Bounds on Maximal Leakage over Bayesian Networks

Anuran Makur* and Japneet Singh[†]
*Department of Computer Science and *[†]School of Electrical and Computer Engineering,
Purdue University, West Lafayette, IN, USA
Email: {amakur, sing1041}@purdue.edu

Abstract—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. **Maximal leakage quantifies the leakage of information from data $X \in \mathcal{X}$ due to an observation Y. While fundamental properties of maximal leakage, such as data processing, sub-additivity, and its connection with mutual information, are well-established, its behavior over Bayesian networks is not well understood and existing bounds are primarily limited to binary $\mathcal{X}$. In this paper, we investigate the behavior of maximal leakage over Bayesian networks with finite alphabets. Our bounds on maximal leakage are established by utilizing coupling-based characterizations which exist for channels satisfying certain conditions. Furthermore, we provide more general conditions under which such coupling characterizations hold for $|\mathcal{X}| = 4$. In the course of our analysis, we also present a new simultaneous coupling result on maximal leakage exponents. Finally, we illustrate the effectiveness of the proposed bounds with some examples.**

I. INTRODUCTION

Characterizing the information flow over a Bayesian network is a fundamental problem in information theory. A key question in this context is: How much information leakage is incurred by a processing step about a source $X \in \mathcal{X}$ as it is processed over a Bayesian network? Consider a system containing sensitive data X and an adversary interacting with the system to receive a noisy observation Y. The composite system between X and Y might be complex and, in many scenarios, can be effectively modeled by a Bayesian network. An important quantity useful in quantifying the "leakage" of information from X to any subset of nodes in the Bayesian network is maximal leakage.

In this paper, we aim to establish bounds on maximal leakage over Bayesian networks that utilize both the structural properties of networks (such as parent node relationships) and the properties of individual channels comprising the network. Maximal leakage, introduced by [1], is defined as the multiplicative increase in the probability of correctly guessing a possibly randomized function of the sensitive data X, given an observation Y. Formally, given a joint distribution $P_{X,Y}$ on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the maximal leakage from X to Y is defined as:

$$\mathcal{L}(X \to Y) = \sup_{U-X-Y-\hat{U}} \log \left(\frac{\mathbb{P}(\{U = \hat{U}\})}{\max_{u \in \mathcal{U}} \mathbb{P}_U(\{u\})} \right), \quad (1)$$

The author ordering is alphabetical.

where $U - X - Y - \hat{U}$ forms a Markov chain and U and $\hat{U}$ take values in the same finite, but arbitrary, alphabet set $\mathcal{U}$. It was shown in [1] that $\mathcal{L}(X \to Y)$ in the above definition reduces to the Sibson's mutual information of order infinity [2], [3]:

$$\mathcal{L}(X \to Y) = \log \left(\sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: P_X(x) > 0} P_{Y|X}(y|x) \right). \quad (2)$$

Maximal leakage has been applied in various domains, including security [4], privacy [5], adaptive data analysis, and learning theory [6]. Its fundamental properties, such as data processing, sub-additivity, and relation to mutual information [1], are well understood. However, despite the basic known properties for maximal leakage, its behavior in the context of Bayesian networks remains far from fully understood.

Bayesian networks [7], which represent joint distributions through directed acyclic graphs, offer a flexible framework for modeling dependencies among a set of variables. In these networks, nodes represent random variables, and edges represent conditional dependencies. The joint probability distribution of a set of random variables $Y_1, \ldots, Y_n$ in a Bayesian network is given by

$$P(Y_1, \ldots, Y_n) = \Pi_{i=1}^n P(Y_i | \text{pa}(Y_i)), \quad (3)$$

where $\text{pa}(Y_i)$ denotes the set of parent nodes of Y_i in the network. Tools for analyzing maximal leakage in the context of Bayesian networks can be useful in extending the applications of maximal leakage to more complex scenarios that involve multiple steps of processing and iterative refinement based on the data, especially in the context of adaptive data analysis [8], [9] and differential privacy [10]. In a simple chain-structured Bayesian network, the leakage from one node to another can be bounded by considering the conditional probabilities along the path. However, analysis of more complex networks presents a significant challenge. Notably, prior works, such as [10], [11], have explored similar problems, but are limited to binary input settings and rely on Dobrushin-type coupling techniques.

Main Contributions: In this work, we leverage recent developments in coupling characterizations of maximal leakage [12] —known to hold for specific channels—to derive bounds for Bayesian networks over finite alphabets. We establish these bounds by constructing a simultaneous coupling for joint probability mass functions (PMFs), which serves as a key tool for deriving these bounds. Additionally, we also provide a

more general condition under which such couplings are known to exist, by extending the characterization for the special case of $|\mathcal{X}| = 4$. Furthermore, we provide examples illustrating the application of our bounds.

A. Related Literature

Maximal leakage is a widely studied quantity in the study of information leakage, with applications spanning information theory [1], computer security [13], and machine learning [14]. The concept of leakage was first formalized in the context of computer security. [15] defined leakage as the logarithm of the multiplicative increase in the probability of correctly guessing a random variable X upon observing Y. Building on this, [13] introduced a worst-case version of this metric, maximizing it over all prior distributions on X. A closed-form expression for maximal leakage in the discrete case was derived by [1], along with analysis of its several variants and their fundamental properties and introduced a conditional version.

The concept of maximal leakage has since been generalized to various forms that unify several leakage measures across different applications. Maximal α-leakage, introduced by [16], incorporates a parameter α that interpolates between mutual information ($\alpha = 1$) and maximal leakage ($\alpha = \infty$). Similarly, maximal (α, β)-leakage [17] unifies several leakage measures, including local differential privacy and Rényi differential privacy. In privacy-related problems, pointwise maximal leakage [5] quantifies leakage about a secret X via individual outcomes. Other variants, such as g-leakage [18], [19], have been proposed to address specific scenarios, particularly in the context of entropy-based bounds on leakage.

Maximal leakage has been widely applied in machine learning and adaptive data analysis. [6] used it to derive statistical guarantees for both adaptive and non-adaptive settings. In learning theory, maximal leakage has been used to analyze the generalization error of iterative algorithms [14]. [20] employed it to upper bound misclassification rates in meta-universal coding strategies. Additionally, [21] examined privacy-utility tradeoffs, using maximal leakage as a privacy metric and as a performance metric in hypothesis testing.

A closely related work [10], presents an adaptive composition theorem for the overall leakage $\mathcal{L}(X \to (Y_1, \ldots, Y_n))$ in a Bayesian network, expressed in terms of individual leakages under the assumption of a binary input alphabet X. Their approach employs Dobrushin-type coupling arguments and the degradation properties of composite channels; however, the analysis is confined to binary input spaces. Furthermore, the bounds derived in this work align with established results on contraction coefficients in Bayesian networks. In particular, they correspond to the bounds formulated for the contraction coefficients of the Kullback-Leibler divergence, and total variation distance in [22], as well as for Doeblin coefficients in [12].

II. BACKGROUND AND PRELIMINARIES

Let X, Y be random variables defined on finite alphabets $\mathcal{X} = [n] \triangleq \{1, \ldots, n\}$ and $\mathcal{Y}$ respectively, and let $P_{X,Y}$ denote the joint PMF of X, Y. Without loss of generality, we will assume that the marginal P_X is such that $P_X(x) > 0$ for all $x \in \mathcal{X}$. Let $P_i \in \mathcal{P}_{\mathcal{Y}}$ (the probability simplex over $\mathcal{Y}$) denote the conditional PMF $P_{Y|X=i}$ for $i \in [n]$. Consider a collection of random variables $Y_1, \ldots, Y_n \in \mathcal{Y}$. A *coupling* of random variables $Y_1, \ldots, Y_n$ distributed as $P_1, \ldots, P_n$, respectively, is defined as a joint distribution $P_{Y_1, \ldots, Y_n}$ that preserves the marginals. For the channel $P_{Y|X} = \left[P_1^{\mathrm{T}}, \ldots, P_n^{\mathrm{T}} \right]^{\mathrm{T}} \in \mathbb{R}_{\mathrm{sto}}^{n \times |\mathcal{Y}|}$ formed by stacking the PMFs $P_1, \ldots, P_n \in \mathcal{P}_{\mathcal{Y}}$, we introduce a quantity called the *leakage exponent* denoted by $\tau_{\max}(P_{Y|X})$. This quantity, referred to as *max-Doeblin* in [12] is defined as

$$\tau_{\max}(P_{Y|X}) \triangleq \sum_{y \in \mathcal{Y}} \max\{P_1(y), \ldots, P_n(y)\}. \tag{4}$$

Under our assumption that $P_X(x) > 0$ for all $x \in \mathcal{X}$, $\tau_{\max}(P_{Y|X})$ corresponds to the exponentiated maximal leakage $\exp(\mathcal{L}(X \to Y))$ by (1). This equivalence motivates referring to $\tau_{\max}(P_{Y|X})$ as the *leakage exponent*. We will interchangeably use the notation $\tau_{\max}(P_{Y|X}) = \tau_{\max}(P_1, \ldots, P_n)$ and similarly for related quantities.

Coupling-based bound: Recently, [12] established a coupling-based bound for the leakage exponent $\tau_{\max}(P_{Y|X})$, which is formally stated in the following proposition:

Proposition 1 (Minimal Coupling Bound for Leakage Exponent [12, Theorem 4])**.** *For any random variables* $Y_1, \ldots, Y_n \in \mathcal{Y}$ *distributed according to the PMFs* $P_1, \ldots, P_n \in \mathcal{P}_{\mathcal{Y}}$, *respectively, we have*

$$\tau_{\max}(P_{Y|X}) \leq \min_{\substack{P_{Y_1, \ldots, Y_n}: \\ P_{Y_1} = P_1, \ldots, P_{Y_n} = P_n}} \sum_{y \in \mathcal{Y}} \mathbb{P}(\cup_{i=1}^n \{Y_i = y\}), \tag{5}$$

where the minimum is taken over all couplings of $P_1, \ldots, P_n$, *and* $\mathbb{P}(\cdot)$ *denotes the probability law corresponding to the coupling. Moreover, if*

$$\tau_{\max_2}(P_{Y|X}) \triangleq \sum_{y \in \mathcal{Y}} \max_2\{P_1(y), \ldots, P_n(y)\} \leq 1, \tag{6}$$

then equality in (5) holds, where $\max_2\{x_1, \ldots, x_n\}$ *denotes the second largest value among* $\{x_1, \ldots, x_n\}$.

The bound in (5) follows by the observation that for any coupling $\mathbb{P}(\cdot)$ we have

$$\sum_{y \in \mathcal{Y}} \mathbb{P}(\cup_{i=1}^n \{Y_i = y\}) \geq \sum_{y \in \mathcal{Y}} \max\{P_1(y), \mathbb{P}(\cup_{i=2}^n \{Y_i = y\})\},$$

$$\geq \sum_{y \in \mathcal{Y}} \max\{P_1(y), P_2(y), \ldots, P_n(y)\}, \tag{7}$$

where the first inequality follows since $\mathbb{P}(A \cup B) \geq \max\{\mathbb{P}(A), \mathbb{P}(B)\}$ and that the coupling preserves marginals, while the second inequality follows by recursively applying the same argument. When the equality in (5) holds, this result provides a coupling characterization for $\tau_{\max}(P_{Y|X})$. For $n = 3$, [12] showed that the condition $\tau_{\max_2}(P_{Y|X}) \leq 1$ is both sufficient and necessary, i.e., for $n = 3$ equality in (5)

if and only if $\tau_{\max_2}(P_{Y|X}) \leq 1$. However, for $n \geq 4$, equality in (5) may hold under more general conditions.

In this work, we primarily focus on channels satisfying $\tau_{\max_2}(P_{Y|X}) \leq 1$, ensuring that equality in (5) is satisfied. For such channels, we leverage this coupling characterization to derive bounds on maximal leakage in Bayesian networks. Examples of such channel include the q-ary symmetric channel W_δ with crossover probability δ, which satisfies $\tau_{\max_2}(W_\delta) \leq 1$ if $\delta \leq 1 - 1/q$, and the erasure channel E_ϵ with erasure probability ϵ, which always satisfies $\tau_{\max_2}(\mathsf{E}_\epsilon) \leq 1$. Furthermore, for $n = 4$, we will establish relaxed conditions under which we establish relaxed conditions under which equality in (5) holds, as detailed in Theorem 3.

III. SIMULTANEOUSLY COUPLING FOR MAXIMAL LEAKAGE

In this section, we establish a result on simultaneous couplings for a finite collection of joint PMFs. Specifically, we show that under the condition $\tau_{\max_2}(\cdot) \leq 1$ applied to one pair of marginal distributions, it is possible to construct a joint coupling that is minimal with respect to that pair of marginals. To formalize, consider m joint probability distributions $P_{X_1,Y_1}, P_{X_2,Y_2}, \ldots, P_{X_m,Y_m}$ over $\mathcal{X} \times \mathcal{Y}$, where the random variables $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y}$ for $i \in [m]$ are defined on finite alphabets $\mathcal{X}$ and $\mathcal{Y}$. Let the marginals of the Y_i's be denoted by $P_{Y_1}, P_{Y_2}, \ldots, P_{Y_m}$. Then, our simultaneous coupling result is stated as follows:

Theorem 1 (Simultaneous Coupling). *Given m joint probability distributions $P_{X_1,Y_1}, P_{X_2,Y_2}, \ldots, P_{X_m,Y_m}$, assume that the marginals $P_{Y_1}, P_{Y_2}, \ldots, P_{Y_m}$ satisfy $\tau_{\max_2}(P_{Y_1}, P_{Y_2}, \ldots, P_{Y_m}) \leq 1$. Then, there exists a coupling $P_{X_1,Y_1,X_2,Y_2,\ldots,X_m,Y_m}$ of $P_{X_1,Y_1}, \ldots, P_{X_m,Y_m}$ such that:*

$$\sum_{y \in \mathcal{Y}} \mathbb{P}(\cup_{i=1}^m \{Y_i = y\}) = \tau_{\max}(P_{Y_1}, \ldots, P_{Y_m}), \quad (8)$$

where $\mathbb{P}(\cdot)$ denotes the probability law under the coupling.

The proof of this theorem is provided in [23]. The construction of the simultaneous coupling utilizes the minimal coupling construction of $P_{Y_1}, P_{Y_2}, \ldots, P_{Y_m}$ while ensuring that the marginals of the joint PMFs P_{X_i,Y_i} for all $i \in [m]$ are preserved. We further note that the condition $\tau_{\max_2}(P_{Y_1}, P_{Y_2}, \ldots, P_{Y_m}) \leq 1$ is both sufficient and necessary for $m = 3$, as discussed in the proof. For larger values, specifically $m \geq 4$, we conjecture that this condition can be further relaxed using a similar recipe for obtaining the simultaneous coupling. In particular, for $m = 4$ the condition can be relaxed to a more general condition in Theorem 3.

IV. BOUNDS OVER BAYESIAN NETWORKS

In this section, we establish bounds on maximal leakage over Bayesian networks. Our objective is to establish upper bounds on the leakage exponent of the composite channel from a single source node to any sink node, based on the properties of the individual channels (such as their leakage exponents) in the Bayesian network. A Bayesian network is a probabilistic
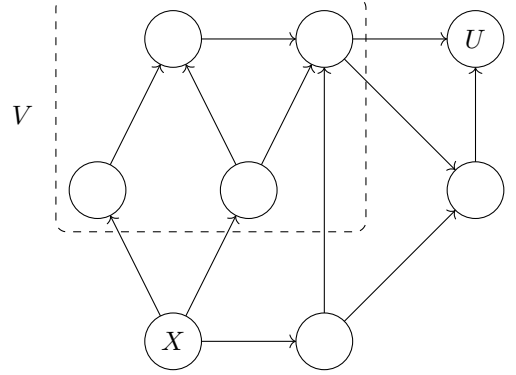

Fig. 1: Illustrative diagram of a Bayesian network. X is the source node, nodes inside the dotted box belong to the set $V \subset \mathsf{V}$ and U is the sink node.

graphical model represented by a directed acyclic graph, where each vertex corresponds to a random variable defined over a finite alphabet. We assume that each variable (vertex) takes values in a finite alphabet of arbitrary length.

Consider a Bayesian network with vertex set V and a source node X taking values in $\mathcal{X} = [n]$. Each vertex U, except the source node, is associated with a conditional distribution $P_{U|\mathrm{pa}(U)}$, where $\mathrm{pa}(U)$ denotes the set of parents of U. These conditional distributions together define the joint probability distribution over the network as in (3) (cf. [7]). We assume that the vertices in the network are topologically sorted. Specifically, for any node U and any non-empty subset of nodes $V \subseteq \mathsf{V}$, we use the notation $U > V$ to indicate that there is no directed path from U to V. Let $V \subset \mathsf{V}$ be any arbitrary non-empty subset of nodes, and suppose that the random variables associated with V take values in $\mathcal{V}$. Let U be a sink node such that $U > V$ in the topological ordering. For clarity, we include an illustration of a Bayesian network in Fig. 1.

The ensuing theorem provides bounds on the leakage exponent over a Bayesian network based on the minimal coupling and simultaneous coupling results presented in Proposition 1 and Theorem 1.

Theorem 2 (Bounds on Maximal Leakage in Bayesian Networks). *Let $V \subseteq \mathsf{V}$ be a non-empty subset of nodes and let $U \in \mathsf{V}$ be any node in V such that $U > V$. Assume that the conditional distributions satisfy*

$$\tau_{\max_2}(P_{U|\mathrm{pa}(U)}) \leq 1 \quad \text{and} \quad \tau_{\max_2}(P_{V|X}) \leq 1.$$

Let the quantity $f(P_{V \cup \mathrm{pa}(U)|X})$ is defined as

$$f(P_{V \cup \mathrm{pa}(U)|X}) \triangleq \sum_{v \in \mathcal{V}} \mathbb{P}\left(Z_1 = \cdots = Z_n, \bigcup_{j=1}^n \{V_j = v\} \right),$$

where $Z_i \sim P_{\mathrm{pa}(U)|X=i}$, $V_i \sim P_{V|X=i}$, and $\mathbb{P}(\cdot)$ is the probability law with respect to the simultaneous coupling of $P_{V_1,Z_1}, \ldots, P_{V_n,Z_n}$ (in Theorem 1). Then, the leakage

exponent of the composite channel $P_{V \cup \{U\}|X}$ satisfies the following bound:

$$\tau_{\max}(P_{V \cup \{U\}|X}) \leq \tau_{\max}(P_{U|\text{pa}(U)}) \cdot \tau_{\max}(P_{V|X}) \\ - (\tau_{\max}(P_{U|\text{pa}(U)}) - 1) \cdot f(P_{V \cup \text{pa}(U)|X}).$$

Proof. Let $Z = \text{pa}(U)$ and let its alphabet set be denoted by $\mathcal{Z}$. Let Q_i denote the conditional distribution of V, Z given $X = i$, i.e., $Q_i = P_{V,Z|X=i}$ for $i \in [n]$. Further, we define the random variables $(V_i, Z_i) \sim Q_i$ for $i \in [n]$. Since $\tau_{\max_2}(P_{V|X}) \leq 1$, and by Theorem 1, there exists a simultaneous coupling $P_{V_1, Z_1, \ldots, V_n, Z_n}$ of joint PMFs $Q_1, \ldots, Q_n$. Let $\mathbb{P}(\cdot)$ denote the corresponding probability operator under the simultaneous coupling, and we have that

$$\sum_{v \in \mathcal{V}} \mathbb{P}(\cup_{i=1}^n \{V_i = v\}) = \tau_{\max}(P_{V|X}).$$

Conditioned on $Z_1 = z_1, \ldots, Z_n = z_n$ for $z_1, \ldots, z_n \in \mathcal{Z}$, define the random variables $U_i \sim P_{U|Z_i=z_i}$. We now construct a minimal coupling of conditional PMFs $P_{U|Z=z_1}, \ldots, P_{U|Z=z_n}$, and let the random variables $(U_1, U_2, \ldots, U_n)$ be distributed according to this coupling (conditioned on $Z_1 = z_1, \ldots, Z_n = z_n$). Then, the minimal coupling of $V_1, Z_1, V_2, Z_2, \ldots, V_n, Z_n$ (earlier) along with the minimal coupling of $U_1, U_2, \ldots, U_n$ define the joint distribution $P_{V_1, U_1, Z_1, \ldots, V_n, Z_n, U_n}$, which satisfies the Markov relation

$$(V_1, V_2, \ldots, V_n) \longrightarrow (Z_1, Z_2, \ldots, Z_n) \longrightarrow (U_1, U_2, \ldots, U_n).$$

Also, by our assumption $\tau_{\max_2}(P_{U|Z}) \leq 1$, we have

$$\sum_{u \in \mathcal{U}} \mathbb{P}(\cup_{i=1}^n \{U_i = u\} \mid Z_1 = z_1, \ldots, Z_n = z_n) - 1$$
$$= \tau_{\max}(P_{U|Z_1=z_1}, \ldots, P_{U|Z_n=z_n}) - 1$$
$$\leq (\tau_{\max}(P_{U|Z}) - 1) \mathbb{1}_{\{Z_1 = \cdots = Z_n\}^c},$$

where $\mathbb{1}_{\mathcal{A}}$ denotes the indicator function of $\mathcal{A}$ and the inequality holds since the alphabet $\mathcal{Z}$ may be larger than $\{z_1, \ldots, z_n\}$ and the two bounds are equal on the set $\{Z_1 = \cdots = Z_n\}$. Now, for any fixed v taking expectation conditioned on the set $\mathcal{A}_v = \{\cup_{j=1}^n \{V_j = v\}\}$ on both sides, gives

$$\sum_{u \in \mathcal{U}} \mathbb{E}[\mathbb{P}(\cup_{i=1}^n \{U_i = u\}|Z_1 = z_1, \ldots, Z_n = z_n)|\mathcal{A}_v] - 1$$
$$\leq (\tau_{\max}(P_{U|Z}) - 1)\mathbb{P}(\{Z_1 = \cdots = Z_n\}^c|\mathcal{A}_v),$$

which simplifies to

$$\sum_{u \in \mathcal{U}} \mathbb{P}(\cup_{i=1}^n \{U_i = u\}|\mathcal{A}_v) - 1$$
$$\leq (\tau_{\max}(P_{U|Z}) - 1)(1 - \mathbb{P}(\{Z_1 = \cdots = Z_n\}|\mathcal{A}_v)).$$

Now, multiplying both sides by $\mathbb{P}(\cup_{j=1}^n \{V_j = v\})$ gives

$$\sum_{u \in \mathcal{U}} \mathbb{P}(\cup_{i=1}^n \{U_i = u\}, \cup_{j=1}^n \{V_j = v\}) - \mathbb{P}(\cup_{j=1}^n \{V_j = v\})$$
$$\leq (\tau_{\max}(P_{U|Z}) - 1)(\mathbb{P}(\cup_{j=1}^n \{V_j = v\}) - \\ \mathbb{P}(Z_1 = \cdots = Z_n, \cup_{j=1}^n \{V_j = v\})).$$

Now observe that

$$\sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{U}} \mathbb{P}(\cup_{i=1}^n \{U_i = u\}, \cup_{j=1}^n \{V_j = v\}) \geq$$
$$\sum_{v \in \mathcal{V}} \sum_{u \in \mathcal{U}} \mathbb{P}(\cup_{i=1}^n \{U_i = u, V_i = v\}) \geq \tau_{\max}(P_{U,V|X})$$

Finally, taking the sum over all $v \in \mathcal{V}$ on both sides and utilizing the above bound and the construction of simultaneous maximal coupling, we have

$$\tau_{\max}(P_{V,U|X}) \leq \tau_{\max}(P_{V|X}) + (\tau_{\max}(P_{U|Z}) - 1) \\ (\tau_{\max}(P_{V|X}) - \mathbb{P}(Z_1 = \cdots = Z_n, \cup_{j=1}^n \{V_j = v\})). \quad (9)$$

Substituting the measure of set under the coupling yields the desired result. $\qquad \square$

The theorem highlights that we can iteratively "peel off" a node (or sets of nodes) in the Bayesian network to obtain an upper bound on the leakage exponent of the composite channel. The existence of a simultaneous coupling for the joint PMFs $P_{V_1, Z_1}, \ldots, P_{V_n, Z_n}$ is central to our approach and is guaranteed by the assumption $\tau_{\max_2}(P_{V|X}) \leq 1$. While the condition may appear restrictive, it arises from the lack of a general characterization when couplings achieving equality in Eqs. (5) and (8) exist for a given set of PMFs. Importantly, the proof of Theorem 2 does not rely on the structure of the coupling but only on its existence and that it achieves the bound in (5). Furthermore, the last term $f(P_{V \cup \text{pa}(U)|X})$ quantifies the "loss in leakage" when the parents $Z_1, \ldots, Z_n$ of U all become equal (conditioned on X). We can utilize the structure of the simultaneous coupling construction presented in Theorem 1 to obtain a simplified bound in terms of the Doeblin coefficient as presented below.

Corollary 1 (Simplified Bounds on Leakage Exponent). *Consider the setting as in Theorem 2, assume that the conditional distributions satisfy $\tau_{\max_2}(P_{U|\text{pa}(U)}) \leq 1$ and $\tau_{\max_2}(P_{V|X}) \leq 1$. Then, the leakage exponent of the composite channel $P_{V \cup \{U\}|X}$ satisfies the following bound:*

$$\tau_{\max}(P_{V \cup \{U\}|X}) \leq \tau_{\max}(P_{U|\text{pa}(U)}) \cdot \tau_{\max}(P_{V|X}) \\ - (\tau_{\max}(P_{U|\text{pa}(U)}) - 1) \cdot \tau(P_{V \cup \text{pa}(U)|X}),$$

where $\tau(P_{Y|X})$ is the Doeblin coefficient of the channel $P_{Y|X}$, defined as

$$\tau(P_{Y|X}) \triangleq \sum_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} P_{Y|X}(y|x).$$

This corollary is a consequence of the structure of our simultaneous coupling in Theorem 1. The resulting bound depends on the Doeblin coefficient of the composite channel $\tau(P_{V \cup \text{pa}(U)|X})$, which can be lower bounded by using the Doeblin coefficient of the individual channels as demonstrated in [12, Theorem 6]. Consequently, our bound in Corollary 1 parallels existing bounds on Bayesian networks for contraction coefficients of the Kullback-Leibler divergence and total variation distance established in [22], as well as for Doeblin coefficients in [12].
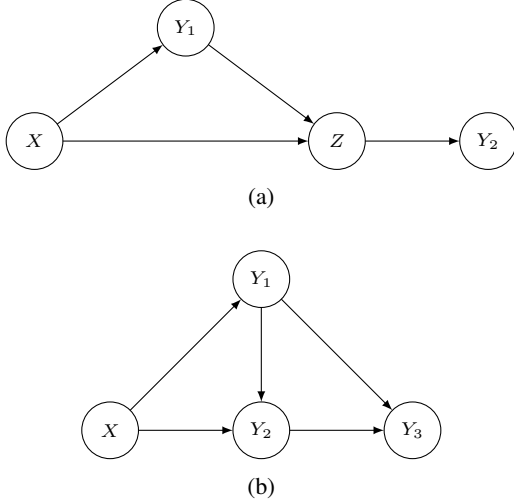
Fig. 2: Examples of Bayesian networks.

A. Application of Bounds on Simple Bayesian Networks

We will now apply the bounds established in Corollary 1 to different examples of Bayesian networks to illustrate their effectiveness.

1) Example 1: Consider the Bayesian network shown in Fig. 2a. Using Corollary 1 under the assumption that $\tau_{\max_2}(P_{Y_2|Z}) \leq 1$ and $\tau_{\max_2}(P_{Y_1|Z}) \leq 1$, the bound on maximal leakage is given by:

$$
\begin{aligned}
\tau_{\max}(P_{Y_1,Y_2|X}) \leq {} & \tau_{\max}(P_{Y_2|Z})\tau_{\max}(P_{Y_1|X}) \\
& - \big(\tau_{\max}(P_{Y_2|Z}) - 1\big)\tau(P_{Y_1,Z|X}).
\end{aligned}
\tag{10}
$$

Taking logarithm on both sides, we obtain

$$
\begin{aligned}
\mathcal{L}(X \to (Y_1, Y_2)) \leq {} & \mathcal{L}(X \to Y_1) + \mathcal{L}(Z \to Y_2) \\
& + \log\left(1 - \frac{(\mathcal{L}(Z \to Y_2) - 1)}{\mathcal{L}(Z \to Y_2)}\frac{\tau(P_{Y_1,Z|X})}{\tau_{\max}(P_{Y_1|Z})}\right).
\end{aligned}
\tag{11}
$$

Observe that when the Doeblin coefficient $\tau(P_{Y_1,Z|X}) > 0$, the bound in (11) improves upon the sub-additivity property in [1]. Intuitively, the improvement arises because the bound accounts for the reduction in leakage due to the "erasure of information by the Doeblin component" [24], which is precisely what our proof technique for Theorem 2 aims to capture. Furthermore, this bound can be applied recursively, provided the required assumptions hold at each step.

2) Example 2: Applying Corollary 1 to the Bayesian network in Fig. 2b, we obtain the bound

$$
\begin{aligned}
\tau_{\max}(P_{(Y_1,Y_2,Y_3)|X}) \leq {} & \tau_{\max}(P_{(Y_1,Y_2)|X}) \cdot \tau_{\max}(P_{Y_3|(Y_1,Y_2)}) \\
& - (\tau_{\max}(P_{Y_3|(Y_1,Y_2)}) - 1) \cdot \tau(P_{(Y_1,Y_2)|X}),
\end{aligned}
\tag{12}
$$

under the assumption that $\tau_{\max_2}(P_{Y_3|(Y_1,Y_2)}) \leq 1$ and $\tau_{\max_2}(P_{(Y_1,Y_2)|X}) \leq 1$. Applying Corollary 1, we can further bound $\tau_{\max}(P_{(Y_1,Y_2)|X})$ as

$$
\tau_{\max}(P_{(Y_1,Y_2)|X}) \leq \tau_{\max}(P_{Y_1|X})\tau_{\max}(P_{Y_2|(X,Y_1)}),
\tag{13}
$$

assuming $\tau_{\max_2}(P_{Y_1|X}) \leq 1$ and $\tau_{\max_2}(P_{Y_2|(X,Y_1)}) \leq 1$.

V. MINIMAL COUPLING CONSTRUCTION FOR LEAKAGE EXPONENT FOR $n = 4$

In this section, we extend the known conditions beyond $\tau_{\max_2}(P_{Y|X}) \leq 1$ under which the coupling achieves the lower bound in (5) exists for the special case of $n = 4$, where $n = |\mathcal{X}|$. Subsequently, we employ this coupling construction to directly extend the previously presented results.

Theorem 3. *For $n = 4$, let $P_1, P_2, P_3, P_4 \in \mathcal{P}_\mathcal{Y}$ be the given PMFs. For $i \neq j$ and $i, j \in [4]$, let $P_{\min}(y) = \min\{P_1(y), P_2(y), P_3(y), P_4(y)\}$ and define N_{ij} as*

$$
\begin{aligned}
N_{ij} \triangleq \sum_{y \in \mathcal{Y}} \bigg(& \min\{P_i(y), P_j(y)\} - \\
& \sum_{k \in [4]\setminus\{i,j\}} \min\{P_i(y), P_j(y), P_k(y)\} + P_{\min}(y) \bigg).
\end{aligned}
\tag{14}
$$

Then, there exists a coupling for $n = 4$ that achieves the lower bound in (5), provided the following condition holds:

$$
\begin{aligned}
\min\{N_{12}, N_{34}\} + \min\{N_{13}, N_{24}\} + \min\{N_{14}, N_{23}\} \\
\geq \tau_{\max_2}(P_1, P_2, P_3, P_4) - 1.
\end{aligned}
\tag{15}
$$

The construction of coupling is presented in [23]. Notably, the quantities N_{ij} are non-negative, and therefore, the condition (15) encompasses the condition $\tau_{\max_2}(\cdot) \leq 1$. The condition (15) arises after solving a system of linear equations while at the same time ensuring the non-negative of the solution. The system of linear equations has been constructed such that the coupling preserves the marginals while at the same time achieving the lower bound in (5). We remark that there indeed exists PMFs that satisfy the condition in (15) (see [23] for an example). Moreover, we can construct a simultaneous coupling of $P_{X_1,Y_1}, \ldots, P_{X_4,Y_4}$ provided that marginals $P_{Y_1}, \ldots, P_{Y_4}$ satisfy the condition in (15). This allows us to directly extend the results in Theorems 1 and 2 to beyond $\tau_{\max_2}(P_{Y|X}) \leq 1$ for $n = 4$.

VI. CONCLUSION AND FUTURE DIRECTIONS

In this work, we investigated the behavior of maximal leakage over Bayesian networks on finite alphabets, establishing bounds using minimal coupling characterizations of leakage exponent and a new simultaneous coupling result. We provided a more general condition under which the coupling characterizations were known for $|\mathcal{X}| = 4$ and demonstrate the effectiveness of the bounds through examples.

Our work also opens several avenues for future research. With minor modifications, we believe that our framework can be generalized to related notions of maximal leakage, such as pointwise maximal leakage [5] and related quantities, as our coupling constructions achieve equality in (7), for every $y \in \mathcal{Y}$. Secondly, relaxing the condition $\tau_{\max_2}(\cdot) \leq 1$ could lead to broader applicability of our results. Moreover, characterizing channels for which a minimal coupling achieves the bound in (5) remains an open problem. Finally, we believe that bounding the gap between the two terms in (5) could enable the extension of our bounds to general channels.

REFERENCES

[1] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2020.

[2] R. Sibson, "Information radius," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, Jun 1969. [Online]. Available: https://doi.org/10.1007/BF00537520

[3] S. Verdú, "α-mutual information," in *2015 Information Theory and Applications Workshop (ITA)*. IEEE, 2015, pp. 1–6.

[4] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," *Electronic Notes in Theoretical Computer Science*, vol. 249, pp. 75–91, 2009, proceedings of the 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1571066109003077

[5] S. Saeidian, G. Cervia, T. J. Oechtering, and M. Skoglund, "Pointwise maximal leakage," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 626–631.

[6] A. R. Esposito, M. Gastpar, and I. Issa, "Learning and adaptive data analysis via maximal leakage," in *2019 IEEE Information Theory Workshop (ITW)*, 2019, pp. 1–5.

[7] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023.

[8] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth, "Generalization in adaptive data analysis and holdout reuse," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/file/bad5f33780c42f2588878a9d07405083-Paper.pdf

[9] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. L. Roth, "Preserving statistical validity in adaptive data analysis," in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 117–126. [Online]. Available: https://doi.org/10.1145/2746539.2746580

[10] I. Issa and A. B. Wagner, "An adaptive composition theorem for maximal leakage for binary inputs," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2851–2855.

[11] R. D. H. Cung, Y. Y. Shkel, and I. Issa, "Binary maximal leakage," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 2748–2753.

[12] A. Makur and J. Singh, "Doeblin coefficients and related measures," *IEEE Transactions on Information Theory*, vol. 70, no. 7, pp. 4667–4692, 2024.

[13] C. Braun, K. Chatzikokolakis, and C. Palamidessi, "Quantitative notions of leakage for one-try attacks," *Electronic Notes in Theoretical Computer Science*, vol. 249, pp. 75–91, 2009, proceedings of the 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1571066109003077

[14] I. Issa, A. R. Esposito, and M. Gastpar, "Generalization error bounds for noisy, iterative algorithms via maximal leakage," in *Proceedings of Thirty Sixth Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, G. Neu and L. Rosasco, Eds., vol. 195. PMLR, 12–15 Jul 2023, pp. 4952–4976.

[15] G. Smith, "On the foundations of quantitative information flow," in *Foundations of Software Science and Computational Structures*, L. de Alfaro, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 288–302.

[16] J. Liao, O. Kosut, L. Sankar, and F. du Pin Calmon, "Tunable measures for information leakage and applications to privacy-utility tradeoffs," *IEEE Transactions on Information Theory*, vol. 65, no. 12, pp. 8043–8066, 2019.

[17] A. Gilani, G. R. Kurri, O. Kosut, and L. Sankar, "Unifying privacy measures via maximal (α, β)-leakage (M$\alpha\beta$eL)," *IEEE Transactions on Information Theory*, vol. 70, no. 6, pp. 4368–4395, 2024.

[18] B. Espinoza and G. Smith, "Min-entropy as a resource," *Information and Computation*, vol. 226, pp. 57–75, 2013, special Issue: Information Security as a Resource. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0890540113000230

[19] M. S. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *2012 IEEE 25th Computer Security Foundations Symposium*, 2012, pp. 265–279.

[20] F. E. Rosas, P. A. M. Mediano, and M. Gastpar, "Learning, compression, and leakage: Minimising classification error via meta-universal compression principles," in *2020 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–5.

[21] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 779–783.

[22] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," in *Convexity and Concentration*, ser. The IMA Volumes in Mathematics and its Applications, E. Carlen, M. Madiman, and E. M. Werner, Eds., vol. 161. New York, NY, USA: Springer, 2017, pp. 211–249.

[23] A. Makur and J. Singh, "Bounds on maximal leakage over bayesian networks," 2024, in preparation. [Link].

[24] A. Gohari, O. Günlü, and G. Kramer, "Coding for positive rate in the source model key agreement problem," *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6303–6323, October 2020.

A. Proof of Theorem 1

In this section, we will outline the construction of the coupling that achieves the bounds mentioned. Before doing so, we introduce the constants c_{XY} and c_Y, which are defined as follows:

$$c_{XY} \triangleq \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \min\{P_{X_1,Y_1}(x,y), \dots, P_{X_m,Y_m}(x,y)\}$$

$$c_Y \triangleq \sum_{y \in \mathcal{Y}} \min\{P_{Y_1}(y), \dots, P_{Y_m}(y)\}.$$

Observe that $c_Y \geq c_{XY}$. Let $x^m \triangleq (x_1, x_2, \dots, x_m)$ and $y^m \triangleq (y_1, y_2, \dots, y_m)$. Additionally, we define $P_{Y_{\min}}(y) \triangleq \min\{P_{Y_1}(y), \dots, P_{Y_m}(y)\}$. We now introduce the following distributions over the product space $\mathcal{X}^m \times \mathcal{Y}^m$:

$$G_1(x^m, y^m) \triangleq \frac{P_{\min}(x,y)}{c_{XY}} \mathbb{1}_{x^m = x, y^m = y}$$

$$G_2(x^m, y^m) \triangleq \frac{P_{Y_{\min}}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)}{c_Y - c_{XY}}$$
$$\times \prod_{i=1}^{m} \frac{P_{X_i,Y_i}(x_i,y) - P_{\min}(x_i,y)}{P_{Y_i}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)} \mathbb{1}_{y^m = y},$$

$$G_3(x^m, y^m) \triangleq \prod_{i=1}^{m} \frac{P_{X_i,Y_i}(x_i,y_i) - P_{\min}(x_i,y_i)}{P_{Y_i}(y_i) - \sum_{x \in \mathcal{X}} P_{\min}(x,y_i)} \times H(y^m)$$

where $H(y^m)$ is defined below and is based on the construction of minimal coupling $P_{Y_1,\dots,Y_m}(y_1, \dots, y_m)$ of the marginals $P_{Y_1}, \dots, P_{Y_m}$ as in [12, Theorem 4] as

$$H(y^m) \triangleq \frac{P_{Y_1,\dots,Y_m}(y^m) - P_{Y_{\min}}(y) \mathbb{1}_{y^m = y}}{1 - c_Y}.$$

The coupling $P_{Y_1,\dots,Y_m}$ exists when $\tau_{\max_2}(P_{Y_1}, \dots, P_{Y_m}) \leq 1$ and by construction $H(y^m)$ is clearly non-negative. Also, note that $H(y^m)$ is, in fact, a PMF, and so does G_1, G_2, G_3. We define the overall coupling $P_{X_1,Y_1,\dots,X_m,Y_m}$ as a convex combination of the the PMFs as:

$$P_{X_1,Y_1,\dots,X_m,Y_m}(x^m, y^m) = c_{XY} G_1(x^m, y^m) +$$
$$(c_Y - c_{XY}) G_2(x^m, y^m) + (1 - c_Y) G_3(x^m, y^m).$$

Now, we will show that the above joint distribution has the right marginals jointly with respect to X and Y. Define $x^{(i)} \triangleq (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ and $y^{(i)} \triangleq (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$. Observe that for all x, y such that $x_i = x$ and $y_i = y$, we have

$$\sum_{x^{(i)}, y^{(i)}} P_{X_1,Y_1,\dots,X_n,Y_n}(x^n, y^n) = \sum_{x^{(i)}, y^{(i)}} c_{XY} G_1(x^m, y^m)$$
$$+ \sum_{x^{(i)}} \sum_{y^{(i)}} (c_Y - c_{XY}) G_2(x^m, y^n)$$
$$+ \sum_{x^{(i)}} \sum_{y^{(i)}} (1 - c_Y) G_3(x^n, y^n)$$
$$= P_{\min}(x,y) + \left(P_{Y_{\min}}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y) \right)$$

$$\times \left(\frac{P_{X_i,Y_i}(x,y) - P_{\min}(x,y)}{P_{Y_i}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)} \right)$$
$$+ \frac{(P_{X_i,Y_i}(x,y) - P_{\min}(x,y))(P_{Y_i}(y) - P_{Y_{\min}}(y))}{P_{Y_i}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)}$$
$$= P_{\min}(x,y) + \frac{(P_{X_i,Y_i}(x,y) - P_{\min}(x,y))}{P_{Y_i}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)}$$
$$\times \left(P_{Y_i}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y) \right)$$
$$= P_{\min}(x,y) + P_{X_i,Y_i}(x,y) - P_{\min}(x,y)$$
$$= P_{X_i,Y_i}(x,y).$$

Now, we will complete the proof by showing that the above coupling satisfies the condition

$$\sum_{y \in \mathcal{Y}} \mathbb{P}(\cup_{i=1}^{m}\{Y_i = y\}) = \tau_{\max}(P_{Y_1}, \dots, P_{Y_n}). \quad (16)$$

This follows directly by observing that marginalizing the joint coupling $\mathbb{P}(x^m, y^m)$ with respect to x^m as

$$\sum_{x^m \in \mathcal{X}^m} \mathbb{P}(x^m, y^m) = \sum_{x \in \mathcal{X}} P_{\min}(x,y) \mathbb{1}_{y^m = y}$$
$$+ (P_{Y_{\min}}(y) - \sum_{x \in \mathcal{X}} P_{\min}(x,y)) \mathbb{1}_{y^m = y} + (1 - c_Y) H(y^m)$$
$$= P_{Y_{\min}}(y) \mathbb{1}_{y^m = y} + P_{Y_1,\dots,Y_m}(y^m) - P_{Y_{\min}}(y) \mathbb{1}_{y^m = y}$$
$$= P_{Y_1,\dots,Y_m}(y^m)$$

Now since the $P_{Y_1,\dots,Y_m}(y^m)$ is constructed based on [12, Theorem 4], therefore it automatically satisfies the condition in (16). This completes the proof. $\square$

We remark that the condition $\tau_{\max_2}(P_{Y_1}, P_{Y_2}, \dots, P_{Y_m}) \leq 1$ is necessary for $m = 3$, since for any coupling of PMFs P_1, P_2, P_3 corresponding to random variables Y_1, Y_2, Y_3, we have the lower bound:

$$\sum_{y \in \mathcal{Y}} \mathbb{P}(\cup_{i=1}^{3}\{Y_i = y\}) \geq \sum_{y \in \mathcal{Y}} \max\{P_1(y), \mathbb{P}(\cup_{i=2}^{3}\{Y_i = y\})\},$$
$$\geq \sum_{y \in \mathcal{Y}} \max\{P_1(y), P_2(y), P_3(y)\},$$

Since by [12], for $n = 3$ the lower bound is achieved if and only if $\tau_{\max_2}(P_1, P_2, \dots, P_3) \leq 1$. The same argument also holds for simultaneous coupling of $P_{X_1,Y_1}, P_{X_2,Y_2}, P_{X_3,Y_3}$ as the coupling needs to preserve the marginals for each of the pairs X_i and Y_i for $i \in [3[$.

B. Proof of Theorem 3

We will now demonstrate a construction of coupling that achieves the lower bound $\tau_{\max}(P_{Y|X})$ in Proposition 1 and for $\tau_{\max_2}(P_1, P_2, P_3, P_4) \geq 1$. For any subset $\mathcal{I} \subset [4]$ with $|\mathcal{I}| \in \{2, 3\}$, we begin by defining the following quantities:

$$\tau \triangleq \tau(P_1, P_2, P_3, P_4), \quad \tau_{\max} \triangleq \tau(P_1, P_2, P_3, P_4),$$
$$\tau_{\max_2} \triangleq \tau_{\max_2}(P_1, P_2, P_3, P_4),$$
$$\tau_{\max_3} \triangleq \tau_{\max_3}(P_1, P_2, P_3, P_4)$$
$$\tau_\mathcal{I} \triangleq \sum_{y \in \mathcal{Y}} \min_{i \in \mathcal{I}} P_i(y), \quad \tau_{\text{pair}} \triangleq \sum_{\mathcal{I}:\mathcal{I} \subset [4], |\mathcal{I}| = 2} \tau_\mathcal{I}$$

$$\tau_{\text{trip}} \triangleq \sum_{\mathcal{I}:\mathcal{I}\subset[4],|\mathcal{I}|=3} \tau_{\mathcal{I}}, \tag{17}$$

where $\max_3\{x_1,x_2,x_3,x_4\}$ denotes the third maximum of x_1,x_2,x_3,x_4. We will utilize the following relations obtained from the maximum-minimums principles:

$$
\begin{aligned}
\tau_{\max_3} &= \tau_{\text{trip}} - 3\tau \\
\tau_{\max_2} &= \tau_{\text{pair}} - 2\tau_{\text{trip}} + 3\tau
\end{aligned}
\tag{18}
$$

Furthermore, we define $y^4 = (y_1,y_2,y_3,y_4)$ and $P_{\min}(y) \triangleq \min\{P_1(y),P_2(y),P_3(y),P_4(y)\}$. Additionally, we introduce the following quantities:

$$
\begin{aligned}
R_1(y) &\triangleq \frac{P_1(y) - \min\{P_1(y),\max\{P_2(y),P_3(y),P_4(y)\}\}}{1-\tau_{12}-\tau_{13}-\tau_{14}+\tau_{123}+\tau_{124}+\tau_{134}-\tau}, \\
R_2(y) &\triangleq \frac{P_2(y) - \min\{P_2(y),\max\{P_1(y),P_3(y),P_4(y)\}\}}{1-\tau_{12}-\tau_{23}-\tau_{24}+\tau_{123}+\tau_{124}+\tau_{234}-\tau}, \\
R_3(y) &\triangleq \frac{P_3(y) - \min\{P_3(y),\max\{P_1(y),P_2(y),P_4(y)\}\}}{1-\tau_{12}-\tau_{23}-\tau_{24}+\tau_{123}+\tau_{124}+\tau_{234}-\tau}, \\
R_4(y) &\triangleq \frac{P_4(y) - \min\{P_4(y),\max\{P_1(y),P_2(y),P_3(y)\}\}}{1-\tau_{14}-\tau_{24}-\tau_{34}+\tau_{124}+\tau_{134}+\tau_{234}-\tau},
\end{aligned}
\tag{19}
$$

We now define the following PMFs where Q_0 corresponds to a PMF whose support is on the points where y_1,y_2,y_3,y_4 are all equal and $Q_{1,i}$ for $i \in [4]$ are the PMFs supported on the points where one of the y_i is different from the remaining three y_j's as:

$$
\begin{aligned}
Q_0(y^4) &\triangleq \frac{P_{\min}(y)}{\tau}\mathbb{1}_{y_1=y_2=y_3=y_4=y}, \\
Q_{1,1}(y^4) &\triangleq R_1(y_1)\times \\
&\quad \frac{\min\{P_2(y),P_3(y),P_4(y)\}-P_{\min}(y)}{\tau_{234}-\tau}\mathbb{1}_{y_2=y_3=y_4=y}, \\
Q_{1,2}(y^4) &\triangleq R_2(y_2)\times \\
&\quad \frac{\min\{P_1(y),P_3(y),P_4(y)\}-P_{\min}(y)}{\tau_{134}-\tau}\mathbb{1}_{y_1=y_3=y_4=y}, \\
Q_{1,3}(y^4) &\triangleq R_3(y_3)\times \\
&\quad \frac{\min\{P_1(y),P_2(y),P_4(y)\}-P_{\min}(y)}{\tau_{124}-\tau}\mathbb{1}_{y_1=y_2=y_4=y}, \\
Q_{1,4}(y^4) &\triangleq R_4(y_4)\times \\
&\quad \frac{\min\{P_1(y),P_2(y),P_3(y)\}-P_{\min}(y)}{\tau_{123}-\tau}\mathbb{1}_{y_1=y_2=y_3=y},
\end{aligned}
$$

where quantities such as τ_{234} corresponds to the definition of $\tau_{\mathcal{I}}$ in (17) with $I=\{2,3,4\}$. Moreover, for any distinct $i,j \in [4]$, we define quantities T_{ij} and their normalizing constants N_{ij}:

$$
\begin{aligned}
T_{ij}(y) &\triangleq \min\{P_i(y),P_j(y)\} - \min\{P_i(y),P_j(y),P_k(y)\} \\
&\quad - \min\{P_i(y),P_j(y),P_l(y)\} + P_{\min}(y) \\
N_{ij} &= \sum_{y\in\mathcal{Y}} T_{ij}(y)
\end{aligned}
\tag{20}
$$

where for each pair (i,j), indices k and l are the remaining two elements of $[4] \setminus \{i,j\}$. We define the following coupling distribution:

$$
\begin{aligned}
Q_{2,12}(y_1,y_2,y_3,y_4) &\triangleq R_1(y_1) \times R_2(y_2)\frac{T_{34}(y)}{N_{34}}\mathbb{1}_{y_3=y_4=y}, \\
Q_{2,13}(y_1,y_2,y_3,y_4) &\triangleq R_1(y_1) \times R_3(y_3)\frac{T_{24}(y)}{N_{24}}\mathbb{1}_{y_2=y_4=y}, \\
Q_{2,14}(y_1,y_2,y_3,y_4) &\triangleq R_1(y_1) \times R_4(y_4)\frac{T_{23}(y)}{N_{23}}\mathbb{1}_{y_2=y_3=y}, \\
Q_{2,23}(y_1,y_2,y_3,y_4) &\triangleq R_2(y_2) \times R_3(y_3)\frac{T_{14}(y)}{N_{14}}\mathbb{1}_{y_1=y_4=y}, \\
Q_{2,24}(y_1,y_2,y_3,y_4) &\triangleq R_2(y_2) \times R_4(y_4)\frac{T_{13}(y)}{N_{13}}\mathbb{1}_{y_1=y_3=y}, \\
Q_{2,34}(y_1,y_2,y_3,y_4) &\triangleq R_3(y_3) \times R_4(y_4)\frac{T_{12}(y)}{N_{12}}\mathbb{1}_{y_1=y_2=y},
\end{aligned}
$$

Define the coupling distribution as

$$
\begin{aligned}
\mathbb{P}(y_1,y_2,y_3,y_4) &= \tau Q_0(y^4) + (\tau_{234}-\tau)Q_{1,1}(y^4)+ \\
&\quad (\tau_{134}-\tau)Q_{1,2}(y^4) + (\tau_{124}-\tau)Q_{1,3}(y^4) \\
&\quad + (\tau_{123}-\tau)Q_{1,4}(y^4) + \beta_{34}Q_{2,12}(y^4) + \beta_{24}Q_{2,13}(y^4) \\
&\quad + \beta_{23}Q_{2,14}(y^4) + \beta_{14}Q_{2,23}(y^4) + \beta_{13}Q_{2,24}(y^4)+ \\
&\quad \beta_{12}Q_{2,34}(y^4) + \alpha_{12}\frac{T_{12}(y)}{N_{12}}\frac{T_{34}(y')}{N_{34}}\mathbb{1}_{y_1=y_2=y,y_3=y_4=y'} \\
&\quad + \alpha_{13}\frac{T_{13}(y)}{N_{13}}\frac{T_{24}(y')}{N_{24}}\mathbb{1}_{y_1=y_3=y,y_2=y_4=y'} \\
&\quad + \alpha_{14}\frac{T_{14}(y)}{N_{14}}\frac{T_{23}(y')}{N_{23}}\mathbb{1}_{y_1=y_4=y,y_2=y_3=y'},
\end{aligned}
\tag{21}
$$

where α_{ij},β_{ij} are non-negative constants which will be found by solving a system of equations such that the coupling preserves the marginals and achieves the lower bound in (5). Observe that our coupling structure is such that for any $\mathcal{I}\subset[4]$ with cardinality $|\mathcal{I}|\geq 3$ and for all $y\in\mathcal{Y}$, we have

$$\mathbb{P}(\cap_{i\in\mathcal{I}}\{Y_i=y\}) = \min_{i\in\mathcal{I}} P_i(y). \tag{22}$$

Now, we impose the following constraints that for any $\mathcal{I}\subset[4]$ we have the same property for $|\mathcal{I}|=2$ and for all $y\in\mathcal{Y}$. This leads us to the following constraints:

$$
\begin{aligned}
\alpha_{12}+\beta_{12} &= N_{12}, \quad \alpha_{12}+\beta_{34} = N_{34}, \\
\alpha_{13}+\beta_{13} &= N_{13}, \quad \alpha_{13}+\beta_{24} = N_{24}, \\
\alpha_{14}+\beta_{14} &= N_{14}, \quad \alpha_{14}+\beta_{23} = N_{23}.
\end{aligned}
\tag{23}
$$

Now observe that if the coupling preserves the marginals then for all $y \in \mathcal{Y}$, we have by maximum-minimums rule $\mathbb{P}(\cup_{i\in[4]}\{Y_i = y\}) = \max\{P_1(y),P_2(y),P_3(y),P_4(y)\}$. Thus, we have the following constraints that yield the fact the coupling preserves marginals

$$
\begin{aligned}
\beta_{23}+\beta_{24}+\beta_{34}+\tau_{234}-\tau &= N_{R_1}, \tag{24} \\
\beta_{13}+\beta_{14}+\beta_{34}+\tau_{134}-\tau &= N_{R_2}, \tag{25} \\
\beta_{12}+\beta_{14}+\beta_{24}+\tau_{124}-\tau &= N_{R_3}, \tag{26} \\
\beta_{12}+\beta_{13}+\beta_{23}+\tau_{123}-\tau &= N_{R_4}. \tag{27}
\end{aligned}
$$

$$\tag{28}$$

where, N_{R_i} is the normalizing constant in the definition of R_i. Writing each of the constraints in terms of α_{ij}, we end up with the following single constraint.

$$\alpha_{12} + \alpha_{13} + \alpha_{14} = \tau_{\max_2} - 1$$

Thus, if we have constants $a, b, c \geq 0$ such that $a + b + c = 1$, and

$$\min\{N_{12}, N_{34}\} \geq a(\tau_{\max_2} - 1),$$
$$\min\{N_{13}, N_{24}\} \geq b(\tau_{\max_2} - 1),$$
$$\min\{N_{14}, N_{23}\} \geq c(\tau_{\max_2} - 1),$$

then we have found a construction of the coupling which preserves the marginals as well as achieves the lower bound in (5). The following condition ensures the existence of such constants a, b, c:

$$\min\{N_{12}, N_{34}\} + \min\{N_{13}, N_{24}\} + \min\{N_{14}, N_{23}\} \geq$$
$$\tau_{\max_2} - 1, \tag{29}$$

This completes the proof. $\qquad\square$