Electronic Colloquium on Computational Complexity, Report No. 341 (2026)

Wyner's Common Information and the Common Entropy are NP-Hard to compute

Roberto Bruno Ugo Vaccaro
Department of Computer Science, University of Salerno
84084 Fisciano (SA), Italy
Email:{rbruno,uvaccaro}@unisa.it

Abstract

Wyner's common information $J(P_{XY})$ and the common entropy $G(P_{XY})$ of a pair of finite random variables (X, Y) with joint distribution P_{XY} minimize, respectively, the mutual information $I(X, Y; W)$ and the entropy $H(W)$ over all auxiliary random variables W that make X and Y conditionally independent. We prove that both quantities are NP-hard to compute, by polynomial-time reductions from the Maximum Edge Biclique problem. Moreover, unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $J(P_{XY})$ within an additive error $1/(8N^2)$, or $G(P_{XY})$ within an additive error $1/(2N)$, where N is the number of positive entries of P_{XY}.

Keywords: common information; common entropy; conditional independence; NP-hardness; computational complexity.

I. Introduction

How much shared randomness is needed for two processors, acting independently once that randomness is given, to generate a prescribed joint distribution? This question leads to two related information measures. For a probability table P_{XY} of finite-valued random variables (X, Y), the first is the Wyner's common information [21] defined as

$$J(P_{XY}) = \inf_{\substack{W:\, X \text{ and } Y \text{ are conditionally} \\ \text{independent given } W}} I(X, Y; W), \tag{1}$$

and the second is the common entropy

$$G(P_{XY}) = \inf_{\substack{W:\, X \text{ and } Y \text{ are conditionally} \\ \text{independent given } W}} H(W), \tag{2}$$

ISSN 1433-8092

introduced by Kumar, Li, and El Gamal [13]. In both infima, the pair of random variables (X, Y) has distribution P_{XY}, the auxiliary random variable W takes finitely many values and makes X and Y conditionally independent.

Wyner gave two interpretations of $J(P_{XY})$. In the Gray–Wyner source-coding network [6], $J(P_{XY})$ is the minimum rate of the message sent to both decoders when the total coding rate approaches $H(X, Y)$ and the reconstruction errors vanish. It also characterizes the asymptotic rate of shared randomness needed by two separate terminals to generate independent copies of (X, Y) with arbitrarily small error. Related generation and synthesis problems are surveyed in [11], [16], [25]. The common entropy $G(P_{XY})$ [13] concerns exact generation of a single pair (X, Y), where the prescribed distribution must be reproduced without error.

These measures also have applications beyond source coding. Kocaoglu *et al*. [12] used common entropy in causal inference, while Jiang and Kocaoglu [10] developed a conditional version for instrumental variable testing and partial identification. Common-information representations also arise in multiview learning and component analysis [9], [17]. A second connection is to nonnegative matrix factorization. If a probability table has nonnegative rank s, a factorization into s nonnegative rank-one matrices supplies an admissible auxiliary with at most s values. Consequently,

$$J(P_{XY}) \le G(P_{XY}) \le \log \operatorname{rank}_+(P_{XY}). \tag{3}$$

Braun and Pokutta [2] used Wyner's common information to obtain lower bounds related to extension complexity, whose connection to nonnegative rank goes back to Yannakakis [22]. Braun, Jain, Lee, and Pokutta [1] characterized J through the asymptotic behavior of nonnegative rank under tensor powers and small perturbations. Earlier variational bounds for J were obtained by Witsenhausen [20]. Other results related to the computation of the Wyner's common information are contained in [8], [18].

A. Our results

In this paper, we prove that computing Wyner's common information and the common entropy is NP-hard. More precisely, we show that the problems of deciding whether $J(P_{XY})$ or $G(P_{XY})$ is at most a given rational threshold are NP-hard, and that both quantities remain hard to approximate within an inverse-polynomial additive error.

All the results are proved by polynomial-time reductions from the MAXIMUM EDGE BICLIQUE problem, which is NP-complete [14], and the two reductions share a common construction, which we describe in Section III. The starting point is a simple observation. An admissible auxiliary W decomposes the table P_{XY} into a sum of matrices, one for each value of W, and, since X and Y are conditionally

independent given W, the positive entries of each matrix form a rectangle, that is, a set of the form $R \times C = \{(x, y) : x \in R,\ y \in C\}$, for some set R of rows and some set C of columns (Section III-A). Hence the admissible auxiliaries of P_{XY} depend on how large the rectangles of positive entries of P_{XY} can be. This naturally leads to a reduction from MAXIMUM EDGE BICLIQUE, which asks whether a bipartite graph has a biclique with many edges, that is, whether its biadjacency matrix has a large rectangle of ones.

The use of large rectangles, or equivalently bicliques, has precedents in the common-information literature. Witsenhausen's convex-geometric treatment of J [20], developed further in [1], studies decompositions into product distributions, whose supports are rectangles. Han and Kobayashi [7] characterized the asymptotic sizes of rectangular subsets of partially jointly typical sets through conditional-independence auxiliaries, thereby connecting large rectangles to Wyner's common information. Yu, Anantharam, and Chen [23] studied biclique rate regions of joint-type graphs and clarified the distinction between exact bicliques and the approximate rectangles arising in the earlier typicality formulation. For the common entropy G, the product-mixture formulation of Kumar, Li, and El Gamal [13], and the one-shot Rényi common-entropy characterizations of Yu and Tan [24], provide related structural connections to nonnegative factorizations and rectangular supports. These works motivated the combinatorial viewpoint adopted here. More precisely, we provide polynomial-time reductions from MAXIMUM EDGE BICLIQUE for the finite-table Shannon objectives J and G, with explicit thresholds and inverse-polynomial gaps.

Both reductions begin by transforming an instance of MAXIMUM EDGE BICLIQUE into a Boolean matrix A with exactly d ones in each row. The original instance is a yes-instance if and only if A contains a rectangle of ones with more than d entries (Section III-D). From A, we then construct a probability table P_{XY} whose rows have equal mass. We use $W = X$ as a reference auxiliary, which corresponds to the decomposition of the table P_{XY} into its rows. If the table P_{XY} has n rows and each has mass $1/n$, then the auxiliary $W = X$ satisfies

$$I(X, Y; W) = H(W) = H(X) = \log n.$$

The question is whether another admissible auxiliary can achieve a smaller value of the objective being minimized.

For Wyner's common information (Section IV), we obtain P_{XY} by dividing the Boolean matrix A by its total number N of ones. Thus every positive entry of P_{XY} equals $1/N$. If every rectangle of ones in A has at most d entries, then we show that the auxiliary $W = X$ is optimal and $J(P_{XY}) = \log n$. Otherwise, a rectangle with more than d entries allows us to construct an admissible auxiliary W such

that

$$I(X,Y;W) < \log n - \frac{1}{4N^2}.$$

This gap proves the NP-hardness of WYNER THRESHOLD and gives an explicit hardness bound for approximation: unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $J(P_{XY})$ within additive error $1/(8N^2)$ (Theorem 1).

For the common entropy (Section V), the preceding construction does not suffice. The entropy $H(W)$ depends only on the probabilities $\Pr(W = w)$. To guarantee a decrease in $H(W)$, our argument requires the new product component supported on a rectangle to have probability mass greater than e times the probability mass of a single row, where $e = \exp(1)$. We therefore modify P_{XY} in two ways. First, we replace each row of A by sufficiently many copies, increasing the probability mass of a large rectangle relative to that of an individual row. Second, for each row, we add a positive entry in a new column whose other entries are zero. This ensures that every admissible mixture of product distributions contains, for each row, at least one component supported entirely on that row. With these modifications, $W = X$ is optimal whenever A contains no rectangle of ones with more than d entries. A sufficiently large rectangle, however, allows us to construct an admissible auxiliary W satisfying

$$H(W) < H(X) - \frac{1}{N}.$$

This proves both the NP-hardness of COMMON ENTROPY THRESHOLD and the hardness of approximating $G(P_{XY})$: unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $G(P_{XY})$ within additive error $1/(2N)$ (Theorem 2).

Finally, in Section VI, we show that the hardness does not depend on the zero entries of the tables produced by the reductions, that is, the hard instances constitute a much larger class. To this end, we replace the zeros by a very small mass, and we show that this changes J and G only slightly. Hence the same thresholds still work, and both problems remain hard on strictly positive tables. Specifically, unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $J(P_{XY})$ within additive error $1/(16L^2)$, or $G(P_{XY})$ within additive error $8/L$, where L is the number of entries of P_{XY} (Theorem 3).

All our results hold in the standard bit model of computation, with rational input distributions and rational thresholds.

B. Organization of the paper

In Section II, we introduce the notation and formally define the decision problems we study. In Section III, we collect the common ingredients of both reductions. Namely, we show that admissible auxiliaries correspond to rank-one decompositions of the probability table, we introduce the auxiliary

random variable W_F, constructed by extracting a rank-one matrix F from the probability table and splitting the remainder into its rows, and we bring the instances of MAXIMUM EDGE BICLIQUE into a regular form. In Section IV, we prove the hardness of Wyner's common information. In Section V, we prove the hardness of the common entropy. In Section VI, we extend these results to strictly positive tables. Section VII contains some concluding remarks and open problems.

II. PRELIMINARIES

A. Notation

All alphabets throughout the paper are finite and nonempty. We denote by $\log$ the logarithm in base 2, by $\ln$ the natural logarithm. Let

$$\phi(t) = -t \log t \quad (t > 0), \qquad \phi(0) = 0, \tag{4}$$

and let $h_2(z) = \phi(z) + \phi(1-z)$, for $z \in [0,1]$, be the binary entropy function.

Let X be a random variable taking values in an alphabet $\mathcal{X}$. We denote by P_X its probability distribution, that is, $P_X(x) = \mathbb{P}(X = x)$ for each $x \in \mathcal{X}$, and by

$$H(X) = \sum_{x \in \mathcal{X}} \phi\big(P_X(x)\big) = -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x)$$

its Shannon entropy. Since $H(X)$ depends only on P_X, we also write $H(P_X)$, and, more generally, $H(Q) = \sum_s \phi(Q(s))$ for any probability distribution Q on a finite set. Moreover, for a probability distribution Q on a finite set S, we denote by $\operatorname{supp}(Q) = \{s \in S : Q(s) > 0\}$ its support.

Let (X, Y) be a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$, with joint probability distribution $P_{XY}(x,y) = \mathbb{P}(X = x, Y = y)$ and marginal distributions P_X and P_Y. We denote by $H(X,Y) = H(P_{XY})$ the joint entropy of (X,Y). For every $x \in \mathcal{X}$ with $P_X(x) > 0$, the conditional distribution of Y given $X = x$ is

$$P_{Y|X}(y \mid x) = \frac{P_{XY}(x,y)}{P_X(x)}, \tag{5}$$

and we write $H(Y \mid X = x) = H\big(P_{Y|X}(\cdot \mid x)\big)$ for its entropy. The conditional entropy of Y given X is the average of these entropies, that is,

$$H(Y \mid X) = \sum_{x \in \mathcal{X}:\, P_X(x) > 0} P_X(x)\, H(Y \mid X = x). \tag{6}$$

The mutual information between X and Y is

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y \mid X). \tag{7}$$

Let W be a third random variable, taking values in a finite set $\mathcal{W}$. All the conditional quantities given W are defined as above, by treating the pair (X,Y) as a single random variable taking values in $\mathcal{X} \times \mathcal{Y}$. In

particular, for every $w \in \mathcal{W}$ with $P_W(w) > 0$, we denote by $P_{XY|W}(\cdot, \cdot \mid w)$ the conditional distribution of (X, Y) given $W = w$, and by $P_{X|W}(\cdot \mid w)$ and $P_{Y|W}(\cdot \mid w)$ its two marginals, that is, the conditional distributions of X and of Y given $W = w$. Moreover, the mutual information between (X, Y) and W can be expressed as follows

$$I(X, Y; W) = H(X, Y) - H(X, Y \mid W) = H(W) - H(W \mid X, Y). \tag{8}$$

We say that the random variables X and Y are *conditionally independent given* W, written $X \perp\!\!\!\perp Y \mid W$, if, for every $w \in \mathcal{W}$ with $P_W(w) > 0$,

$$P_{XY|W}(x, y \mid w) = P_{X|W}(x \mid w)\, P_{Y|W}(y \mid w) \qquad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{9}$$

Throughout the paper, it will be convenient to regard joint distributions as matrices. Thus, we identify a joint distribution P_{XY} with the nonnegative matrix whose rows are indexed by $\mathcal{X}$ and whose columns are indexed by $\mathcal{Y}$, and we call it a *probability table*. Moreover, the row x and column y of P_{XY} are the vectors $P_{XY}(x, \cdot)$ and $P_{XY}(\cdot, y)$, respectively. The *support* of P_{XY} is the set $\operatorname{supp}(P_{XY}) = \{(x, y) \in \mathcal{X} \times \mathcal{Y} : P_{XY}(x, y) > 0\}$ of its positive entries. Two size parameters of a table will be used throughout:

$$N = |\operatorname{supp}(P_{XY})|, \qquad L = |\mathcal{X}||\mathcal{Y}|. \tag{10}$$

Here, N counts the positive entries of P_{XY} and L counts all the entries, including zeros.

For a nonnegative matrix F indexed by $\mathcal{X} \times \mathcal{Y}$, we let $\mu(F) = \sum_{x,y} F(x, y)$ be its *mass* and $\operatorname{supp}(F) = \{(x, y) : F(x, y) > 0\}$ its support. For two such matrices F and F', we write $F \leq F'$ if $F(x, y) \leq F'(x, y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Given probability distributions Q on $\mathcal{X}$ and Q' on $\mathcal{Y}$, we denote by $Q \otimes Q'$ the product distribution on $\mathcal{X} \times \mathcal{Y}$, defined by $(Q \otimes Q')(x, y) = Q(x)Q'(y)$, and we recall that $H(Q \otimes Q') = H(Q) + H(Q')$. For a set S, we denote by $\mathbb{1}_S$ its indicator function.

Finally, a *rectangle* is a set of positions of the form $R \times C \subseteq \mathcal{X} \times \mathcal{Y}$, with R and C nonempty, and its *size* is $|R||C|$. For a nonnegative matrix F indexed by $\mathcal{X} \times \mathcal{Y}$, we say that $R \times C$ is a *rectangle of* F if $R \times C \subseteq \operatorname{supp}(F)$, that is, if $F(x, y) > 0$ for all $x \in R$ and $y \in C$. Moreover, we denote by

$$\beta(F) = \max\big\{|R||C| : \ R \times C \text{ is a rectangle of } F\big\} \tag{11}$$

the largest size of a rectangle of F with the convention that $\beta(F) = 0$ if $\operatorname{supp}(F) = \emptyset$, in which case F has no rectangles. In particular, since P_{XY} is a nonnegative matrix, $\beta(P_{XY})$ is the largest size of a rectangle of the probability table P_{XY}.

B. Auxiliaries, Wyner's common information, and the common entropy

Both Wyner's common information and the common entropy are defined as infima over auxiliary random variables that make X and Y conditionally independent. We formalize an auxiliary random variable as follows.

Definition 1. Let (X, Y) be a pair of random variables with joint distribution P_{XY}, and let $\mathcal{W}$ be a finite set. An *auxiliary* of (X, Y) is a random variable W taking values in $\mathcal{W}$ such that the joint distribution P_{XYW} of (X, Y, W) on $\mathcal{X} \times \mathcal{Y} \times \mathcal{W}$ satisfies

$$\sum_{w \in \mathcal{W}} P_{XYW}(x, y, w) = P_{XY}(x, y) \qquad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{12}$$

An auxiliary W is *admissible* if $X \perp\!\!\!\perp Y \mid W$.

Without loss of generality, we assume that $P_W(w) > 0$ for every $w \in \mathcal{W}$, since values of probability zero can be omitted. Then, *Wyner's common information* of P_{XY} is defined as

$$J(P_{XY}) = \inf_{W \text{ admissible}} I(X, Y; W), \tag{13}$$

and the *common entropy* of P_{XY} as

$$G(P_{XY}) = \inf_{W \text{ admissible}} H(W). \tag{14}$$

The auxiliaries $W = X$ and $W = Y$ are always admissible. Indeed, given $W = x$, the random variable X is constant, and a constant random variable is independent of every random variable; the same holds for $W = Y$. Moreover, every admissible W satisfies $I(X, Y; W) = H(W) - H(W \mid X, Y) \leq H(W)$. Taking infima over all admissible auxiliaries W, choosing $W = X$ and $W = Y$ in (14), and recalling that the entropy of a random variable is at most the logarithm of the size of its alphabet, we obtain the bounds

$$0 \leq J(P_{XY}) \leq G(P_{XY}) \leq \min\{H(X), H(Y)\} \leq \log(\min\{|\mathcal{X}|, |\mathcal{Y}|\}). \tag{15}$$

C. Decision problems

In this section, we state the decision versions of the optimization problems (13) and (14), which we use to prove NP-hardness[1]. Throughout, a probability distribution (or table) is called *rational* if all its probabilities are rational numbers.

WYNER THRESHOLD

[1]We refer to [5] for the basics about NP-hardness.

INPUT: A rational probability table P_{XY} on $\mathcal{X} \times \mathcal{Y}$ and a rational number θ.

QUESTION: Does there exist an admissible auxiliary W of P_{XY} such that $I(X, Y; W) \leq \theta$?

COMMON ENTROPY THRESHOLD

INPUT: A rational probability table P_{XY} on $\mathcal{X} \times \mathcal{Y}$ and a rational number θ.

QUESTION: Does there exist an admissible auxiliary W of P_{XY} such that $H(W) \leq \theta$?

In other words, WYNER THRESHOLD and COMMON ENTROPY THRESHOLD compare $J(P_{XY})$ and $G(P_{XY})$, defined in (13) and (14), with the threshold θ. Indeed, since both infima are attained (see [13], [21]), the two problems above ask whether $J(P_{XY}) \leq \theta$ and whether $G(P_{XY}) \leq \theta$, respectively.

We prove the NP-hardness of these problems by reductions from the MAXIMUM EDGE BICLIQUE problem [14], which we now recall. A *bipartite graph* Γ is a triple (U, V, E_Γ), where U and V are two disjoint finite sets, whose elements are called left and right vertices, respectively, and $E_\Gamma \subseteq U \times V$ is a set of pairs (x, y), called edges, each joining a left vertex x to a right vertex y. The *degree* $\deg_\Gamma(v)$ of a vertex v is the number of edges incident to v, and a vertex of degree zero is called *isolated*. A *biclique* of Γ is a set $R \times C \subseteq E_\Gamma$, with $\emptyset \neq R \subseteq U$ and $\emptyset \neq C \subseteq V$; that is, every left vertex in R is joined to every right vertex in C. A biclique $R \times C$ has $|R||C|$ edges.

It is convenient to describe the bipartite graph Γ by its *biadjacency matrix* A, that is, the Boolean matrix with rows indexed by U, columns indexed by V, and $A(x, y) = 1$ if and only if $(x, y) \in E_\Gamma$. In particular, the number of ones in row x of A is $\deg_\Gamma(x)$. Since $\operatorname{supp}(A) = E_\Gamma$, the bicliques of Γ correspond exactly to the rectangles of A. Thus, by the definition of $\beta(\cdot)$ in (11), the largest number of edges of a biclique of Γ is $\beta(A)$, and we also denote it by $\beta(\Gamma) = \beta(A)$. The problem from which we reduce is the MAXIMUM EDGE BICLIQUE defined below.

MAXIMUM EDGE BICLIQUE

INPUT: A bipartite graph Γ, given by its biadjacency matrix A, and a positive integer k.

QUESTION: Is $\beta(\Gamma) \geq k$, that is, does Γ contain a biclique with at least k edges?

It is well known that MAXIMUM EDGE BICLIQUE is NP-complete [14].

III. COMMON INGREDIENTS OF THE REDUCTIONS

In this section, we collect the arguments shared by the reductions for WYNER THRESHOLD in Section IV and for COMMON ENTROPY THRESHOLD in Section V. The section is organized as follows. In Section III-A, we show that admissible auxiliaries are equivalent to decompositions of the probability table P_{XY} into nonnegative matrices of rank one, and that the support of each of these matrices is a rectangle of P_{XY}. In Section III-B, we use this correspondence to derive a lower bound on Wyner's common information in terms of $\beta(P_{XY})$, that is, the largest size of a rectangle of P_{XY}. In Section III-C, we

construct an admissible auxiliary by extracting a single rank-one component from P_{XY}, and we compute its entropy and its mutual information with (X, Y). Finally, in Section III-D, we show how to transform, in polynomial time and without changing its answer, any instance of MAXIMUM EDGE BICLIQUE into a regular form, which is the starting point of both reductions.

A. Auxiliaries as rank-one decompositions

We now describe admissible auxiliaries in terms of matrices. To this end, we define the components of an auxiliary and recall the notion of rank-one decomposition of P_{XY}.

Definition 2. Let W be an auxiliary of (X, Y) taking values in $\mathcal{W}$. For every $w \in \mathcal{W}$, the *component* of W at w is the nonnegative matrix F_w indexed by $\mathcal{X} \times \mathcal{Y}$ defined by

$$F_w(x, y) = P_{XYW}(x, y, w), \qquad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{16}$$

We notice that the component F_w is the part of the table P_{XYW} produced when $W = w$. By (12), the components of W sum to P_{XY}; moreover, the mass of F_w is $\mu(F_w) = \sum_{(x,y)} F_w(x, y) = P_W(w)$, and normalizing F_w by its mass gives the conditional distribution of the pair (X, Y) given $W = w$, that is,

$$\frac{F_w}{\mu(F_w)} = P_{XY|W}(\cdot, \cdot \mid w). \tag{17}$$

We recall that a matrix F indexed by $\mathcal{X} \times \mathcal{Y}$ has *rank one* if $F \neq 0$ and there exist vectors $u \in \mathbb{R}^{\mathcal{X}}$ and $v \in \mathbb{R}^{\mathcal{Y}}$ such that

$$F(x, y) = u(x)\, v(y) \qquad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \tag{18}$$

that is, $F = uv^{\top}$, where $v^{\top}$ denotes the transpose of v. A *rank-one decomposition* of a nonnegative matrix M indexed by $\mathcal{X} \times \mathcal{Y}$ is a finite family $(F_i)_{i \in \mathcal{I}}$ of nonzero nonnegative matrices of rank one such that

$$\sum_{i \in \mathcal{I}} F_i = M. \tag{19}$$

The next lemma shows that a nonzero nonnegative matrix has rank one if and only if it is a positive multiple of a product distribution, and that, consequently, its support is a rectangle. We notice that some of the main steps of the following proof also appear in the proofs of Theorems 3.1 and 4.1 of [3].

Lemma 1. *A nonzero nonnegative matrix F indexed by $\mathcal{X} \times \mathcal{Y}$ has rank one if and only if*

$$F = \mu(F)\,(Q \otimes Q') \tag{20}$$

for some probability distributions Q on $\mathcal{X}$ and Q' on $\mathcal{Y}$. In this case, $\operatorname{supp}(F) = \operatorname{supp}(Q) \times \operatorname{supp}(Q')$ *is a rectangle.*

Proof. If F has the form (20), then $F(x,y) = u(x)v(y)$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$, with $u = \mu(F)Q$ and $v = Q'$. Hence, by (18), F has rank one.

Conversely, assume that F has rank one. By (18), $F(x,y) = u(x)v(y)$ for some $u \in \mathbb{R}^{\mathcal{X}}$ and $v \in \mathbb{R}^{\mathcal{Y}}$. We first show that u and v can be chosen to be nonnegative. Since $F \neq 0$, there exists an entry (x_0, y_0) such that

$$F(x_0, y_0) = u(x_0)\, v(y_0) > 0, \tag{21}$$

so that $u(x_0)$ and $v(y_0)$ are nonzero and have the same sign. Since replacing u and v by $-u$ and $-v$ does not change their product, we may assume that $u(x_0) > 0$ and $v(y_0) > 0$. Now, since F is nonnegative, for every $x \in \mathcal{X}$ it holds that $u(x)\, v(y_0) = F(x, y_0) \geq 0$, which, together with $v(y_0) > 0$, gives $u(x) \geq 0$. In the same way, for every $y \in \mathcal{Y}$ it holds that $u(x_0)\, v(y) = F(x_0, y) \geq 0$, which, together with $u(x_0) > 0$, gives $v(y) \geq 0$. Hence, u and v can be chosen to be nonnegative.

We now normalize u and v. Let $s_u = \sum_{x \in \mathcal{X}} u(x)$ and $s_v = \sum_{y \in \mathcal{Y}} v(y)$. Since u and v are nonnegative, and $u(x_0) > 0$ and $v(y_0) > 0$, both s_u and s_v are positive, and therefore $Q = u/s_u$ and $Q' = v/s_v$ are probability distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively. By construction, $F(x,y) = s_u s_v\, Q(x)Q'(y)$ for all (x,y), that is, $F = s_u s_v\, (Q \otimes Q')$. Summing this identity over all the entries, and recalling that Q and Q' sum to one, we obtain that $\mu(F) = s_u s_v$, which proves (20). Finally, by (20), every entry of F is the product $F(x,y) = \mu(F)\, Q(x)\, Q'(y)$, where $\mu(F) > 0$ and $Q(x), Q'(y) \geq 0$. Hence $F(x,y) > 0$ if and only if $Q(x) > 0$ and $Q'(y) > 0$, that is, $\operatorname{supp}(F) = \operatorname{supp}(Q) \times \operatorname{supp}(Q')$. Since Q and Q' are probability distributions, $\operatorname{supp}(Q)$ and $\operatorname{supp}(Q')$ are nonempty, and therefore $\operatorname{supp}(F)$ is a rectangle. □

We now show that the components of an admissible auxiliary form a rank-one decomposition of P_{XY}, and that, conversely, every rank-one decomposition of P_{XY} is obtained from some admissible auxiliary. This correspondence is a well known result, see [3, Section 6]. We give a proof for completeness in Appendix A.

Lemma 2. *Let P_{XY} be a probability table. Then the following properties hold:*

1) *If W is an admissible auxiliary of $(X,Y) \sim P_{XY}$, then its components $(F_w)_{w \in \mathcal{W}}$ form a rank-one decomposition of P_{XY}. Moreover, for every $w \in \mathcal{W}$, the component F_w satisfies $F_w \leq P_{XY}$, its support is a rectangle of P_{XY}, and*

$$F_w = P_W(w)\, \big(P_{X|W}(\cdot \mid w) \otimes P_{Y|W}(\cdot \mid w)\big). \tag{22}$$

2) *Conversely, if $(F_w)_{w \in \mathcal{W}}$ is a rank-one decomposition of P_{XY}, indexed by a finite set $\mathcal{W}$, then setting*

$$P_{XYW}(x,y,w) = F_w(x,y), \qquad (x,y,w) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{W}, \tag{23}$$

defines an admissible auxiliary W of (X,Y) taking values in $\mathcal{W}$, with $P_W(w) = \mu(F_w) > 0$ for every $w \in \mathcal{W}$.

3) For every admissible auxiliary W with components $(F_w)_{w\in\mathcal{W}}$, it holds that

$$H(W) = \sum_{w\in\mathcal{W}} \phi\big(\mu(F_w)\big), \tag{24}$$

$$H(X,Y \mid W) = \sum_{w\in\mathcal{W}} \mu(F_w)\, H\left(\frac{F_w}{\mu(F_w)}\right) = \sum_{w\in\mathcal{W}} \mu(F_w)\Big[H\big(P_{X|W}(\cdot \mid w)\big) + H\big(P_{Y|W}(\cdot \mid w)\big)\Big], \tag{25}$$

$$I(X,Y;W) = H(X,Y) - \sum_{w\in\mathcal{W}} \mu(F_w)\, H\left(\frac{F_w}{\mu(F_w)}\right). \tag{26}$$

By property 2 of Lemma 2, every rank-one decomposition of P_{XY} defines an admissible auxiliary. Since a nonzero matrix with a single nonzero row, or column, has rank one, splitting P_{XY} into its nonzero rows, each viewed as a matrix vanishing outside that row, gives a rank-one decomposition, corresponding to $W = X$. Similarly, splitting P_{XY} into its nonzero columns corresponds to $W = Y$. More generally, if $F_1, \ldots, F_s$ are nonzero nonnegative matrices of rank one with $\sum_i F_i \le P_{XY}$, then $F_1, \ldots, F_s$, together with the nonzero rows (or columns) of $P_{XY} - \sum_i F_i$, form a rank-one decomposition of P_{XY}, which we say is obtained by *extracting* $F_1, \ldots, F_s$ from P_{XY}.

B. A lower bound from the largest rectangle

By Lemma 2, the support of every component of an admissible auxiliary is a rectangle of P_{XY}, and hence it contains at most $\beta(P_{XY})$ entries. Therefore, the conditional distribution of $(X,Y) \sim P_{XY}$ given any value of an admissible auxiliary has entropy at most $\log \beta(P_{XY})$. The next lemma shows that this yields a lower bound on Wyner's common information.

Lemma 3. *Every admissible auxiliary W of $(X,Y) \sim P_{XY}$ satisfies $H(X,Y \mid W) \le \log \beta(P_{XY})$. Consequently, it holds that*

$$J(P_{XY}) \ge H(X,Y) - \log \beta(P_{XY}). \tag{27}$$

Proof. Let W be an admissible auxiliary of (X,Y) with components $(F_w)_{w\in\mathcal{W}}$. By property 1 of Lemma 2, for every $w \in \mathcal{W}$ the support of F_w is a rectangle of P_{XY}, so that $|\operatorname{supp}(F_w)| \le \beta(P_{XY})$. Moreover, by (17), it holds that $P_{XY|W}(\cdot,\cdot \mid w) = F_w/\mu(F_w)$, which is a probability distribution supported on $\operatorname{supp}(F_w)$. Thus, since the entropy of a distribution is at most the logarithm of the size of its support, we have that

$$H(X,Y \mid W = w) = H\left(\frac{F_w}{\mu(F_w)}\right) \le \log|\operatorname{supp}(F_w)| \le \log \beta(P_{XY}) \qquad \forall w \in \mathcal{W}. \tag{28}$$

Averaging (28) with weights $P_W(w) = \mu(F_w)$, we obtain that $H(X, Y \mid W) \leq \log \beta(P_{XY})$. Substituting this bound into (26), we get that $I(X, Y; W) \geq H(X, Y) - \log \beta(P_{XY})$, and taking the infimum over all admissible W proves (27). □

C. Extracting one component

In this subsection, we show how a rank-one matrix contained in a probability table, and hence supported on a rectangle of the table, can be used to construct admissible auxiliaries. Specifically, in both reductions, we use this construction to improve on the auxiliary $W = X$. The idea of the construction is the following. We start from the auxiliary $W = X$, which corresponds to the decomposition of P_{XY} into its rows and satisfies $H(W) = I(X, Y; W) = H(X)$. We then remove from P_{XY} a nonnegative matrix F of rank one, which may take mass from several rows, and split the remainder $P_{XY} - F$ into its rows. This yields a new admissible auxiliary W_F, whose components are F and the nonzero rows of $P_{XY} - F$. Moreover, when all the rows of P_{XY} have the same mass, as in both reductions, we give conditions on F under which the auxiliary W_F improves on $W = X$.

We now describe formally this construction. Let P_{XY} be a probability table with $n = |\mathcal{X}|$ rows, each of mass $\omega = 1/n$, so that $H(X) = \log n$. Let $F \leq P_{XY}$ be a nonzero nonnegative matrix of rank one, and, for every $x \in \mathcal{X}$, let $s_x = \sum_{y \in \mathcal{Y}} F(x, y)$ be the mass that F takes from row x, so that

$$\sum_{x \in \mathcal{X}} s_x = \mu(F), \qquad 0 \leq s_x \leq \omega. \tag{29}$$

Since F has rank one and every nonzero row of $P_{XY} - F$ has rank one, F together with the nonzero rows of $P_{XY} - F$ form a rank-one decomposition of P_{XY}. We denote by W_F the corresponding admissible auxiliary, given by property 2 of Lemma 2. The auxiliary W_F takes a new value $\star \notin \mathcal{X}$, whose component is F, and the value x, whose component is row x of $P_{XY} - F$. More formally,

$$P_{XYW_F}(x, y, \star) = F(x, y), \qquad P_{XYW_F}(x, y, x') = \big(P_{XY}(x, y) - F(x, y)\big)\mathbb{1}_{\{x'=x\}}, \quad x' \in \mathcal{X}. \tag{30}$$

In other words, having observed $(X, Y) = (x, y)$, the auxiliary W_F equals $\star$ with probability $F(x, y)/P_{XY}(x, y)$, and x otherwise. In particular, $W_F = X$ whenever $(X, Y) \notin \operatorname{supp}(F)$, while, if $(X, Y) = (x, y) \in \operatorname{supp}(F)$, then $W_F = \star$ with probability $F(x, y)/P_{XY}(x, y)$, and $W_F = x$ otherwise.

The next lemma computes the entropy of the auxiliary W_F defined in (30) and its mutual information with (X, Y), in terms of the mass $\mu(F)$ of F and of the masses s_x that F takes from the rows of P_{XY}. Moreover, it also bounds the entropy of W_F in terms of $\mu(F)$ and of the mass ω of a row. We recall from (4) that $\phi(t) = -t \log t$ for $t > 0$.

Lemma 4. *Let P_{XY} be a probability table with $n = |\mathcal{X}|$ rows, each of mass $\omega = 1/n$, let $F \leq P_{XY}$ be a nonzero nonnegative matrix of rank one, and let W_F be the admissible auxiliary of (X, Y) defined in* (30)*. Then the following properties hold:*

1) The entropy of W_F is

$$H(W_F) = \phi\big(\mu(F)\big) + \sum_{x \in \mathcal{X}} \phi(\omega - s_x). \tag{31}$$

2) The mutual information between (X, Y) and W_F is

$$I(X, Y; W_F) = H(W_F) - \sum_{(x,y) \in \mathrm{supp}(P_{XY})} P_{XY}(x, y)\, h_2\!\left(\frac{F(x, y)}{P_{XY}(x, y)}\right). \tag{32}$$

3) The entropy of W_F satisfies

$$H(W_F) \leq \log n + \mu(F) \log \frac{e\omega}{\mu(F)} \leq \log n - \big(\mu(F) - e\omega\big) \log e. \tag{33}$$

Proof. We first prove property 1. By construction, the components of W_F are F, with mass $\mu(F)$, and, for every $x \in \mathcal{X}$ with $s_x = \sum_{y \in \mathcal{Y}} F(x, y) < \omega$, row x of $P_{XY} - F$, with mass $\omega - s_x$. Hence (31) follows from (24), where the rows with $s_x = \omega$ contribute $\phi(0) = 0$.

We now prove property 2. By (8), it holds that $I(X, Y; W_F) = H(W_F) - H(W_F \mid X, Y)$. Fix $(x, y) \in \mathrm{supp}(P_{XY})$. By (30), given $(X, Y) = (x, y)$, the auxiliary W_F equals $\star$ with probability $F(x, y)/P_{XY}(x, y)$, and x otherwise. Hence $H(W_F \mid X = x, Y = y) = h_2\big(F(x, y)/P_{XY}(x, y)\big)$, and averaging over $(x, y) \in \mathrm{supp}(P_{XY})$ with weights $P_{XY}(x, y)$ gives (32).

Finally, we prove property 3. Since ϕ is concave on $[0, \infty)$ and differentiable at $\omega > 0$, it lies below its tangent at ω, that is, $\phi(t) \leq \phi(\omega) + \phi'(\omega)(t - \omega)$ for every $t \geq 0$. Applying this inequality with $t = \omega - s_x$, and recalling that $\phi'(\omega) = -\log(e\omega)$, we obtain that, for every $x \in \mathcal{X}$,

$$\phi(\omega - s_x) \leq \phi(\omega) - s_x\, \phi'(\omega) = \phi(\omega) + s_x \log(e\omega). \tag{34}$$

Summing (34) over the n rows, and using (29) and $n\, \phi(\omega) = n \cdot \frac{1}{n} \log n = \log n$, we obtain that

$$\sum_{x \in \mathcal{X}} \phi(\omega - s_x) \leq \log n + \mu(F) \log(e\omega). \tag{35}$$

Combining (31) and (35) with $\phi(\mu(F)) = \mu(F) \log \frac{1}{\mu(F)}$, we get the leftmost inequality in (33). For the rightmost inequality, it suffices to show that $\mu(F) \log \frac{e\omega}{\mu(F)} \leq -\big(\mu(F) - e\omega\big) \log e$, that is,

$$\mu(F) \log \frac{\mu(F)}{e\omega} \geq \big(\mu(F) - e\omega\big) \log e. \tag{36}$$

To this end, we use the inequality $\ln v \geq 1 - 1/v$, which holds for every $v > 0$. Evaluating it at $v = \mu(F)/(e\omega)$, we obtain that

$$\ln \frac{\mu(F)}{e\omega} \geq 1 - \frac{e\omega}{\mu(F)}.$$

Multiplying both sides by $\mu(F)/\ln 2$, and recalling that $1/\ln 2 = \log e$, we obtain that

$$\mu(F)\log\frac{\mu(F)}{e\omega} \geq \mu(F)\Big(1-\frac{e\omega}{\mu(F)}\Big)\log e = \big(\mu(F)-e\omega\big)\log e,$$

which proves (36), and consequently the rightmost inequality in (33). □

One can see that, by (33), the auxiliary W_F satisfies $H(W_F) < \log n = H(X)$ whenever $\mu(F) > e\omega$, that is, whenever the mass of F exceeds e times the mass of a row. This is the condition that we will use in the reduction for COMMON ENTROPY THRESHOLD. In the reduction for WYNER THRESHOLD, instead, (32) gives an additional gain, since every entry (x, y) with $0 < F(x, y) < P_{XY}(x, y)$ contributes a positive term to the sum in (32). The next lemma makes this gain explicit when the probability table P_{XY} is constant on a rectangle $R \times C$ and F takes a fraction t of the mass of every entry of $R \times C$.

Lemma 5. *Let P_{XY} be a probability table with $n = |\mathcal{X}|$ rows, each of mass $\omega = 1/n$, and let $R \times C$ be a rectangle of P_{XY} such that $P_{XY}(x, y) = c$ for every $(x, y) \in R \times C$, for some $c > 0$. Let $m = \sum_{(x,y)\in R\times C} P_{XY}(x, y) = c\,|R||C|$ be the mass of P_{XY} on $R \times C$. For $0 < t \leq 1$, let F_t be the matrix defined by $F_t(x, y) = tc$ for $(x, y) \in R \times C$, and $F_t(x, y) = 0$ otherwise. Then the admissible auxiliary W_{F_t} obtained from F_t, according to* (30)*, satisfies the inequality*

$$I(X, Y; W_{F_t}) \leq \log n - tm\big(\log(nm) - t\log e\big). \tag{37}$$

Proof. We first check that F_t satisfies the assumptions of Lemma 4. Since $F_t(x, y) = \big(tc\,\mathbb{1}_R(x)\big)\mathbb{1}_C(y)$ for every $(x, y) \in \mathcal{X}\times\mathcal{Y}$, the matrix F_t is nonnegative and nonzero, and it has rank one by (18). Moreover, $F_t \leq P_{XY}$, since $F_t = tc \leq c = P_{XY}$ on $R \times C$ and $F_t = 0$ elsewhere. Therefore, by property 2 of Lemma 4, we have that

$$I(X, Y; W_{F_t}) = H(W_{F_t}) - \sum_{(x,y)\in\mathrm{supp}(P_{XY})} P_{XY}(x, y)\, h_2\left(\frac{F_t(x, y)}{P_{XY}(x, y)}\right). \tag{38}$$

We now compute the sum in the right-hand side of (38). The ratio F_t/P_{XY} equals t on $R \times C$ and 0 elsewhere on $\mathrm{supp}(P_{XY})$. Thus, since $h_2(0) = 0$, it follows that

$$\sum_{(x,y)\in\mathrm{supp}(P_{XY})} P_{XY}(x, y)\, h_2\left(\frac{F_t(x, y)}{P_{XY}(x, y)}\right) = \sum_{(x,y)\in R\times C} c\, h_2(t) = m\, h_2(t). \tag{39}$$

Substituting (39) into (38), and bounding $H(W_{F_t})$ by the first inequality in (33), with $\mu(F_t) = tc\,|R||C| = tm$ and $\omega = 1/n$, we obtain that

$$I(X, Y; W_{F_t}) \leq \log n + tm\log\frac{e}{ntm} - m\, h_2(t). \tag{40}$$

We now bound $h_2(t)$ from below. For $0 < t < 1$, the inequality $\ln \frac{1}{1-t} \geq t$ gives $(1-t)\log \frac{1}{1-t} \geq t(1-t)\log e$, and hence

$$h_2(t) = t \log \frac{1}{t} + (1-t) \log \frac{1}{1-t} \geq t \log \frac{1}{t} + t(1-t) \log e. \tag{41}$$

Inequality (41) also holds for $t = 1$, since both sides vanish. Substituting (41) into (40), and expanding the logarithm, we obtain that

$$\begin{aligned} I(X, Y; W_{F_t}) &\leq \log n + tm \log \frac{e}{ntm} - tm \log \frac{1}{t} - t(1-t)m \log e \\ &= \log n + tm \log e - tm \log(nm) + tm \log \frac{1}{t} - tm \log \frac{1}{t} - t(1-t)m \log e \\ &= \log n - tm \log(nm) + t^2 m \log e \\ &= \log n - tm\big(\log(nm) - t \log e\big), \end{aligned} \tag{42}$$

which proves (37). □

Lemmas 4 and 5 give two different conditions for improving on the auxiliary $W = X$, whose entropy and mutual information are both equal to $\log n$. On the one hand, the right-hand side of (37) is smaller than $\log n$ if and only if $\log(nm) > t \log e$, that is, $t < \ln(nm)$. Hence, if the mass m of the rectangle is larger than the mass $\omega = 1/n$ of a row, that is, $nm > 1$, then $I(X, Y; W_{F_t}) < \log n$ for every $0 < t < \min\{1, \ln(nm)\}$. On the other hand, as observed above, $H(W_F) < \log n$ is guaranteed when F is a rank-one matrix with $F \leq P_{XY}$ and $\mu(F) > e\omega$, that is, when the mass of F exceeds e times the mass of a row. This is the reason why the reductions for WYNER THRESHOLD and COMMON ENTROPY THRESHOLD use different tables.

D. Regularization of the source instances

Both reductions start from an instance (Γ, k) of MAXIMUM EDGE BICLIQUE brought into a regular form. Specifically, from the biadjacency matrix of Γ we construct a Boolean matrix A without zero columns, with n rows, where n is a power of two, such that every row of A contains the same number d of ones. In both reductions, the probability table is then constructed from A. The properties of A ensure that all the rows of the constructed probability table have the same mass, and that the number of rows is a power of two, so that the entropy $H(X)$ of the auxiliary $W = X$ is an integer and the thresholds used in the reductions are rational. Moreover, since every row of A is a rectangle of size d, the question of MAXIMUM EDGE BICLIQUE reduces to whether A has a rectangle of size larger than d. The following lemma summarizes these properties and shows that such a matrix A can be constructed in polynomial time.

Lemma 6. *From an instance (Γ, k) of* MAXIMUM EDGE BICLIQUE*, one can construct, in polynomial time, a Boolean matrix A with n rows and m columns, none of which is zero, and an integer $d \geq 1$, such that the following properties hold:*

1) *n is a power of two, and every row of A contains exactly d ones.*
2) *the size $\beta(\Gamma)$ of the largest biclique of Γ is greater or equal to k if and only if the size $\beta(A)$ of the largest rectangle of A is strictly greater than d.*

Proof. Let $\Gamma = (U, V, E_\Gamma)$. We consider three cases.

Case 1: $k > |E_\Gamma|$. Since every biclique of Γ is a set of edges, it has at most $|E_\Gamma|$ edges, so that $\beta(\Gamma) \leq |E_\Gamma| < k$. Hence, (Γ, k) is a no-instance of MAXIMUM EDGE BICLIQUE. In this case, we output $d = 1$ and the 2×2 identity matrix A, which satisfies property 1. Moreover, every rectangle of A with two rows or two columns contains a zero entry, so that the only rectangles of A are its single positive entries, and $\beta(A) = 1 = d$. Therefore, both $\beta(\Gamma) \geq k$ and $\beta(A) > d$ are false, and property 2 is satisfied.

Case 2: $k \leq |E_\Gamma|$, and some vertex has degree at least k. If this vertex is a left vertex x, then $\{x\} \times C$, where C is the set of neighbours of x, is a biclique with at least k edges. Similarly, if it is a right vertex y, then $R \times \{y\}$, where R is the set of neighbours of y, is a biclique with at least k edges. In both cases, (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. Thus, we output $d = 2$ and the 2×2 matrix A with all entries equal to one, which satisfies property 1 and has $\beta(A) = 4 > d$. Therefore, both $\beta(\Gamma) \geq k$ and $\beta(A) > d$ are true, and property 2 is satisfied.

Case 3: $k \leq |E_\Gamma|$, and every vertex has degree at most $k - 1$. The idea is to enlarge Γ by adding new vertices, so that every left vertex has the same degree $d = k - 1$ and the number of left vertices is a power of two. Moreover, we add only vertices that cannot belong to a biclique with more than d edges, so that the answer to the question of MAXIMUM EDGE BICLIQUE does not change.

Let us construct the matrix A. Since Γ has at least one edge, some left vertex has degree at least 1 and, by assumption, at most $k - 1$. Therefore, $k \geq 2$, and we let $d = k - 1 \geq 1$. Next, we remove the isolated vertices of Γ, and we still denote by Γ the resulting graph. This does not change $\beta(\Gamma)$, since no biclique contains an isolated vertex, and it does not change the degrees of the other vertices. Let n_0 be the number of remaining left vertices of Γ. Since every edge has exactly one left endpoint, and every left vertex is the endpoint of at least one edge, we have that $n_0 \leq |E_\Gamma|$. Moreover, $n_0 \geq 2$. Indeed, if $n_0 = 1$, then all the edges would have the same left endpoint, and the degree of this vertex would be $|E_\Gamma| \geq k$, contradicting the assumption of Case 3.

Let n be the least power of two with $n \geq n_0$. Then, it holds that

$$n_0 \leq n < 2n_0 \leq 2|E_\Gamma|. \tag{43}$$

We now construct the matrix A. For every left vertex x of Γ, we add $d - \deg_\Gamma(x) \geq 0$ new right vertices, each adjacent only to x. Then we add $n - n_0$ new left vertices, each adjacent to d new right vertices of its own. Let Γ' be the resulting graph, and let A be its biadjacency matrix, so that $\beta(A) = \beta(\Gamma')$. By construction, Γ' contains Γ, and the new vertices satisfy the following two properties: every new right vertex has a single neighbour, and every new left vertex is adjacent only to new right vertices.

We first show that A satisfies property 1. Every left vertex x of Γ has now degree $\deg_\Gamma(x) + \big(d - \deg_\Gamma(x)\big) = d$ in Γ', and every new left vertex has degree d in Γ' by construction. Hence, A has n rows, where n is a power of two, and every row contains exactly d ones. Moreover, A has no zero column, since no right vertex of Γ' is isolated. Indeed, the right vertices of Γ are not isolated, and every new right vertex has exactly one neighbour.

We now show that A satisfies property 2. Recall that Γ' contains Γ, that every left vertex of Γ' has degree $d = k - 1$, and that the vertices added to Γ satisfy the following two properties: every new right vertex has a single neighbour, and every new left vertex is adjacent only to new right vertices.

Assume first that $\beta(\Gamma) \geq k$. Since Γ is contained in Γ', every biclique of Γ is also a biclique of Γ', and hence $\beta(A) = \beta(\Gamma') \geq \beta(\Gamma) \geq k > d$.

Assume now that $\beta(A) > d$, and let $R \times C$ be a biclique of Γ' with more than d edges. We show that $R \times C$ involves only vertices of Γ. If C contained a new right vertex, then, since this vertex has a single neighbour, R would consist of a single left vertex, and $R \times C$ would have at most d edges, since every left vertex of Γ' has degree d. This is a contradiction, and hence C contains only right vertices of Γ. As a consequence, R cannot contain a new left vertex, since such a vertex is adjacent only to new right vertices. Therefore $R \times C$ involves only vertices of Γ, and, since no edge was added between vertices of Γ, it is a biclique of Γ with more than $d = k - 1$ edges. Hence $\beta(\Gamma) \geq k$. This proves property 2.

Finally, we conclude by showing that the construction takes polynomial time in the size of (Γ, k). Deciding which case applies only requires comparing k with $|E_\Gamma|$ and with the degrees of the left vertices, which can be done in polynomial time. In Cases 1 and 2, the output has constant size. In Case 3, we have that $d = k - 1 < |E_\Gamma|$, and, by (43), the matrix A has $n < 2|E_\Gamma|$ rows. Moreover, since A has no zero columns, every column of A contains at least one of the nd ones of A, so that the number of columns of A is $m \leq nd < 2|E_\Gamma|^2$. Hence A has $nm < 4|E_\Gamma|^3$ entries. Thus, since $|E_\Gamma|$ is at most the size of the biadjacency matrix of Γ, the matrix A has polynomial size, and can be computed in polynomial time. □

IV. Hardness of Wyner's common information

In this section, we prove that Wyner Threshold is NP-hard, and that, unless $\mathrm{P} = \mathrm{NP}$, Wyner's common information cannot be approximated in polynomial time within an additive error of order $1/N^2$, where N is the number of positive entries of the probability table P_{XY}. These results are formalized in the following theorem.

Theorem 1. Wyner Threshold *is NP-hard. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a rational probability table* P_{XY} *on a finite alphabet* $\mathcal{X} \times \mathcal{Y}$*, outputs a value* v *such that*

$$\big|v - J(P_{XY})\big| \leq \frac{1}{8N^2}, \tag{44}$$

where $N = |\operatorname{supp}(P_{XY})|$ *is the number of positive entries of* P_{XY}*.*

The rest of this section is organized as follows. In Section IV-A, we describe how to construct a probability table P_{XY} from a Boolean matrix A whose rows all contain the same number d of ones, and we show that the auxiliary $W = X$ is optimal for this probability table when A has no rectangle of size larger than d, and that it can be improved otherwise. In Section IV-B, we combine these results with Lemma 6 to prove Theorem 1.

A. A criterion for uniform-support tables

Let A be a Boolean matrix whose rows all contain exactly d ones. In this subsection, we define a probability table P^A_{XY} from A, and we show that the auxiliary $W = X$ is optimal for P^A_{XY} if and only if every rectangle of A has size at most d.

Definition 3. Let A be a Boolean matrix with rows indexed by $\mathcal{X}$, $|\mathcal{X}| = n$, and columns indexed by $\mathcal{Y}$, such that every row contains exactly $d \geq 1$ ones, and let $N = nd$. The *uniform-support table* of A is

$$P^A_{XY}(x, y) = \frac{A(x, y)}{N}, \qquad (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{45}$$

We first observe some elementary properties of P^A_{XY}, which follow directly from (45). The positive entries of P^A_{XY} are exactly the entries of A equal to one. Hence a set $R \times C$ is a rectangle of P^A_{XY} if and only if it is a rectangle of A, and

$$\beta(P^A_{XY}) = \beta(A) \geq d, \tag{46}$$

where the inequality holds because, for every row x of A, the set $\{x\} \times \{y \in \mathcal{Y} : A(x, y) = 1\}$ is a rectangle of A of size d. In addition, every row of P^A_{XY} contains exactly d positive entries, each equal

to $1/N$, so that $P_X(x) = d/N = 1/n$ for every $x \in \mathcal{X}$, that is, X is uniform on $\mathcal{X}$. Moreover, P^A_{XY} is uniform on its N positive entries. Therefore, we have that

$$H(X) = \log n, \qquad H(X,Y) = \log N = \log n + \log d. \tag{47}$$

Consequently, the auxiliary $W = X$ satisfies $I(X,Y;X) = H(X) = \log n$, so that

$$J(P^A_{XY}) \leq \log n. \tag{48}$$

The next lemma shows that, if the matrix A has no rectangle of size larger than d, then no admissible auxiliary of $(X,Y) \sim P^A_{XY}$ improves on $W = X$. Intuitively, by Lemma 2, the support of every component of an admissible auxiliary is a rectangle of P^A_{XY}, that is, a rectangle of A, and hence it contains at most d entries. Therefore no component can be more spread than a single row of P^A_{XY}, which is already uniform on d entries.

Lemma 7. *Let A be a Boolean matrix with n rows, each containing exactly $d \geq 1$ ones, and let P^A_{XY} be its uniform-support table. If $\beta(A) \leq d$, then $J(P^A_{XY}) = \log n$.*

Proof. By Lemma 3, it holds that $J(P^A_{XY}) \geq H(X,Y) - \log \beta(P^A_{XY})$. Since $H(X,Y) = \log N$, with $N = nd$, and $\beta(P^A_{XY}) = \beta(A) \leq d$, we obtain that

$$J(P^A_{XY}) \geq \log N - \log d = \log n. \tag{49}$$

On the other hand, the auxiliary $W = X$ satisfies $I(X,Y;X) = H(X) = \log n$, so that $J(P^A_{XY}) \leq \log n$. Therefore $J(P^A_{XY}) = \log n$. □

We now show that, if A has a rectangle of size larger than d, then the auxiliary W_{F_t} of Lemma 5, where F_t takes a fraction t of the corresponding entries of P^A_{XY}, improves on $W = X$ by an amount of order $1/N^2$.

Lemma 8. *Let A be a Boolean matrix with n rows, each containing exactly $d \geq 1$ ones, and let P^A_{XY} be its uniform-support table. If $\beta(A) > d$, then*

$$J(P^A_{XY}) < \log n - \frac{1}{4N^2}, \tag{50}$$

where $N = nd$.

Proof. Since $\beta(A) > d$, there exists a rectangle $R \times C$ of A, and therefore of P^A_{XY}, of size $\sigma = |R||C| \geq d + 1$. On this rectangle, P^A_{XY} is constant and equal to $1/N$, so that the mass of $R \times C$ is $m = \sigma/N$. Moreover, all the rows of P^A_{XY} have the same mass $1/n$. We can therefore apply Lemma 5. For $0 < t \leq 1$, let F_t be the matrix equal to t/N on $R \times C$ and to 0 elsewhere, and let W_{F_t} be the admissible auxiliary

obtained by removing F_t from P^A_{XY} and splitting the remainder into its rows, as in (30). Since $J(P^A_{XY})$ is the infimum of $I(X,Y;W)$ over all admissible auxiliaries W, Lemma 5 gives

$$J(P^A_{XY}) \le I(X,Y;W_{F_t}) \le \log n - tm\big(\log(nm) - t\log e\big).$$

Since $N = nd$, we have that $nm = \sigma/d$. Choosing $t = 1/(2(d+1))$, which belongs to $(0,1]$, we obtain that

$$J(P^A_{XY}) \le \log n - \frac{\sigma t}{N}\Big(\log\frac{\sigma}{d} - t\log e\Big). \tag{51}$$

We now bound from below the amount subtracted from $\log n$ in (51). Since $\sigma \ge d+1$, and since $\ln(1+v) \ge v/(1+v)$ for every $v \ge 0$, and recalling that $\log z = \ln z \cdot \log e$, we obtain, with $v = 1/d$, that

$$\log\frac{\sigma}{d} \ge \log\Big(1+\frac{1}{d}\Big) \ge \frac{\log e}{d+1} = 2t\log e. \tag{52}$$

Using (52) in (51), we have that

$$\begin{aligned} J(P^A_{XY}) &\le \log n - \frac{\sigma t}{N}\big(2t\log e - t\log e\big) = \log n - \frac{\sigma t^2 \log e}{N} \\ &\overset{(a)}{\le} \log n - \frac{\log e}{4N(d+1)} \overset{(b)}{<} \log n - \frac{1}{4N^2}, \end{aligned}$$

where (a) follows from $\sigma t^2 = \sigma/(4(d+1)^2) \ge 1/(4(d+1))$, since $\sigma \ge d+1$, and (b) follows from $\log e > 1$ and $d+1 \le \sigma \le N$, since the rectangle $R \times C$ is contained in the support of P^A_{XY}, which has N entries. This proves (50). □

Combining Lemmas 7 and 8, we obtain the following criterion, which distinguishes the cases $\beta(A) \le d$ and $\beta(A) > d$.

Lemma 9. *Let A be a Boolean matrix with n rows, each containing exactly $d \ge 1$ ones, and let P^A_{XY} be its uniform-support table, with $N = nd$. Then, it holds that*

$$J(P^A_{XY}) \begin{cases} = \log n, & \text{if } \beta(A) \le d, \\ < \log n - \dfrac{1}{4N^2}, & \text{if } \beta(A) > d. \end{cases} \tag{53}$$

In particular, $J(P^A_{XY}) < \log n$ if and only if $\beta(A) > d$.

B. The reduction and the proof of Theorem 1

We now describe in full the reduction from MAXIMUM EDGE BICLIQUE to WYNER THRESHOLD. Let (Γ, k) be an instance of MAXIMUM EDGE BICLIQUE, and let A and d be the Boolean matrix and the integer given by Lemma 6 applied to (Γ, k). Thus A has n rows, where n is a power of two, each

containing exactly d ones, and $N = nd$ ones in total. Let $P^A_{XY} = A/N$ be the uniform-support table of A, and let $r = \log n$, which is an integer. By Lemma 9, Wyner's common information of P^A_{XY} is equal to r if $\beta(A) \leq d$, and smaller than $r - \frac{1}{4N^2}$ if $\beta(A) > d$. We place the threshold in the middle of this gap, that is, we let

$$\theta = r - \frac{1}{8N^2}. \tag{54}$$

Therefore, the instance of WYNER THRESHOLD associated with (Γ, k) is the pair (P^A_{XY}, θ).

Proof of Theorem 1. We start by showing that WYNER THRESHOLD is NP-hard. To this end, we show that the map $(\Gamma, k) \mapsto (P^A_{XY}, \theta)$ described above is a polynomial-time reduction from MAXIMUM EDGE BICLIQUE to WYNER THRESHOLD.

We first show that the reduction runs in polynomial time in the size of (Γ, k). By Lemma 6, the matrix A is constructed in polynomial time, and hence it has polynomial size. In particular, its number n of rows and its number $N = nd$ of ones are polynomial in the size of (Γ, k). The table P^A_{XY} has the same size as A, and each of its entries is equal to 0 or to $1/N$, so that it is a rational number with $O(\log N)$ bits. It remains to evaluate θ. Since n is a power of two, $r = \log n$ is an integer, which is the binary exponent of n, and, by (54), $\theta = r - 1/(8N^2)$ is a rational number with $O(\log N)$ bits. Hence, the whole reduction runs in polynomial time.

We now show that the reduction is correct, that is, (P^A_{XY}, θ) is a yes-instance of WYNER THRESHOLD if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. We recall that (P^A_{XY}, θ) is a yes-instance of WYNER THRESHOLD if there exists an admissible auxiliary W of (X, Y) such that $I(X, Y; W) \leq \theta$, and that (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE if Γ contains a biclique with at least k edges, that is, $\beta(\Gamma) \geq k$. Assume first that (Γ, k) is a yes-instance, that is, $\beta(\Gamma) \geq k$. Then $\beta(A) > d$ by property 2 of Lemma 6, and hence, by Lemma 9 and by the choice of θ in (54), we have that

$$J(P^A_{XY}) < r - \frac{1}{4N^2} < r - \frac{1}{8N^2} = \theta. \tag{55}$$

Since $J(P^A_{XY})$ is the infimum of $I(X, Y; W)$ over all admissible auxiliaries W of (X, Y), it follows from (55) that there exists an admissible auxiliary W such that $I(X, Y; W) < \theta$. Hence (P^A_{XY}, θ) is a yes-instance of WYNER THRESHOLD.

Assume now that (Γ, k) is a no-instance, that is, $\beta(\Gamma) < k$. Then $\beta(A) \leq d$ by property 2 of Lemma 6, and hence, by Lemma 9 and by the choice of θ in (54), we have that

$$J(P^A_{XY}) = r > r - \frac{1}{8N^2} = \theta. \tag{56}$$

Since $J(P_{XY}^A)$ is the infimum of $I(X,Y;W)$ over all admissible auxiliaries W of (X,Y), it follows from (56) that every admissible auxiliary W satisfies $I(X,Y;W) \geq J(P_{XY}^A) > \theta$. Hence (P_{XY}^A, θ) is a no-instance of WYNER THRESHOLD. This proves that WYNER THRESHOLD is NP-hard.

It remains to prove the inapproximability result. Suppose, by contradiction, that there exists a polynomial-time algorithm that, given a rational probability table P_{XY}, outputs a value v such that $|v - J(P_{XY})| \leq \frac{1}{8N^2}$, as in (44). We apply this algorithm to the table P_{XY}^A, and we show that comparing its output v with θ decides MAXIMUM EDGE BICLIQUE. If (Γ, k) is a yes-instance, then, by (44) and (55), we obtain that

$$v \leq J(P_{XY}^A) + \frac{1}{8N^2} < r - \frac{1}{4N^2} + \frac{1}{8N^2} = r - \frac{1}{8N^2} = \theta. \tag{57}$$

If (Γ, k) is a no-instance, then, by (44) and (56), we obtain that

$$v \geq J(P_{XY}^A) - \frac{1}{8N^2} = r - \frac{1}{8N^2} = \theta. \tag{58}$$

By (57) and (58), we have that $v < \theta$ if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. Therefore, running the reduction, then the algorithm on P_{XY}^A, and finally comparing v with θ would decide MAXIMUM EDGE BICLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

V. HARDNESS OF THE COMMON ENTROPY

In this section, we prove that COMMON ENTROPY THRESHOLD is NP-hard, and that, unless $\mathrm{P} = \mathrm{NP}$, the common entropy cannot be approximated in polynomial time within an additive error of order $1/N$, where N is the number of positive entries of the probability table P_{XY}. These results are formalized in the following theorem.

Theorem 2. COMMON ENTROPY THRESHOLD *is NP-hard. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a rational probability table* P_{XY} *on a finite alphabet* $\mathcal{X} \times \mathcal{Y}$*, outputs a value* v *such that*

$$\big|v - G(P_{XY})\big| \leq \frac{1}{2N}, \tag{59}$$

where $N = |\operatorname{supp}(P_{XY})|$ *is the number of positive entries of* P_{XY}.

The table used in the reduction for WYNER THRESHOLD cannot be used for COMMON ENTROPY THRESHOLD, as we show with an example in Appendix B. Therefore, in Section V-A, we construct, from a Boolean matrix A whose rows all contain the same number d of ones, a different probability table $\widehat{P}_{XY}^A$, which we will use in the reduction. In Section V-B, we show that the auxiliary $W = X$ is optimal for $\widehat{P}_{XY}^A$ if $\beta(A) \leq d$, and that it can be improved by more than $1/N$ otherwise. Finally, in Section V-C, we combine these results with Lemma 6 to prove Theorem 2.

A. The table of the reduction

Throughout this section, let A be a Boolean matrix with $n \geq 2$ rows, where n is a power of two, and m columns, none of which is zero, such that every row of A contains exactly $d \geq 1$ ones.

We first explain the idea behind the construction of the probability table $\widehat{P}^A_{XY}$ from the Boolean matrix A. As in Section IV, we use the auxiliary $W = X$, whose components are the rows of the table, as a baseline, and we compare every admissible auxiliary with it. By Lemma 4, if all the rows of a probability table have the same mass ω, then extracting a component of mass larger than $e\omega$ yields an auxiliary whose entropy is smaller than the entropy of $W = X$. Following this idea, we build the probability table $\widehat{P}^A_{XY}$ from A in two steps.

First, we replace each row of A by T identical copies. In this way, a rectangle of A with $d+1$ entries yields a rectangle with $T(d+1)$ entries, whose mass exceeds e times the mass of a row if T is large enough. However, the T copies of a row of A also form a rectangle, with Td entries, whose mass is T times the mass of a row. Therefore, when $T > e$, by Lemma 4, extracting the T copies of a row as a single component would lower the entropy, even if A has no rectangle with more than d entries. To prevent this, we add, for every row, a new column that is positive only in that row, and we put a large mass on its single positive entry. By construction, these entries cannot be covered by components spread over several rows. Indeed, the support of such a component would be a rectangle containing a new column and at least two rows. This is impossible, since each new column is positive in a single row. Therefore, a component spread over several rows can only take mass from the original columns of A. Moreover, since the new entries increase the mass of every row, the T copies of a row of A no longer have mass larger than e times the mass of a row. In this way, we show that the auxiliary $W = X$ of $(X, Y) \sim \widehat{P}^A_{XY}$ is optimal whenever $\beta(A) \leq d$, that is, whenever A has no rectangle with more than d entries, and it can be improved otherwise.

We now give the full details of the construction of $\widehat{P}^A_{XY}$. Let T be the least power of two such that $T \geq 16(d+1)$, and let $q = nT$. Let B be the Boolean $q \times m$ matrix obtained by replacing each row of A by T identical copies. We index the rows of B by the pairs $r = (x, j)$, with $1 \leq x \leq n$ and $1 \leq j \leq T$, where (x, j) is the j-th copy of row x of A, so that

$$B\big((x,j),y\big) = A(x,y) \qquad \text{for every column } y \text{ of } A. \tag{60}$$

In particular, every row of B is a copy of a row of A, so that it contains exactly d ones. For a positive integer λ, which will be chosen below, let

$$\kappa = d + \lambda \quad \text{and} \quad M = q\kappa, \tag{61}$$

and let I_q be the $q \times q$ identity matrix. We define $\widehat{P}^A_{XY}$ as the $q \times (m+q)$ matrix obtained by placing the matrix λI_q to the right of B, and by dividing all the entries by M, that is,

$$\widehat{P}^A_{XY} = \frac{1}{M}\left(B \mid \lambda I_q\right). \tag{62}$$

Thus, the rows of $\widehat{P}^A_{XY}$ are indexed by the rows $r = (x, j)$ of B, where $r = (x, j)$ is the j-th copy of row x of A, and its columns are the m columns of B, indexed by the columns y of A, followed by q new columns, one for each row r of B, which we denote by σ_r. We call the columns of the first kind *core columns*, since they come from A, and the columns σ_r *private columns*, since each column σ_r has a single positive entry, which lies in row r, and the column is therefore private to row r. More explicitly, the entries of $\widehat{P}^A_{XY}$ are given, for every row r of B, by

$$\widehat{P}^A_{XY}(r, y) = \frac{B(r, y)}{M} \quad \text{for every column } y \text{ of } A, \qquad \widehat{P}^A_{XY}(r, \sigma_{r'}) = \begin{cases} \lambda/M, & \text{if } r' = r, \\ 0, & \text{otherwise.} \end{cases} \tag{63}$$

Hence, every row r of $\widehat{P}^A_{XY}$ contains d entries equal to $1/M$ in the core columns, and one entry equal to λ/M in the private column σ_r. In particular, every row has mass $(d+\lambda)/M = \kappa/(q\kappa) = 1/q$. Hence, $\widehat{P}^A_{XY}$ is a probability table for every positive integer λ.

Moreover, the private columns of $\widehat{P}^A_{XY}$ have the following property. Let F be a component of an admissible auxiliary of $(X, Y) \sim \widehat{P}^A_{XY}$ that is positive at the entry (r, σ_r), for some row r. By Lemma 2, the support of F is a rectangle $R \times C$ of $\widehat{P}^A_{XY}$, and $\sigma_r \in C$. Thus all the entries (r', σ_r) with $r' \in R$ are positive. Since the only positive entry of the column σ_r lies in row r, it follows that $R = \{r\}$. Therefore the mass of the private columns can only be carried by components lying in a single row.

Now, in order to complete the construction of $\widehat{P}^A_{XY}$, it remains to specify the value of λ. The value of λ must satisfy two conditions. Indeed, as we will show in Section V-B, the private entries λ/M must be large enough to make $W = X$ optimal when $\beta(A) \le d$, which holds if $Td \le e\lambda$. At the same time, to allow an improvement on $W = X$ when $\beta(A) > d$, it must hold that $T(d+1) > e\kappa$, where $\kappa = d + \lambda$. The next technical lemma, whose proof is given in Appendix C, shows that both conditions hold for a suitable choice of λ.

Lemma 10. *Let T be the least power of two such that $T \ge 16(d+1)$, and let*

$$\ell = \lceil \log(2Td) \rceil, \qquad e_\ell = \sum_{i=0}^{\ell} \frac{1}{i!}, \qquad \lambda = \left\lceil \frac{Td}{e_\ell} \right\rceil, \qquad \kappa = d + \lambda. \tag{64}$$

Then λ can be computed in polynomial time in the bit-length of d, and the following properties hold:

1) *$Td \le e\lambda$ and $e\kappa < Td + T/2$.*
2) *$\kappa < T(d+1)/2$ and $\lambda > 1$.*

In the rest of this section, we fix λ and κ as in Lemma 10, so that the table $\widehat{P}^A_{XY}$ is completely determined by the Boolean matrix A. Moreover, since all the rows of $\widehat{P}^A_{XY}$ have mass $1/q$, the auxiliary $W = X$ of $(X, Y) \sim \widehat{P}^A_{XY}$ has entropy $H(X) = \log q$. We denote this value by

$$r_A = \log q = \log(nT) = \log n + \log T, \tag{65}$$

and we note that r_A is an integer, since n and T are powers of two. The next lemma collects the properties of $\widehat{P}^A_{XY}$ that we will use in the rest of this section.

Lemma 11. *The probability table $\widehat{P}^A_{XY}$ defined in* (62)*, with λ and κ as in Lemma 10, satisfies the following properties:*

1) *The number of positive entries and the total number of entries of $\widehat{P}^A_{XY}$ are $N = q(d+1)$ and $L = q(m+q)$, respectively, and $L \geq 32N$.*
2) *Every rectangle of B is a rectangle of $\widehat{P}^A_{XY}$, and every rectangle of $\widehat{P}^A_{XY}$ with at least two rows is a rectangle of B.*
3) *$\beta(B) = T\beta(A)$.*

Proof. We first prove property 1. By (63), every row of $\widehat{P}^A_{XY}$ contains exactly $d+1$ positive entries. Since $\widehat{P}^A_{XY}$ has q rows and $m+q$ columns, it follows that $N = q(d+1)$ and $L = q(m+q)$. Moreover, since $n \geq 2$, $T \geq 16(d+1)$ and $q = nT$, we have that $q \geq 32(d+1)$. Hence, $L/N = (m+q)/(d+1) \geq q/(d+1) \geq 32$.

We now prove property 2. By (63), the positive entries of $\widehat{P}^A_{XY}$ in the core columns are exactly the entries where B equals one, so that every rectangle of B is a rectangle of $\widehat{P}^A_{XY}$. Conversely, let $R \times C$ be a rectangle of $\widehat{P}^A_{XY}$ with at least two rows, that is, $|R| \geq 2$. If C contained a private column σ_r, then all the entries (r', σ_r) with $r' \in R$ would be positive. This is impossible, since R contains a row $r' \neq r$, while the only positive entry of σ_r lies in row r. Hence, C contains only core columns, which are the columns of B, and therefore $R \times C$ is a rectangle of B.

Finally, we prove property 3. To this end, we show that $\beta(B) \leq T\beta(A)$ and that $\beta(B) \geq T\beta(A)$. We first show that $\beta(B) \leq T\beta(A)$. Let $Z_0 \times Y_0$ be a rectangle of B, and let X_0 be the set of rows x of A that have at least one copy in Z_0, that is, such that $(x, j) \in Z_0$ for some $1 \leq j \leq T$. We claim that $X_0 \times Y_0$ is a rectangle of A. Indeed, let $x \in X_0$ and $y \in Y_0$. By the definition of X_0, there exists j such that $(x, j) \in Z_0$. Since $Z_0 \times Y_0$ is a rectangle of B, we have that $B((x, j), y) = 1$, and hence, by (60), $A(x, y) = B((x, j), y) = 1$. This proves the claim, so that $|X_0||Y_0| \leq \beta(A)$. Moreover, since each row of A has T copies among the rows of B, we have that $|Z_0| \leq T|X_0|$. Therefore

$$|Z_0||Y_0| \leq T|X_0||Y_0| \leq T\beta(A).$$

Since $Z_0 \times Y_0$ is an arbitrary rectangle of B, it follows that $\beta(B) \leq T\beta(A)$.

We now show that $\beta(B) \geq T\beta(A)$. Let $X_0 \times Y_0$ be a rectangle of A of size $\beta(A)$. By (60), the set $\big(X_0 \times \{1, \ldots, T\}\big) \times Y_0$, formed by all the copies of the rows in X_0 and by the columns in Y_0, is a rectangle of B of size $T|X_0||Y_0| = T\beta(A)$. Hence $\beta(B) \geq T\beta(A)$. Together, the two inequalities prove that $\beta(B) = T\beta(A)$. □

B. A criterion for the table of the reduction

In this subsection, we show that the auxiliary $W = X$ of $(X, Y) \sim \widehat{P}^A_{XY}$, where $\widehat{P}^A_{XY}$ is the probability table defined in (62), reaches $G(\widehat{P}^A_{XY})$ if $\beta(A) \leq d$, and that it can be improved by more than $1/N$ if $\beta(A) > d$, where $N = q(d+1)$ is the number of positive entries of $\widehat{P}^A_{XY}$. We prove these two results in Lemmas 12 and 13, respectively.

We first consider the case $\beta(A) \leq d$. Recall that B is the intermediate matrix obtained from A by replacing each row with T identical copies, defined in (60). By properties 2 and 3 of Lemma 11, every component of an admissible auxiliary of $(X, Y) \sim \widehat{P}^A_{XY}$ with positive entries in at least two different rows lies on a rectangle of B, and hence its support has at most $T\beta(A) \leq Td$ entries. Using this property, the next lemma shows that no admissible auxiliary improves on $W = X$.

Lemma 12. *Let A be a Boolean matrix with $n \geq 2$ rows, where n is a power of two, and no zero columns, such that every row contains exactly $d \geq 1$ ones, and let $\widehat{P}^A_{XY}$ be the probability table defined in* (62)*, with λ and κ as in Lemma 10. If $\beta(A) \leq d$, then*

$$G(\widehat{P}^A_{XY}) = r_A, \tag{66}$$

where $r_A = \log q$ is defined in (65)*.*

Proof. Since the auxiliary $W = X$ has entropy $H(X) = \log q = r_A$, it suffices to show that every admissible auxiliary W of $(X, Y) \sim \widehat{P}^A_{XY}$ satisfies $H(W) \geq \log q$. Let W be such an auxiliary. We first simplify W, producing another admissible auxiliary $\widetilde{W}$ such that $H(\widetilde{W}) \leq H(W)$, and then we prove that $H(\widetilde{W}) \geq \log q$.

Let us simplify W. The components of W are of two kinds, namely the components whose support lies in a single row, and the components whose support meets at least two distinct rows. For every row r, we merge all the components of W whose support lies in row r into a single matrix, which we denote by G_r, while we leave the other components unchanged, and we denote them by $F_1, \ldots, F_s$, with $s \geq 0$. Let $\mathcal{I} = \{1, \ldots, s\}$, which is empty if $s = 0$.

Each matrix G_r lies in a single row, so it has rank one. Moreover, G_r is nonzero, since, by construction of $\widehat{P}^A_{XY}$, the private column σ_r is positive only in row r. Hence, the entry (r, σ_r) can only be covered by

components whose support lies in row r. Consequently the matrices G_r, for every row r, together with the matrices F_i, for $i \in \mathcal{I}$, form a rank-one decomposition of $\widehat{P}^A_{XY}$. Let $\widetilde{W}$ be the corresponding admissible auxiliary, given by property 2 of Lemma 2. For $\phi(v) = -v \log v$, it holds that $\phi(a+b) \leq \phi(a) + \phi(b)$ for all $a, b \geq 0$. Therefore, merging components does not increase the entropy of the auxiliary (see (24)). Hence, $H(\widetilde{W}) \leq H(W)$. It therefore suffices to show that $H(\widetilde{W}) \geq \log q$.

We now bound the masses of the components of $\widetilde{W}$. We call the matrices G_r the *row components* of $\widetilde{W}$, and the matrices F_i its *cross components*.

We first bound the mass of every cross component. Let $i \in \mathcal{I}$. By Lemma 2, the support of F_i is a rectangle of $\widehat{P}^A_{XY}$ with at least two rows. Hence, by property 2 of Lemma 11, it is a rectangle of B. Moreover, by property 3 of Lemma 11, and since $\beta(A) \leq d$, this rectangle, and therefore the support of F_i, has at most $\beta(B) = T\beta(A) \leq Td$ entries. By (63), $\widehat{P}^A_{XY}$ equals $1/M$ on these entries. Hence, since $F_i \leq \widehat{P}^A_{XY}$, all the entries of F_i are at most $1/M$. Thus, we obtain that

$$\mu(F_i) \leq \frac{Td}{M} \qquad \forall i \in \mathcal{I}. \tag{67}$$

We next bound the masses of every row component. For every row r, let $\tau_r = \sum_{i \in \mathcal{I}} \sum_y F_i(r, y)$ be the total mass that the cross components take from row r. Since row r has mass $\omega = 1/q$, and it is covered by G_r and by the cross components, we have that $\mu(G_r) = \omega - \tau_r$. Moreover, summing over all the rows the masses τ_r, we obtain the total mass of the cross components, that is,

$$\sum_r \tau_r = \sum_{i \in \mathcal{I}} \mu(F_i). \tag{68}$$

As observed before, by construction of $\widehat{P}^A_{XY}$, the entry (r, σ_r), which is equal to λ/M by (63), can only be covered by components whose support lies in row r. Since all these components have been merged into G_r, we have that $G_r(r, \sigma_r) = \lambda/M$. The mass of G_r is the sum of its entries, which are nonnegative, and hence it is at least $G_r(r, \sigma_r)$. Therefore, it holds that

$$\mu(G_r) = \omega - \tau_r \geq \frac{\lambda}{M} \qquad \text{for every row } r. \tag{69}$$

We can now prove that $H(\widetilde{W}) \geq \log q$. By (24), the entropy of $\widetilde{W}$ is

$$H(\widetilde{W}) = \sum_r \phi\big(\mu(G_r)\big) + \sum_{i \in \mathcal{I}} \phi\big(\mu(F_i)\big) = \sum_r \phi(\omega - \tau_r) + \sum_{i \in \mathcal{I}} \phi\big(\mu(F_i)\big). \tag{70}$$

We bound the two sums in the right-hand side of (70) separately. For the first sum, we compare $\phi(\omega - \tau_r)$ with $\phi(\omega)$. Since $\phi(v) = -v \log v$, its derivative is $\phi'(v) = -\log(ev)$. Hence, by the fundamental theorem of calculus, for every row r we have that

$$\phi(\omega - \tau_r) - \phi(\omega) = -\int_{\omega - \tau_r}^{\omega} \phi'(v)\, dv = \int_{\omega - \tau_r}^{\omega} \log(ev)\, dv.$$

On the interval of integration, we have that $v \geq \omega - \tau_r \geq \lambda/M$ by (69). Hence, it holds that $\log(ev) \geq \log \frac{e\lambda}{M}$. Moreover, since the interval has length τ_r, it follows that

$$\phi(\omega - \tau_r) - \phi(\omega) \geq \tau_r \log \frac{e\lambda}{M}.$$

Summing over the q rows, and using $q\,\phi(\omega) = q \cdot \frac{1}{q} \log q = \log q$, we obtain that

$$\sum_r \phi(\omega - \tau_r) \geq \sum_r \left(\phi(\omega) + \tau_r \log \frac{e\lambda}{M} \right) = \log q + \Big(\sum_r \tau_r \Big) \log \frac{e\lambda}{M}. \tag{71}$$

For the second sum, by (67), we have that, for every $i \in \mathcal{I}$,

$$\phi\big(\mu(F_i)\big) = \mu(F_i) \log \frac{1}{\mu(F_i)} \geq \mu(F_i) \log \frac{M}{Td}.$$

Summing over $i \in \mathcal{I}$, and using (68), we obtain that

$$\sum_{i \in \mathcal{I}} \phi\big(\mu(F_i)\big) \geq \Big(\sum_r \tau_r \Big) \log \frac{M}{Td}. \tag{72}$$

Substituting (71) and (72) into (70), we obtain that

$$H(\widetilde{W}) \geq \log q + \Big(\sum_r \tau_r \Big) \Big(\log \frac{e\lambda}{M} + \log \frac{M}{Td} \Big) = \log q + \Big(\sum_r \tau_r \Big) \log \frac{e\lambda}{Td} \geq \log q,$$

where the last inequality follows from $Td \leq e\lambda$, by property 1 of Lemma 10. Hence $H(W) \geq H(\widetilde{W}) \geq \log q$ for every admissible auxiliary W, which proves that $G(\widehat{P}^A_{XY}) = \log q = r_A$. □

We now consider the case $\beta(A) > d$. In this case, by property 3 of Lemma 11, B has a rectangle with at least $T(d+1)$ entries, each equal to $1/M$ in $\widehat{P}^A_{XY}$. Its mass is thus at least $T(d+1)/M$, which, by the choice of λ, exceeds e times the mass κ/M of a row. Using this, the next lemma shows that the auxiliary W_F of Lemma 4, where F is supported on this rectangle, strictly improves on the auxiliary $W = X$.

Lemma 13. *Let A be a Boolean matrix with $n \geq 2$ rows, where n is a power of two, and no zero columns, such that every row contains exactly $d \geq 1$ ones, and let $\widehat{P}^A_{XY}$ be the probability table defined in* (62)*, with λ and κ as in Lemma 10. If $\beta(A) > d$, then*

$$G(\widehat{P}^A_{XY}) < r_A - \frac{1}{N}, \tag{73}$$

where r_A is defined in (65) *and $N = q(d+1)$.*

Proof. Since $G(\widehat{P}^A_{XY})$ is the infimum of $H(W)$ over all admissible auxiliaries W, it suffices to exhibit an admissible auxiliary whose entropy is smaller than $r_A - 1/N$. We obtain it from Lemma 4, by extracting from $\widehat{P}^A_{XY}$ a component supported on a rectangle with at least $T(d+1)$ entries.

Since $\beta(A) > d$ and $\beta(A)$ is an integer, we have that $\beta(A) \geq d+1$. Hence, by property 3 of Lemma 11, B has a rectangle $R \times C$ of size

$$\sigma = \beta(B) = T\beta(A) \geq T(d+1). \tag{74}$$

By property 2 of Lemma 11, $R \times C$ is also a rectangle of $\widehat{P}^A_{XY}$, and, by (63), $\widehat{P}^A_{XY}$ equals $1/M$ on all the entries of $R \times C$.

We now define the matrix F to which we apply Lemma 4. Let F be the matrix equal to $1/M$ on $R \times C$ and to 0 elsewhere. Then F is nonnegative and nonzero, and it has rank one, since $F(x,y) = \big(\frac{1}{M}\mathbb{1}_R(x)\big)\mathbb{1}_C(y)$. Moreover, $F \leq \widehat{P}^A_{XY}$ and its mass is $\mu(F) = \sigma/M$.

We now apply Lemma 4, since every row of $\widehat{P}^A_{XY}$ has the same mass $\omega = (d+\lambda)/M = \kappa/M = 1/q$. Hence, by the second inequality in (33), with $\mu(F) = \sigma/M$ and $\omega = \kappa/M$, the auxiliary W_F satisfies

$$G(\widehat{P}^A_{XY}) \leq H(W_F) \leq \log q - \Big(\frac{\sigma}{M} - \frac{e\kappa}{M}\Big)\log e = r_A - \frac{(\sigma - e\kappa)\log e}{q\kappa}, \tag{75}$$

where the last equality follows from $\log q = r_A$ and $M = q\kappa$.

To complete the proof, it remains to bound from below the amount subtracted from r_A in (75). By (74) and property 1 of Lemma 10, which gives $e\kappa < Td + T/2$, we have that

$$\sigma - e\kappa > T(d+1) - \Big(Td + \frac{T}{2}\Big) = \frac{T}{2}. \tag{76}$$

Therefore, we have that

$$\frac{(\sigma - e\kappa)\log e}{q\kappa} \overset{(a)}{>} \frac{T\log e}{2q\kappa} \overset{(b)}{>} \frac{\log e}{q(d+1)} \overset{(c)}{=} \frac{\log e}{N} \overset{(d)}{>} \frac{1}{N}, \tag{77}$$

where (a) follows from (76), (b) follows from $\kappa < T(d+1)/2$, by property 2 of Lemma 10, (c) follows from $N = q(d+1)$, by property 1 of Lemma 11, and (d) follows from $\log e > 1$. Finally, combining (77) with (75) proves (73), which concludes the proof. □

Combining Lemmas 12 and 13, we obtain the following criterion, which distinguishes the cases $\beta(A) \leq d$ and $\beta(A) > d$ by a gap of size at least $1/N$.

Lemma 14. *Let A be a Boolean matrix with $n \geq 2$ rows, where n is a power of two, and no zero columns, such that every row contains exactly $d \geq 1$ ones, let $\widehat{P}^A_{XY}$ be the probability table defined in* (62)*, with λ and κ as in Lemma 10, and let $N = q(d+1)$. Then the following properties hold:*

1) *If $\beta(A) \leq d$, then $G(\widehat{P}^A_{XY}) = r_A$.*
2) *If $\beta(A) > d$, then $G(\widehat{P}^A_{XY}) < r_A - \dfrac{1}{N}$.*

C. The reduction and the proof of Theorem 2

We now describe the reduction from MAXIMUM EDGE BICLIQUE to COMMON ENTROPY THRESHOLD. Let (Γ, k) be an instance of MAXIMUM EDGE BICLIQUE, and let A and d be the Boolean matrix and the integer given by Lemma 6 applied to (Γ, k). Let $\widehat{P}^A_{XY}$ be the probability table defined in (62), constructed from A as in Section V-A, and let $N = q(d+1)$ be its number of positive entries. By Lemma 14, the common entropy of $\widehat{P}^A_{XY}$ is equal to r_A if $\beta(A) \leq d$, and smaller than $r_A - \frac{1}{N}$ if $\beta(A) > d$. We place the threshold in the middle of this gap, that is, we let

$$\theta = r_A - \frac{1}{2N}. \tag{78}$$

Therefore, the instance of COMMON ENTROPY THRESHOLD associated with (Γ, k) is the pair $(\widehat{P}^A_{XY}, \theta)$.

Proof of Theorem 2. We start by showing that COMMON ENTROPY THRESHOLD is NP-hard. To this end, we show that the map $(\Gamma, k) \mapsto (\widehat{P}^A_{XY}, \theta)$ described above is a polynomial-time reduction from MAXIMUM EDGE BICLIQUE to COMMON ENTROPY THRESHOLD.

We first show that the reduction runs in polynomial time in the size of (Γ, k). To this end, we show that all the parameters of the construction are polynomial in the size of (Γ, k), and that they can be computed in polynomial time.

We start with the matrix A. By Lemma 6, the matrix A is constructed in polynomial time, and hence its number n of rows, its number m of columns, and the number d of ones in each row are polynomial in the size of (Γ, k). We next consider the parameters of the table $\widehat{P}^A_{XY}$. Since T is the least power of two with $T \geq 16(d+1)$, we have that $T < 32(d+1)$, and hence $q = nT$ is polynomial in d and n. By Lemma 10, λ is computed in polynomial time, and, by (140), $\lambda < Td/e + 2$, so that λ, κ and $M = q\kappa$ are polynomial as well. Hence the table $\widehat{P}^A_{XY}$ has $q(m+q)$ entries, each equal to 0, $1/M$ or λ/M, which are rational numbers with $O(\log M)$ bits. It remains to evaluate θ. Since n and T are powers of two, we have that $r_A = \log q = \log n + \log T$ is an integer. Moreover, since $N = q(d+1)$ is polynomial, by (78), $\theta = r_A - 1/(2N)$ is a rational number with $O(\log N)$ bits. Hence the whole reduction runs in polynomial time.

We now show that the reduction is correct, that is, $(\widehat{P}^A_{XY}, \theta)$ is a yes-instance of COMMON ENTROPY THRESHOLD if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. We recall that $(\widehat{P}^A_{XY}, \theta)$ is a yes-instance of COMMON ENTROPY THRESHOLD if there exists an admissible auxiliary W of (X, Y) such that $H(W) \leq \theta$, and that (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE if $\beta(\Gamma) \geq k$.

Assume first that (Γ, k) is a yes-instance, that is, $\beta(\Gamma) \geq k$. Then $\beta(A) > d$ by property 2 of Lemma 6, and hence, by Lemma 14 and by the choice of θ in (78), we have that

$$G(\widehat{P}^A_{XY}) < r_A - \frac{1}{N} < r_A - \frac{1}{2N} = \theta. \tag{79}$$

Since $G(\widehat{P}^A_{XY})$ is the infimum of $H(W)$ over all admissible auxiliaries W of (X, Y), it follows from (79) that there exists an admissible auxiliary W such that $H(W) < \theta$. Hence $(\widehat{P}^A_{XY}, \theta)$ is a yes-instance of COMMON ENTROPY THRESHOLD.

Assume now that (Γ, k) is a no-instance, that is, $\beta(\Gamma) < k$. Then $\beta(A) \leq d$ by property 2 of Lemma 6, and hence, by Lemma 14 and by the choice of θ in (78), we have that

$$G(\widehat{P}^A_{XY}) = r_A > r_A - \frac{1}{2N} = \theta. \tag{80}$$

Since $G(\widehat{P}^A_{XY})$ is the infimum of $H(W)$ over all admissible auxiliaries W of (X, Y), it follows from (80) that every admissible auxiliary W satisfies $H(W) \geq G(\widehat{P}^A_{XY}) > \theta$. Hence $(\widehat{P}^A_{XY}, \theta)$ is a no-instance of COMMON ENTROPY THRESHOLD. This proves that the reduction is correct and that COMMON ENTROPY THRESHOLD is NP-hard.

It remains to prove the inapproximability result. Suppose, by contradiction, that there exists a polynomial-time algorithm that, given a rational probability table P_{XY}, outputs a value v such that $|v - G(P_{XY})| \leq \frac{1}{2N}$, as in (59). We apply this algorithm to the table $\widehat{P}^A_{XY}$, and we show that comparing its output v with θ decides MAXIMUM EDGE BICLIQUE. If (Γ, k) is a yes-instance, then, by (59) and (79), we obtain that

$$v \leq G(\widehat{P}^A_{XY}) + \frac{1}{2N} < r_A - \frac{1}{N} + \frac{1}{2N} = r_A - \frac{1}{2N} = \theta. \tag{81}$$

If (Γ, k) is a no-instance, then, by (59) and (80), we obtain that

$$v \geq G(\widehat{P}^A_{XY}) - \frac{1}{2N} = r_A - \frac{1}{2N} = \theta. \tag{82}$$

By (81) and (82), we have that $v < \theta$ if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. Therefore, running the reduction, then the algorithm on $\widehat{P}^A_{XY}$, and finally comparing v with θ would decide MAXIMUM EDGE BICLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

VI. STRICTLY POSITIVE TABLES

The reductions of Sections IV and V produce probability tables with many zero entries. Tables with this property might be unlikely to arise in practice. We therefore ask whether the hardness depends on the presence of zero entries, or persists even when every pair of outcomes has positive probability. In this section, we show that WYNER THRESHOLD and COMMON ENTROPY THRESHOLD remain NP-hard even

on strictly positive probability tables. To prove this, we use a natural idea, namely we remove the zero entries of the tables by filling them with a small amount of probability. If this amount is small enough, one may expect both Wyner's common information and the common entropy of the table to change only slightly, so that the reductions of the previous sections still work. In what follows, we show that this is indeed the case.

More precisely, we add one step at the end of each of the reductions described in Sections IV and V. Let P_{XY} be the table produced by either reduction, namely P^A_{XY} for WYNER THRESHOLD and $\widehat{P}^A_{XY}$ for COMMON ENTROPY THRESHOLD, and let N and L be its numbers of positive entries and of entries, respectively. We fill the zeros of P_{XY} with a small total mass η, that is, we multiply every positive entry by $1-\eta$, and we replace every zero entry by $\eta/(L-N)$. In this way, we obtain a strictly positive probability table, which we denote by $\widetilde{P}_{XY}$, and the new reduction outputs the pair $(\widetilde{P}_{XY}, \theta)$, with the same threshold θ as before.

The new reduction is correct as long as filling the zeros changes Wyner's common information and the common entropy by less than half of the gaps of Lemmas 9 and 14, which are $1/(4N^2)$ and $1/N$, respectively. Indeed, in this case the threshold θ, which lies in the middle of each gap, still separates yes-instances from no-instances. The main technical result of this section shows that, if η is a sufficiently small inverse polynomial in L, both quantities change by at most $1/(16L^2)$, which is smaller than half of each gap. This leads to the following theorem.

Theorem 3. WYNER THRESHOLD *and* COMMON ENTROPY THRESHOLD *are NP-hard even on strictly positive rational tables. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a strictly positive rational probability table* P_{XY} *on a finite alphabet* $\mathcal{X} \times \mathcal{Y}$*, approximates* $J(P_{XY})$ *within the additive error* $1/(16L^2)$*, or* $G(P_{XY})$ *within the additive error* $8/L$*, where* $L = |\mathcal{X}||\mathcal{Y}|$*.*

The rest of this section is organized as follows. In Section VI-A, we show that filling the zeros of a table with a small mass η changes Wyner's common information and the common entropy only slightly. In Section VI-B, we use this result to prove Theorem 3.

A. Filling the zeros of a table

Let P_{XY} be a probability table on $\mathcal{X} \times \mathcal{Y}$, with N positive entries and $L = |\mathcal{X}||\mathcal{Y}|$ entries, and assume that $N < L$, that is, that P_{XY} has at least one zero entry. For $0 < \eta < 1$, let P^η_{XY} be the table obtained from P_{XY} by filling its zeros with total mass η, that is,

$$P^\eta_{XY}(x,y) = \begin{cases} (1-\eta)P_{XY}(x,y), & \text{if } P_{XY}(x,y) > 0, \\ \eta/(L-N), & \text{if } P_{XY}(x,y) = 0. \end{cases} \tag{83}$$

Equivalently, $P_{XY}^{\eta} = (1-\eta)P_{XY} + \eta U_{XY}$, where U_{XY} is the uniform distribution on the set $(\mathcal{X} \times \mathcal{Y}) \setminus \operatorname{supp}(P_{XY})$.

We now bound $J(P_{XY}^{\eta})$ and $G(P_{XY}^{\eta})$ from above and from below in terms of $J(P_{XY})$ and $G(P_{XY})$, respectively. Since admissible auxiliaries correspond to rank-one decompositions, by Lemma 2, we obtain both bounds by transforming a decomposition of one table into a decomposition of the other, with almost the same entropy and mutual information.

We first collect some elementary estimates, which we will use in the proofs of the next lemmas. The proof of Lemma 15 is in Appendix D.

Lemma 15. *The following properties hold:*

1) *For every $0 < t \le 1/e$, it holds that $h_2(t) \le 2t\log(1/t)$.*
2) *For every $0 \le b \le c \le 1$, it holds that $\phi(c-b) \le \phi(c) + b\log e$.*
3) *For every $c > 0$, the function $t \mapsto t\log(c/t)$ is increasing on $(0, c/e]$.*
4) *If $z_1, \dots, z_s \ge 0$ have sum $v > 0$, then $\sum_{i=1}^{s} \phi(z_i) \le v\log(s/v)$.*
5) *Let Q be a probability distribution on a finite set $\mathcal{Z}$, let $S \subseteq \mathcal{Z}$, and let $\varepsilon = Q(\mathcal{Z} \setminus S) \in (0,1)$. Let Q_S and $Q_{\bar{S}}$ be the conditional distributions of Q on S and on $\mathcal{Z} \setminus S$, respectively. Then*

$$H(Q) = h_2(\varepsilon) + (1-\varepsilon)H(Q_S) + \varepsilon H(Q_{\bar{S}}), \tag{84}$$

and, consequently, $(1-\varepsilon)H(Q_S) \ge H(Q) - h_2(\varepsilon) - \varepsilon\log|\mathcal{Z}|$.

The next lemma shows that filling the zeros of P_{XY} increases Wyner's common information and the common entropy by at most $h_2(\eta) + \eta\log a$, where $a = \min\{|\mathcal{X}|, |\mathcal{Y}|\}$.

Lemma 16. *Let P_{XY} be a probability table on $\mathcal{X} \times \mathcal{Y}$ with at least one zero entry, let $0 < \eta < 1$, and let P_{XY}^{η} be the table defined in* (83)*. Let $a = \min\{|\mathcal{X}|, |\mathcal{Y}|\}$. Then the following properties hold:*

1) *$J(P_{XY}^{\eta}) \le J(P_{XY}) + h_2(\eta) + \eta\log a$.*
2) *$G(P_{XY}^{\eta}) \le G(P_{XY}) + h_2(\eta) + \eta\log a$.*

Proof. Let $(X,Y) \sim P_{XY}$ and $(X',Y') \sim P_{XY}^{\eta}$. Recall that $P_{XY}^{\eta} = (1-\eta)P_{XY} + \eta U_{XY}$, where U_{XY} is the uniform distribution on the zero entries of P_{XY}, that is, $U_{XY}(x,y) = 1/(L-N)$ if $P_{XY}(x,y) = 0$, and $U_{XY}(x,y) = 0$ otherwise. Moreover, P_{XY} and U_{XY} have disjoint supports. Let W be an admissible auxiliary of (X,Y). We construct from W an admissible auxiliary W' of (X',Y'), and we show that

$$H(W') \le H(W) + h_2(\eta) + \eta\log a, \qquad I(X',Y';W') \le I(X,Y;W) + h_2(\eta) + \eta\log a. \tag{85}$$

Since $G(P_{XY}^{\eta}) \le H(W')$ and $J(P_{XY}^{\eta}) \le I(X',Y';W')$, taking the infimum over W in (85) proves both properties.

Let us construct the auxiliary W'. Let $(F_w)_{w\in\mathcal{W}}$ be the components of W, so that $\sum_w F_w = P_{XY}$. We need to handle the new mass ηU_{XY}. To this end, we assume that $a = |\mathcal{X}|$, since the other case follows by interchanging rows and columns. Under this assumption, we split ηU_{XY} into its nonzero rows, so that we use at most a new components. Let $F'_1, \dots, F'_k$ be these rows, each viewed as a matrix vanishing outside that row, so that $k \le |\mathcal{X}| = a$. Each F'_j lies in a single row, and therefore has rank one. Moreover, it holds that

$$\sum_{w\in\mathcal{W}} (1-\eta)F_w + \sum_{j=1}^{k} F'_j = (1-\eta)P_{XY} + \eta U_{XY} = P^{\eta}_{XY}. \tag{86}$$

Hence, the matrices $(1-\eta)F_w$, for $w \in \mathcal{W}$, and F'_j, for $1 \le j \le k$, form a rank-one decomposition of P^{η}_{XY}. By property 2 of Lemma 2, this decomposition defines an admissible auxiliary W' of (X', Y'). More explicitly, the alphabet $\mathcal{W}'$ of W' consists of the old values $w \in \mathcal{W}$ and of k new values, say $1, \dots, k$. The component of W' at an old value w is $(1-\eta)F_w$, and the component at a new value j is F'_j. Consequently, it holds that

$$P_{W'}(w) = (1-\eta)\mu(F_w) \quad \forall w \in \mathcal{W} \quad \text{and} \quad P_{W'}(j) = \mu(F'_j) \quad \forall j \in \{1, \dots, k\}. \tag{87}$$

We now prove the leftmost inequality in (85). Let $B = \mathbb{1}_{W'\in\mathcal{W}}$ be the indicator of the event that W' takes an old value in $\mathcal{W}$. Since the old components have total mass $1-\eta$ and the new components have total mass η, we have that

$$P_B(1) = 1-\eta \quad \text{and} \quad P_B(0) = \eta. \tag{88}$$

Moreover, conditionally on $B = 1$, the distribution of W' is the same as that of W. Indeed, for every $w \in \mathcal{W}$, we have that

$$P_{W'|B}(w \mid 1) = \frac{P_{W'}(w)P_{B|W'}(1 \mid w)}{P_B(1)} = \frac{P_{W'}(w)}{1-\eta} = \frac{(1-\eta)\mu(F_w)}{1-\eta} = P_W(w). \tag{89}$$

Conditionally on $B = 0$, the distribution of W' is instead given by

$$P_{W'|B}(j \mid 0) = \frac{P_{W'}(j)P_{B|W'}(0 \mid j)}{P_B(0)} = \frac{P_{W'}(j)}{\eta} = \frac{\mu(F'_j)}{\eta}, \qquad 1 \le j \le k. \tag{90}$$

Hence, we have that $P_{W'|B=1} = P_W$ and $P_{W'|B=0} = (\mu(F'_1)/\eta, \dots, \mu(F'_k)/\eta)$. Therefore, by (24) and applying property 5 of Lemma 15 to the distribution of W', we obtain that

$$\begin{aligned} H(W') &= h_2(\eta) + (1-\eta)H(W) + \eta H\left(\frac{\mu(F'_1)}{\eta}, \dots, \frac{\mu(F'_k)}{\eta}\right) \\ &\le h_2(\eta) + H(W) + \eta \log a, \end{aligned} \tag{91}$$

where the inequality follows since $H\left(\frac{\mu(F'_1)}{\eta}, \dots, \frac{\mu(F'_k)}{\eta}\right) \le \log a$ and $(1-\eta)H(W) \le H(W)$. Hence (91) proves the leftmost inequality in (85).

We now prove the rightmost inequality in (85). We recall that

$$I(X', Y'; W') = H(X', Y') - H(X', Y' \mid W'). \tag{92}$$

We compute the two terms on the right-hand side of (92) separately. We first compute $H(X', Y' \mid W')$. To this end, from (17) we have that

$$P_{X'Y'|W'}(\cdot, \cdot \mid w) = \frac{(1-\eta)F_w}{(1-\eta)\mu(F_w)} = \frac{F_w}{\mu(F_w)} \quad \forall w \in \mathcal{W} \tag{93}$$

and

$$P_{X'Y'|W'}(\cdot, \cdot \mid j) = \frac{F'_j}{\mu(F'_j)} \quad \forall 1 \le j \le k. \tag{94}$$

Using these identities together with the definition of conditional entropy, $P_{W'}(w) = (1-\eta)\mu(F_w)$ and $P_{W'}(j) = \mu(F'_j)$, we obtain that

$$\begin{aligned} H(X', Y' \mid W') &= \sum_{w \in \mathcal{W}} P_{W'}(w) H(X', Y' \mid W' = w) + \sum_{j=1}^{k} P_{W'}(j) H(X', Y' \mid W' = j) \\ &= (1-\eta) \sum_{w \in \mathcal{W}} \mu(F_w) H\left(\frac{F_w}{\mu(F_w)}\right) + \sum_{j=1}^{k} \mu(F'_j) H\left(\frac{F'_j}{\mu(F'_j)}\right) \\ &= (1-\eta) H(X, Y \mid W) + \sum_{j=1}^{k} \mu(F'_j) H\left(\frac{F'_j}{\mu(F'_j)}\right), \end{aligned} \tag{95}$$

where the last equality follows from property 3 of Lemma 2.

We next compute $H(X', Y')$. Let $B = \mathbb{1}_{(X',Y') \in \mathrm{supp}(P_{XY})}$ be the indicator of the event that $(X', Y') \in \mathrm{supp}(P_{XY})$. Recalling that $P^{\eta}_{XY} = (1-\eta)P_{XY} + \eta U_{XY}$, and since P_{XY} and U_{XY} have disjoint supports, we have that $P_B(1) = 1 - \eta$ and $P_B(0) = \eta$. Moreover, conditionally on $B = 1$, the pair (X', Y') has distribution P_{XY}, while conditionally on $B = 0$, it has distribution U_{XY}. Then, applying property 5 of Lemma 15, we obtain that

$$H(X', Y') = h_2(\eta) + (1-\eta) H(X, Y) + \eta H(U_{XY}). \tag{96}$$

We now express $H(U_{XY})$ in terms of the components $F'_1, \ldots, F'_k$ of W'. Let $(X'', Y'') \sim U_{XY}$, and let R be the index j of the row of F'_j that contains (X'', Y''). Since F'_j coincides with ηU_{XY} on its row and vanishes elsewhere, we have that

$$P_R(j) = \frac{\mu(F'_j)}{\eta} \quad \text{and} \quad P_{X''Y''|R}(\cdot, \cdot \mid j) = \frac{F'_j}{\mu(F'_j)} \qquad \forall\, 1 \le j \le k. \tag{97}$$

Since R is a function of (X'', Y''), the chain rule gives $H(X'', Y'') = H(R) + H(X'', Y'' \mid R)$, that is,

$$H(U_{XY}) = H\Big(\frac{\mu(F'_1)}{\eta}, \ldots, \frac{\mu(F'_k)}{\eta}\Big) + \sum_{j=1}^{k} \frac{\mu(F'_j)}{\eta} H\Big(\frac{F'_j}{\mu(F'_j)}\Big). \tag{98}$$

Substituting (98) into (96), we obtain that

$$H(X',Y') = h_2(\eta) + (1-\eta)H(X,Y) + \eta H\Big(\frac{\mu(F_1')}{\eta},\dots,\frac{\mu(F_k')}{\eta}\Big) + \sum_{j=1}^{k}\mu(F_j')H\Big(\frac{F_j'}{\mu(F_j')}\Big). \tag{99}$$

Finally, substituting (99) and (95) into (92), the sums over j cancel, and we obtain that

$$\begin{aligned} I(X',Y';W') &= h_2(\eta) + (1-\eta)\big(H(X,Y) - H(X,Y \mid W)\big) + \eta H\Big(\frac{\mu(F_1')}{\eta},\dots,\frac{\mu(F_k')}{\eta}\Big) \\ &\le h_2(\eta) + I(X,Y;W) + \eta\log a, \end{aligned}$$

where the inequality follows from $H(X,Y) - H(X,Y \mid W) = I(X,Y;W)$, $(1-\eta)I(X,Y;W) \le I(X,Y;W)$, and $H\big(\frac{\mu(F_1')}{\eta},\dots,\frac{\mu(F_k')}{\eta}\big) \le \log k \le \log a$. This proves the rightmost inequality in (85), which completes the proof. □

We now bound $J(P_{XY}^{\eta})$ and $G(P_{XY}^{\eta})$ from below. The idea is to turn every admissible auxiliary of P_{XY}^{η} into an admissible auxiliary of P_{XY} with almost the same costs. In this case, the difficulty lies in the fact that a decomposition of P_{XY}^{η} is not necessarily a valid decomposition of P_{XY}, since some of its components may put mass on the zeros of P_{XY}. Despite this, we first show, in the next lemma, whose proof is in Appendix E, that a product distribution that puts a small mass on a set of entries can be restricted to a set of the form $R \times C$ that avoids these entries, losing only a small part of its mass. Since, by Lemma 1, every component of an admissible auxiliary is a positive multiple of a product distribution, this allows us to restrict every component of a decomposition of P_{XY}^{η} to a set avoiding the zeros of P_{XY}. We then use this restriction to prove the lower bounds in Lemma 18.

Lemma 17. *Let $a = \min\{|\mathcal{X}|,|\mathcal{Y}|\}$, let $D \subseteq \mathcal{X}\times\mathcal{Y}$, and let Q and Q' be probability distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\delta = (Q\otimes Q')(D)$. Then there exists a set $S = R\times C$, with $R\subseteq\mathcal{X}$ and $C\subseteq\mathcal{Y}$ possibly empty, that does not intersect D and such that*

$$(Q\otimes Q')\big((\mathcal{X}\times\mathcal{Y})\setminus S\big) \le 2\sqrt{a\delta}. \tag{100}$$

The next lemma gives the lower bounds on $J(P_{XY}^{\eta})$ and $G(P_{XY}^{\eta})$.

Lemma 18. *Let P_{XY} be a probability table on $\mathcal{X}\times\mathcal{Y}$ with at least one zero entry, let $0<\eta<1$, and let P_{XY}^{η} be the table defined in* (83)*. Let $a = \min\{|\mathcal{X}|,|\mathcal{Y}|\}$ and $\gamma = \sqrt{a\eta}$, and assume that $\gamma \le 1/4$. Then the following properties hold:*

1) $J(P_{XY}) \le J(P_{XY}^{\eta}) + h_2(2\gamma) + (2\gamma+\eta)\log L$.

2) $G(P_{XY}) \le G(P_{XY}^{\eta}) + 2\gamma\log\dfrac{ea}{2\gamma}$.

Proof. Let $(X,Y) \sim P_{XY}$ and $(X',Y') \sim P_{XY}^{\eta}$, and let W' be an admissible auxiliary of (X',Y'). We show that we can construct from W' an admissible auxiliary W of (X,Y) such that

$$H(W) \leq H(W') + 2\gamma \log \frac{ea}{2\gamma} \tag{101}$$

and

$$I(X,Y;W) \leq I(X',Y';W') + h_2(2\gamma) + (2\gamma+\eta)\log L. \tag{102}$$

Then, since $G(P_{XY}) \leq H(W)$ and $J(P_{XY}) \leq I(X,Y;W)$, taking the infimum over W' in (101) and (102) proves properties 2 and 1, respectively.

We first construct W. Let $D = (\mathcal{X}\times\mathcal{Y}) \setminus \operatorname{supp}(P_{XY})$ be the set of zero entries of P_{XY}. For every value $w \in \mathcal{W}'$ of W', let F_w be the component of W' at w, let $\mu_w = \mu(F_w)$ be its mass, and let

$$d_w = \sum_{(x,y)\in D} F_w(x,y) \tag{103}$$

be the mass that F_w puts on D. The components F_w cannot be used directly for P_{XY} when $d_w > 0$, since every component of a decomposition of P_{XY} must vanish on D. We therefore proceed in two steps. First, we restrict each component F_w to a set that avoids D. Second, we cover what is still missing from P_{XY} by its rows or columns.

We start with the first step. Fix a value $w \in \mathcal{W}'$. By Lemma 1, we can write $F_w = \mu_w\,(Q_w \otimes Q'_w)$, for some probability distributions Q_w on $\mathcal{X}$ and Q'_w on $\mathcal{Y}$. Moreover, the product distribution $Q_w \otimes Q'_w$ puts mass d_w/μ_w on D. Hence, by Lemma 17, there exists a set $S_w = R_w \times C_w$ that does not intersect D and such that $Q_w \otimes Q'_w$ puts mass at most $2\sqrt{ad_w/\mu_w}$ outside S_w. Since $F_w = \mu_w\,(Q_w \otimes Q'_w)$, every mass computed under F_w is μ_w times the corresponding mass computed under $Q_w \otimes Q'_w$. Hence, multiplying by μ_w the bound given by Lemma 17, we obtain that the mass that F_w puts outside S_w, which we denote by r_w, satisfies

$$r_w = \sum_{(x,y)\notin S_w} F_w(x,y) \leq 2\mu_w\sqrt{\frac{ad_w}{\mu_w}} = 2\sqrt{a\mu_w d_w}. \tag{104}$$

Now, we let $\widetilde{F}_w$ be the restriction of F_w to S_w, that is, the matrix equal to F_w on S_w and to 0 elsewhere, and we call it the restricted component at w. Since $F_w = \mu_w\,(Q_w \otimes Q'_w)$ and $S_w = R_w \times C_w$, we have that

$$\widetilde{F}_w(x,y) = \mu_w\left(Q_w(x)\mathbb{1}_{R_w}(x)\right)\left(Q'_w(y)\mathbb{1}_{C_w}(y)\right), \tag{105}$$

so that $\widetilde{F}_w$ is, if nonzero, of rank one. Moreover, its mass is $\mu(\widetilde{F}_w) = \mu_w - r_w$. We now observe that the restricted components $(\widetilde{F}_w)_{w\in\mathcal{W}'}$ do not exceed P_{XY}. Indeed, by construction, they vanish on the zero

entries of P_{XY}, and, on the positive entries of P_{XY}, they are at most the components F_w, whose sum is $P^{\eta}_{XY} = (1-\eta)P_{XY} \le P_{XY}$ by (83). Hence, the remainder

$$F^{\mathrm{rem}} = P_{XY} - \sum_{w \in \mathcal{W}'} \widetilde{F}_w \tag{106}$$

is a nonnegative matrix.

We now carry out the second step. We assume that $a = |\mathcal{X}|$, the other case being symmetric with rows and columns interchanged, and we split the remainder F^{rem} into its nonzero rows. In this way, we obtain at most $|\mathcal{X}| = a$ matrices of rank one, which we denote by $F^{\mathrm{rem}}_1, \dots, F^{\mathrm{rem}}_k$ and call them the remainder components, and which sum to F^{rem}. Hence, the nonzero restricted components, together with the remainder components, are nonnegative matrices of rank one that sum to P_{XY} by (106). Therefore, they form a rank-one decomposition of P_{XY}. By property 2 of Lemma 2, this decomposition defines an admissible auxiliary W of (X, Y). Thus, we constructed from W' an admissible auxiliary W of (X, Y).

We now show that the auxiliary W satisfies (101) and (102). To this end, we first show that the remainder F^{rem} has small mass. Recall that μ_w is the mass of the component F_w of W', that r_w is the mass removed from F_w by the restriction to S_w, and that d_w is the mass that F_w puts on the zero entries of P_{XY}. Since the restricted component $\widetilde{F}_w$ has mass $\mu_w - r_w$, and since P_{XY} has mass 1, the remainder $F^{\mathrm{rem}} = P_{XY} - \sum_w \widetilde{F}_w$ has mass

$$\mu(F^{\mathrm{rem}}) = 1 - \sum_w (\mu_w - r_w) = \sum_w r_w, \tag{107}$$

where we used $\sum_w \mu_w = 1$, since the components of W' sum to the probability table P^{η}_{XY}. In other words, the remainder consists exactly of the mass removed by the restrictions. We now bound this mass. By (104), each r_w is at most $2\sqrt{a\mu_w d_w}$. Hence, by the Cauchy–Schwarz inequality, we get that

$$\sum_w r_w \le 2\sqrt{a} \sum_w \sqrt{\mu_w}\sqrt{d_w} \le 2\sqrt{a} \Big(\sum_w \mu_w \Big)^{1/2} \Big(\sum_w d_w \Big)^{1/2} = 2\sqrt{a\eta}, \tag{108}$$

where the last equality follows because $\sum_w \mu_w = 1$, and $\sum_w d_w$ is the total mass that the components of W' put on the zero entries of P_{XY}, that is, the mass of P^{η}_{XY} on the zero entries of P_{XY}, which equals to η by (83). Combining (107) and (108), and recalling that $\gamma = \sqrt{a\eta}$, we obtain that

$$\mu(F^{\mathrm{rem}}) = \sum_w r_w \le 2\gamma. \tag{109}$$

We now prove (101). We recall that the components of W' are the matrices F_w with masses μ_w. The components of W are the nonzero restricted components $\widetilde{F}_w$, with masses $\mu_w - r_w$, and the remainder components $F^{\mathrm{rem}}_1, \dots, F^{\mathrm{rem}}_k$, with $k \le a$, whose masses sum to $\mu(F^{\mathrm{rem}})$. Thus, by (24), we get that

$$H(W') = \sum_w \phi(\mu_w), \qquad H(W) = \sum_w \phi(\mu_w - r_w) + \sum_{j=1}^{k} \phi\big(\mu(F^{\mathrm{rem}}_j)\big). \tag{110}$$

We bound the two sums in the expression of $H(W)$ separately. For the first sum, property 2 of Lemma 15, with $c = \mu_w$ and $b = r_w$, gives $\phi(\mu_w - r_w) \le \phi(\mu_w) + r_w \log e$ for every w. Summing over w, and using (110) and (107), we obtain that

$$\sum_w \phi(\mu_w - r_w) \le H(W') + \mu(F^{\mathrm{rem}}) \log e. \tag{111}$$

For the second sum, property 4 of Lemma 15, applied to the $k \le a$ masses $\mu(F_1^{\mathrm{rem}}), \dots, \mu(F_k^{\mathrm{rem}})$, whose sum is $\mu(F^{\mathrm{rem}})$, gives

$$\sum_{j=1}^{k} \phi\big(\mu(F_j^{\mathrm{rem}})\big) \le \mu(F^{\mathrm{rem}}) \log \frac{k}{\mu(F^{\mathrm{rem}})} \le \mu(F^{\mathrm{rem}}) \log \frac{a}{\mu(F^{\mathrm{rem}})}. \tag{112}$$

Substituting (111) and (112) into the expression of $H(W)$ in (110), we obtain that

$$H(W) \le H(W') + \mu(F^{\mathrm{rem}}) \log \frac{ea}{\mu(F^{\mathrm{rem}})} \le H(W') + 2\gamma \log \frac{ea}{2\gamma}, \tag{113}$$

where the last inequality follows from (109), since the function $t \mapsto t \log(ea/t)$ is increasing for $t \le a$, by property 3 of Lemma 15, and $2\gamma \le 1 \le a$. This proves (101).

We now prove (102). Since $I(X, Y; W) = H(X, Y) - H(X, Y \mid W)$ and $I(X', Y'; W') = H(X', Y') - H(X', Y' \mid W')$, it suffices to bound $H(X, Y \mid W)$ from below in terms of $H(X', Y' \mid W')$, and $H(X, Y)$ from above in terms of $H(X', Y')$.

We first bound $H(X, Y \mid W)$. Applying (25) to W', whose components are the matrices F_w with masses μ_w, we get that

$$H(X', Y' \mid W') = \sum_w \mu_w \, H\Big(\frac{F_w}{\mu_w}\Big). \tag{114}$$

Applying (25) to W, and discarding the nonnegative contribution of the remainder components $F_1^{\mathrm{rem}}, \dots, F_k^{\mathrm{rem}}$, we get that

$$H(X, Y \mid W) \ge \sum_w (\mu_w - r_w) \, H\Big(\frac{\widetilde{F}_w}{\mu_w - r_w}\Big), \tag{115}$$

where the terms with $\mu_w = r_w$ are zero. We now compare, for every w, the term indexed by w in (115) with the term indexed by w in (114). The idea is that $\widetilde{F}_w$ is obtained from F_w by removing only a small fraction of its mass, so that their normalized entropies are close.

Fix w, and let $\theta_w = r_w/\mu_w$ be the fraction of the mass of F_w removed by the restriction to S_w, so that $\mu_w - r_w = (1 - \theta_w)\mu_w$. The normalized component F_w/μ_w is a probability distribution on $\mathcal{X} \times \mathcal{Y}$, which puts mass $1 - \theta_w$ on S_w and mass θ_w outside S_w. Since $\widetilde{F}_w$ coincides with F_w on S_w and vanishes elsewhere, the conditional distribution of F_w/μ_w on S_w is exactly $\widetilde{F}_w/(\mu_w - r_w)$. Therefore, since $\mathcal{X} \times \mathcal{Y}$ has L elements, property 5 of Lemma 15 gives

$$(1 - \theta_w) \, H\Big(\frac{\widetilde{F}_w}{\mu_w - r_w}\Big) \ge H\Big(\frac{F_w}{\mu_w}\Big) - h_2(\theta_w) - \theta_w \log L. \tag{116}$$

In other words, removing the mass θ_w decreases the entropy by at most $h_2(\theta_w)+\theta_w \log L$. Inequality (116) also holds in the extreme cases $\theta_w = 0$, where the two normalized matrices coincide, and $\theta_w = 1$, where its left-hand side is read as 0 and its right-hand side is at most $H(F_w/\mu_w) - \log L \le 0$.

Multiplying (116) by μ_w, so that its left-hand side becomes the term indexed by w in (115), summing over w, and using (114), we obtain from (115) that

$$H(X,Y \mid W) \ge H(X',Y' \mid W') - \sum_w \mu_w h_2(\theta_w) - \Big(\sum_w r_w\Big) \log L. \tag{117}$$

It remains to bound the two sums on the right-hand side of (117). By (109), we have that $\sum_w r_w \le 2\gamma$. Moreover, since the weights μ_w sum to 1 and h_2 is concave, Jensen's inequality gives $\sum_w \mu_w h_2(\theta_w) \le h_2\big(\sum_w \mu_w \theta_w\big)$. Since $\sum_w \mu_w \theta_w = \sum_w r_w \le 2\gamma \le 1/2$, and h_2 is increasing on $[0, 1/2]$, we obtain that

$$\sum_w \mu_w h_2(\theta_w) \le h_2(2\gamma). \tag{118}$$

Substituting (109) and (118) into (117), we obtain that

$$H(X,Y \mid W) \ge H(X',Y' \mid W') - h_2(2\gamma) - 2\gamma \log L. \tag{119}$$

We next bound $H(X,Y)$. Recall that $P^\eta_{XY} = (1-\eta)P_{XY} + \eta U_{XY}$, where P_{XY} and U_{XY} have disjoint supports. Hence, by property 5 of Lemma 15, applied to P^η_{XY} and to the support of P_{XY}, we have that

$$H(X',Y') = h_2(\eta) + (1-\eta)H(X,Y) + \eta H(U_{XY}) \ge (1-\eta)H(X,Y). \tag{120}$$

Since $H(X,Y) \le \log L$, (120) gives

$$H(X,Y) \le H(X',Y') + \eta H(X,Y) \le H(X',Y') + \eta \log L. \tag{121}$$

Finally, combining (119) and (121), we obtain that

$$I(X,Y;W) = H(X,Y) - H(X,Y \mid W) \le H(X',Y') - H(X',Y' \mid W') + h_2(2\gamma) + (2\gamma+\eta)\log L. \tag{122}$$

Since $H(X',Y') - H(X',Y' \mid W') = I(X',Y';W')$, (122) proves (102). □

We now combine the bounds of Lemmas 16 and 18 showing that, for an appropriate value of η, $J(P^\eta_{XY})$ and $G(P^\eta_{XY})$ are close to $J(P_{XY})$ and $G(P_{XY})$, respectively.

Lemma 19. *Let P_{XY} be a probability table on $\mathcal{X} \times \mathcal{Y}$ with at least one zero entry, let $L = |\mathcal{X}||\mathcal{Y}|$, and let P^η_{XY} be the table defined in* (83)*. If $L \ge 2$ and $0 < \eta \le 2^{-16}L^{-9}$, then*

$$|J(P^\eta_{XY}) - J(P_{XY})| \le \frac{1}{16L^2} \qquad \textit{and} \qquad |G(P^\eta_{XY}) - G(P_{XY})| \le \frac{1}{16L^2}. \tag{123}$$

Proof. Let $a = \min\{|\mathcal{X}|, |\mathcal{Y}|\}$ and $\gamma = \sqrt{a\eta}$. Since $a \le L$ and $\eta \le 2^{-16}L^{-9}$, we have that

$$\eta \le \gamma \le \sqrt{L\eta} \le 2^{-8}L^{-4} \le \frac{1}{2e}, \tag{124}$$

where the first inequality follows from $\eta^2 \le a\eta$. Moreover, $\gamma \le 1/4$ so that Lemma 18 applies. Combining Lemma 18 with Lemma 16 we obtain that

$$J(P_{XY}) - h_2(2\gamma) - (2\gamma+\eta)\log L \;\le J(P_{XY}^{\eta}) \;\le\; J(P_{XY}) + h_2(\eta) + \eta\log a, \tag{125}$$

$$G(P_{XY}) - 2\gamma\log\frac{ea}{2\gamma} \;\le G(P_{XY}^{\eta}) \;\le\; G(P_{XY}) + h_2(\eta) + \eta\log a. \tag{126}$$

Hence, to prove the lemma, it suffices to show that the terms

$$h_2(\eta) + \eta\log a, \qquad h_2(2\gamma) + (2\gamma+\eta)\log L, \qquad 2\gamma\log\frac{ea}{2\gamma} \tag{127}$$

appearing in (125) and (126) are at most $1/(16L^2)$. To this end, we first show that each of them is at most $4\gamma\log\frac{2L}{\gamma}$, and then that $4\gamma\log\frac{2L}{\gamma}$ is at most $1/(16L^2)$.

By properties 1 and 3 of Lemma 15, we have that $h_2(t) \le 2t\log(1/t)$ and that $t\log(1/t)$ is increasing for $t \le 1/e$. Moreover, since $\eta \le \gamma \le 1/e$ and $2\gamma \le 1/e$ by (124), and since $a \le L$ and $ea/2 \le (e/2)L \le 2L$, we obtain that

$$h_2(\eta) + \eta\log a \le 2\eta\log\frac{1}{\eta} + \eta\log L \le 2\gamma\log\frac{1}{\gamma} + \gamma\log L,$$

$$h_2(2\gamma) + (2\gamma+\eta)\log L \le 4\gamma\log\frac{1}{2\gamma} + 3\gamma\log L \le 4\gamma\log\frac{1}{\gamma} + 3\gamma\log L,$$

$$2\gamma\log\frac{ea}{2\gamma} \le 2\gamma\log\frac{2L}{\gamma} = 2\gamma\log\frac{1}{\gamma} + 2\gamma\log(2L).$$

Since $\log\frac{1}{\gamma} \ge 0$ and $\log L \le \log(2L)$, each of the right-hand sides of the above inequalities is at most

$$4\gamma\log\frac{1}{\gamma} + 4\gamma\log(2L) = 4\gamma\log\frac{2L}{\gamma}. \tag{128}$$

It remains to show that $4\gamma\log\frac{2L}{\gamma} \le 1/(16L^2)$. By property 3 of Lemma 15, applied with $c = 2L$, the function $t \mapsto t\log(2L/t)$ is increasing for $t \le 2L/e$. Thus, since $\gamma \le 2^{-8}L^{-4} \le 2L/e$ by (124), we can replace γ by $2^{-8}L^{-4}$ in (128), and we obtain that

$$4\gamma\log\frac{2L}{\gamma} \le 4\cdot 2^{-8}L^{-4}\log\left(2\cdot 2^{8}L^{5}\right) = \frac{9+5\log L}{64L^4}. \tag{129}$$

Finally, since $\log L \le L-1$ for every $L \ge 2$, we have that $9 + 5\log L \le 5L + 4 \le 4L^2$. Hence, the right-hand side of (129) is at most $\frac{4L^2}{64L^4} = \frac{1}{16L^2}$. Therefore all the terms in (127) are at most $1/(16L^2)$, and, by (125) and (126), this proves (123). □

B. Proof of Theorem 3

In this subsection, we prove Theorem 3. Let (Γ, k) be an instance of MAXIMUM EDGE BICLIQUE, and let (P_{XY}, θ) be the instance produced from (Γ, k) by the reduction of Section IV or of Section V, so that $P_{XY} = P_{XY}^{A}$ and $\theta = r - \frac{1}{8N^2}$ for WYNER THRESHOLD, as in (54), and $P_{XY} = \widehat{P}_{XY}^{A}$ and

$\theta = r_A - \frac{1}{2N}$ for COMMON ENTROPY THRESHOLD, as in (78). Let N and L be the numbers of positive entries and of entries of P_{XY}, respectively. If P_{XY} is strictly positive, then the original reduction already produces a strictly positive table, and there is nothing to prove. Hence, we assume that P_{XY} has at least one zero entry, that is, $N < L$. Let

$$\eta = 2^{-16}L^{-9}, \tag{130}$$

and let $\widetilde{P}_{XY} = P^{\eta}_{XY}$ be the table obtained from P_{XY} by filling its zeros with total mass η, as in (83). Therefore, the instance of WYNER THRESHOLD or of COMMON ENTROPY THRESHOLD associated with (Γ, k) is the pair $(\widetilde{P}_{XY}, \theta)$, with the same threshold θ as in the original reduction.

Proof of Theorem 3. The proof is along the same lines as those of Theorems 1 and 2. We start by showing that WYNER THRESHOLD and COMMON ENTROPY THRESHOLD are NP-hard on strictly positive tables. To this end, we show that the map $(\Gamma, k) \mapsto (\widetilde{P}_{XY}, \theta)$ described above is a polynomial-time reduction from MAXIMUM EDGE BICLIQUE to WYNER THRESHOLD, and from MAXIMUM EDGE BICLIQUE to COMMON ENTROPY THRESHOLD, respectively.

We first show that the reduction runs in polynomial time in the size of (Γ, k). By the proofs of Theorems 1 and 2, the pair (P_{XY}, θ) is computed in polynomial time. Thus, L is polynomial in the size of (Γ, k), and $\eta = 2^{-16}L^{-9}$ is a rational number with $O(\log L)$ bits. Moreover, by (83), every entry of $\widetilde{P}_{XY}$ is either $(1-\eta)P_{XY}(x,y)$ or $\eta/(L-N)$, where $L - N \geq 1$. Hence, every entry is a rational number of polynomial size, computable in polynomial time. Therefore, since $\widetilde{P}_{XY}$ has L entries, the whole reduction runs in polynomial time.

We now show that the reduction is correct, that is, $(\widetilde{P}_{XY}, \theta)$ is a yes-instance of WYNER THRESHOLD or of COMMON ENTROPY THRESHOLD if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. We first observe that filling the zeros changes J and G only slightly. Since P_{XY} has at least one zero entry, we have that $L \geq 2$, and hence Lemma 19 gives

$$|J(\widetilde{P}_{XY}) - J(P_{XY})| \leq \zeta \qquad \text{and} \qquad |G(\widetilde{P}_{XY}) - G(P_{XY})| \leq \zeta, \qquad \text{where } \zeta = \frac{1}{16L^2}. \tag{131}$$

We prove the correctness for WYNER THRESHOLD. In this case, $P_{XY} = P^{A}_{XY}$ and $\theta = r - \frac{1}{8N^2}$. Since $N \leq L$, we have that $\zeta \leq \frac{1}{16N^2}$. Assume first that (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE, that is, $\beta(\Gamma) \geq k$. We prove that $(\widetilde{P}_{XY}, \theta)$ is a yes-instance of WYNER THRESHOLD. By property 2 of Lemma 6, we have that $\beta(A) > d$, and hence, by Lemma 9, it follows that $J(P^{A}_{XY}) < r - \frac{1}{4N^2}$. Thus, by (131) and recalling that $\zeta \leq \frac{1}{16N^2}$ and $\theta = r - \frac{1}{8N^2}$, we get that

$$J(\widetilde{P}_{XY}) \leq J(P^{A}_{XY}) + \zeta < r - \frac{1}{4N^2} + \frac{1}{16N^2} = r - \frac{3}{16N^2} < \theta. \tag{132}$$

Since $J(\widetilde{P}_{XY})$ is the infimum of $I(X,Y;W)$ over all admissible auxiliaries W, it follows from (132) that there exists an admissible auxiliary W such that $I(X,Y;W) < \theta$. Hence, $(\widetilde{P}_{XY}, \theta)$ is a yes-instance of WYNER THRESHOLD.

Assume now that (Γ, k) is a no-instance of MAXIMUM EDGE BICLIQUE, that is, $\beta(\Gamma) < k$. We prove that $(\widetilde{P}_{XY}, \theta)$ is a no-instance of WYNER THRESHOLD. By property 2 of Lemma 6, we have that $\beta(A) \leq d$, and hence, by Lemma 9, it follows that $J(P^A_{XY}) = r$. Thus, by (131) and recalling that $\zeta \leq \frac{1}{16N^2}$ and $\theta = r - \frac{1}{8N^2}$, we get that

$$J(\widetilde{P}_{XY}) \geq J(P^A_{XY}) - \zeta \geq r - \frac{1}{16N^2} > \theta. \tag{133}$$

Since $J(\widetilde{P}_{XY})$ is the infimum of $I(X,Y;W)$ over all admissible auxiliaries W, it follows from (133) that every admissible auxiliary W satisfies $I(X,Y;W) \geq J(\widetilde{P}_{XY}) > \theta$. Hence, $(\widetilde{P}_{XY}, \theta)$ is a no-instance of WYNER THRESHOLD.

The correctness for COMMON ENTROPY THRESHOLD follows in the same way, with G, r_A and the gap $1/N$ of Lemma 14 in place of J, r and the gap $1/(4N^2)$ of Lemma 9. In this case, $P_{XY} = \widehat{P}^A_{XY}$, $\theta = r_A - \frac{1}{2N}$. Moreover, since $L \geq N \geq 1$, we have that $L^2 \geq N^2 \geq N$, and hence $\zeta \leq \frac{1}{16N}$. Then, by property 2 of Lemma 6 and Lemma 14, we have that $G(\widehat{P}^A_{XY}) < r_A - \frac{1}{N}$ if $\beta(\Gamma) \geq k$, and $G(\widehat{P}^A_{XY}) = r_A$ if $\beta(\Gamma) < k$. Hence, by (131), we get that

$$G(\widetilde{P}_{XY}) < r_A - \frac{1}{N} + \frac{1}{16N} < \theta \quad \text{if } \beta(\Gamma) \geq k, \qquad G(\widetilde{P}_{XY}) \geq r_A - \frac{1}{16N} > \theta \quad \text{if } \beta(\Gamma) < k, \tag{134}$$

and the conclusion follows as for WYNER THRESHOLD. This proves that WYNER THRESHOLD and COMMON ENTROPY THRESHOLD are NP-hard on strictly positive tables.

It remains to prove the inapproximability results. We first consider WYNER THRESHOLD. Suppose, by contradiction, that there exists a polynomial-time algorithm that, given a strictly positive rational probability table, outputs a value v within the additive error $\frac{1}{16L^2}$ of its Wyner's common information. We apply this algorithm to $\widetilde{P}_{XY}$, and we show that comparing its output v with θ decides MAXIMUM EDGE BICLIQUE. Since $N \leq L$, we have that $\frac{1}{16L^2} \leq \frac{1}{16N^2}$. If (Γ, k) is a yes-instance, then (132) gives

$$v \leq J(\widetilde{P}_{XY}) + \frac{1}{16N^2} < r - \frac{3}{16N^2} + \frac{1}{16N^2} = r - \frac{1}{8N^2} = \theta. \tag{135}$$

If (Γ, k) is a no-instance, then (133) gives

$$v \geq J(\widetilde{P}_{XY}) - \frac{1}{16N^2} \geq r - \frac{1}{16N^2} - \frac{1}{16N^2} = r - \frac{1}{8N^2} = \theta. \tag{136}$$

By (135) and (136), we have that $v < \theta$ if and only if (Γ, k) is a yes-instance of MAXIMUM EDGE BICLIQUE. Therefore, running the reduction, then the algorithm on $\widetilde{P}_{XY}$, and finally comparing v with θ would decide MAXIMUM EDGE BICLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$.

The argument for COMMON ENTROPY THRESHOLD is the same, with an algorithm that outputs a value v such that $|v - G(P_{XY})| \leq 8/L$. Since $L \geq 32N$ by property 1 of Lemma 11, we have that $\frac{8}{L} \leq \frac{1}{4N}$, and hence (134) gives

$$v < r_A - \frac{11}{16N} < r_A - \frac{1}{2N} = \theta \quad \text{if } \beta(\Gamma) \geq k, \qquad v \geq r_A - \frac{5}{16N} > r_A - \frac{1}{2N} = \theta \quad \text{if } \beta(\Gamma) < k. \tag{137}$$

Again, comparing v with θ would decide MAXIMUM EDGE BICLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

VII. CONCLUSIONS AND OPEN PROBLEMS

We have proved that Wyner's common information and the common entropy are NP-hard to compute, with explicit hardness bounds for additive approximation. These results hold even for strictly positive rational probability tables. Both reductions exploit the relation between conditional independence and rectangular supports, while accounting for the different objectives $I(X, Y; W)$ and $H(W)$.

The main question left by these results is how accurately the two quantities can be approximated in polynomial time. Our hardness bounds decrease with the alphabet sizes. Can one prove, for each quantity, that approximation within some fixed additive error $\varepsilon > 0$ is already NP-hard, independently of those sizes? On the algorithmic side, an equally important goal is to design efficient approximation algorithms with explicit error guarantees. Ideally, these algorithms would also construct an admissible auxiliary whose objective value is provably close to the optimum.

The extension to strictly positive tables suggests a further question. Does hardness persist when every entry lies between c/L and C/L, where L is the number of entries and $0 < c < 1 < C$ are fixed constants? Our perturbation argument allows much smaller probabilities and does not settle this case. Understanding this restriction could help identify which features of a distribution make the optimization difficult.

ACKNOWLEDGEMENTS

U. Vaccaro wants to acknowledge useful discussions with F. Cicalese and M. Kocaoglu about the computational complexity of computing the common entropy.

APPENDIX A
PROOF OF LEMMA 2

Proof. We first prove property 1. Let W be an admissible auxiliary of (X, Y). We first show that each component of W has rank one, and then that its components form a rank-one decomposition of P_{XY}. Fix $w \in \mathcal{W}$. By the definition (16) of F_w, the chain rule, and the conditional independence (9), for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ we have that

$$F_w(x, y) = P_W(w)\, P_{XY|W}(x, y \mid w) = P_W(w)\, P_{X|W}(x \mid w)\, P_{Y|W}(y \mid w),$$

which proves (22). Since $\mu(F_w) = P_W(w) > 0$, (22) is of the form (20) with $Q = P_{X|W}(\cdot \mid w)$ and $Q' = P_{Y|W}(\cdot \mid w)$. Thus, Lemma 1 implies that F_w is a nonzero matrix of rank one whose support is a rectangle. Moreover, by (12), $\sum_{w' \in \mathcal{W}} F_{w'} = P_{XY}$, so that the components of W form a rank-one decomposition of P_{XY}. Finally, since the components of W are all nonnegative, we also have that $F_w \leq P_{XY}$. Hence $\operatorname{supp}(F_w) \subseteq \operatorname{supp}(P_{XY})$, and $\operatorname{supp}(F_w)$ is a rectangle of P_{XY}.

We now prove property 2. Let $(F_w)_{w \in \mathcal{W}}$ be a rank-one decomposition of P_{XY}, and let $P_{XYW}(x, y, w) = F_w(x, y)$. Since the matrices F_w are nonnegative and sum to P_{XY}, the numbers $P_{XYW}(x, y, w)$ are nonnegative and sum to one, so that P_{XYW} is a joint distribution of (X, Y, W) satisfying (12). Moreover, $P_W(w) = \mu(F_w) > 0$, since each F_w is nonzero. Hence, W is an auxiliary of (X, Y), and its components are the matrices F_w by construction. It remains to show that W is admissible. Fix $w \in \mathcal{W}$. By Lemma 1, $F_w = \mu(F_w)\,(Q_w \otimes Q'_w)$ for some probability distributions Q_w on $\mathcal{X}$ and Q'_w on $\mathcal{Y}$, so that, by (17), we have that

$$P_{XY|W}(\cdot, \cdot \mid w) = Q_w \otimes Q'_w. \tag{138}$$

Summing (138) over $y \in \mathcal{Y}$, and using that Q'_w sums to one, we obtain that the conditional distribution of X given $W = w$ satisfies

$$P_{X|W}(x \mid w) = \sum_{y \in \mathcal{Y}} Q_w(x)\, Q'_w(y) = Q_w(x),$$

and, similarly, $P_{Y|W}(y \mid w) = Q'_w(y)$. Hence, it follows from (138) that $P_{XY|W}(x, y \mid w) = P_{X|W}(x \mid w)\, P_{Y|W}(y \mid w)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and all $w \in \mathcal{W}$, that is, $X \perp\!\!\!\perp Y \mid W$ by (9), and W is admissible.

Finally, we prove property 3. Equation (24) is the definition of $H(W)$, since $P_W(w) = \mu(F_w)$. Equation (25) follows from the definition of conditional entropy and from (17), together with $H(Q \otimes Q') = H(Q) + H(Q')$ applied to (22). Substituting (25) into the first expression in (8), we obtain (26). □

APPENDIX B

In this section, we show that the same table used for the reduction for WYNER THRESHOLD cannot be used for COMMON ENTROPY THRESHOLD. Indeed, by Lemma 9, a rectangle of A of size $d+1$ suffices to lower Wyner's common information of the uniform-support table P^A_{XY} below $\log n$, whereas the following example shows that the same rectangle may fail to lower the common entropy.

Example 1. Let A be the Boolean matrix with rows $(1,1,1,0)$ and $(1,1,0,1)$, so that $n=2$ and $d=3$. Its uniform-support table is

$$P^A_{XY} = \frac{1}{6}\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}.$$

The matrix A has the rectangle $\{1,2\}\times\{1,2\}$, of size $4 > d$, and hence $J(P^A_{XY}) < \log n = 1$ by Lemma 9. We now show that, nevertheless, $G(P^A_{XY}) = 1$. Since the auxiliary $W = X$, whose two components are the rows of P^A_{XY}, has entropy $H(X) = 1$, it suffices to show that every admissible auxiliary W of (X,Y) satisfies $H(W) \geq 1$.

Let W be an admissible auxiliary of (X,Y), with components $(F_w)_{w\in\mathcal{W}}$, where $\mathcal{W}$ is any finite set. By Lemma 2, the support of each component of W is a rectangle of P^A_{XY}. Hence, each component is of one of the following three types: its support is contained in row 1, its support is contained in row 2, or its support meets both rows. In the last case, the columns of the support must be positive in both rows, and hence the component lies in columns 1 and 2. We group the components of W according to their type, and we denote by S_1, S_2 and S_3 the sums of all the components of the first, second and third type, respectively, where a sum is zero if there is no component of that type. Since S_3 lies in columns 1 and 2, and $S_1+S_2+S_3 = P^A_{XY}$, these three matrices have the form

$$S_1 = \begin{pmatrix} \frac{1}{6}-u_1 & \frac{1}{6}-u_2 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad S_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{6}-v_1 & \frac{1}{6}-v_2 & 0 & \frac{1}{6} \end{pmatrix}, \quad S_3 = \begin{pmatrix} u_1 & u_2 & 0 & 0 \\ v_1 & v_2 & 0 & 0 \end{pmatrix},$$

for some numbers $u_1,u_2,v_1,v_2 \in [0,\frac{1}{6}]$. Letting $c_1 = u_1+u_2$ and $c_2 = v_1+v_2$, so that $0 \leq c_1, c_2 \leq \frac{1}{3}$, the masses of the three matrices are

$$\mu(S_1) = \frac{1}{2} - c_1, \qquad \mu(S_2) = \frac{1}{2} - c_2, \qquad \mu(S_3) = c_1 + c_2.$$

We now bound $H(W)$ from below in terms of these masses. By (24), $H(W)$ is the sum of $\phi(\mu(F_w))$ over all $w \in \mathcal{W}$. Since $\phi(a+b) \leq \phi(a)+\phi(b)$ for all $a,b \geq 0$, we get that

$$\begin{aligned} H(W) &= \sum_{w\in\mathcal{W}} \phi\big(\mu(F_w)\big) \geq \phi\big(\mu(S_1)\big) + \phi\big(\mu(S_2)\big) + \phi\big(\mu(S_3)\big) \\ &= \phi\Big(\frac{1}{2}-c_1\Big) + \phi\Big(\frac{1}{2}-c_2\Big) + \phi(c_1+c_2). \end{aligned} \tag{139}$$

Finally, we show that the right-hand side of (139) is at least 1 for all $c_1, c_2 \in [0, \frac{1}{3}]$. Since ϕ is concave, the right-hand side is a concave function of (c_1, c_2), and hence the minimum is attained at one of its four vertices. If $c_1 = c_2 = 0$, the right-hand side equals $\phi(\frac{1}{2})+\phi(\frac{1}{2}) = 1$. If one of c_1, c_2 equals $\frac{1}{3}$ and the other one equals 0, it equals $\phi(\frac{1}{6}) + \phi(\frac{1}{2}) + \phi(\frac{1}{3}) \approx 1.46$. If $c_1 = c_2 = \frac{1}{3}$, it equals $\phi(\frac{1}{6}) + \phi(\frac{1}{6}) + \phi(\frac{2}{3}) \approx 1.25$. Therefore $H(W) \geq 1$, which proves that $G(P^A_{XY}) = 1$.

APPENDIX C
PROOF OF LEMMA 10

Proof. We observe that both properties of the lemma follow from the fact that λ is close to Td/e, namely that

$$\frac{Td}{e} \leq \lambda < \frac{Td}{e} + 2. \tag{140}$$

Therefore, we first prove (140). We start by bounding the error $e - e_\ell = \sum_{i \geq \ell+1} 1/i!$ of the partial sum e_ℓ. The proof is a standard estimate of the tail of the series of e (see, e.g., [15, Chapter 3]). For every $i \geq \ell + 1$, we can write $i! = (\ell+1)! \cdot (\ell+2) \cdots i$, where the product $(\ell+2) \cdots i$ has $i - \ell - 1$ factors, each at least 2. Hence $i! \geq (\ell+1)!\, 2^{i-\ell-1}$, with strict inequality for $i \geq \ell + 2$. Therefore, summing the resulting geometric series, we obtain that

$$0 < e - e_\ell < \frac{1}{(\ell+1)!} \sum_{k \geq 0} 2^{-k} = \frac{2}{(\ell+1)!} \overset{(a)}{\leq} \frac{2}{2^\ell} \overset{(b)}{\leq} \frac{1}{Td}, \tag{141}$$

where (a) follows from $(\ell+1)! \geq 2^\ell$, since $(\ell+1)!$ is the product of the ℓ factors $2, 3, \ldots, \ell+1$, each at least 2, and (b) follows from $2^\ell \geq 2Td$, which holds by the definition of ℓ.

We now compare λ with Td/e. Since $e > e_\ell$, and $e_\ell \geq 1 + 1 = 2$ because $\ell \geq 1$, it follows from (141) that

$$0 < \frac{Td}{e_\ell} - \frac{Td}{e} = \frac{Td\,(e - e_\ell)}{e e_\ell} < \frac{1}{e e_\ell} < \frac{1}{4}. \tag{142}$$

Moreover, since λ is the ceiling of Td/e_ℓ, we have that $Td/e_\ell \leq \lambda < Td/e_\ell + 1$. Combining these inequalities with (142), we obtain that

$$\frac{Td}{e} < \frac{Td}{e_\ell} \leq \lambda < \frac{Td}{e_\ell} + 1 < \frac{Td}{e} + \frac{1}{4} + 1 < \frac{Td}{e} + 2,$$

which proves (140).

We now prove property 1. The inequality $Td \leq e\lambda$ is the lower bound in (140). For the second inequality, the upper bound in (140) gives $e\kappa = ed + e\lambda < ed + Td + 2e = Td + e(d+2)$. Since $e < 3$ and $T \geq 16(d+1)$, we have that $e(d+2) < 3(d+2) < 8(d+1) \leq T/2$, and hence $e\kappa < Td + T/2$.

We next prove property 2. By the upper bound in (140) and $e > 2$, we have that $\kappa = d + \lambda < d + Td/2 + 2$. Since $T \geq 16(d+1)$, we have that $d + 2 < T/2$, and hence $\kappa < Td/2 + T/2 = T(d+1)/2$. Moreover, by the lower bound in (140) and $Td \geq 16(d+1) \geq 32$, we have that $\lambda \geq Td/e \geq 32/e > 1$.

Finally, we show that λ can be computed in polynomial time in d. By definition, λ is obtained from e_ℓ, which is a sum of $\ell + 1$ fractions. Hence it suffices to show that $\ell = O(\log(d+1))$, and that e_ℓ is a ratio of two integers whose bit lengths are polynomial in $\log(d+1)$. Since T is the least power of two such that $T \geq 16(d+1)$, we have that $T < 32(d+1)$, and hence $\ell = \lceil \log(2Td) \rceil = O(\log(d+1))$. Moreover, since $i!$ divides $\ell!$ for every $i \leq \ell$, we can write all the terms of e_ℓ with the common denominator $\ell!$, namely

$$e_\ell = \sum_{i=0}^{\ell} \frac{1}{i!} = \frac{1}{\ell!} \sum_{i=0}^{\ell} \frac{\ell!}{i!},$$

where $\sum_{i=0}^{\ell} \ell!/i!$ is an integer at most $(\ell+1)\,\ell!$. Since $\ell! \leq \ell^\ell$, both the numerator and the denominator of e_ℓ have $O(\ell \log \ell)$ bits, that is, $O(\log(d+1) \log\log(d+1))$ bits, and they can be computed with $O(\ell)$ arithmetic operations. Therefore, $\lambda = \lceil Td/e_\ell \rceil$ is the ceiling of a ratio of two integers of polynomial bit length, and it can be computed in polynomial time in the bit length of d. □

APPENDIX D
PROOF OF LEMMA 15

Proof. We prove the properties in order. We first prove property 1. We recall that $h_2(t) = t \log \frac{1}{t} + (1-t) \log \frac{1}{1-t}$. Since $\ln(1+v) \leq v$ for every $v \geq 0$, choosing $v = \frac{t}{1-t}$ gives $\ln \frac{1}{1-t} \leq \frac{t}{1-t}$. Multiplying this inequality by $(1-t) \log e$, and using the fact that $\log e \leq \log \frac{1}{t}$, since $t \leq 1/e$, we get that

$$(1-t) \log \frac{1}{1-t} \leq t \log e \leq t \log \frac{1}{t}. \tag{143}$$

Substituting the above inequality into the expression of $h_2(t)$ proves property 1.

We now prove property 2. If $c = 0$, then $b = 0$, and the property trivially holds. Thus, we assume that $c > 0$. Since ϕ is concave, it lies below its tangent at c, that is,

$$\phi(c-b) \leq \phi(c) + \phi'(c)\big((c-b) - c\big). \tag{144}$$

Since $\phi'(c) = -\log(ec)$ and since $\log(ec) \leq \log e$ because $c \leq 1$, it follows that

$$\phi(c-b) \leq \phi(c) + b \log(ec) \leq \phi(c) + b \log e, \tag{145}$$

which proves property 2.

Property 3 holds because the derivative of $t \log(c/t)$, which is $\log \frac{c}{et}$, is nonnegative for $t \leq c/e$.

We now prove property 4. For every i, let $p_i = z_i/v$, so that $p = (p_1, \dots, p_s)$ is a probability vector with s entries and $z_i = v p_i$. Then, we have that

$$\phi(z_i) = -v p_i \log(v p_i) = v p_i \log \frac{1}{v} + v\, \phi(p_i). \tag{146}$$

Summing (146) over i, and recalling that $\sum_i p_i = 1$ and that $\sum_i \phi(p_i) = H(p) \le \log s$, we obtain that

$$\sum_{i=1}^{s} \phi(z_i) = v \log \frac{1}{v} + v\, H(p) \le v \log \frac{1}{v} + v \log s = v \log \frac{s}{v}, \tag{147}$$

which proves property 4.

Finally, we prove property 5. Let $Z \sim Q$, and let B be the indicator of the event $Z \notin S$, so that $\Pr(B = 1) = \varepsilon$. Given $B = 0$, the random variable Z has distribution Q_S, and, given $B = 1$, it has distribution $Q_{\bar{S}}$. Since B is a function of Z, the chain rule [4, Theorem 2.2.1] gives $H(Z) = H(B) + H(Z \mid B)$, which is (84). Finally, bounding $H(Q_{\bar{S}}) \le \log |\mathcal{Z}|$ in (84) and rearranging gives the last inequality of property 5. □

APPENDIX E
PROOF OF LEMMA 17

Proof. We assume that $a = |\mathcal{X}|$, since the other case follows by interchanging rows and columns. We first treat the case $\delta = 0$. In this case, let $R = \operatorname{supp}(Q)$ and $C = \operatorname{supp}(Q')$, so that $S = R \times C$ is the support of $Q \otimes Q'$. Since $(Q \otimes Q')(D) = 0$, the set S does not intersect D, and the mass outside S is zero, which proves (100).

Thus, assume that $\delta > 0$. For every row $x \in \mathcal{X}$, let $D_x = \{y \in \mathcal{Y} : (x, y) \in D\}$ be the set of columns where row x meets D, and let $\rho(x) = Q'(D_x)$. Then, it holds that

$$\sum_{x \in \mathcal{X}} Q(x)\rho(x) = \sum_{x \in \mathcal{X}} \sum_{y \in D_x} Q(x) Q'(y) = (Q \otimes Q')(D) = \delta. \tag{148}$$

If S contains row x, then it cannot contain any column of D_x. Hence, keeping row x in S forces us to remove columns of probability $\rho(x)$. Therefore, for a threshold $t > 0$, which we choose below, we keep only the rows with $\rho(x) \le t$, and we remove the columns that meet D in one of the kept rows, that is, we let

$$R = \{x \in \mathcal{X} : \rho(x) \le t\}, \qquad C = \mathcal{Y} \setminus \bigcup_{x \in R} D_x, \qquad S = R \times C. \tag{149}$$

We observe that if $x \in R$ and $y \in C$, then $y \notin D_x$, that is, $(x, y) \notin D$. Hence, S does not intersect D.

We now bound the mass outside S. An entry outside S has its row outside R or its column outside C, so that

$$(Q \otimes Q')\big((\mathcal{X} \times \mathcal{Y}) \setminus S\big) \le Q(\mathcal{X} \setminus R) + Q'(\mathcal{Y} \setminus C). \tag{150}$$

We bound the two terms on the right-hand side of (150) separately. For the first term, every row outside R satisfies $\rho(x) > t$ by construction. Hence, by (148), we have that

$$t\,Q(\mathcal{X} \setminus R) \le \sum_{x \notin R} Q(x)\rho(x) \le \delta, \tag{151}$$

which implies that

$$Q(\mathcal{X} \setminus R) \le \frac{\delta}{t} \tag{152}$$

For the second term, by (149), the removed columns are exactly those that meet D in one of the kept rows, that is, $\mathcal{Y} \setminus C = \bigcup_{x \in R} D_x$. Hence, by the union bound, and since $Q'(D_x) = \rho(x) \le t$ for every $x \in R$, we have that

$$Q'(\mathcal{Y} \setminus C) = Q'\Big(\bigcup_{x \in R} D_x\Big) \le \sum_{x \in R} Q'(D_x) = \sum_{x \in R} \rho(x) \le |R| t \le at, \tag{153}$$

where the last inequality follows from $|R| \le |\mathcal{X}| = a$. Substituting (152) and (153) into (150), we obtain that

$$(Q \otimes Q')\big((\mathcal{X} \times \mathcal{Y}) \setminus S\big) \le \frac{\delta}{t} + a\,t. \tag{154}$$

Finally, the right-hand side of (154) is minimized over $t > 0$ by choosing $t = \sqrt{\delta/a}$, which gives

$$(Q \otimes Q')\big((\mathcal{X} \times \mathcal{Y}) \setminus S\big) \le \frac{\delta}{\sqrt{\delta/a}} + a\sqrt{\frac{\delta}{a}} = \sqrt{a\delta} + \sqrt{a\delta} = 2\sqrt{a\delta}, \tag{155}$$

and proves (100). □

REFERENCES

[1] G. Braun, R. Jain, T. Lee, and S. Pokutta, Information-theoretic approximations of the nonnegative rank, *Computational Complexity*, vol. **26**, pp. 147–197, 2017.

[2] G. Braun and S. Pokutta, Common information and unique disjointness, *Algorithmica*, vol. **76**, pp. 597–629, 2016.

[3] J. E. Cohen and U. G. Rothblum, Nonnegative ranks, decompositions, and factorizations of nonnegative matrices, *Linear Algebra and its Applications*, vol. **190**, pp. 149–168, 1993.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[5] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, 1979.

[6] R. M. Gray and A. D. Wyner, Source coding for a simple network, *The Bell System Technical Journal*, vol. **53**, pp. 1681–1721, 1974.

[7] T. S. Han and K. Kobayashi, Maximal rectangular subsets contained in the set of partially jointly typical sequences for dependent random variables, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. **70**, pp. 15–32, 1985.

[8] S.-L. Huang, X. Xu, L. Zheng and G. W. Wornell, A local characterization for Wyner common information, *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 2252–2257, 2020.

[9] T.-H. Huang and H. El Gamal, Efficient solvers for Wyner common information with application to multi-modal clustering, *IEEE Transactions on Information Theory*, vol. **71**, pp. 2054–2074, 2025.

[10] Z. Jiang and M. Kocaoglu, Conditional common entropy for instrumental variable testing and partial identification, in *Proceedings of the 41st International Conference on Machine Learning (ICML)*, Proceedings of Machine Learning Research, vol. **235**, pp. 21824–21843, 2024.

[11] P.A. Karakasis and W. Saad, From common information to canonical correlation analysis and beyond: Bridging Information Theory, Signal Processing, and Machine Learning, *Entropy* 2026, 28, 1161.

[12] M. Kocaoglu, S. Shakkottai, A. G. Dimakis, C. Caramanis, and S. Vishwanath, Applications of common entropy for causal inference, in *Advances in Neural Information Processing Systems*, vol. **33**, pp. 17514–17525, 2020.

[13] G. R. Kumar, C. T. Li, and A. El Gamal, Exact common information, in: *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, pp. 161–165, 2014.

[14] R. Peeters, The maximum edge biclique problem is NP-complete, *Discrete Applied Mathematics*, vol. **131**, pp. 651–654, 2003.

[15] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw–Hill, 1976.

[16] M. Sudan, H. Tyagi, and S. Watanabe, Communication for generating correlation: a unifying survey, *IEEE Transactions on Information Theory*, vol. **66**, pp. 5–37, 2020.

[17] E. Sula and M. Gastpar, Common information components analysis, *Entropy*, vol. **23**, no. 2, article 151, 2021.

[18] E. Sula and M. Gastpar, The Gray-Wyner network and Wyner's common information for gaussian sources, *IEEE Transactions on Information Theory*, vol. **68**, no. 2, pp. 1369–1384, 2022.

[19] S. A. Vavasis, On the complexity of nonnegative matrix factorization, *SIAM Journal on Optimization*, vol. **20**, pp. 1364–1377, 2009.

[20] H. S. Witsenhausen, Values and bounds for the common information of two discrete random variables, *SIAM Journal on Applied Mathematics*, vol. **31**, pp. 313–333, 1976.

[21] A. D. Wyner, The common information of two dependent random variables, *IEEE Transactions on Information Theory*, vol. **21**, pp. 163–179, 1975.

[22] M. Yannakakis, Expressing combinatorial optimization problems by linear programs, *Journal of Computer and System Sciences*, vol. **43**, pp. 441–466, 1991.

[23] L. Yu, V. Anantharam, and J. Chen, Graphs of joint types, noninteractive simulation, and stronger hypercontractivity, *IEEE Transactions on Information Theory*, vol. **70**, pp. 2287–2308, 2024.

[24] L. Yu and V. Y. F. Tan, On exact and ∞-Rényi common informations, *IEEE Transactions on Information Theory*, vol. **66**, pp. 3366–3406, 2020.

[25] L. Yu and V. Y. F. Tan, Common information, noise stability, and their extensions, *Foundations and Trends in Communications and Information Theory*, vol. **19**, pp. 107–389, 2022.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il