Similarly, the *P_4-free cover number* of a bipartite graph $G = (L, R, E)$, represented by $\mathsf{P_4\text{-}fc}(G)$, is the minimum number m such that there exist P_4-free graphs $G_i = (L, R, E_i)$, for $1 \leqslant i \leqslant m$, and the edge set E is the union of $E_1, E_2, \ldots, E_m$. Note that $\mathsf{P_4\text{-}fc}(G) \leqslant \mathsf{P_4\text{-}fp}(G)$, because every partition is also a cover.

Figure 4 illustrates the P_4-free partition of the graph corresponding to the function INEQ_N, where $N = 4$.

Random variables, entropy, and mutual information. A random variable X on sample space $\mathcal{X}$ is a real-valued function on $X : \mathcal{X} \to \mathbb{R}$. A discrete random variable is a random variable that takes only a finite or countably infinite number of values.

Let X be a discrete random variable on a sample space $\mathcal{X}$ and probability mass function $p(x) = \Pr[X = x]$, for all $x \in \mathcal{X}$. The entropy is a measure of uncertainty of a random variable and is defined formally below.

Definition 1 (Entropy)**.** *The entropy $H(X)$ of a discrete random variable X is defined by*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \ .$$

The relative entropy or Kullback-Leibler distance is a measure of the distance between two distributions.

Definition 2 (Kullback-Leibler distance)**.** *For probability mass functions $p(x)$ and $q(x)$, the relative entropy is*

$$D(p \| q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \ .$$

We can now define mutual information as the relative entropy of two random variables between their joint distribution and their product distribution.

Definition 3 (Mutual Information)**.** *For two random variables X and Y with joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$, the mutual information is defined as*

$$I(X; Y) = D(p(x, y) \| p(x)p(y)) \ .$$

Information-theoretic Measures as Graph Properties. Let p_{XY} define a joint distribution (X, Y) over a sample space $\Omega_X \times \Omega_Y$. A distribution is *flat* if the probability of sampling any element in the sample space is either zero or an appropriate positive constant. In the sequel, we consider only flat probability distributions.

Observe that flat probability distributions over the sample space $\Omega_X \times \Omega_Y$ are equivalent to bipartite graphs over partite sets Ω_X and Ω_Y (with non-empty edge-set). For example, any bipartite graph $G(\Omega_X, \Omega_Y, E)$ (uniquely) corresponds to the joint distribution p_{XY} that samples a uniformly random element from the set E. Consequently, given a flat distribution p_{XY}, one defines the unique bipartite graph corresponding to it $G(p_{XY})$, and, vice-versa.

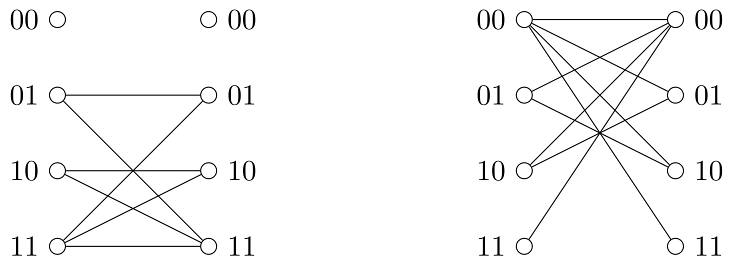
Let $I(X; Y)$ represent the *mutual information* of the random variables X and Y. Note that $I(X; Y) = 0$ if and only if X and Y are independent of each other. Interestingly, one can characterize the independence of random variables as an equivalent graph property.

Proposition 4. *A flat distribution p_{XY} satisfying $I(X; Y) = 0$ implies that the bipartite graph $G(p_{XY})$ is a biclique.*

Suppose G has $c \in \mathbb{N}$ connected components, and, w.l.o.g., assume that the components are named $\{1, 2, \ldots, c\}$. Let C be the function $E \to \{1, 2, \ldots, c\}$ that outputs the component's name containing an edge. One can equivalently interpret C as a random variable over the sample space $\{1, 2, \ldots, c\}$ such that $C = k$ with probability e_k/e, where e_k is the number of edges in the k-th component of G, and $e = |E|$. The Markov chain $X \leftrightarrow C \leftrightarrow Y$, an essential concept in information theory, communication complexity, and cryptography, has an equivalent characterization in graph properties.

Proposition 5. *For a flat p_{XY}, the Markov chain $X \leftrightarrow C \leftrightarrow Y$ is equivalent to the graph $G(p_{XY})$ being P_4-free.*

Suppose Alice gets x and Bob gets y sampled from a P_4-free flat p_{XY}, Section 1.1 argues that they always agree on their shared key s, if and only if the secret key is a function of $C(x, y)$. Furthermore, the fact that Alice and Bob's samples are independent of each other conditioned on the secret key s, implies that $X \leftrightarrow S \leftrightarrow Y$, and the shared key is identical to C.



(a) Graph of INT_n, for $n = 2$. (b) Graph of DISJ_n, for $n = 2$.

Figure 7: Bipartite graphs corresponding to the distributions INT_n and DISJ_n, for $n = 2$.

Communication complexity as Graph properties. Let $f\colon X \times Y \to \{0,1\}$ be a Boolean function. The bipartite graph $G(f) \coloneqq (X, Y, E)$, where E is the set of all input-pairs (x, y) satisfying $f(x, y) = 1$, is a unique encoding of the function f. Observe that the complement of the graph $G(f)$, represented by $G(f)^c$, is identical to $G(1-f)$, where $1-f$ is the complement of the function f. Figure 7 presents the graph corresponding to the functions INT_n and DISJ_n, for $n = 2$, which are defined below.

In deterministic communication complexity, the set of input-pairs of the parties consistent with a particular transcript is a *combinatorial rectangle*. That is, there exist $X' \subseteq X$ and $Y' \subseteq Y$ such that any $x \in X'$ and $y \in Y'$ results in that particular transcript. Suppose the output of the function corresponding to this transcript is 1. Then, one concludes that X' and Y' induce a biclique in the bipartite graph $G(f)$. Otherwise, if the output of the function corresponding to the transcript is 0, the vertex sets X' and Y' induce a biclique in $G(1-f)$.

We shall study the following graphs encoding well-studied functions from communication theory.

1. 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. 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. For $N \in \mathbb{N}$, let $\mathsf{EQ}_N = (\{1, 2, \ldots, N\}, \{1, 2, \ldots, N\}, E)$ 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 $u = v$.

4. 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 $u \neq v$.

Note that INT_n and DISJ_n are complements of each other, and EQ_N and INEQ_N are complements of each other. Figure 7 illustrates the graph corresponding to INT_n and DISJ_n for $n = 2$.

Remark 2. *It is well-known that the non-deterministic communication complexity of computing a function f is equivalent to the problem of* covering *the bipartite graph $G(f)$ using the minimum number of bicliques (a.k.a., the* biclique cover number *of G) [Juk12].*

B Modeling the Motivating Problem A as P_4-free Partition Number

Let Ω_X and Ω_Y be the sample space of Alice and Bob's samples, respectively. Let p_{XY} be a uniform distribution over an arbitrary subset of $\Omega_X \times \Omega_Y$. For this probability distribution, consider a bipartite graph G with partite set Ω_X and Ω_Y. The edge set of G contains (x, y) such that the probability of sampling (x, y) according to p_{XY} is positive.

Recall that the genie observes the sample (x, y) (that is, an edge in the graph G) and computes the assistance $z \in \{0, 1\}^k$. Conditioned on the assistance z that the genie provides, let $G_z := (\Omega_X, \Omega_Y, E_z)$, where $E_z \subseteq E$ is the set of all samples (x, y) where the (deterministic) genie provides z as assistance. Observe that the edge-sets in $\{E_z\}_{z\in\{0,1\}^k}$ partition the edge-set E.

Fix z, the assistance that the genie provides. Conditioned on z, the samples (x, y) of Alice and Bob are distributed according to the flat joint distribution $p_{XY|Z=z}$, where Z denotes the random variable for genie's assistance. This distribution is identical to the distribution corresponding to the graph G_z. Henceforth, the flat distribution $p_{XY|Z=z}$ is equivalent to the graph G_z.

Next, consider Alice receiving her sample x and Bob receiving his sample y from the joint distribution $p_{XY|Z=z} \equiv G_z$. Alice partitions her sample space Ω_X (the partition possibly depends on z) to obtain the secret key s_A. Similarly, Bob partitions his sample space Ω_Y to determine the secret key s_B. The fact that Alice and Bob *always* agree on the key $s = s_A = s_B$, implies that both Alice's and Bob's partitions of Ω_X and Ω_Y respect the connected components of G_z.[7] Consequently, the secret key s is a function of $C(x, y)$, (the identifier of) the connected component where their respective samples belong.

Now, the parties use the left-over entropy in their respective samples (after agreeing on the shared key s). That is, Alice and Bob, respectively, use the conditional distributions $(X|Z = z, S = s)$ and $(Y|Z = z, S = s)$ as their left-over sources of independent randomness. However, unless $X \leftrightarrow (S, Z) \leftrightarrow Y$, their randomness is not independent; that is, the shared secret key and genie's suggestion annihilate the correlation between Alice and Bob's samples. This constraint implies that $I(X; Y|S = s, Z = z) = 0$, which is equivalent to the graph corresponding to the flat $(X, Y|S = s, Z = z)$ being a biclique. That is, the random variables S and C are identical, and the graph G_z is P_4-free (and contains at least two connected components so that s has non-trivial entropy).

Consequently, our objective is to find the minimum $k \in \mathbb{N}$ such that there exists a partition of G into $\{G_z\}_{z\in\{0,1\}^k}$, where each G_z is P_4-free; that is, determine $k = \log_2 (\mathsf{P_4\text{-}fp}\,(G))$.

C Modeling the Motivating Problem B as P_4-free Cover Number

Given a Boolean function $f : X \times Y \to \{0, 1\}$, we encode it as the bipartite graph $G(f)$ with partite sets X and Y such that the edge set of $G(f)$ contains all $(x, y) \in X \times Y$ such that $f(x, y) = 1$.

First, let us begin by observing that a function f that may be evaluated by making one call to the EQ oracle without any non-deterministic input z if and only if the bipartite graph $G(f)$ is P_4-free.[8]

[7] Suppose not. Then, assume that Alice outputs secret key s_A for some vertex $x \in \Omega_X$ in a connected component, and outputs a different secret key s'_A for some other vertex $x' \in \Omega_X$ in the same connected component. Consider a path in G_z connecting the vertices x and x'. There will exist x_1 on this path where Alice outputs secret key $x_{A,1}$, and x_2 ($\neq x_1$) on this path where Alice outputs secret key $s_{A,2}$ such that $s_{A,1} \neq s_{A,1}$ and the distance between x_1 and x_2 is two. Let $y \in \Omega_Y$ be a sample of Bob that is at distance one from both x_1 and x_2 in the graph G_z. Obviously, the secret key output by Bob for sample y disagrees with $s_{A,1}$ or $s_{A,2}$. A similar argument also holds for Bob.

[8] Note that if f is P_4-free then Alice computes the connected component C_x of the bipartite graph $G(f)$ her

Next, consider any f that has a non-deterministic communication protocol using EQ oracle once. For every (x, y) such that $f(x, y) = 1$, then the edge $(x, y) \in E(G(f))$ is covered in some graph G_z, where z is the non-deterministic input. Following the discussion above G_z is P_4-free. Furthermore, if (x, y) is such that $f(x, y) = 0$, then the edges $(x, y) \in E(G(f))$ is not covered in any edge-set of G_z. Consequently, the set of all possible G_z covers $G(f)$.

So, the motivating problem is to find a covering of $G(f)$ with the minimum number of P_4-free bipartite graphs, the P_4-free cover number.

D Common Information: Discussion

Wyner's Common Information. Wyner's common information [Wyn75] is one of the several measures (for example, Shannon's mutual information, Gács-Körner [GK73] common information being some other prominent notions) of information that is common to X and Y. Formally, the quantity is defined as follows.

$$J(X; Y) := \min_{Z\colon X \leftrightarrow Z \leftrightarrow Y} H(Z),$$

where $H(Z)$ is the entropy of the random variable Z. Intuitively, it is the smallest entropy random variable that annihilates the correlation between X and Y. As an approximation, one can consider Z with the smallest support and p_{XY} being a flat. In this case, it is easy to see that $p_{XY|Z=z}$ is a biclique. Consequently, this Wyner's common information, intuitively, corresponds to partitioning the graph $G(p_{XY})$ the smallest number of bicliques, a.k.a. biclique partition number. Observe that Wyner's common information specifically kills the possibility of establishing a shared secret key; consequently, it is an inappropriate measure for our motivating problem. Furthermore, Wyner's common information can be non-zero while our genie needs to provide no assistance (refer to forward or flip distribution in Figure 2). The length of our genie's assistance may be exponentially smaller than Wyner's common information as well (refer to the noisy typewriter distribution in Figure 2).

Non-interactive Joint Simulation of Distributions. Information theory studies the possibility of simulating a sample from a joint distribution (U, V) given multiples samples from the joint distribution (X, Y), namely, *non-interactive simulation of joint distributions*. This line of research starts with the seminal works of Gács and Körner [GK73], Witsenhausen [Wit75], and Wyner [Wyn75]. In this setting, $Z = \emptyset$ (that is, the genie does not provide any assistance). The objective of the parties is to generate samples u and v from their local views such that the joint distribution of the samples (u, v) emulates a fixed joint distribution (U, V). Note that in our problem statement, we distilling out the shared secret key and the independent randomness. This problem is more general and (U, V) can be any arbitrary distribution. Even the decision version of the problem where one has to determine whether samples from one joint distribution may be non-interactively simulated from the samples of another joint distribution, in its full generality, is a difficult problem [GKS16, DMN18]. Technically, reverse hypercontractivity [AG76, Bor82, MOR+06, MOS13, KA16, DMN18, BG15, MO05], and maximal correlation [Hir35, Wit75, AG76, Rén59, AGKN13] are few of the most prominent techniques employed to prove the impossibility of non-interactive simulations. We refer the interested reader to an exceptional survey by Sudan, Tyagi, and Watanabe [STW20] for a thorough introduction to this field.

input is. Similarly, Bob also computes the connected component C_y where his input y belongs. Finally, they output $\mathsf{EQ}(C_x, C_y)$.

For the other direction, suppose Alice feeds $A(x)$ to the EQ oracle, and Bob feeds $B(y)$ to the EQ oracle. Conditioned on $A(x) = B(y) = \lambda$, the set of all (x, y) forms a combinatorial rectangle R_λ. Note that R_λ are disjoint, because A and B are deterministic functions. So, $G(f)$ is a disjoint union of combinatorial rectangles, a.k.a., it is P_4-free.

Non-interactive Correlation Distillation. This problem is a special case of non-interactive joint simulation of distributions where the target samples of Alice and Bob are identical, that is, $U = V$. The end objective is to emulate a shared secret key that the parties agree on [MOR+06, MO05, Yan04, BM11, CMN14].

Secure Non-interactive Joint Simulation. The recent work [KMN20] initiates the study of secure non-interactive joint simulations with the stronger objective of being cryptographically secure. For example, a difference of setting from non-interactive joint simulation is that information cannot be erased. This study is motivated by defining the achievable rate of the efficiency for secure computation protocols, and characterizing the rate-achieving secure protocol constructions.

Assisted Common Information (and Variants). A sequence of works develops "monotone properties" for interactive protocols, which refine and generalize the notions of common information [Wyn75, GK73] discussed above. For example, [WW05] proposes *monotones* for cryptographic protocols. Recently, generalizations of common information were explored in [PP10, PP11, PP14, RP14]. These works, in general, study how well the dependence between a pair of random variables can be resolved by a piece of common information. These notions of dependence satisfy the invariant that an interactive protocol cannot reduce this quantity. Consequently, they find applications in proving rate lower bounds in interactive protocols.

Leakage attacks in Cryptography. The work of [BMN17] studied P_4-free partition number of some interesting graphs. They studied this property in the context of upper-bounding the leakage resilience of setups in the cryptographic setting. They considered using the joint samples from probability distributions p_{XY} to perform two-party general secure computation in the presence of leakage. That is, the adversarial party obtains the leakage $L(X, Y)$ in addition to its local sample, where $L(\cdot, \cdot)$ is an arbitrary leakage function. Despite this leakage, the objective of the parties is to perform general secure computation using an interactive protocol. They showed that $\lceil \log_2 (G(p_{XY})) \rceil$ bits of leakage suffices to make the setup entirely useless for secure computation. They also demonstrated that the bound obtained by this technique is significantly tighter than the bound Wyner's common information entails, which is relevant to ruling out shared key agreement only, a significantly simpler task than two-party general secure computation [Kil00].

E Relation to Other Graph Properties

In this section, we explore the connection of P_4-free partition and cover numbers to graph properties such as star arboricity, biclique partition, and biclique cover number.

Star Arboricity. A *tree* is a graph where any two vertices are connected by a unique path. A *forest* is a disjoint union of trees. The *arboricity* of a graph, represented by $\mathsf{a}(G)$, is the minimum number of forests into which its edges can be partitioned. Observe that if there exists a covering of a graph with m forests then there also exists a partitioning of that graph with (at most) m forests. Consequently, partitioning into and covering with the minimal number of forests are identical graph properties. One can efficiently compute the star arboricity of a graph using a greedy strategy because it is expressible as a matroid partitioning problem [GW88, GW92]. The arboricity of a graph measures how dense the graph is. A graph with many edges has high arboricity, and graphs with high arboricity contain a dense subgraph.

A *star* is a tree with one internal node, or, equivalently, is $K_{1,r}$ a biclique with where one vertex connects to r vertices in the other partite set. 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. Similar to the previous case, partitioning and covering a graph into the minimum number of star forests are equivalent. By separating the odd and the even level edges of a forest one can form two star forest partitioning its edges. Consequently, we have

$$\mathsf{a}\left(G\right) \leqslant \mathsf{sa}\left(G\right) \leqslant 2\mathsf{a}\left(G\right).$$

Note that a star forest is a P_4-free graph. Therefore, we conclude the following result.

Proposition 6. *For any bipartite graph G, the following bound holds.*

$$\mathsf{P_4\text{-}fc}\left(G\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right) \leqslant \mathsf{sa}\left(G\right).$$

However, this bound is poor for dense graphs, for example, the biclique $K_{N,N}$.

The following result by Algor and Alon [AA89] upper bounds the star arboricity of degree-bounded graphs.

Imported Theorem 6 (Consequence of [AA89])**.** *For any graph G with maximum degree Δ, the following bound holds.*

$$\mathsf{sa}\left(G\right) \leqslant \frac{1}{2} \cdot \Delta \cdot (1 + o(1)).$$

This result already yields non-trivial upper-bounds for $\mathsf{P_4\text{-}fc}\left(G\right)$ and $\mathsf{P_4\text{-}fp}\left(G\right)$ by upper-bounding its star-arboricity for several interesting functions (for example INT_n). Note, however, Theorem 3 provides an upper bound of $\mathsf{P_4\text{-}fc}\left(\mathsf{DISJ}_n\right)$ that is exponentially better than the upper bound entailed by [AA89].

Biclique Partition Number. Recall that a *biclique* is a complete bipartite graph. The *biclique partition number* of a graph, represented by $\mathsf{bp}\left(G\right)$, is the minimum number of bicliques needed to partition its edges. Graham and Pollak introduced this problem motivated by the network addressing problem and graph storage problem [GP71, GP72] (see also [BF88, VL85, YY06, vLW01]). The celebrated Graham-Pollak Theorem states that $\mathsf{bp}\left(K_N\right) = (N-1)$ [GP72, Tve82, Pec84, Vis08, Vis13]. However, all proofs are algebraic, and no purely combinatorial proof is known. In general, $\mathsf{bp}\left(G\right) \geqslant \max\left\{n_+(G), n_-(G)\right\}$ [GP72, Hof72, Tve82, Pec84], where $n_+(\cdot)$ and $n_-(\cdot)$, respectively, represent the number of positive and negative eigenvalues of the adjacency matrix of the graph.

Observe that the biclique partition number admits a trivial upper bound, $\mathsf{bp}\left(G\right) \leqslant$ the size of the smallest vertex cover of G. Determining the $\mathsf{bp}\left(G\right)$ of a general graph is a hard problem [KRW88] (even for bipartite graphs [Orl77]) and is also hard to approximate [CHHK14].

Section 1.1 establishes the connection between Wyner's common information of p_{XY} [Wyn75] with the biclique partition number of the bipartite graph $G(p_{XY})$.

Since a biclique is P_4-free, we naturally have the following bound.

Proposition 7. *For any graph G, the following bound holds:*

$$\mathsf{P_4\text{-}fc}\left(G\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right) \leqslant \mathsf{bp}\left(G\right).$$

Biclique partition number entirely ignores the potential of compressing multiple bicliques into one graph. Consequently, for most graphs, the upper-bound above is loose. For example, a matching has high biclique partition number; however, its P_4-free partition number is one.

Let M_G represent the adjacency matrix of the bipartite graph G. Algebraically, the notion of binary matrix factorization of M_G is identical to $\mathsf{bp}\left(G\right)$ [KPRW19]. The outer product of two binary vectors represents the adjacency matrix of a biclique. That is, *Boolean rank-one matrices* are bicliques. So, the minimum r such that M_G is the sum of r Boolean rank-one matrices represents the Boolean rank of M_G. Note that $\mathsf{bp}\left(G\right) = r$.

Biclique Cover Number. Covering a graph with the minimum number of bicliques has received significant attention in theoretical computer science [Orl77, Sim90, FMPS07, GH07, JK09, CHHK14] due to widespread application. Representative applications, for example, as [EU18] indicates, span computational biology [NMWA78, NHC$^+$12, NED$^+$13], data mining [MMG$^+$08], machine learning [SW03], automata theory [GH07], communication complexity [JK09], and graph drawing [HMR06]. Let alone computing the biclique cover number exactly (which is hard even for bipartite graphs [Orl77] and chordal bipartite graphs [Mül96]), approximating it is hard as well [Sim90, GH07, CHHK14].

The biclique cover number is at most the biclique partition number; however, it can be exponentially smaller. For example, $\mathsf{bc}\left(K_N\right) = \lceil \log_2 N \rceil$ [Pin13]; but, the Graham-Pollak Theorem states that $\mathsf{bp}\left(K_N\right) = (N-1)$ [GP72, Tve82, Pec84, Vis08, Vis13]. In general, Pinto [Pin13] proved that $\mathsf{bp}\left(G\right) \leqslant (3^{\mathsf{bc}(G)} - 1)/2$, and presented a graph family achieving equality in this bound.

Observe that $\mathsf{bc}\left(\mathsf{EQ}_N\right) \geqslant N$, where EQ_N is a equality function with size-N domain for the input of both parties. Intuitively, the graph corresponding to EQ_N is a matching, and no combinatorial rectangle can cover two edges of this matching. The "fooling set argument" relies on this observation to show lower bounds on $\mathsf{bc}\left(G\right)$, for a general G. It identifies a subset of vertices that induce a matching in the graph G. Therefore, the size of this matching lower-bounds $\mathsf{bc}\left(G\right)$.

Let M_G be the adjacency graph of the bipartite graph G. Then, there are *algebraic* matrix properties of M_G that help estimate the biclique cover number of G. For example, the non-negative rank of M_G[9] upper-bounds $\mathsf{bc}\left(G\right)$ [Yan91].[10]

F Proof of Theorem 1

Our proof of hardness for both partition and cover number is based on a result from [GO09], which shows that computing the edge partition of a bipartite planar graph into two star forests is NP-complete. For a definition of star forests and star arboricity, see Appendix E.

Theorem 7 (Gonçalves and Ochem [GO09])**.** *For any $g > 3$, deciding whether a bipartite planar graph G with girth* [11] *at least g and maximum degree 3 satisfies $\mathsf{sa}\left(G\right) \leqslant 2$ is NP-complete.*

Proof of Theorem 1 . First we show the decision problem is in NP, that is, given a partition of the edge set of G into $\leqslant 2$ components we can verify in polynomial time whether it is a P_4-free partition of size $\leqslant 2$ 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 7 is polynomial-time reducible to the P_4-free partition and cover number on bipartite graphs. The decision problem in Theorem 7 is NP-complete for any bipartite planar graph of girth at least $g > 3$; in particular, it holds for $g \geqslant 6$. Suppose we have a bipartite planar graph G with girth $g \geqslant 6$ 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 $\mathsf{sa}\left(G\right) = \mathsf{P_4\text{-}fp}\left(G\right) = \mathsf{P_4\text{-}fc}\left(G\right)$, 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 less than or equal to 2 if and only if so does the biclique partition number of G. □

[9] A non-negative rank-one matrix can be written as the outer product of two vectors whose entries are non-negative. A matrix M has non-negative rank r, if there exist r non-negative rank-one matrices that add to M.

[10] Consider the decomposition of M_G into minimum number of non-negative rank-one matrices. Consider a biclique cover that indicates whether the entries of these non-negative rank-one matrices are positive or not. This reduction provides a biclique cover of G.

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

G Proof of Theorem 2

Let H be a fixed graph, a classical problem in graph theory is finding the maximum number of edges in a graph on N vertices which does not contain a copy of H.

Definition 4 (Turan number)**.** *Turan number denoted by $ex(N,H)$ is the maximum number of edges in a graph on N vertices which 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 5 (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 $|L|=M$, $|R|=N$ which does not contain a sub-graph of the form $K_{s,t}$.*

Imported Theorem 8. *[ES74] $ex(N,K_{a,b}) \geqslant C'N^{2-\frac{a+b-2}{ab-1}}$, where C' is a positive absolute constant.*

Considering the adjacency matrix of a $K_{a,b}$-free graph on n vertices, we get $z(N,N,a,b) \geqslant 2ex(N,K_{a,b})$.

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 1. *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 $\mathsf{P_4\text{-}fp}\,(G) \geqslant \frac{e(G)}{2Nt}$.*

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. For any combinatorial rectangle in G', *length* $\leqslant 2N$ and *width* $\leqslant t$, since if *width* $= t+1 \leqslant$ *length*, then there exists a $K_{t+1,t+1}$-subgraph in G. This implies that $e(G') < 2Nt$, and consequently $\mathsf{P_4\text{-}fp}\,(G) \geqslant \frac{e(G)}{2Nt}$. □

The proof of Theorem 2 follows from the fact about Zarankiewicz function of $K_{t+1,t+1}$-free bipartite graphs and Lemma 1.

Proof of Theorem 2 . 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 Lemma 8,

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

where C is a positive absolute constant. By Lemma 1, we get that

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

□

G.1 Erdős-Rényi graphs do not have Large Bicliques

In this section, we will show that Erdős-Rényi graphs do not have *large* bicliques with *high* probability. We follow the standard outline for first moment techniques, see, for example, [FK16] Chapter 7.2. Let $G \leftarrow \mathsf{ER}(N,N,p)$, where $p \in (0,1)$ is a constant. 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.

Therefore, we have

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

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

H Estimates for INT_n, DISJ_n, and INEQ_N

In this section, we establish upper bounds for DISJ_n and INT_n in terms of P_4-free partition/cover number (see Theorem 11 and Theorem 12). 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 11).

H.1 P_4-free Partition/Cover Number and Graph Products

In this section, we introduce the notion of a graph product, and we prove 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 6 (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 9 (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.*

Proof. Let $(u_1,a_1),(u_2,a_2)$ be two distinct vertices in the left partite set of $G\times H$. Let $(v_1,b_1),(v_2,b_2)$ be two distinct vertices in the right partite set of $G \times H$. We emphasize that the vertices, for example, u_1, u_2 need not be distinct.

Consider the subgraph S induced by these four vertices. If $e(S) \leqslant 2$, then S is P_4-free. In the sequel, we shall prove that if $e(S) \geqslant 3$ implies that $e(S) = 4$, which proves that the graph S is P_4-free.

Suppose, without loss of generality, we have $(u_1,a_1) \sim (v_1,b_1) \sim (u_2,a_2) \sim (v_2,b_2)$, where $x \sim y$ denotes an edge between the two vertices x and y. We will call this assumption as the (*)-assumption in the sequel. Our objective is to prove that $(u_1,a_1) \sim (v_2,b_2)$.

The first case. Suppose, we have $u_1 \neq u_2$, $v_1 \neq v_2$, $a_1 \neq a_2$, and $b_1 \neq b_2$. Now, the (*)-assumption implies that $u_1 \sim v_1 \sim u_2 \sim v_2$ and $a_1 \sim b_1 \sim a_2 \sim b_2$. Since, the graphs G and H are themselves P_4-free, we have $u_1 \sim v_2$ and $a_1 \sim b_2$. Therefore, we also have $(u_1,a_1) \sim (v_2,b_2)$ in the product graph $G \times H$.

The remaining case. Without loss of generality, assume that $u_1 = u_2$. Similar to the above case, the (*)-assumption implies that $u_1 \sim v_1 \sim u_2 \sim v_2$ and $a_1 \sim b_1 \sim a_2 \sim b_2$. Then, the fact that $u_2 \sim v_2$ is equivalent to $u_1 \sim v_2$. Similarly, irrespective of whether $a_1 = a_2$ or not, or $b_1 = b_2$ or not, we have the fact that $a_1 \sim b_1 \sim a_2 \sim b_2$ implies $a_1 \sim b_2$. Therefore, we also have $(u_1,a_1) \sim (v_2,b_2)$.

This exhaustive case analysis completes the proof. □

Claim 10 (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*

$$\mathsf{P_4\text{-}fp}\left(G \times H\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right) \cdot \mathsf{P_4\text{-}fp}\left(H\right) \ .$$

Proof. Suppose $\mathsf{P_4\text{-}fp}\left(G\right) = k$, and the graph G partitions into graphs $G_1, \ldots, G_k$, such that the graph G_i, for every $1 \leqslant i \leqslant k$, is a P_4-free graph. Similarly, suppose $\mathsf{P_4\text{-}fp}\left(H\right) = \ell$, and the graph H partitions into $H_1, \ldots, H_\ell$, such that H_j, for every $1 \leqslant j \leqslant \ell$, is a P_4-free graph. Therefore, one can partition $G \times H$ as follows.

$$(G \times H) = \left(\sum_{i=1}^{k} G_i\right) \times \left(\sum_{j=1}^{\ell} H_j\right) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} G_i \times H_j.$$

By Claim 9, each $G_i \times H_j$ graph is P_4-free.

Furthermore, every edge in the graph $G \times H$ occurs exactly once in a unique graph $G_i \times H_j$. For example, consider an edge $e = (\ (u,a),(v,b)\ ) \in E(G \times H)$. Let $1 \leqslant i \leqslant k$ be the unique index such that $(u,v) \in E(G_i)$. Let $1 \leqslant j \leqslant \ell$ be the unique index such that $(a,b) \in E(H_j)$. Note that $e \in E(G_i \times H_j)$, and $e \notin E(G_{i'} \times H_{j'})$ for any other $i \neq i' \in [k]$ and $j \neq j' \in [\ell]$.

Therefore, $G_i \times H_j$, for $1 \leqslant i \leqslant k$, and $1 \leqslant j \leqslant \ell$, is a P_4-free partition of the graph $G \times H$. Consequently, we have $\mathsf{P_4\text{-}fp}\left(G \times H\right) \leqslant k\ell$. □

For any bipartite graph G, since $\mathsf{P_4\text{-}fc}\left(G\right) \leqslant \mathsf{P_4\text{-}fp}\left(G\right)$, the claim below follows.

Corollary 4 (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.*

$$\mathsf{P_4\text{-}fc}\left(G \times H\right) \leqslant \mathsf{P_4\text{-}fc}\left(G\right) \cdot \mathsf{P_4\text{-}fc}\left(H\right)$$

H.2 Bound on DISJ_n

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

Theorem 11. *For any $n \in \mathbb{N}$, the following bounds hold on the disjointness graph D_n.*

$$\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_n\right) = \mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_1^n\right) \leqslant 2^{\lceil n/2 \rceil}, \tag{1}$$

$$\mathsf{sa}\left(\mathsf{DISJ}_n\right) > \lceil (3/2)^n \rceil = \left\lceil 2.25^{n/2} \right\rceil. \tag{2}$$

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

$$\begin{aligned}\mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_n\right) = \mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_{n-2} \times \mathsf{DISJ}_2\right) &\leqslant \mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_{n-2}\right) \cdot \mathsf{P_4\text{-}fp}\left(\mathsf{DISJ}_2\right), && \text{(using Claim 10)}\\ &\leqslant 2^{\lceil n-2/2 \rceil} \cdot 2, && \text{using the inductive hypothesis}\\ &= 2^{\lceil n/2 \rceil}.\end{aligned}$$

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. □

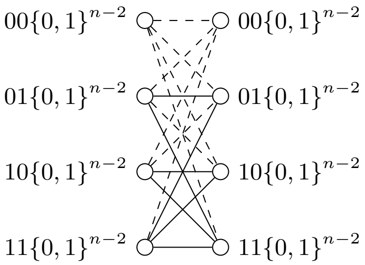


Figure 8: Partition of edges of INT_n into two sets.

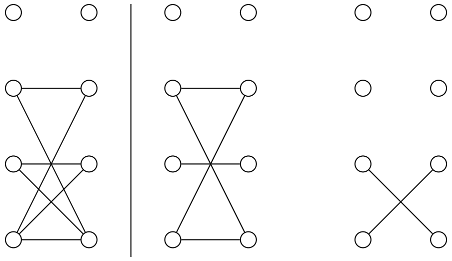


Figure 9: Partition of G_1 in Lemma 2 in two P_4-free graphs.

H.3 Bound on INT_n

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

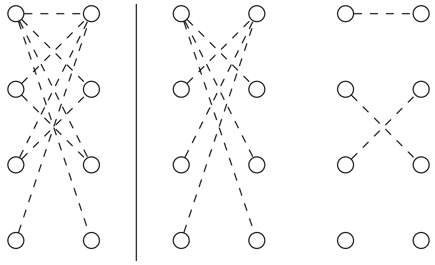


Figure 10: Partition of H_1 in Lemma 2 in two P_4-free graphs.

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

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 8). 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 9 for an illustration. Similarly, $H = \mathsf{INT}_{n-2} \times H_1$ where H_1 has P_4-free partition number 2. See Figure 10 for an illustration. Combing the above observations, we get that

$$\begin{aligned}\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) &\leqslant \mathsf{P_4\text{-}fp}\left(G\right) + \mathsf{P_4\text{-}fp}\left(H\right)\\ &\leqslant \mathsf{P_4\text{-}fp}\left(K_{2^{n-2},2^{n-2}} \times G_1\right) + \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2} \times H_1\right)\\ &\leqslant \mathsf{P_4\text{-}fp}\left(K_{2^{n-2},2^{n-2}}\right) \cdot \mathsf{P_4\text{-}fp}\left(G_1\right) + \mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) \cdot \mathsf{P_4\text{-}fp}\left(H_1\right) && \text{(By Claim 10)}\\ &\leqslant 2 + 2\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right)\end{aligned}$$

□

The main theorem of this section is presented below.

Theorem 12. *For all even $n \in \mathbb{N}$,*

$$\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) \leqslant 2 \cdot 2^{n/2} - 2.$$

For all odd $n \in \mathbb{N}$,

$$\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) \leqslant 3 \cdot 2^{(n-1)/2} - 2.$$

Proof. The following holds for all even n.

$$\begin{aligned}\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) &\leqslant 2\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2}\right) + 2\\ &\leqslant 2^k\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_{n-2k}\right) + 2 + 2^2 + \ldots + 2^k && \forall k \in [n-2, n/2-1]\\ &\leqslant 2^{n/2-1}\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_2\right) + \sum_{i=1}^{n/2-1} 2^i\\ &= 2^{n/2} + 2 \cdot (2^{n/2-1} - 1) && \because \mathsf{P_4\text{-}fp}\left(I_2\right) = 2\\ &= 2 \cdot 2^{n/2} - 2\end{aligned}$$

Similar to the analysis above, when n is odd the result below follows

$$\mathsf{P_4\text{-}fp}\left(\mathsf{INT}_n\right) \leqslant 3 \cdot 2^{(n-1)/2} - 2\ .$$

This completes the proof. □

H.4 P_4-free partition number of complement of Matching

Let $N(v)$ denote the neighbours of vertex v in a graph. We use the kernalization technique used in [FMPS09] presented below.

Definition 7. *For any given graph $G : (V, E)$, let $K(G)$ be the graph such that when the following two rules are applied on $K(G)$, the graph does not change. The rules are as follows:*

1. *If the degree of any vertex is 0, then we remove the vertex.*
2. *If $\exists\ u, v \in V$, such that $N(u) = N(v)$, then we remove u from the graph.*

[FMPS09] note that for any graph G the biclique partition number of G and $K(G)$ are equal. The next proposition gives a similar relationship for P_4-free partition number.

Proposition 8. *For any graph G, $\mathsf{P_4\text{-}fp}\left(G\right) = \mathsf{P_4\text{-}fp}\left(K(G)\right)$.*

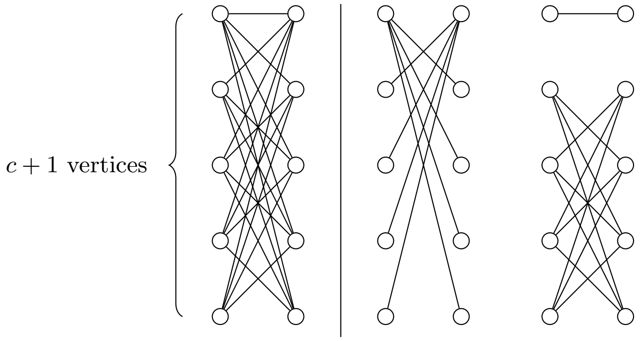


Figure 11: G' and partition of edges of G' in two sets, as in Claim 5, shown here for $c = 4$.

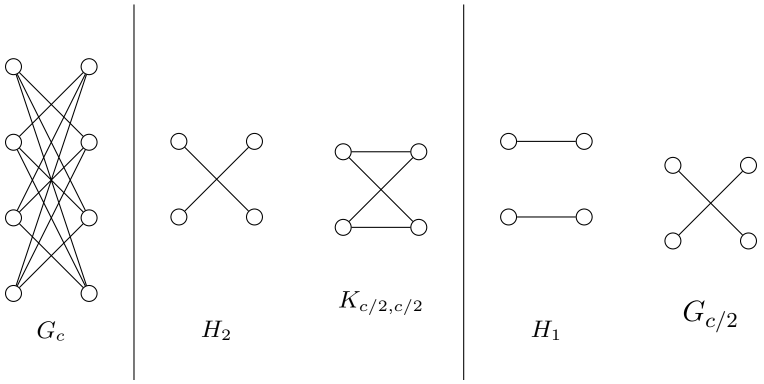


Figure 12: Representation of decomposition of G_c in Claim 5, shown here for $c = 4$. Edges of G_c are partitioned into the following two sets: $H_2 \times K_{c/2,c/2}$ and $H_1 \times G_{c/2}$.

In the sequel, we use this observation to show an upper bound on the P_4-free partition number of the complement of a P_4-free graph.

Proof of Claim 5 . Consider the graph $G' = K(K_{N,N} \setminus G)$. When G contains isolated vertices, observe that G' is isomorphic to the first graph in Figure 11 i.e. G' is isomorphic to a complete bipartite graph with $c+1$ vertices in each partite set and a matching of size c removed. The edge set of G' can be partitioned as follows: remove all the edges incident to the vertices with degree $c+1$ (note that there is only one vertex in each partite set of degree $c+1$) *except* the edge connecting them to each other. These removed edges form a P_4-free graph, call it S. We analyze the P_4-free partition number of the remaining graph separately.

Let G_c denote the remaining graph. G_c is a bipartite graph with c vertices in each partite set and a perfect matching removed. Observe that $E(G_c)$ is the union of $H_1 \times G_{\lceil c/2 \rceil}$ and $H_2 \times K_{\lceil c/2 \rceil, \lfloor c/2 \rfloor}$ (as demonstrated in Figure 12). Therefore, we get the following

$$\begin{aligned} \mathsf{P_4\text{-}fp}\left(G_c\right) &\leqslant \mathsf{P_4\text{-}fp}\left(H_1 \times G_{\lceil c/2 \rceil}\right) + \mathsf{P_4\text{-}fp}\left(H_2 \times K_{\lceil c/2 \rceil, \lceil c/2 \rceil}\right) \\ &\leqslant \mathsf{P_4\text{-}fp}\left(G_{c/2}\right) + 1 \end{aligned}$$

Solving the recursion we get $\mathsf{P_4\text{-}fp}\left(G_c\right) \leqslant \lceil \log c \rceil$, therefore

$$\mathsf{P_4\text{-}fp}\left(G\right) \leqslant \mathsf{P_4\text{-}fp}\left(G_c\right) + \mathsf{P_4\text{-}fp}\left(S\right) \leqslant \lceil \log c \rceil + 1 \; .$$

Furthermore, when G is a perfect matching, $K_{N,N} \setminus G$ is isomorphic to G_N (biclique with perfect matching removed), hence $\mathsf{P_4\text{-}fp}\left(K_{N,N} \setminus G\right) \leqslant \lceil \log N \rceil$. □

I Representative Scheduling Problem

Covering a graph with the minimum number of bicliques has received significant attention in theoretical computer science [Orl77, Sim90, FMPS07, GH07, JK09, CHHK14] due to widespread applications. Representative applications, for example, as [EU18] indicates, span computational biology [NMWA78, NHC+12, NED+13], data mining [MMG+08], machine learning [SW03], automata theory [GH07], communication complexity [JK09], security and access control [SLY06, EHM+08], and graph drawing [HMR06]. This graph property is referred to as the *biclique cover number* (also known as, bipartite dimension, and rectangle cover number).

A representative template. Let U be the set of users and D be the set of sensitive data. A Boolean matrix G defines which user has access to which data. That is, $G_{u,d} = 1$ implies that the user $u \in U$ should have access to the data $d \in D$; otherwise, if $G_{u,d} = 0$, then the user $u \in U$ should not have access to the data $d \in D$. It is possible to *many-to-many multicast* a subset of data $D' \subseteq D$ to a subset of users $U' \subseteq U$. Consequently, all the users in U' simultaneously receive all the data in D'. What is the *minimum number* of multicast necessary to help each user to receive all the data of its choice?

Note that each multicast above induces a biclique/combinatorial rectangle. Consequently, this combinatorial problem is equivalent to the biclique/rectangle cover number of the graph G, the minimum number of bicliques/rectangles to cover the bipartite graph/matrix G. Several applications mentioned above, for example, [NMWA78, SLY06, EHM+08, NHC+12, NED+13], fall into this template. If the users insist on receiving every data only once then this problem is equivalent to the biclique partition number of the graph G.

Leveraging parallelism. Observe that it may be possible to schedule multiple of these multicast instances simultaneously (refer to Figure 6). For example, two multicast instances above are non-conflicting if their sets of users and the set of data are disjoint. Clearly, non-conflicting multicast instances can be scheduled in parallel. In general, let $U_1, U_2, \dots, U_c$ be disjoint subsets of users, for arbitrary $c \in \mathbb{N}$, and $D_1, D_2, \dots, D_c$ be disjoint subsets of data. We shall enable the many-to-many multicast of the data in D_i to all users in U_i, for $1 \leqslant i \leqslant c$. Intuitively, we have parallelized multiple non-conflicting multicast instances. What is the *minimum number* of such parallelized multicast instances necessary to help each user to receive all the data of its choice?

This problem is equivalent to the P_4-free cover number of the graph G. If each user insists on receiving each data only once, then the problem is equivalent to the P_4-free partition number of the graph G.