Electronic Colloquium on Computational Complexity, Report No. 307 (2026)

The Information Bottleneck and the Privacy Funnel are NP-hard

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

Abstract

Given a joint distribution P_{XY} of random variables X, Y on finite alphabets, the information bottleneck method seeks a representation T of X that is as informative as possible about Y while compressing X; the privacy funnel asks for a representation that discloses much about X while leaking little about Y. In both scenarios T and Y are conditionally independent given X. We show that computing the optimal value $\mathrm{IB}(P_{XY}, R) = \sup_T\{I(Y;T) : I(X;T) \leq R\}$ of the information bottleneck problem, as well as its Lagrangian relaxation $L_\beta(P_{XY}) = \sup_T[I(Y;T) - \beta I(X;T)]$, is NP-hard. Moreover, unless $\mathrm{P} = \mathrm{NP}$, neither $\mathrm{IB}(P_{XY}, R)$ nor $L_\beta(P_{XY})$ can be approximated in polynomial time within an additive error of order $|\mathcal{X}|^{-(1+\alpha)}$, for every fixed $\alpha > 0$. We also prove that computing the privacy funnel $\mathrm{PF}(P_{XY}, r) = \inf_T\{I(Y;T) : I(X;T) \geq r\}$ and its Lagrangian relaxation $\Lambda_\beta(P_{XY}) = \inf_T[I(Y;T) - \beta I(X;T)]$ is NP-hard, and that, unless $\mathrm{P} = \mathrm{NP}$, neither can be approximated in polynomial time within an additive error of order $|\mathcal{X}|^{-2}$.

Index Terms— Information Bottleneck, Privacy Funnel, NP-hardness

I. Introduction

Extracting compact representations that preserve information relevant to a task is a central problem in information theory, statistics, machine learning, and beyond. The information bottleneck (IB) method, introduced by Tishby, Pereira and Bialek in 1999 [27][1], formulates this problem as a trade-off between compressing an observed random variable X and retaining information about a target variable Y.

Applications of IB include clustering and document classification [27], as well as the design of representations for neural classifiers and generative models [13]. Its operational interpretation through remote

[1]We should note that an equivalent optimization problem has previously been studied in the literature on information theory [1], [30], [29], [28].

ISSN 1433-8092

source coding under logarithmic loss, together with extensions to distributed learning and communication, connects these applications to information-theoretic performance limits [31]. The IB trade-off thus serves both as an objective for constructing representations and as a benchmark for assessing their quality. The vast literature on the information bottleneck problem and its numerous applications cannot be synthesized here; we refer the reader to the review papers [3], [13], [14], [18], [19], [22], [31], [24] for a partial overview of the area.

Formally, for a joint distribution P_{XY} on finite alphabets $\mathcal{X} \times \mathcal{Y}$ and a budget $R \geq 0$, the IB problem corresponds to the following optimization problem:

$$\mathrm{IB}(P_{XY}, R) = \sup_{\substack{P_{T|X}\colon T-X-Y \\ I(X;T)\leq R}} I(Y;T). \tag{1}$$

The supremum ranges over all finite representation alphabets $\mathcal{T}$ and stochastic encoders $P_{T|X}$; the Markov condition $T - X - Y$ means that T and Y are conditionally independent given X. We also consider the Lagrangian formulation

$$L_\beta(P_{XY}) = \sup_{P_{T|X}\colon T-X-Y} \big[I(Y;T) - \beta I(X;T)\big], \qquad \beta > 0. \tag{2}$$

All information quantities are measured in bits.

In general, computing the optimal values in (1) and (2) is computationally challenging [2], [15]. Both problems are non-convex optimization problems [3], [13], [27]. The iterative algorithm proposed in [27], as well as its later variants, is only guaranteed to converge to a local optimum [11], [25], [15]. A basic question is therefore whether the optimal values in (1) and (2) admit efficient computation or approximation with provable accuracy for arbitrary finite input distributions.

To the best of our knowledge, exact characterizations of the information bottleneck curve are known in a handful of cases only. For binary sources observed through a binary symmetric channel, the exact trade-off follows from Mrs. Gerber's lemma [30]; extensions cover binary-input memoryless symmetric channels with uniform input [7], [26]. For nonbinary Hamming channels with uniform input, Benger, Asoodeh and Chen [4] characterize the entire curve as the upper concave envelope of an explicit scalar function. Other exact cases include discrete deterministic targets $Y = f(X)$, for which $\mathrm{IB}(P_{XY}, R) = \min\{R, H(Y)\}$ for all $R \geq 0$ [17, Sec. 3], and the case in which Y is obtained from X through an erasure channel with erasure probability ϵ, independently of the input symbol. In this case, $\mathrm{IB}(P_{XY}, R) = (1-\epsilon)R$ for $0 \leq R \leq H(X)$, as stated in [3, Lemma 1]. In the continuous setting, Chechik *et al.* [8] obtain an analytic solution for jointly Gaussian vectors, with optimal representations given by noisy linear projections onto canonical-correlation directions. Dikshtein, Weinberger and Shamai [9] determine the optimal asymptotic relevance rate for jointly stationary Gaussian processes through spectral water filling.

A closely related problem is the privacy funnel (PF), introduced by Makhdoumi et al. [20], in which one seeks a representation T of X that reveals at least a prescribed amount of information about X, while leaking as little information as possible about Y. Formally, for a rate $0 \leq r \leq H(X)$, the PF problem corresponds to the following optimization problem:

$$\mathrm{PF}(P_{XY}, r) = \inf_{\substack{P_{T|X}:\, T-X-Y \\ I(X;T)\geq r}} I(Y;T), \tag{3}$$

and its Lagrangian formulation is

$$\Lambda_\beta(P_{XY}) = \inf_{P_{T|X}:\, T-X-Y} \big[I(Y;T) - \beta I(X;T)\big], \qquad \beta > 0. \tag{4}$$

Problems (3) and (4) share the same computational difficulties as (1) and (2).

A. Our results

Our results provide a complexity-theoretic explanation for the computational difficulties discussed above, by establishing a fundamental obstacle to efficient computation in the worst case. For the information bottleneck, we prove that the threshold decision problems associated with both (1) and (2) are NP-hard. Thus, unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can solve these problems for all instances.

The NP-hardness of (2) is first obtained with the Lagrange multiplier fixed at $\beta = 1/2$, and then extended to every fixed rational $\beta \in (0,1)$, and to every fixed real $\beta \in (0,1)$ computable in polynomial time. Moreover, unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $L_{1/2}(P_{XY})$ within additive error $1/(54|\mathcal{X}|^{3/2})$, or $\mathrm{IB}(P_{XY}, R)$ within additive error $1/(144|\mathcal{X}|^{3/2})$ (Theorems 1 and 2). By suitably modifying the source instances of the decision problem from which we perform the reduction, we successively improve the exponent $3/2$ as follows: for every fixed $\alpha > 0$, unless $\mathrm{P} = \mathrm{NP}$, no polynomial-time algorithm can approximate $L_{1/2}(P_{XY})$ within $1/(36|\mathcal{X}|^{1+\alpha})$, or $\mathrm{IB}(P_{XY}, R)$ within $1/(96|\mathcal{X}|^{1+\alpha})$, and the same holds, up to a constant factor, for every fixed rational $\beta \in (0,1)$ (Theorem 3).

Building on the same methodology, we show that computing the optimal value of the privacy funnel is NP-hard as well. This hardness applies not only to the exact computation, but also to approximation within an additive error of order $|\mathcal{X}|^{-2}$ (Theorem 4). Moreover, the same results hold for the Lagrangian formulation (4), for every fixed rational multiplier $\beta \in (0,1)$ (Proposition 3).

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

B. Overview of the proofs

All the results are proved by polynomial-time reductions from the CLIQUE problem, and the reductions share a common construction. Given an instance (G, k) of CLIQUE, we construct a larger graph $\bar{G}$ by

inserting, for each edge e of G, a clique C_e of size $k-1$ containing only the edge e among the edges of G. This construction ensures that the cliques C_e identify a partition of the edges of $\bar{G}$, and that no clique of size larger than $k-1$ is created, apart from those already present in G. In all our reductions, the random variable X is uniformly distributed on the oriented edges of $\bar{G}$. We then consider two encoders of X: a deterministic encoder T_0, which maps each oriented edge to the clique C_e containing it, and, if G contains a k-clique K, a randomized encoder T_1 which uses K.

For the information bottleneck, given $X = x$, the random variable Y is one of the two endpoints of the oriented edge x, chosen uniformly at random. If G has no k-clique, we show that $L_{1/2}(P_{XY})$ is achieved by the encoder T_0. If, instead, G has a k-clique, we show that the encoder T_1 strictly improves over T_0, and that the improvement is at least inverse-polynomial in $|\mathcal{X}|$. This yields both the NP-hardness of the Lagrangian problem and the hardness of approximating $L_{1/2}(P_{XY})$. The result for the constrained formulation (1) follows from the same construction, by a suitable choice of the rate budget R. Moreover, by modifying the instance (G, k) of CLIQUE without changing its answer, we further improve the exponent $3/2$ in Theorems 1 and 2 to $1 + \alpha$, for every fixed $\alpha > 0$.

For the privacy funnel, we keep the same X, T_0 and T_1 as for the IB problem, but we change the channel from X to Y: given X, the random variable Y either reveals X, or reveals a pair of oriented edges that cannot belong to a common clique. The reduction relies on the lower bound $I(Y;T) \geq I(X;T) - H(X \mid Y)$, which holds for every encoder T. More precisely, we show that, if G does not contain a k-clique, then the lower bound cannot be attained at any rate smaller than $\log |E(G)|$, while, if G contains a k-clique, it is attained at a strictly smaller rate by combining the encoders T_0 and T_1.

C. Organization of the paper

In Section II, we set the notation, recall some basic facts on encoders, and formally define the decision problems we study. In Section III, we describe the extended graph $\bar{G}$ and the encoders T_0 and T_1, which are common to all our reductions. In Section IV, we prove the hardness of the information bottleneck, both in its Lagrangian and in its constrained formulation. In Section V, we prove the hardness of the privacy funnel. Section VI contains some concluding remarks and open problems.

II. PRELIMINARIES

A. Notation

All alphabets throughout the paper are finite, and we denote by $\log$ the logarithm in base 2 and by $\ln$ the natural logarithm. Let X be a random variable taking values in an alphabet $\mathcal{X}$. We denote by P_X

the probability distribution of X, that is, $P_X(x) = \mathbb{P}(X = x)$ for each $x \in \mathcal{X}$, and by

$$H(X) = -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x) \tag{5}$$

the Shannon entropy of X, with the convention $0 \log 0 = 0$. Since $H(X)$ depends only on the distribution P_X, we also write $H(P_X)$ to denote the entropy of the distribution P_X. We denote by $h_2(z) = -z \log z - (1-z)\log(1-z)$, for $z \in [0,1]$, the binary entropy function.

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. Without loss of generality, we assume that $P_X(x) > 0$ for all $x \in \mathcal{X}$; otherwise, we simply remove from $\mathcal{X}$ the values with $P_X(x) = 0$. Then, for every $x \in \mathcal{X}$, the conditional distribution $P_{Y|X}(\cdot \mid x)$ of Y given $X = x$ is

$$P_{Y|X}(y \mid x) = \frac{P_{XY}(x,y)}{P_X(x)}, \qquad \forall y \in \mathcal{Y}, \tag{6}$$

and we denote by $H(Y \mid X = x) = H\big(P_{Y|X}(\cdot \mid x)\big)$ its entropy. The conditional entropy of Y given X is the average of $H(Y \mid X = x)$ with respect to P_X, that is,

$$H(Y \mid X) = \sum_{x \in \mathcal{X}} P_X(x)\, H(Y \mid X = x), \tag{7}$$

and the mutual information between X and Y is

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

Conditional entropies and mutual information given a third random variable Z are defined analogously; we recall that $I(X;Y \mid Z) = H(X \mid Z) - H(X \mid Y, Z)$, and the chain rules

$$H(X,Y) = H(X) + H(Y \mid X) \qquad \text{and} \qquad I(X;Y,Z) = I(X;Z) + I(X;Y \mid Z). \tag{9}$$

Let (X, Y, T) be a triple of random variables with joint distribution P_{XYT} on $\mathcal{X} \times \mathcal{Y} \times \mathcal{T}$. The random variables Y and T are conditionally independent given X if

$$P_{YT|X}(y,t \mid x) = P_{Y|X}(y \mid x)\, P_{T|X}(t \mid x), \qquad \forall (x,y,t) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{T}. \tag{10}$$

Let P and Q be two probability distributions on a finite set S, and let $\mathrm{supp}(P) = \{s \in S : P(s) > 0\}$ denote the support of P. If $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$, the Kullback–Leibler divergence and the χ^2-divergence of P from Q are defined as

$$D(P \,\|\, Q) = \sum_{s \in \mathrm{supp}(P)} P(s) \log \frac{P(s)}{Q(s)} \qquad \text{and} \qquad \chi^2(P \,\|\, Q) = \sum_{s \in \mathrm{supp}(Q)} \frac{\big(P(s) - Q(s)\big)^2}{Q(s)}. \tag{11}$$

Both divergences are non-negative, and they vanish if and only if $P = Q$. Moreover, the two divergences are related by the following well-known inequality.

Proposition 1 ([23])**.** *Let P and Q be two probability distributions on a finite set S such that* $\operatorname{supp}(P) \subseteq \operatorname{supp}(Q)$*. Then, it holds that*

$$D(P \,\|\, Q) \leq \log\left(1 + \chi^2(P \,\|\, Q)\right) \leq \frac{\chi^2(P \,\|\, Q)}{\ln 2}. \tag{12}$$

Finally, given a probability distribution P on an alphabet S, we denote by $P \otimes P$ the product distribution on $S \times S$, defined by $(P \otimes P)(s, s') = P(s)P(s')$ for all $(s, s') \in S \times S$.

B. Encoders

Let $(X, Y) \sim P_{XY}$ be a pair of random variables taking values in $\mathcal{X} \times \mathcal{Y}$. In the optimization problems (1)–(4), one seeks a representation (or encoder) T of X that is conditionally independent of Y given X. This type of encoder is formalized as follows.

Definition 1. Let $\mathcal{T}$ be a finite set. An encoder of X is a conditional distribution (or channel) $P_{T|X}$ from $\mathcal{X}$ to $\mathcal{T}$, that is, a collection $\{P_{T|X}(\cdot \mid x)\}_{x \in \mathcal{X}}$ of probability distributions on $\mathcal{T}$. The encoder $P_{T|X}$ induces a random variable T taking values in $\mathcal{T}$ such that (X, Y, T) has joint distribution

$$P_{XYT}(x, y, t) = P_X(x)\, P_{Y|X}(y \mid x)\, P_{T|X}(t \mid x), \qquad \forall (x, y, t) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{T}. \tag{13}$$

In particular, T and Y are conditionally independent given X. Moreover, we refer to the mutual information $I(X; T)$ as the *rate* of the encoder T.

To simplify the notation, we identify an encoder with the random variable T it induces, and we refer to T itself as the encoder whenever the channel $P_{T|X}$ is clear from the context. Moreover, without loss of generality, we assume that $P_T(t) > 0$ for all $t \in \mathcal{T}$, and we call the distributions $\{P_{X|T}(\cdot \mid t)\}_{t \in \mathcal{T}}$ the *posteriors* of T.

Let W denote the channel from $\mathcal{X}$ to $\mathcal{Y}$ induced by P_{XY}, that is, $W(y \mid x) = P_{Y|X}(y \mid x)$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. The channel W maps every probability distribution Q on $\mathcal{X}$ to the probability distribution WQ on $\mathcal{Y}$ defined by

$$(WQ)(y) = \sum_{x \in \mathcal{X}} W(y \mid x)\, Q(x), \qquad \forall y \in \mathcal{Y}. \tag{14}$$

In particular, $WP_X = P_Y$. The next lemma shows that the same relation holds conditionally on T. Indeed, since T and Y are conditionally independent given X, applying W to the posterior of X given $T = t$ gives the posterior of Y given $T = t$.

Lemma 1. *Let T be an encoder of X. Then, it holds that*

$$P_{Y|T}(\cdot \mid t) = W P_{X|T}(\cdot \mid t), \qquad \forall t \in \mathcal{T}. \tag{15}$$

Proof. Let $y \in \mathcal{Y}$ and $t \in \mathcal{T}$. By (13), and since $P_X(x)\, P_{T|X}(t \mid x) = P_T(t)\, P_{X|T}(x \mid t)$, we have that

$$P_{Y|T}(y \mid t) = \frac{\sum_{x \in \mathcal{X}} P_X(x)\, P_{T|X}(t \mid x)\, W(y \mid x)}{P_T(t)} = \sum_{x \in \mathcal{X}} W(y \mid x)\, P_{X|T}(x \mid t) = \big(W P_{X|T}(\cdot \mid t)\big)(y),$$

where the last equality follows from (14). □

The next simple lemma gives a lower bound on the information that an encoder leaks about Y, in terms of the information it reveals about X. It plays a useful role in Section V.

Lemma 2. *For every joint distribution P_{XY} and every encoder T of X,*

$$I(Y;T) \geq I(X;T) - H(X \mid Y), \tag{16}$$

with equality if and only if $H(X \mid Y,T) = 0$.

Proof. Since Y and T are conditionally independent given X, we have that $I(X;T) = I(X,Y;T) = I(Y;T) + I(X;T \mid Y)$. Moreover, $I(X;T \mid Y) = H(X \mid Y) - H(X \mid Y,T) \leq H(X \mid Y)$, with equality if and only if $H(X \mid Y,T) = 0$. Putting these together, we get that

$$I(Y;T) = I(X;T) - I(X;T \mid Y) \geq I(X;T) - H(X \mid Y),$$

which proves (16), with equality if and only if $H(X \mid Y,T) = 0$. □

Finally, we show that two encoders can be combined into a new encoder, whose rate interpolates between their rates and whose posteriors are those of the two encoders. We will use this fact in Section V.

Lemma 3. *Let T_a and T_b be two encoders of X taking values in $\mathcal{T}_a$ and $\mathcal{T}_b$, respectively, and let $\theta \in [0,1]$. Let T_θ be the encoder of X taking values in $(\{a\} \times \mathcal{T}_a) \cup (\{b\} \times \mathcal{T}_b)$ defined by*

$$P_{T_\theta|X}\big((a,t) \mid x\big) = \theta\, P_{T_a|X}(t \mid x), \qquad P_{T_\theta|X}\big((b,t) \mid x\big) = (1-\theta)\, P_{T_b|X}(t \mid x). \tag{17}$$

In other words, T_θ applies T_a with probability θ and T_b with probability $1-\theta$, independently of X, and reveals which of the two encoders has been applied. Then, the posteriors of T_θ are posteriors of T_a or of T_b, that is,

$$P_{X|T_\theta}\big(\cdot \mid (a,t)\big) = P_{X|T_a}(\cdot \mid t) \quad \textit{and} \quad P_{X|T_\theta}\big(\cdot \mid (b,t)\big) = P_{X|T_b}(\cdot \mid t), \tag{18}$$

whenever the left-hand sides are defined. Moreover,

$$I(X;T_\theta) = \theta\, I(X;T_a) + (1-\theta)\, I(X;T_b). \tag{19}$$

Proof. Let J be a random variable taking values in $\{a,b\}$, independent of (X,Y), with $\mathbb{P}(J=a) = \theta$. Given X and $J=j$, we generate T_j through the channel $P_{T_j|X}$, and we let $T_\theta = (J, T_J)$. Then T_θ has

the channel defined in (17), and it is an encoder of X, since it depends only on X and on J, which is independent of (X, Y).

We first prove (18). Let $j \in \{a, b\}$ with $\mathbb{P}(J = j) > 0$, and let t be such that $P_{T_j}(t) > 0$. By (17), we have that $P_{T_\theta}\big((j, t)\big) = \mathbb{P}(J = j)\, P_{T_j}(t)$, and hence, by Bayes' rule, for every $x \in \mathcal{X}$, it holds that

$$P_{X|T_\theta}\big(x \mid (j, t)\big) = \frac{P_X(x)\, \mathbb{P}(J = j)\, P_{T_j|X}(t \mid x)}{\mathbb{P}(J = j)\, P_{T_j}(t)} = P_{X|T_j}(x \mid t),$$

which proves (18).

We now prove (19). Since J is independent of X, we have that $I(X; J) = 0$. Moreover, given $J = j$, the pair (X, T_J) has the joint distribution induced by the encoder T_j, so that $I(X; T_J \mid J = j) = I(X; T_j)$. Hence, by the chain rule (9), we obtain that

$$I(X; T_\theta) = I(X; J) + I(X; T_J \mid J) = \theta\, I(X; T_a) + (1 - \theta)\, I(X; T_b). \qquad \square$$

C. Decision problems

In this section, we state the decision versions of the optimization problems (1)–(4), which we use to prove NP-hardness[2]. Throughout, a probability distribution is called rational if all its probabilities are rational numbers.

IB-THRESHOLD

INPUT: A rational probability distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, and non-negative rational numbers R and ϑ.

QUESTION: Does there exist an encoder T of X such that $I(X; T) \leq R$ and $I(Y; T) \geq \vartheta$?

IB-LAGRANGIAN$_\beta$, for a fixed $\beta > 0$

INPUT: A rational probability distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, and a rational number τ.

QUESTION: Does there exist an encoder T of X such that $I(Y; T) - \beta I(X; T) \geq \tau$?

PF-THRESHOLD

INPUT: A rational probability distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, and non-negative rational numbers r and ϑ.

QUESTION: Does there exist an encoder T of X such that $I(X; T) \geq r$ and $I(Y; T) \leq \vartheta$?

PF-LAGRANGIAN$_\beta$, for a fixed $\beta > 0$

INPUT: A rational probability distribution P_{XY} on a finite alphabet $\mathcal{X} \times \mathcal{Y}$, and a rational number τ.

QUESTION: Does there exist an encoder T of X such that $I(Y; T) - \beta I(X; T) \leq \tau$?

[2]We refer to [12] for the basics about NP-hardness.

We write IB-LAGRANGIAN for IB-LAGRANGIAN$_{1/2}$. We observe that the optima in (1), (2) and (4), as well as the optimum in (3) whenever $r \leq H(X)$, are attained (see, e.g., [3], [28]). Moreover, if $r > H(X)$, then no encoder satisfies $I(X;T) \geq r$, and, in this case, we adopt the convention $\mathrm{PF}(P_{XY}, r) = +\infty$. Hence, the above decision problems are equivalent to comparing the optimal values of the corresponding optimization problems with a threshold. Namely, IB-THRESHOLD asks whether $\mathrm{IB}(P_{XY}, R) \geq \vartheta$, IB-LAGRANGIAN$_\beta$ asks whether $L_\beta(P_{XY}) \geq \tau$, and, similarly, PF-THRESHOLD and PF-LAGRANGIAN$_\beta$ ask whether $\mathrm{PF}(P_{XY}, r) \leq \vartheta$ and $\Lambda_\beta(P_{XY}) \leq \tau$, respectively.

We prove the NP-hardness of these problems by reductions from the well-known CLIQUE decision problem [16], which we now recall. A graph G is a pair $(V(G), E(G))$, where $V(G)$ is a finite set of vertices and $E(G)$ is a set of unordered pairs $\{u, v\}$ of distinct vertices $u, v \in V(G)$, called edges. A clique of G is a subset $C \subseteq V(G)$ such that $\{u, v\} \in E(G)$ for all distinct $u, v \in C$. The size of a clique C is its cardinality $|C|$, and a clique of size k is called a k-clique. The clique number of G is

$$\omega(G) = \max\{|C| : C \subseteq V(G) \text{ is a clique of } G\}. \tag{20}$$

CLIQUE

INPUT: A graph G and a positive integer k.

QUESTION: Does G contain a k-clique, that is, is $\omega(G) \geq k$?

It is well known that CLIQUE is NP-complete [16].

III. THE EXTENDED GRAPH AND THE ENCODERS T_0 AND T_1

In this section, we describe the construction shared by all our reductions: the extended graph $\bar{G}$ of an instance (G, k) of CLIQUE, the random variable X, which is uniformly distributed on the oriented edges of $\bar{G}$, and the two encoders T_0 and T_1 of X. The random variable Y, and hence the joint distribution P_{XY}, will be defined differently in the reductions for the information bottleneck (Section IV) and for the privacy funnel (Section V). However, since the encoders T_0 and T_1 are channels from X, the results of this section do not depend on the random variable Y.

A. Oriented edges and cliques

Let G be a graph with at least one edge, and let

$$A(G) = \big\{(u, v) \in V(G) \times V(G) : \{u, v\} \in E(G)\big\} \tag{21}$$

denote the set of the $2|E(G)|$ oriented edges of G. For a clique C of G of size $j \geq 2$, let

$$A(C) = \{(u, v) \in C \times C : u \neq v\} \subseteq A(G) \tag{22}$$

be the set of its $j(j-1)$ oriented edges, and let U_C denote the probability distribution on $A(G)$ that is uniform on $A(C)$, that is,

$$U_C(x) = \begin{cases} \dfrac{1}{j(j-1)}, & x \in A(C), \\ 0, & \text{otherwise.} \end{cases} \tag{23}$$

Clearly, $H(U_C) = \log\big(j(j-1)\big)$.

B. The extended graph

Let (G, k) be an instance of CLIQUE, and let

$$n = |V(G)|, \qquad N = |E(G)|, \qquad B = \binom{k}{2} = \frac{k(k-1)}{2}. \tag{24}$$

We assume that G is represented by its adjacency matrix, so that the size of (G, k) is at least n. For the sake of simplicity, we restrict to instances of CLIQUE satisfying

$$k \geq 4 \qquad \text{and} \qquad N \geq B. \tag{25}$$

This does not affect the NP-hardness. Indeed, if $k \leq 3$, CLIQUE can be solved in $O(n^3)$ time by examining all subsets of at most three vertices. On the other hand, if $N < B$, then G has no k-clique, since a k-clique has B edges.

Let $m = k-1 \geq 3$. We construct the extended graph $\bar{G}$ from G as follows: for each edge $e = \{u, v\} \in E(G)$, we add a set S_e of $m-2$ new vertices, and we add all the edges needed to make $C_e = \{u, v\} \cup S_e$ a clique of size m. No other vertices or edges are added. Throughout the paper, we let

$$s = m(m-1), \qquad \mu = \frac{1}{s}, \qquad M = |A(\bar{G})|. \tag{26}$$

The next lemma collects the properties of $\bar{G}$ on which our reductions rely.

Lemma 4. *The graph $\bar{G}$ satisfies the following properties.*

1) *For every $e \in E(G)$, the only edge of G in $E(C_e)$ is e. Moreover, the collection $\{E(C_e)\}_{e \in E(G)}$ is a partition of $E(\bar{G})$. Consequently, every oriented edge of $\bar{G}$ belongs to $A(C_e)$ for exactly one $e \in E(G)$.*
2) *The number of oriented edges of $\bar{G}$ is*

$$M = |A(\bar{G})| = Nm(m-1) = Ns. \tag{27}$$

3) $\omega(\bar{G}) = \max\{\omega(G), m\}$.
4) $|V(\bar{G})| \leq n + n^3$, $M \leq n^4$, *and $\bar{G}$ can be constructed in time polynomial in n.*

Proof. 1. Let $e \in E(G)$. Since $C_e \cap V(G) = e$, the only edge of G with both endpoints in C_e is e, which proves the first claim. We now show that $\{E(C_e)\}_{e \in E(G)}$ is a partition of $E(\bar{G})$. By construction, every edge of $\bar{G}$ belongs to $E(C_e)$ for some $e \in E(G)$. Thus, it suffices to show that this e is unique. Let $x = \{u, v\} \in E(\bar{G})$. If $x \in E(G)$, then, by the first claim, $x \in E(C_e)$ if and only if $x = e$. If $x \notin E(G)$, then x was added in the construction of $\bar{G}$, and at least one of its endpoints, say u, belongs to S_e for some $e \in E(G)$, so that $x \in E(C_e)$. Since the sets S_e are pairwise disjoint, $u \notin C_{e'}$ for every $e' \neq e$, and hence $x \notin E(C_{e'})$ for every $e \neq e'$. In both cases, x belongs to exactly one set $E(C_e)$. Therefore, each edge of $\bar{G}$ lies in exactly one of the sets $E(C_e)$, with $e \in E(G)$. Hence, $\{E(C_e)\}_{e \in E(G)}$ forms a partition of $E(\bar{G})$. Finally, since $(u, v) \in A(C_e)$ if and only if $\{u, v\} \in E(C_e)$, every oriented edge of $\bar{G}$ belongs to $A(C_e)$ for exactly one $e \in E(G)$.

2. By property 1, $E(\bar{G})$ is the disjoint union of the N sets $E(C_e)$, each of cardinality $\binom{m}{2}$. Hence, the total number of oriented edges is $|A(\bar{G})| = 2N\binom{m}{2} = Nm(m-1) = Ns$.

3. We first show that $\omega(\bar{G}) \leq \max\{\omega(G), m\}$. Let K be a clique of $\bar{G}$. If K contains a vertex $z \in S_e$ for some $e \in E(G)$, then, by construction of $\bar{G}$, all the neighbours of z belong to C_e. Therefore, $K \subseteq C_e$ and $|K| \leq m$. Otherwise, if K does not contain a vertex $z \in S_e$ for any $e \in E(G)$, then we have that $K \subseteq V(G)$. Therefore, K is a clique of G, and $|K| \leq \omega(G)$. Conversely, every clique of G is a clique of $\bar{G}$, and each C_e is a clique of $\bar{G}$ of size m. Thus, $\omega(\bar{G}) \geq \max\{\omega(G), m\}$. Putting everything together we have that $\omega(\bar{G}) = \max\{\omega(G), m\}$.

4. We have that $N \leq \binom{n}{2} \leq n^2$, and, by (25), $\binom{k}{2} = B \leq N \leq \binom{n}{2}$, which implies $k \leq n$ and hence $m < n$. Since $\bar{G}$ is obtained from G by adding $m - 2$ new vertices for each edge of G, together with the edges of the cliques C_e, we get that $|V(\bar{G})| = n + N(m-2) \leq n + n^3$, and, by property 2, $M = Nm(m-1) \leq n^4$. Therefore, $\bar{G}$ can be constructed in time polynomial in n. □

C. The random variable X and the encoder T_0

In all our reductions, the random variable X is uniformly distributed on $\mathcal{X} = A(\bar{G})$, that is,

$$P_X(x) = \frac{1}{M} = \frac{\mu}{N}, \qquad \forall x \in A(\bar{G}). \tag{28}$$

By property 1 of Lemma 4, we can define the deterministic encoder T_0 of X, taking values in $E(G)$, by

$$P_{T_0|X}(e \mid x) = \begin{cases} 1, & \text{if } x \in A(C_e), \\ 0, & \text{otherwise,} \end{cases} \qquad \forall x \in A(\bar{G}),\ e \in E(G). \tag{29}$$

In other words, T_0 maps each oriented edge x to the unique edge $e \in E(G)$ such that $x \in A(C_e)$, that is, it reveals the clique C_e containing x. The following lemma computes the posteriors and the rate of T_0.

Lemma 5. *The encoder T_0 defined in* (29) *satisfies*

$$P_{T_0}(e) = \frac{1}{N} \qquad \textit{and} \qquad P_{X|T_0}(\cdot \mid e) = U_{C_e}, \qquad \forall e \in E(G). \tag{30}$$

Consequently, $H(X \mid T_0) = \log s$ *and*

$$I(X;T_0) = \log N. \tag{31}$$

Proof. We first prove (30). For every $e \in E(G)$, the set $A(C_e)$ contains $s = m(m-1)$ oriented edges, each of probability $1/M$ under P_X. Hence, by (29) and property 2 of Lemma 4, we have that

$$P_{T_0}(e) = \sum_{x \in A(\bar{G})} P_X(x)\, P_{T_0|X}(e \mid x) = \frac{|A(C_e)|}{M} = \frac{s}{Ns} = \frac{1}{N}.$$

Moreover, by Bayes' rule, $P_{X|T_0}(x \mid e) = N/M = 1/s$ if $x \in A(C_e)$, and $P_{X|T_0}(x \mid e) = 0$ otherwise, that is, $P_{X|T_0}(\cdot \mid e) = U_{C_e}$, where U_{C_e} is the probability distribution on $A(\bar{G})$ that is uniform on $A(C_e)$ (cf. (23)).

We now compute the rate of T_0. Since each U_{C_e} is uniform on s oriented edges, by (7) we get that $H(X \mid T_0) = \sum_{e \in E(G)} \frac{1}{N} \log s = \log s$. Moreover, since X is uniform on the M elements of $A(\bar{G})$, it follows that $I(X;T_0) = H(X) - H(X \mid T_0) = \log M - \log s = \log N$, where the last equality holds by property 2 of Lemma 4. □

D. The encoder T_1

We now assume that G contains a k-clique K, and we define an encoder T_1 that uses K. The encoder depends on a parameter $p \in (0,1)$, which will be chosen differently for the information bottleneck and for the privacy funnel. Let T_1 be the encoder taking values in $\{\star\} \cup E(G)$, where $\star$ is a new symbol, defined by

$$P_{T_1|X}(\star \mid x) = \begin{cases} p, & \text{if } x \in A(K), \\ 0, & \text{otherwise,} \end{cases} \qquad P_{T_1|X}(e \mid x) = \big(1 - P_{T_1|X}(\star \mid x)\big)\, P_{T_0|X}(e \mid x), \tag{32}$$

for every $x \in A(\bar{G})$ and $e \in E(G)$. In other words, if $X \notin A(K)$, then $T_1 = T_0$, while, if $X \in A(K)$, then $T_1 = \star$ with probability p and $T_1 = T_0$ with probability $1-p$. Thus, T_1 differs from T_0 only on the oriented edges of the k-clique K. Let

$$\varepsilon = 2\mu p \qquad \text{and} \qquad \lambda = \frac{B\varepsilon}{N}. \tag{33}$$

Note that $\varepsilon < 1$, since $\mu \leq 1/6$ and $p < 1$. For every $e = \{u, v\} \in E(K)$, let $A(e) = \{(u, v), (v, u)\}$ denote the set of the two orientations of e, and let V_e be the probability distribution on $A(\bar{G})$ defined by

$$V_e(x) = \begin{cases} a_0 = \dfrac{\mu}{1-\varepsilon}, & \text{if } x \in A(C_e) \setminus A(e), \\ a_1 = \dfrac{\mu(1-p)}{1-\varepsilon}, & \text{if } x \in A(e), \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

The distribution V_e is a slight perturbation of the uniform distribution U_{C_e} (30). Indeed, the probability of the two orientations of e is slightly decreased, and the probability of the other oriented edges of C_e is slightly increased. The next lemma computes the output distribution and the posteriors of T_1, which we need to compare T_1 with T_0.

Lemma 6. *The encoder T_1 defined in* (32) *satisfies*

$$P_{T_1}(t) = \begin{cases} \lambda, & t = \star, \\ (1-\varepsilon)/N, & t \in E(K), \\ 1/N, & t \in E(G) \setminus E(K), \end{cases} \qquad P_{X|T_1}(\cdot \mid t) = \begin{cases} U_K, & t = \star, \\ V_t, & t \in E(K), \\ U_{C_t}, & t \in E(G) \setminus E(K). \end{cases} \tag{35}$$

Proof. Recall that $P_X(x) = 1/M = \mu/N$ for every $x \in A(\bar{G})$, by (28). We consider the cases $T_1 = \star$ and $T_1 = e$ with $e \in E(G)$, separately.

We first consider the case $T_1 = \star$. Since K is a clique of size k, it has $k(k-1) = 2B$ oriented edges. Hence, by the definition (32) of the encoder T_1, and since $\varepsilon = 2\mu p$ and $\lambda = \frac{B\varepsilon}{N}$ by (33), we have that

$$P_{T_1}(\star) = \sum_{x \in A(\bar{G})} P_X(x)\, P_{T_1|X}(\star \mid x) = \sum_{x \in A(K)} \frac{\mu}{N}\, p = \frac{2B\mu p}{N} = \frac{B\varepsilon}{N} = \lambda,$$

and, by Bayes' rule,

$$P_{X|T_1}(x \mid \star) = \frac{P_X(x)\, P_{T_1|X}(\star \mid x)}{P_{T_1}(\star)} = \frac{\mu p/N}{2B\mu p/N} = \frac{1}{k(k-1)}$$

for $x \in A(K)$, and $P_{X|T_1}(x \mid \star) = 0$ otherwise. By definition (23) of U_K, this means that $P_{X|T_1}(\cdot \mid \star) = U_K$.

We now consider the case $T_1 = e$ for $e \in E(G)$. By (29) and (32), we have that

$$P_{T_1|X}(e \mid x) = \begin{cases} 1-p, & \text{if } x \in A(C_e) \cap A(K), \\ 1, & \text{if } x \in A(C_e) \setminus A(K), \\ 0, & \text{if } x \notin A(C_e). \end{cases} \tag{36}$$

Thus, T_1 differs from T_0 only on the oriented edges in $A(C_e) \cap A(K)$. We claim that

$$A(C_e) \cap A(K) = \begin{cases} A(e), & \text{if } e \in E(K), \\ \emptyset, & \text{otherwise.} \end{cases} \tag{37}$$

Indeed, let $x = (u, v) \in A(K)$. Since K is a clique of G, $\{u, v\}$ is an edge of G, and, by property 1 of Lemma 4, it holds that $\{u, v\} \in E(C_e)$ if and only if $\{u, v\} = e$. Hence, $A(C_e) \cap A(K)$ consists of the orientations of e that belong to $A(K)$, which are both orientations if $e \in E(K)$, and none otherwise, proving (37).

Now, if $e \in E(G) \setminus E(K)$, then, by (37) and (36), $P_{T_1|X}(e \mid \cdot) = P_{T_0|X}(e \mid \cdot)$, and, by Lemma 5, $P_{T_1}(e) = 1/N$ and $P_{X|T_1}(\cdot \mid e) = U_{C_e}$. If $e \in E(K)$, then, by (37), it holds that $A(C_e) \cap A(K) = A(e)$. Notice that $A(e)$ contains 2 elements, while $A(C_e) \setminus A(e)$ contains $s - 2$ elements. Hence, by (36), and using $\mu s = 1$, we obtain that

$$P_{T_1}(e) = \sum_{x \in A(\bar{G})} P_X(x)\, P_{T_1|X}(e \mid x) = \frac{\mu}{N}\big((s-2) + 2(1-p)\big) = \frac{1 - 2\mu p}{N} = \frac{1-\varepsilon}{N}.$$

Moreover, applying Bayes' rule, one can see that the following equalities hold: $P_{X|T_1}(x \mid e) = \frac{\mu/N}{(1-\varepsilon)/N} = a_0$ if $x \in A(C_e) \setminus A(e)$, $P_{X|T_1}(x \mid e) = \frac{\mu(1-p)/N}{(1-\varepsilon)/N} = a_1$ if $x \in A(e)$, and $P_{X|T_1}(x \mid e) = 0$ otherwise; that is, $P_{X|T_1}(\cdot \mid e) = V_e$ by (34). □

Lemma 6 shows that the encoder T_1 replaces part of the posteriors U_{C_e} of T_0 with the posterior U_K. In doing so, however, it perturbs the posterior U_{C_e} into V_e (cf. (34)) for every $e \in E(K)$. The next lemma shows that the perturbed posteriors V_e are close to U_{C_e}, namely, their divergence from U_{C_e} is quadratic in ε.

Lemma 7. *For every $e \in E(K)$, we have that $H(V_e) = \log s - D(V_e \,\|\, U_{C_e})$. Moreover, the Kullback–Leibler divergence $D(V_e \,\|\, U_{C_e})$ does not depend on e, and satisfies*

$$D(V_e \,\|\, U_{C_e}) \le \frac{\chi^2(V_e \,\|\, U_{C_e})}{\ln 2} \le \frac{\varepsilon^2}{2\mu(1-\varepsilon)^2 \ln 2}. \tag{38}$$

Proof. We first prove the identity for $H(V_e)$. Since $\operatorname{supp}(V_e) \subseteq A(C_e)$ and $U_{C_e}(x) = \mu = 1/s$ for every $x \in A(C_e)$, we have that

$$D(V_e \,\|\, U_{C_e}) = \sum_{x \in \operatorname{supp}(V_e)} V_e(x) \log \frac{V_e(x)}{\mu} = -H(V_e) + \log \frac{1}{\mu} = \log s - H(V_e).$$

By (34), V_e takes the value a_0 on $s - 2$ oriented edges and the value a_1 on 2 oriented edges, and these values do not depend on e. Hence, $H(V_e)$, and thus $D(V_e \,\|\, U_{C_e})$, does not depend on e.

We now prove (38). The first inequality in (38) follows from Proposition 1, since $\mathrm{supp}(V_e) \subseteq A(C_e) = \mathrm{supp}(U_{C_e})$. For the second one, we compute $\chi^2(V_e \,\|\, U_{C_e})$. Using $\mu p = \varepsilon/2$, we have that

$$a_0 - \mu = \frac{\mu\varepsilon}{1-\varepsilon} \qquad \text{and} \qquad a_1 - \mu = \frac{\mu(1-p) - \mu(1-\varepsilon)}{1-\varepsilon} = -\frac{\varepsilon(1/2-\mu)}{1-\varepsilon}. \tag{39}$$

Since $U_{C_e}(x) = \mu$ for every $x \in A(C_e)$, the definition in (11) gives

$$\chi^2(V_e \| U_{C_e}) = \sum_{x \in A(C_e)} \frac{(V_e(x)-\mu)^2}{\mu}.$$

We split this sum into two groups. The $s-2$ oriented edges in $A(C_e) \setminus A(e)$ each contribute $(a_0-\mu)^2/\mu$, while the two orientations of e, which form $A(e)$, each contribute $(a_1-\mu)^2/\mu$. Thus,

$$\chi^2(V_e \| U_{C_e}) = \frac{(s-2)(a_0-\mu)^2 + 2(a_1-\mu)^2}{\mu}.$$

Substituting the expressions for $a_0 - \mu$ and $a_1 - \mu$ from (39), we obtain

$$\begin{aligned}
\chi^2(V_e \,\|\, U_{C_e}) &= \sum_{x \in A(C_e)} \frac{\big(V_e(x)-\mu\big)^2}{\mu} \\
&= \frac{1}{\mu}\Big((s-2)\frac{\mu^2\varepsilon^2}{(1-\varepsilon)^2} + 2\frac{\varepsilon^2(1/2-\mu)^2}{(1-\varepsilon)^2}\Big) = \frac{\varepsilon^2}{\mu(1-\varepsilon)^2}\Big(\mu - 2\mu^2 + \frac{1}{2} - 2\mu + 2\mu^2\Big) \\
&= \frac{\varepsilon^2(1/2-\mu)}{\mu(1-\varepsilon)^2} \le \frac{\varepsilon^2}{2\mu(1-\varepsilon)^2},
\end{aligned} \tag{40}$$

where the third equality uses $\mu s = 1$. Dividing (40) by $\ln 2 > 0$, we obtain

$$\frac{\chi^2(V_e \| U_{C_e})}{\ln 2} \le \frac{\varepsilon^2}{2\mu(1-\varepsilon)^2 \ln 2},$$

which is the second inequality in (38). Together with the first inequality, already established using Proposition 1, this completes the proof. □

In what follows, we denote by D_V the common value of $D(V_e \,\|\, U_{C_e})$ for $e \in E(K)$. The next lemma computes the rate of the encoder T_1 defined in (32), and shows that, compared to T_0, the rate of T_1 differs by the quantity η defined in (41).

Lemma 8. *The encoder T_1 defined in* (32) *satisfies*

$$I(X;T_1) = \log N - \eta, \qquad \textit{where} \qquad \eta = \lambda \log \frac{m+1}{m-1} - \frac{B(1-\varepsilon)}{N} D_V. \tag{41}$$

Proof. Since X is uniform on $A(\bar{G})$, we have that $I(X;T_1) = \log M - H(X \mid T_1)$, and, since $\log M - \log s = \log N$, the claim (41) is equivalent to prove that

$$H(X \mid T_1) = \log s + \eta. \tag{42}$$

Moreover, since $H(X \mid T_0) = \log s$ by Lemma 5, this implies that η is the increase in conditional entropy obtained by replacing T_0 with T_1. We now prove (42).

By Lemma 6 and (7), and since $|E(K)| = B$ and $|E(G) \setminus E(K)| = N - B$, we have that

$$\begin{aligned} H(X \mid T_1) &= \lambda\, H(U_K) + \frac{1-\varepsilon}{N} \sum_{e \in E(K)} H(V_e) + \frac{1}{N} \sum_{e \in E(G) \setminus E(K)} H(U_{C_e}) \\ &= \lambda\, H(U_K) + \frac{1-\varepsilon}{N} \sum_{e \in E(K)} H(V_e) + \frac{N-B}{N}\, H(U_{C_e}). \end{aligned} \tag{43}$$

We now substitute the entropies of the posteriors. Since U_K and U_{C_e} are uniform on $k(k-1)$ and s oriented edges, respectively, we have that $H(U_K) = \log\big(k(k-1)\big)$ and $H(U_{C_e}) = \log s$, and, by Lemma 7, $H(V_e) = \log s - D_V$. Hence, (43) becomes

$$H(X \mid T_1) = \lambda \log\big(k(k-1)\big) + \frac{B(1-\varepsilon)}{N}\big(\log s - D_V\big) + \frac{N-B}{N} \log s. \tag{44}$$

Finally, by Lemma 6, the probabilities of the values of T_1 sum to 1, that is,

$$\lambda + \frac{B(1-\varepsilon)}{N} + \frac{N-B}{N} = 1. \tag{45}$$

Multiplying (45) by $\log s$ and subtracting the obtained result from (44), the terms $\frac{B(1-\varepsilon)}{N} \log s$ and $\frac{N-B}{N} \log s$ cancel, and we obtain that

$$H(X \mid T_1) - \log s = \lambda \log \frac{k(k-1)}{s} - \frac{B(1-\varepsilon)}{N}\, D_V = \eta,$$

where the last equality follows since $\frac{k(k-1)}{s} = \frac{(m+1)m}{m(m-1)} = \frac{m+1}{m-1}$. This proves (42). □

IV. Hardness of the information bottleneck

In this section, we prove our results for the information bottleneck, which are formalized in the following theorems.

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

$$\big|v - L_{1/2}(P_{XY})\big| \leq \frac{1}{54\,|\mathcal{X}|^{3/2}}. \tag{46}$$

Theorem 2. IB-Threshold *is NP-hard. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a rational joint distribution* P_{XY} *on a finite alphabet* $\mathcal{X} \times \mathcal{Y}$ *and a non-negative rational number* R*, outputs a value* v *such that*

$$\big|v - \mathrm{IB}(P_{XY}, R)\big| \leq \frac{1}{144\,|\mathcal{X}|^{3/2}}. \tag{47}$$

The rest of this section is organized as follows. In Section IV-A, given a channel W from $\mathcal{X}$ to $\mathcal{Y}$, we define the cost of a probability distribution Q as $g_W(Q) = H(WQ) - \frac{1}{2}H(Q)$, and we show that the objective of the Lagrangian depends on an encoder only through the average cost of its posteriors. In Section IV-B, we introduce the channel that maps each oriented edge of a graph to one of its endpoints, and we provide, for this channel, a lower bound on the cost of every distribution on the oriented edges of the graph in terms of its clique number. In Section IV-C, we define the joint distribution of the reduction, and we show that, if G has no k-clique, the encoder T_0 is optimal, while, if G contains a k-clique, the encoder T_1 strictly improves on it. Successively, we use these results to prove Theorem 1 in Section IV-D. In Section IV-E, we extend Theorem 1 to every fixed multiplier $\beta \in (0, 1)$, and, in Section IV-F, we prove Theorem 2. Finally, in Section IV-G, we improve the exponent $3/2$ in Theorems 1 and 2 and in Proposition 2 to $1+\alpha$, for every fixed $\alpha > 0$ (Theorem 3). We made no attempt to optimize the constants in these results.

A. The Lagrangian in terms of the posteriors

We first show that the objective of $L_{1/2}(P_{XY})$ depends on an encoder only through its posteriors. Given a channel W from $\mathcal{X}$ to $\mathcal{Y}$, we define the *cost* of a probability distribution Q on $\mathcal{X}$ as

$$g_W(Q) = H(WQ) - \frac{1}{2}H(Q). \tag{48}$$

Using (48), we can express the objective of the Lagrangian in terms of the posteriors of the encoder T, as shown in the following lemma.

Lemma 9. *Let W be the channel from $\mathcal{X}$ to $\mathcal{Y}$ induced by the joint distribution P_{XY}, and let T be an encoder of X. Then, it holds that*

$$I(Y;T) - \frac{1}{2}I(X;T) = g_W(P_X) - \sum_{t\in\mathcal{T}} P_T(t)\, g_W\big(P_{X|T}(\cdot \mid t)\big). \tag{49}$$

In particular, $L_{1/2}(P_{XY}) = g_W(P_X) - \inf_T \sum_{t\in\mathcal{T}} P_T(t)\, g_W\big(P_{X|T}(\cdot \mid t)\big)$.

Proof. By (8) and (7), we have that

$$I(X;T) = H(P_X) - \sum_{t\in\mathcal{T}} P_T(t)\, H\big(P_{X|T}(\cdot \mid t)\big) \tag{50}$$

and, similarly,

$$I(Y;T) = H(P_Y) - \sum_{t\in\mathcal{T}} P_T(t)\, H\big(P_{Y|T}(\cdot \mid t)\big). \tag{51}$$

We first rewrite (51) in terms of the channel W. Since $P_Y = WP_X$ and, by Lemma 1, $P_{Y|T}(\cdot \mid t) = WP_{X|T}(\cdot \mid t)$, we have

$$I(Y;T) = H(WP_X) - \sum_{t\in\mathcal{T}} P_T(t)\, H\big(WP_{X|T}(\cdot \mid t)\big).$$

Subtracting one half of (50) and grouping the unconditional and conditional entropy terms separately, we obtain

$$\begin{aligned} I(Y;T) - \frac{1}{2}I(X;T) &= \left[H(WP_X) - \frac{1}{2}H(P_X)\right] \\ &\quad - \sum_{t\in\mathcal{T}} P_T(t)\left[H\big(WP_{X|T}(\cdot \mid t)\big) - \frac{1}{2}H\big(P_{X|T}(\cdot \mid t)\big)\right] \\ &= g_W(P_X) - \sum_{t\in\mathcal{T}} P_T(t)\, g_W\big(P_{X|T}(\cdot \mid t)\big), \end{aligned}$$

where the last equality uses the definition of g_W in (48). This proves (49).

Finally, the marginal P_X and the channel W are fixed, so $g_W(P_X)$ does not depend on the encoder. Thus, maximizing the Lagrangian objective is equivalent to minimizing the average cost of the posterior distributions. Taking the supremum over all encoders in (49) therefore gives

$$L_{1/2}(P_{XY}) = g_W(P_X) - \inf_T \sum_{t\in\mathcal{T}} P_T(t)\, g_W\big(P_{X|T}(\cdot \mid t)\big),$$

which proves the last claim. □

By Lemma 9, the objective of the Lagrangian depends on the encoder T only through the average cost of its posteriors. Moreover, Lemma 9 also provides a natural strategy to lower and upper bound $L_{1/2}(P_{XY})$. Indeed, if $g_W(Q) \geq c$ for every probability distribution Q on $\mathcal{X}$, then, since each posterior $P_{X|T}(\cdot \mid t)$ is a probability distribution on $\mathcal{X}$, the infimum in Lemma 9 is at least c, and hence

$$L_{1/2}(P_{XY}) \leq g_W(P_X) - c. \tag{52}$$

On the other hand, for every fixed encoder T of X, Lemma 9 yields the following lower bound:

$$L_{1/2}(P_{XY}) \geq g_W(P_X) - \sum_{t\in\mathcal{T}} P_T(t)\, g_W\big(P_{X|T}(\cdot \mid t)\big). \tag{53}$$

B. The endpoint channel and an entropic Turán bound

Let G be a graph with at least one edge. We consider the channel W_G from $A(G)$ to $V(G)$ that maps each oriented edge of G to one of its two endpoints, chosen uniformly at random, that is,

$$W_G\big(w \mid (u,v)\big) = \begin{cases} 1/2, & w \in \{u,v\}, \\ 0, & \text{otherwise}, \end{cases} \tag{54}$$

for every $(u, v) \in A(G)$ and $w \in V(G)$. In what follows, we write g_G for the cost g_{W_G} associated with the channel W_G, as defined in (48), that is, we set $g_G(Q) = H(W_G Q) - \frac{1}{2}H(Q)$ for every probability distribution Q on $A(G)$. For every integer $j \geq 2$, let

$$c_j = -\frac{1}{2}\log\Big(1 - \frac{1}{j}\Big), \tag{55}$$

and note that c_j is strictly decreasing in j.

The next lemma computes the cost of the uniform distribution on the oriented edges of a clique, which, by Lemma 5, is the form of the posteriors of T_0.

Lemma 10. *Let C be a clique of G of size $j \geq 2$. Then $W_G U_C$ is the uniform distribution on C, and $g_G(U_C) = c_j$.*

Proof. We first compute $W_G U_C$. For every $v \in V(G)$, by (23) and (54) we have that

$$(W_G U_C)(v) = \frac{1}{j(j-1)} \sum_{(w,u)\in A(C)} W_G\big(v \mid (w,u)\big),$$

where each term of the sum is $1/2$ if $v \in \{w, u\}$, and 0 otherwise. Therefore, if $v \in C$, exactly $2(j-1)$ oriented edges of $A(C)$ have v as an endpoint, so that $(W_G U_C)(v) = \frac{1}{j(j-1)} \cdot 2(j-1) \cdot \frac{1}{2} = \frac{1}{j}$; if $v \notin C$, no oriented edge of $A(C)$ has v as an endpoint, so that $(W_G U_C)(v) = 0$. Hence, $W_G U_C$ is uniform on C, and $H(W_G U_C) = \log j$. Moreover, since $H(U_C) = \log\big(j(j-1)\big)$, we finally get that

$$g_G(U_C) = \log j - \frac{1}{2}\log\big(j(j-1)\big) = -\frac{1}{2}\log\Big(1 - \frac{1}{j}\Big) = c_j,$$

which concludes the proof. □

We now show that no probability distribution Q on $A(G)$ has a cost $g_G(Q)$ smaller than $c_{\omega(G)}$, where $\omega(G)$ is the clique number of G and c_j is as in (55). To this end, we first recall the following result of Motzkin and Straus [21].

Lemma 11 ([21])**.** *For every probability distribution P on $V(G)$,*

$$\sum_{(u,v)\in A(G)} P(u)\,P(v) \leq 1 - \frac{1}{\omega(G)}. \tag{56}$$

Lemma 11 allows us to prove the following useful result.

Lemma 12. *For every probability distribution Q on $A(G)$, it holds that*

$$g_G(Q) \geq c_{\omega(G)}, \tag{57}$$

with equality if $Q = U_C$ for a clique C of size $\omega(G)$.

Proof. The idea of the proof is to express $g_G(Q)$ as a relative entropy, and then to bound it from below by Jensen's inequality and Lemma 11.

Let $\pi = W_G Q$, and let Q_1 and Q_2 be the distributions on $V(G)$ defined by

$$Q_1(v) = \sum_{u:(v,u)\in A(G)} Q(v,u) \qquad \text{and} \qquad Q_2(v) = \sum_{u:(u,v)\in A(G)} Q(u,v).$$

Since, by (54), an oriented edge is mapped to each of its two endpoints with probability $1/2$, we have that

$$\begin{aligned} \pi(v) &= \sum_{(w,u)\in A(G)} W_G(v \mid (w,u)) Q(w,u) \\ &= \frac{1}{2}\big(Q_1(v) + Q_2(v)\big), \qquad \forall v \in V(G). \end{aligned} \tag{58}$$

We extend Q to a distribution on $V(G) \times V(G)$ by setting it to zero outside $A(G)$. We notice that if $Q(u,v) > 0$, then $\pi(u) \geq \frac{1}{2} Q_1(u) > 0$ and $\pi(v) \geq \frac{1}{2} Q_2(v) > 0$. Hence $\operatorname{supp}(Q) \subseteq \operatorname{supp}(\pi \otimes \pi)$, and we can write

$$\begin{aligned} D(Q \,\|\, \pi \otimes \pi) &= -H(Q) - \sum_{(u,v)\in A(G)} Q(u,v) \log \pi(u) - \sum_{(u,v)\in A(G)} Q(u,v) \log \pi(v) \\ &= -H(Q) - \sum_{v \in V(G)} \big(Q_1(v) + Q_2(v)\big) \log \pi(v) \\ &= -H(Q) - 2 \sum_{v \in V(G)} \pi(v) \log \pi(v) = 2H(\pi) - H(Q) = 2 g_G(Q), \end{aligned} \tag{59}$$

where the second equality follows from the definitions of Q_1 and Q_2, and the third one from (58).

We now bound $D(Q \,\|\, \pi \otimes \pi)$ from below. By Jensen's inequality applied to the concave function $\log$, and by Lemma 11, we have that

$$\begin{aligned} D(Q \,\|\, \pi \otimes \pi) &= - \sum_{(u,v)\in \operatorname{supp}(Q)} Q(u,v) \log \frac{\pi(u)\pi(v)}{Q(u,v)} \geq -\log\Big(\sum_{(u,v)\in A(G)} \pi(u)\pi(v) \Big) \\ &\geq -\log\Big(1 - \frac{1}{\omega(G)}\Big) = 2 c_{\omega(G)}. \end{aligned} \tag{60}$$

Combining (60) with (59) proves (57). The equality case, instead, follows from Lemma 10, since for a clique C of size $\omega(G)$ we have that $g_G(U_C) = c_{\omega(G)}$. □

The bound of Lemma 12 can be seen as an entropic form of Turán's theorem. For symmetric distributions, that is, when $Q(u,v) = Q(v,u)$ for all $(u,v) \in A(G)$, it was obtained by Chao and Yu [6]. Finally, we bound the decrease of c_j when j increases by one.

Lemma 13. *For every integer $m \geq 2$, $c_m - c_{m+1} \geq \frac{1}{2m^2 \ln 2}$.*

Proof. By (55), $c_m - c_{m+1} = \frac{1}{2}\log\left(\frac{m}{m-1}\cdot\frac{m}{m+1}\right) = -\frac{1}{2}\log\left(1-\frac{1}{m^2}\right) \geq \frac{1}{2m^2\ln 2}$, where the inequality follows from $-\ln(1-z) \geq z$ for $0 \leq z < 1$. □

C. The reduction

Let (G,k) be an instance of CLIQUE satisfying (25), and let $\bar{G}$ be the extended graph of Section III-B. We let X be uniformly distributed on $\mathcal{X} = A(\bar{G})$, as in (28), and, given that $X = (u,v)$, we let the value of the random variable Y be one of the two endpoints of (u,v), chosen uniformly at random. Hence $\mathcal{Y} = V(\bar{G})$, the channel from X to Y is $W_{\bar{G}}$, defined in (54), and the joint distribution of (X,Y) is

$$P_{XY}\big((u,v),w\big) = \begin{cases} \dfrac{1}{2M}, & \text{if } w \in \{u,v\}, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall (u,v) \in A(\bar{G}),\ w \in V(\bar{G}). \tag{61}$$

Let

$$\Pi_0 = g_{\bar{G}}(P_X) - c_m. \tag{62}$$

The next lemma shows that the encoder T_0 achieves exactly the value Π_0 defined in (62).

Lemma 14. *The encoder T_0 defined in* (29) *satisfies $I(Y;T_0) - \frac{1}{2}I(X;T_0) = \Pi_0$. In particular, $L_{1/2}(P_{XY}) \geq \Pi_0$.*

Proof. By Lemma 5, the posteriors of T_0 are the distributions U_{C_e}, with probability $1/N$ each. Hence, by Lemma 9, we have that

$$I(Y;T_0) - \frac{1}{2}I(X;T_0) = g_{\bar{G}}(P_X) - \frac{1}{N}\sum_{e\in E(G)} g_{\bar{G}}(U_{C_e}). \tag{63}$$

Since each C_e is a clique of $\bar{G}$ of size m, Lemma 10 gives $g_{\bar{G}}(U_{C_e}) = c_m$, and hence (63) equals $g_{\bar{G}}(P_X) - c_m = \Pi_0$. Finally, $L_{1/2}(P_{XY}) \geq \Pi_0$ follows from the definition (2) of $L_{1/2}(P_{XY})$ as a supremum over all encoders. □

We now show that, if G has no k-clique, then the encoder T_0 is optimal, that is, $L_{1/2}(P_{XY}) = \Pi_0$. While, if G contains a k-clique, then the encoder T_1 strictly improves on T_0, that is, $L_{1/2}(P_{XY}) > \Pi_0$. This is the conceptual step to solve the CLIQUE problem through the estimation of $L_{1/2}(P_{XY})$, and it will prove the hardness of IB-LAGRANGIAN.

1) No-instances: G does not contain a k-clique: We demonstrate that, if G does not contain a k-clique, then no encoder of X improves on T_0.

Lemma 15. *Assume that $\omega(G) < k$. Then every encoder T of X satisfies*

$$I(Y;T) - \frac{1}{2}I(X;T) \leq \Pi_0. \tag{64}$$

Consequently, $L_{1/2}(P_{XY}) = \Pi_0$*, and* $\mathrm{IB}(P_{XY}, R) \le \frac{R}{2} + \Pi_0$ *for every* $R \ge 0$*.*

Proof. Since $\omega(G) \le k-1 = m$, property 3 of Lemma 4 gives $\omega(\bar{G}) = m$. Hence, by Lemma 12 applied to $\bar{G}$, we have that $g_{\bar{G}}(Q) \ge c_m$ for every probability distribution Q on $A(\bar{G})$, and in particular this holds for every posterior of an encoder T of X. Thus, by Lemma 9, every encoder T of X satisfies

$$I(Y;T) - \frac{1}{2} I(X;T) = g_{\bar{G}}(P_X) - \sum_{t \in \mathcal{T}} P_T(t)\, g_{\bar{G}}\big(P_{X|T}(\cdot \mid t)\big) \le g_{\bar{G}}(P_X) - c_m = \Pi_0,$$

which proves (64). Taking the supremum over all encoders gives $L_{1/2}(P_{XY}) \le \Pi_0$, and, together with Lemma 14, $L_{1/2}(P_{XY}) = \Pi_0$. Finally, if $I(X;T) \le R$, then (64) gives $I(Y;T) \le \frac{1}{2} I(X;T) + \Pi_0 \le \frac{R}{2} + \Pi_0$, and taking the supremum over all such encoders yields $\mathrm{IB}(P_{XY}, R) \le \frac{R}{2} + \Pi_0$. □

2) Yes-instances: G does contain a k-clique: We now assume that G contains a k-clique K, and we show that the encoder T_1 of Section III-D strictly improves on T_0. To this end, we choose

$$p = \frac{1}{4m^2}, \qquad \text{so that} \qquad \varepsilon = 2\mu p = \frac{\mu}{2m^2} \quad \text{and} \quad \lambda = \frac{B\varepsilon}{N} = \frac{B\mu}{2Nm^2}. \tag{65}$$

By Lemma 6, the encoder T_1 replaces part of the posteriors U_{C_e} of T_0, whose cost is c_m, with the posterior U_K, whose cost is $c_k < c_m$. In doing so, however, it perturbs U_{C_e} into V_e for every $e \in E(K)$. The next lemma shows that this perturbation increases the cost by at most a quantity quadratic in ε.

Lemma 16. *For every* $e \in E(K)$*,*

$$g_{\bar{G}}(V_e) - c_m \le \frac{1}{2} D(V_e \| U_{C_e}) \le \frac{\varepsilon^2}{4\mu(1-\varepsilon)^2 \ln 2}. \tag{66}$$

Proof. The distribution $W_{\bar{G}} U_{C_e}$ is uniform on C_e, whereas $W_{\bar{G}} V_e$ is supported on the same set. Thus, replacing U_{C_e} by V_e cannot increase the output entropy. Any increase in cost is therefore bounded by one half of the decrease in input entropy, which, by Lemma 7, equals $D(V_e \| U_{C_e})$. We now make this comparison explicit. By (55), we can write

$$c_m = \frac{1}{2} \log \frac{m}{m-1} = \log m - \frac{1}{2} \log s. \tag{67}$$

Combining (67) with the definition (48) of $g_{\bar{G}}$, and using $H(V_e) = \log s - D(V_e \| U_{C_e})$ from Lemma 7, we obtain that

$$\begin{aligned} g_{\bar{G}}(V_e) - c_m &= \big(H(W_{\bar{G}} V_e) - \log m\big) + \frac{1}{2}\big(\log s - H(V_e)\big) \\ &= \big(H(W_{\bar{G}} V_e) - \log m\big) + \frac{1}{2} D(V_e \| U_{C_e}). \end{aligned} \tag{68}$$

Since $\mathrm{supp}(V_e) \subseteq A(C_e)$, by (54) the distribution $W_{\bar{G}} V_e$ is supported on the m vertices of C_e. This implies that $H(W_{\bar{G}} V_e) \le \log m$. Hence, the first term in (68) is non-positive, which proves the first inequality in (66). The second one immediately follows from (38) of Lemma 7. □

We now show that the encoder T_1 strictly improves on T_0.

Lemma 17. *Assume that $\omega(G) \geq k$, let T_1 be the encoder defined in* (32) *with p as in* (65)*, and let*

$$\delta = \frac{1}{8Nm^4}. \tag{69}$$

Then $I(Y;T_1) - \frac{1}{2}I(X;T_1) > \Pi_0 + \delta$. Consequently, $L_{1/2}(P_{XY}) > \Pi_0 + \delta$.

Proof. The idea of the proof is to compare the average cost of the posteriors of T_1 with the cost c_m of the posteriors of T_0. The encoder T_0 has average posterior cost c_m. By Lemma 9, to show that T_1 improves the Lagrangian objective over T_0 by more than δ, it suffices to prove that its average posterior cost is less than $c_m - \delta$. Let

$$\Psi = \sum_t P_{T_1}(t)\, g_{\bar{G}}\big(P_{X|T_1}(\cdot \mid t)\big)$$

be the average cost of the posteriors of T_1. By Lemma 6, and since $g_{\bar{G}}(U_K) = c_k$ and $g_{\bar{G}}(U_{C_e}) = c_m$ by Lemma 10, we have that

$$\Psi = \lambda c_k + \frac{1-\varepsilon}{N} \sum_{e \in E(K)} g_{\bar{G}}(V_e) + \frac{N-B}{N}\, c_m.$$

Since the probabilities in (35) sum to 1, we can write c_m as

$$c_m = \lambda c_m + \frac{1-\varepsilon}{N} \sum_{e \in E(K)} c_m + \frac{N-B}{N}\, c_m. \tag{70}$$

Subtracting (70) from Ψ, the terms corresponding to $e \in E(G) \setminus E(K)$ cancel out, and we obtain that

$$\Psi - c_m = -\lambda(c_m - c_k) + \frac{1-\varepsilon}{N} \sum_{e \in E(K)} \big(g_{\bar{G}}(V_e) - c_m\big). \tag{71}$$

The negative term $-\lambda(c_m - c_k)$ represents the decrease in average posterior cost contributed by U_K, relative to the baseline c_m. The second term is the total weighted change in cost caused by perturbing U_{C_e} into V_e. We bound $\lambda(c_m - c_k)$ from below and the second term from above.

We first bound $\lambda(c_m - c_k)$ from below. Since $k = m + 1$, Lemma 13 gives

$$\lambda(c_m - c_k) \geq \frac{\lambda}{2m^2 \ln 2}. \tag{72}$$

We now bound the perturbation term from above. We first note that ε is small. Indeed, since $m \geq 3$, we have that $\mu \leq \frac{1}{6}$ and $m^2 \geq 9$, so that $\varepsilon = \frac{\mu}{2m^2} \leq \frac{1}{108}$, and hence $\frac{1}{1-\varepsilon} \leq \frac{108}{107}$. Therefore, since $|E(K)| = B$, by Lemma 16 and using $\lambda = \frac{B\varepsilon}{N}$ and $\frac{\varepsilon}{\mu} = \frac{1}{2m^2}$, we get that

$$\frac{1-\varepsilon}{N} \sum_{e \in E(K)} \big(g_{\bar{G}}(V_e) - c_m\big) \leq \frac{B\varepsilon^2}{4N\mu(1-\varepsilon)\ln 2} = \frac{\lambda}{m^2 \ln 2} \cdot \frac{1}{8(1-\varepsilon)} \leq \frac{\lambda}{m^2 \ln 2} \cdot \frac{27}{214}. \tag{73}$$

Substituting (72) and (73) into (71), we get that

$$\Psi - c_m \leq -\frac{\lambda}{m^2 \ln 2}\Big(\frac{1}{2} - \frac{27}{214}\Big) = -\frac{40}{107} \cdot \frac{\lambda}{m^2 \ln 2} < -\frac{\lambda}{2m^2},$$

where the last inequality follows from $\ln 2 < \frac{7}{10}$, which gives $\frac{40}{107 \ln 2} > \frac{400}{749} > \frac{1}{2}$. Since $\lambda = \frac{B\mu}{2Nm^2}$ by (65) and $B\mu = \frac{k(k-1)}{2} \cdot \frac{1}{(k-1)(k-2)} = \frac{k}{2(k-2)} \geq \frac{1}{2}$, we obtain that

$$\Psi - c_m < -\frac{B\mu}{4Nm^4} \leq -\frac{1}{8Nm^4} = -\delta. \tag{74}$$

Finally, by Lemma 9 and (74), $I(Y;T_1) - \frac{1}{2} I(X;T_1) = g_{\bar{G}}(P_X) - \Psi > g_{\bar{G}}(P_X) - c_m + \delta = \Pi_0 + \delta$. The last statement $L_{1/2}(P_{XY}) > \Pi_0 + \delta$ follows from the definition of $L_{1/2}(P_{XY})$ as a supremum over all encoders. □

D. Proof of Theorem 1

We now combine the results of the previous sections to prove Theorem 1. Let (G, k) be an instance of CLIQUE satisfying (25), and let P_{XY} be the joint distribution defined in (61). By Lemmas 15 and 17, we have that $L_{1/2}(P_{XY}) = \Pi_0$ if $\omega(G) < k$, and $L_{1/2}(P_{XY}) > \Pi_0 + \delta$ if $\omega(G) \geq k$. Hence, the two cases are separated by a gap of size δ, and we place the threshold in the middle of this gap. However, since Π_0 involves logarithms, and is therefore in general irrational, we choose the threshold τ as a rational number such that

$$\Big|\tau - \Big(\Pi_0 + \frac{\delta}{2}\Big)\Big| \leq \frac{\delta}{12}. \tag{75}$$

Therefore, the instance of IB-LAGRANGIAN associated with (G, k) is the pair (P_{XY}, τ), where P_{XY} is defined in (61) and τ in (75). Before presenting the proof, since the inapproximability results are stated in terms of the size of the input, the next lemma relates δ defined in (69) to the size $M = |\mathcal{X}|$ of the alphabet of X.

Lemma 18. *Let δ be as in* (69)*. Then $\delta \geq \frac{1}{18M^{3/2}}$.*

Proof. We first express δ in terms of M and m. By property 2 of Lemma 4, $N = M/(m(m-1))$, and substituting this into (69) we obtain that

$$\delta = \frac{m(m-1)}{8Mm^4} = \frac{1 - 1/m}{8Mm^2}. \tag{76}$$

We now bound m from above in terms of M. Since $B = \frac{m(m+1)}{2}$ and $N \geq B$ by (25), we have that $M = Nm(m-1) \geq \frac{m^2(m^2-1)}{2}$. Since $m \geq 3$, we have that $m^2 - 1 \geq \frac{8}{9}m^2$, and hence $M \geq \frac{4}{9}m^4$, that is,

$$m^2 \leq \frac{3}{2}\sqrt{M}. \tag{77}$$

Finally, using (77) and $1 - \frac{1}{m} \geq \frac{2}{3}$ in (76), we conclude that $\delta \geq \frac{2/3}{8M \cdot \frac{3}{2}\sqrt{M}} = \frac{1}{18M^{3/2}}$. □

Proof of Theorem 1. We start by showing that IB-LAGRANGIAN is NP-hard. To this end, we show that the construction of the instance (P_{XY}, τ) described above is a polynomial-time reduction from CLIQUE to IB-LAGRANGIAN. Given an instance (G, k) of CLIQUE satisfying (25), the reduction constructs the extended graph $\bar{G}$, the joint distribution P_{XY} defined in (61), and a rational threshold τ satisfying (75), and outputs the instance (P_{XY}, τ) of IB-LAGRANGIAN.

We first show that the reduction runs in polynomial time. Since G is represented by its adjacency matrix, the size of (G, k) is at least n, so it suffices to show that the construction runs in time polynomial in n. By property 4 of Lemma 4, $\bar{G}$ can be constructed in polynomial time, with $|V(\bar{G})| \leq n + n^3$ and $M \leq n^4$. Hence, P_{XY} is a table with $|A(\bar{G})| \cdot |V(\bar{G})| \leq n^4(n+n^3)$ entries, each equal to 0 or to $1/(2M)$ by (61), which are rational numbers with $O(\log n)$ bits. It remains to compute τ. Note that this does not require computing $L_{1/2}(P_{XY})$, but only approximating $\Pi_0 + \delta/2$. The value δ in (69) is a rational number with $O(\log n)$ bits, while Π_0 is an explicit expression involving polynomially many logarithms of rational numbers. Indeed, $H(P_X) = \log M$, $c_m = \frac{1}{2}\log\frac{m}{m-1}$, and, denoting by d_v the degree of a vertex v in $\bar{G}$, each vertex v is an endpoint of $2d_v$ oriented edges, each of probability $1/M$ and mapped to v with probability $1/2$, so that $(W_{\bar{G}}P_X)(v) = \frac{d_v}{M}$ and

$$H(W_{\bar{G}}P_X) = \sum_{v \in V(\bar{G}) : d_v > 0} \frac{d_v}{M} \log \frac{M}{d_v}.$$

Hence, by Lemma 25 in Appendix A with additive error $\delta/12$, a rational τ satisfying (75) can be computed in polynomial time. Hence, the whole reduction runs in polynomial time.

We now show that the reduction is correct, that is, (P_{XY}, τ) is a yes-instance of IB-LAGRANGIAN *if and only if* (G, k) is a yes-instance of CLIQUE. Recall that (P_{XY}, τ) is a yes-instance of IB-LAGRANGIAN if there exists an encoder T of X such that $I(Y;T) - \frac{1}{2}I(X;T) \geq \tau$, and that (G, k) is a yes-instance of CLIQUE if $\omega(G) \geq k$, that is, G contains a clique of size k. By (75), the threshold τ satisfies

$$\Pi_0 + \frac{5\delta}{12} \leq \tau \leq \Pi_0 + \frac{7\delta}{12}, \tag{78}$$

that is, τ lies strictly between Π_0 and $\Pi_0 + \delta$. Assume first that (G, k) is a yes-instance, that is, $\omega(G) \geq k$. By Lemma 17 and the right inequality in (78), the encoder T_1 satisfies

$$I(Y;T_1) - \frac{1}{2}I(X;T_1) > \Pi_0 + \delta \geq \tau - \frac{7\delta}{12} + \delta = \tau + \frac{5\delta}{12}.$$

Hence, (P_{XY}, τ) is a yes-instance of IB-LAGRANGIAN, and $L_{1/2}(P_{XY}) > \tau + \frac{5\delta}{12}$. Assume now that (G, k) is a no-instance, that is, $\omega(G) < k$. By Lemma 15 and the left inequality in (78), every encoder

T of X satisfies $I(Y;T) - \frac{1}{2}I(X;T) \le \Pi_0 \le \tau - \frac{5\delta}{12} < \tau$. Hence, (P_{XY}, τ) is a no-instance of IB-LAGRANGIAN, and $L_{1/2}(P_{XY}) \le \tau - \frac{5\delta}{12}$. This proves that the reduction is correct, and hence that IB-LAGRANGIAN is NP-hard.

It remains to prove the inapproximability result. By the two cases above, $L_{1/2}(P_{XY})$ lies above $\tau + \frac{5\delta}{12}$ if $\omega(G) \ge k$, and below $\tau - \frac{5\delta}{12}$ if $\omega(G) < k$. Suppose that a polynomial-time algorithm $\mathcal{A}lg$ outputs, given P_{XY}, a value v with $|v - L_{1/2}(P_{XY})| \le \frac{1}{54M^{3/2}}$. By Lemma 18, this error is at most $\frac{\delta}{3} < \frac{5\delta}{12}$. Hence, if $\omega(G) \ge k$, then $v \ge L_{1/2}(P_{XY}) - \frac{\delta}{3} > \tau$, while, if $\omega(G) < k$, then $v \le L_{1/2}(P_{XY}) + \frac{\delta}{3} < \tau$. Thus, running the reduction, then the algorithm $\mathcal{A}lg$, and comparing v with τ would decide CLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

E. General fixed Lagrange multipliers

The hardness result of Theorem 1 can be extended to any other fixed multiplier $\beta \in (0,1)$. Note that, for $\beta \ge 1$, the data-processing inequality gives $L_\beta(P_{XY}) = 0$.

Proposition 2. *Let $\beta \in \mathbb{Q} \cap (0,1)$ be fixed, and let $\kappa_\beta = 2\min\{\beta, 1-\beta\}$. Given a rational joint distribution P_{XY}, one can construct in polynomial time a rational joint distribution $\tilde{P}_{XZ}$ with the same marginal P_X such that*

$$L_\beta(\tilde{P}_{XZ}) = \kappa_\beta \, L_{1/2}(P_{XY}). \tag{79}$$

Consequently, IB-LAGRANGIAN$_\beta$ *is NP-hard, and, unless* $\mathrm{P} = \mathrm{NP}$*, no polynomial-time algorithm approximates $L_\beta(P_{XY})$ within an additive error of $\kappa_\beta/(54|\mathcal{X}|^{3/2})$.*

Proof. We first prove (79). Let $\rho_\beta = \beta - \frac{\kappa_\beta}{2}$ and $e_\beta = 1 - \kappa_\beta - \rho_\beta$. These three numbers form a probability distribution. Indeed, they sum to 1 by definition, and, if $\beta \le \frac{1}{2}$, then $\kappa_\beta = 2\beta$, $\rho_\beta = 0$ and $e_\beta = 1 - 2\beta \ge 0$, while, if $\beta \ge \frac{1}{2}$, then $\kappa_\beta = 2(1-\beta)$, $\rho_\beta = 2\beta - 1 \ge 0$ and $e_\beta = 0$. Let J be a random variable independent of (X,Y), taking values 0, 1 and 2 with probabilities κ_β, ρ_β and e_β, respectively, and let Z be the random variable on $\mathcal{Z} = (\{0\} \times \mathcal{Y}) \cup (\{1\} \times \mathcal{X}) \cup \{\perp\}$, where $\perp$ is a new symbol, defined by

$$Z = \begin{cases} (0, Y), & \text{if } J = 0, \\ (1, X), & \text{if } J = 1, \\ \perp, & \text{if } J = 2. \end{cases} \tag{80}$$

In other words, Z reveals Y with probability κ_β, reveals X with probability ρ_β, and reveals nothing with probability e_β. Let $\tilde{P}_{XZ}$ be the joint distribution of (X,Z). Its marginal on X is P_X, so that the encoders of X for P_{XY} and for $\tilde{P}_{XZ}$ coincide.

Let T be an encoder of X. Since J is a function of Z and is independent of (X, Y, T), the chain rule gives

$$I(Z;T) = I(Z,J;T) = I(J;T) + I(Z;T \mid J) = \kappa_\beta\, I(Y;T) + \rho_\beta\, I(X;T), \tag{81}$$

since $I(J;T) = 0$, and, given J, the random variable Z is equivalent to Y, to X, or constant. Subtracting $\beta I(X;T)$ from both sides of (81), and using $\beta - \rho_\beta = \kappa_\beta/2$, we get that

$$I(Z;T) - \beta I(X;T) = \kappa_\beta\Big(I(Y;T) - \frac{1}{2} I(X;T)\Big). \tag{82}$$

Since (82) holds for every encoder, taking the supremum over all encoders proves (79).

We now show that IB-LAGRANGIAN$_\beta$ is NP-hard by reducing IB-LAGRANGIAN to it. Given an instance (P_{XY}, τ) of IB-LAGRANGIAN, consider the instance $(\tilde{P}_{XZ}, \kappa_\beta\tau)$ of IB-LAGRANGIAN$_\beta$. By (82), an encoder satisfies $I(Y;T) - \frac{1}{2} I(X;T) \geq \tau$ if and only if it satisfies $I(Z;T) - \beta I(X;T) \geq \kappa_\beta\tau$, so the two instances have the same answer. Moreover, $|\mathcal{Z}| = |\mathcal{X}| + |\mathcal{Y}| + 1$, and, since κ_β, ρ_β and e_β are fixed rational numbers, by (80) each entry of $\tilde{P}_{XZ}$ is a fixed rational multiple of an entry of P_{XY} or of P_X. Hence, the instance $(\tilde{P}_{XZ}, \kappa_\beta\tau)$ can be computed in polynomial time, and, by Theorem 1, IB-LAGRANGIAN$_\beta$ is NP-hard.

Finally, suppose that a polynomial-time algorithm approximates L_β within an additive error of $\kappa_\beta/(54|\mathcal{X}|^{3/2})$. Running it on $\tilde{P}_{XZ}$, which has the same input alphabet $\mathcal{X}$ as P_{XY}, and dividing its output by κ_β, we obtain, by (79), an approximation of $L_{1/2}(P_{XY})$ within $1/(54|\mathcal{X}|^{3/2})$, which, by Theorem 1, is impossible unless $\mathrm{P} = \mathrm{NP}$. □

Remark 1. If β is irrational, then κ_β, ρ_β and e_β may be irrational, and the joint distribution $\tilde{P}_{XZ}$, constructed in the proof of Proposition 2, is not a rational distribution. Nevertheless, we can still extend Proposition 2, with a different constant, to every fixed real $\beta \in (0,1)$ that is computable in polynomial time, that is, such that a rational number within distance 2^{-r} of β can be computed in time polynomial in r. The idea is to apply the construction (80) with a rational multiplier b close to β, chosen depending on the instance. Replacing β with b changes the objective of every encoder by $|\beta - b|\, I(X;T) \leq |\beta - b| \log|\mathcal{X}|$, so that $|L_\beta(\tilde{P}_{XZ}) - L_b(\tilde{P}_{XZ})| \leq |\beta - b| \log|\mathcal{X}|$. The instances of Theorem 1 are separated from the threshold by a margin $\gamma = \frac{5\delta}{12}$, which becomes $\kappa_b\gamma$ after the construction (80). Hence, if $|\beta - b| \log|\mathcal{X}|$ is a small fraction of $\kappa_b\gamma$, the threshold $\kappa_b\tau$ still separates yes- and no-instances. By Lemma 18, γ is of order $|\mathcal{X}|^{-3/2}$, so the required precision is of order $(|\mathcal{X}|^{3/2} \log|\mathcal{X}|)^{-1}$, which corresponds to $O(\log|\mathcal{X}|)$ bits. Hence, b can be computed in polynomial time, and the arguments in the proof of Proposition 2 show that, for every fixed polynomial-time computable real $\beta \in (0,1)$, IB-LAGRANGIAN$_\beta$ is NP-hard, and that L_β cannot be approximated in polynomial time within an additive error of order $|\mathcal{X}|^{-3/2}$, unless $\mathrm{P} = \mathrm{NP}$.

F. Proof of Theorem 2

We now prove Theorem 2, using the same joint distribution P_{XY} defined in (61). However, in this case, an instance of IB-THRESHOLD is a triple (P_{XY}, R, ϑ). Thus, we need to define the budget R and the threshold ϑ. By Lemma 15, if $\omega(G) < k$, then $\mathrm{IB}(P_{XY}, R) \le \frac{R}{2} + \Pi_0$ for every $R \ge 0$. On the other hand, if $\omega(G) \ge k$, then, by Lemma 17, $I(Y;T_1) > \frac{1}{2} I(X;T_1) + \Pi_0 + \delta$.

We must choose the budget R to satisfy two requirements. Whenever G contains a k-clique, we need $R \ge I(X;T_1)$ so that T_1 satisfies the rate constraint. We also need R to be sufficiently close to $I(X;T_1)$ so that the upper bound $\frac{R}{2} + \Pi_0$ for the no-case given by Lemma 15 remains below $I(Y;T_1)$ by a gap of order δ.

The difficulty is that the reduction must choose R without knowing whether G contains a k-clique, while T_1 is defined only when such a clique exists. The next lemma addresses this difficulty by giving an explicit expression, depending only on N and k, that is defined for every instance and equals $I(X;T_1)$ whenever G contains a k-clique. The reduction can therefore use this expression to choose R from (G,k) alone, without knowing whether such a clique exists.

Lemma 19. *Let p, ε and λ be as in* (65)*, let a_0 and a_1 be as in* (34)*, and let*

$$R^* = \log N - \lambda \log \frac{m+1}{m-1} + \frac{B(1-\varepsilon)}{N}\Big(\log s - (s-2)\, a_0 \log \frac{1}{a_0} - 2a_1 \log \frac{1}{a_1} \Big). \tag{83}$$

Then R^* depends only on N and k, and, if $\omega(G) \ge k$, the encoder T_1 defined in (32)*, with p as in* (65)*, satisfies $I(X;T_1) = R^*$.*

Proof. We first show that R^* depends only on $N = |E(G)|$ and k. Recall that $m = k-1$, $B = k(k-1)/2$, $s = m(m-1)$ and $\mu = 1/s$. Hence, by (34) and (65), the quantities p, ε, λ, a_0 and a_1, and thus all the quantities in (83), depend only on N and k. Moreover, $a_0, a_1 > 0$, so the logarithms are well defined.

Now assume that $\omega(G) \ge k$, and fix a clique K of size k in G. Let T_1 be the encoder defined in (32) using this clique, with p chosen as in (65). We show that $I(X;T_1) = R^*$. By Lemma 8, we have that

$$I(X;T_1) = \log N - \lambda \log \frac{m+1}{m-1} + \frac{B(1-\varepsilon)}{N}\, D_V, \tag{84}$$

and, by Lemma 7, $D_V = \log s - H(V_e)$. Since, by (34), V_e takes the value a_0 on $s-2$ oriented edges and the value a_1 on 2 oriented edges, we have that $H(V_e) = (s-2)\, a_0 \log \frac{1}{a_0} + 2a_1 \log \frac{1}{a_1}$. Substituting this into (84) gives $I(X;T_1) = R^*$. □

We can now define the budget R and the threshold ϑ. Since R^* in (83) is in general irrational, we take a rational number $\tilde{R}$ such that $|\tilde{R} - R^*| \le \frac{\delta}{4}$, and we set

$$R = \max\Big\{ \tilde{R} + \frac{\delta}{4},\ 0 \Big\}. \tag{85}$$

The additive term $\frac{\delta}{4}$ ensures that $R \geq R^*$, so that the encoder T_1 satisfies the rate constraint whenever G contains a k-clique, while the maximum ensures that $R \geq 0$. To choose the threshold ϑ, we compare the two cases. If $\omega(G) \geq k$, then, by Lemma 19, $R^* = I(X;T_1) \geq 0$, so that the maximum in (85) equals $\tilde{R} + \frac{\delta}{4}$, and, since $|\tilde{R} - R^*| \leq \frac{\delta}{4}$, we have that

$$R^* \leq R \leq R^* + \frac{\delta}{2}. \tag{86}$$

Hence, by Lemma 17 and $R^* \geq R - \delta/2$ from the rightmost inequality in (86), we obtain that

$$I(Y;T_1) > \frac{R^*}{2} + \Pi_0 + \delta \geq \frac{1}{2}\Big(R - \frac{\delta}{2}\Big) + \Pi_0 + \delta = \frac{R}{2} + \Pi_0 + \frac{3\delta}{4}. \tag{87}$$

On the other hand, if $\omega(G) < k$, then $\mathrm{IB}(P_{XY}, R) \leq \frac{R}{2} + \Pi_0$ by Lemma 15. Thus, as in the proof of Theorem 1, we place the threshold in the middle of this gap, and, since this value is in general irrational, we take ϑ as a rational number such that

$$\Big|\vartheta - \Big(\frac{R}{2} + \Pi_0 + \frac{3\delta}{8}\Big)\Big| \leq \frac{\delta}{8}. \tag{88}$$

Therefore, the instance of IB-THRESHOLD associated with (G, k) is (P_{XY}, R, ϑ), where P_{XY} is defined in (61), R in (85) and ϑ in (88).

Proof of Theorem 2. We start by showing that IB-THRESHOLD is NP-hard. To this end, we show that the construction of (P_{XY}, R, ϑ) is a polynomial-time reduction from CLIQUE to IB-THRESHOLD.

We first show that the reduction runs in polynomial time. As shown in the proof of Theorem 1, $\bar{G}$ and P_{XY} can be constructed in polynomial time, so it remains to show that R and ϑ can be computed in polynomial time. By (83), R^* is a linear combination, with rational coefficients, of the logarithms of N, $m+1$, $m-1$, s, $1/a_0$ and $1/a_1$, which are rational numbers with $O(\log n)$ bits. Hence, by Lemma 25 in Appendix A, $\tilde{R}$ can be computed in polynomial time, and so can R by (85). Similarly, $\frac{R}{2} + \Pi_0 + \frac{3\delta}{8}$ is the sum of a rational number and of Π_0, which, as shown in the proof of Theorem 1, is an explicit expression involving polynomially many logarithms. Hence, again by Lemma 25, a rational ϑ satisfying (88) can be computed in polynomial time. Therefore, the whole reduction runs in polynomial time.

We now show that the reduction is correct, that is, (P_{XY}, R, ϑ) is a yes-instance of IB-THRESHOLD if and only if (G, k) is a yes-instance of CLIQUE. Recall that (P_{XY}, R, ϑ) is a yes-instance of IB-THRESHOLD if there exists an encoder T of X such that $I(X;T) \leq R$ and $I(Y;T) \geq \vartheta$, and that (G, k) is a yes-instance of CLIQUE if $\omega(G) \geq k$. We notice that, by (88), the threshold ϑ satisfies

$$\frac{R}{2} + \Pi_0 + \frac{\delta}{4} \leq \vartheta \leq \frac{R}{2} + \Pi_0 + \frac{\delta}{2}. \tag{89}$$

Assume first that (G, k) is a yes-instance, that is, $\omega(G) \geq k$. We show that the encoder T_1 satisfies both $I(X;T_1) \leq R$ and $I(Y;T_1) \geq \vartheta$. By Lemma 19 and the leftmost inequality in (86), we have that

$I(X;T_1) = R^* \leq R$, and therefore the encoder T_1 satisfies the rate constraint. Moreover, by (87) and the rightmost inequality in (89), we get that $I(Y;T_1) > \frac{R}{2} + \Pi_0 + \frac{3\delta}{4} \geq \vartheta + \frac{\delta}{4}$. Hence, T_1 satisfies both $I(X;T_1) \leq R$ and $I(Y;T_1) \geq \vartheta$, and (P_{XY}, R, ϑ) is a yes-instance of IB-THRESHOLD. Moreover, $\mathrm{IB}(P_{XY}, R) \geq I(Y;T_1) > \vartheta + \frac{\delta}{4}$.

Assume now that (G,k) is a no-instance, that is, $\omega(G) < k$. By Lemma 15 and the leftmost inequality in (89), we get that $\mathrm{IB}(P_{XY}, R) \leq \frac{R}{2} + \Pi_0 \leq \vartheta - \frac{\delta}{4}$. Hence, every encoder with $I(X;T) \leq R$ satisfies $I(Y;T) < \vartheta$, and (P_{XY}, R, ϑ) is a no-instance of IB-THRESHOLD. This proves that the reduction is correct, and hence that IB-THRESHOLD is NP-hard.

It remains to prove the inapproximability result. By the two cases above, $\mathrm{IB}(P_{XY}, R)$ lies above $\vartheta + \frac{\delta}{4}$ if $\omega(G) \geq k$, and below $\vartheta - \frac{\delta}{4}$ otherwise. Suppose that a polynomial-time algorithm outputs v with $|v - \mathrm{IB}(P_{XY}, R)| \leq \frac{1}{144M^{3/2}}$. By Lemma 18, this error is at most $\frac{\delta}{8} < \frac{\delta}{4}$. Hence, if $\omega(G) \geq k$, then $v \geq \mathrm{IB}(P_{XY}, R) - \frac{\delta}{8} > \vartheta$, while, if $\omega(G) < k$, then $v \leq \mathrm{IB}(P_{XY}, R) + \frac{\delta}{8} < \vartheta$. Thus, running the reduction, then the algorithm, and comparing v with ϑ would decide CLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

G. A better exponent

In this section, we show that the exponent $3/2$ in Theorems 1 and 2 and Proposition 2 can be lowered to $1+\alpha$, for every fixed $\alpha > 0$. Recall that a smaller exponent means a larger additive error, and hence a stronger inapproximability result.

To explain where the exponent $3/2$ comes from, recall from (76) that the gap of the reduction is $\delta = \frac{1-1/m}{8Mm^2}$. The factor $\frac{1}{M}$ is of the desired order, while the factor $\frac{1}{m^2}$ has to be bounded in terms of M. In the proof of Lemma 18, the only relation between m and M comes from the assumption $N \geq B \approx m^2/2$ in (25), which gives $m^2 \leq \frac{3}{2}\sqrt{M}$, and hence $\delta \geq \frac{1}{18M^{3/2}}$. The idea to obtain a better result is therefore to replace the instance (G,k) with an equivalent instance (G',k), that is, one with the same answer to CLIQUE, in which G' has many more edges than G. In this way, m^2 becomes small with respect to $M = |A(\bar{G}')|$. This is formalized in the following theorem.

Theorem 3. *Let $\alpha > 0$ be fixed. Unless $\mathrm{P} = \mathrm{NP}$, there is no polynomial-time algorithm that, given a rational joint distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$, approximates $L_{1/2}(P_{XY})$ within an additive error of $\frac{1}{36|\mathcal{X}|^{1+\alpha}}$, nor one that, given also a non-negative rational number R, approximates $\mathrm{IB}(P_{XY}, R)$ within $\frac{1}{96|\mathcal{X}|^{1+\alpha}}$. Moreover, for every fixed $\beta \in \mathbb{Q} \cap (0,1)$, there is no polynomial-time algorithm that approximates $L_\beta(P_{XY})$ within $\frac{\kappa_\beta}{36|\mathcal{X}|^{1+\alpha}}$.*

Proof. We first construct the new instance of CLIQUE. Let $c = \lceil 1/\alpha \rceil$, so that c is an integer and $1/c \leq \alpha$, and let (G, k) be an instance of CLIQUE satisfying (25), with $n = |V(G)|$. Let $N' = n^{2c}$, and let $u_1, v_1, \ldots, u_{N'}, v_{N'}$ be $2N'$ new vertices, not in $V(G)$. We define the graph G' by

$$V(G') = V(G) \cup \{u_1, v_1, \ldots, u_{N'}, v_{N'}\} \quad \text{and} \quad E(G') = E(G) \cup \big\{\{u_i, v_i\} : 1 \leq i \leq N'\big\},$$

that is, G' is the disjoint union of G and of N' isolated edges. We claim that G' contains a k-clique if and only if G does. Indeed, every clique of G is a clique of G'. Conversely, let C be a clique of G' that contains a new vertex, say u_i. Since the only neighbour of u_i in G' is v_i, we have that $C \subseteq \{u_i, v_i\}$, and hence $|C| \leq 2 < k$. The same holds for v_i. Therefore, every k-clique of G' is contained in $V(G)$, and it is a k-clique of G. Moreover, (G', k) satisfies (25), since $|E(G')| = N + N' \geq N \geq B$. Finally, G' has $n + 2n^{2c}$ vertices, and, since c is fixed, its adjacency matrix can be constructed in time polynomial in n.

We now apply the reductions of Theorems 1 and 2 to (G', k) instead of (G, k). All the results of Sections III and IV hold for every instance satisfying (25), and hence for (G', k) as well. Moreover, since the size of G' is polynomial in n, both reductions still run in time polynomial in n. The only change concerns the lower bound on δ. Since $m = k - 1 < n$, the extended graph of G' has

$$M = |E(G')|\, m(m-1) \geq N' = n^{2c} > m^{2c}$$

oriented edges, that is, $m^2 \leq M^{1/c}$. Using this bound in (76) in place of (77), together with $1 - \frac{1}{m} \geq \frac{2}{3}$ and $1/c \leq \alpha$, we get that

$$\delta \geq \frac{2/3}{8\, M \cdot M^{1/c}} = \frac{1}{12\, M^{1+1/c}} \geq \frac{1}{12\, M^{1+\alpha}}. \tag{90}$$

The proof of Theorem 1 rules out polynomial-time approximation of $L_{1/2}(P_{XY})$ within $\frac{\delta}{3}$, and the proof of Theorem 2 rules out polynomial-time approximation of $\mathrm{IB}(P_{XY}, R)$ within $\frac{\delta}{8}$. Since $M = |\mathcal{X}|$, by (90) we have that $\frac{\delta}{3} \geq \frac{1}{36|\mathcal{X}|^{1+\alpha}}$ and $\frac{\delta}{8} \geq \frac{1}{96|\mathcal{X}|^{1+\alpha}}$, which proves the first two claims. The last claim follows from the first one exactly as in the proof of Proposition 2, since the construction (80) does not change the alphabet $\mathcal{X}$ and multiplies the Lagrangian by κ_β. □

Note that Theorem 3 improves on Theorems 1 and 2 only for $0 < \alpha < 1/2$, that is, for $c \geq 3$, and that it becomes stronger as α decreases, at the price of a larger graph G'. The exponent cannot be lowered to 1 within this construction. Indeed, the improvement δ of T_1 over T_0 is of order $\frac{1}{Nm^4} = \Theta\big(\frac{1}{Mm^2}\big)$, and m cannot be bounded, since CLIQUE can be solved in polynomial time for every fixed k.

V. HARDNESS OF THE PRIVACY FUNNEL

In this section, we prove our results for the privacy funnel. For the constrained formulation (3), we prove the following theorem.

Theorem 4. PF-THRESHOLD *is NP-hard. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a rational joint distribution* P_{XY} *on* $\mathcal{X} \times \mathcal{Y}$ *and a rational number* $0 \le r \le H(X)$*, outputs a value* v *such that*

$$\big|v - \mathrm{PF}(P_{XY}, r)\big| \le \frac{1}{4|\mathcal{X}|^2}. \tag{91}$$

Theorem 4 is proved by a reduction from CLIQUE, which reuses the extended graph $\bar{G}$, the random variable X, and the encoders T_0 and T_1 of Section III. Only the channel from X to Y, and hence the joint distribution P_{XY}, changes. A key tool for the reduction is the lower bound of Lemma 2, namely $I(Y;T) \ge I(X;T) - H(X \mid Y)$, which holds for every encoder T of X.

The rest of this section is organized as follows. In Section V-A, we introduce compatible sets of oriented edges, that is, sets of oriented edges of a common clique, and we define a function Φ of the distributions on the oriented edges, which plays a similar role to that of the cost g_W of Section IV. We then bound Φ in terms of the clique number of $\bar{G}$. In Section V-B, we define the joint distribution of the reduction, and we show that an encoder attains the lower bound of Lemma 2 if its posteriors are supported on compatible sets. In Section V-C, we show that, if G has no k-clique, the lower bound of Lemma 2 cannot be attained at any rate smaller than $\log N$, while, if G contains a k-clique, the encoder T_1 attains the lower bound at a strictly smaller rate. Finally, we use these results to prove Theorem 4 in Section V-D, and in Section V-E, we extend our results to the Lagrangian formulation (4) (Proposition 3). As for the information bottleneck, we made no attempt to optimize the constants in these results.

A. Compatible sets and an entropic bound

Let (G, k) be an instance of CLIQUE satisfying (25), and let $\bar{G}$ be the extended graph of Section III-B. We say that two distinct oriented edges $x, x' \in A(\bar{G})$ are *compatible* if the set of their endpoints is a clique of $\bar{G}$, that is, if x and x' can be oriented edges of the same clique; otherwise, they are *incompatible*. A set $S \subseteq A(\bar{G})$ is *compatible* if any two of its elements are compatible. We denote by $\mathcal{I}(\bar{G})$ the set of unordered pairs $\{x, x'\}$ of incompatible oriented edges of $\bar{G}$.

The next lemma shows that a set of oriented edges is compatible if and only if it is contained in a clique of $\bar{G}$. Consequently, the size of a compatible set is bounded in terms of the clique number of $\bar{G}$.

Lemma 20. *A set* $S \subseteq A(\bar{G})$ *is compatible if and only if* $S \subseteq A(C)$ *for some clique* C *of* $\bar{G}$*. Moreover, every compatible set* S *satisfies* $|S| \le \omega(\bar{G})(\omega(\bar{G}) - 1)$*.*

Proof. We first prove that, if $S \subseteq A(C)$ for some clique C of $\bar{G}$, then S is compatible. Let $x, x' \in S$ be distinct. Since $x, x' \in A(C)$, all their endpoints belong to C. Moreover, since C is a clique and every

subset of a clique is still a clique, it follows that the set of the endpoints of x and x' is a clique of $\bar{G}$, which means that x and x' are compatible. Hence, S is compatible.

Conversely, assume that S is compatible. We show that $S \subseteq A(C)$ for some clique C of $\bar{G}$. Let

$$C = \{u \in V(\bar{G}) : u \text{ is an endpoint of some } x \in S\}. \tag{92}$$

By definition of C, we have that $S \subseteq A(C)$, so it suffices to show that C is a clique of $\bar{G}$. Let $u, v \in C$ be distinct, and let $x, x' \in S$ be oriented edges such that u is an endpoint of x and v is an endpoint of x'. If $x = x'$, then u and v are the two endpoints of x, that is, $x = (u, v)$ or $x = (v, u)$. Since $x \in A(\bar{G})$, we have that $\{u, v\} \in E(\bar{G})$, so u and v are adjacent. If $x \neq x'$, then x and x' are compatible, since S is compatible, and the set of their endpoints, which contains u and v, is a clique of $\bar{G}$. Hence, u and v are adjacent also in this case. This implies that C is a clique of $\bar{G}$. Hence, we proved that a set $S \subseteq A(\bar{G})$ is compatible if and only if $S \subseteq A(C)$ for some clique C of $\bar{G}$.

Moreover, if S is compatible, then $S \subseteq A(C)$ for some clique C of $\bar{G}$, and hence it follows that

$$|S| \leq |A(C)| = |C|(|C| - 1) \leq \omega(\bar{G})(\omega(\bar{G}) - 1)$$

by definition of ω in (20). □

For a probability distribution Q on $A(\bar{G})$, let

$$\Phi(Q) = H(Q) - \sum_{\{x,x'\} \in \mathcal{I}(\bar{G})} \big(Q(x) + Q(x')\big)\, h_2\Big(\frac{Q(x)}{Q(x) + Q(x')}\Big), \tag{93}$$

where we assume that each term of the sum with $Q(x) + Q(x') = 0$ is equal to zero. In words, $\Phi(Q)$ is the entropy of Q minus a penalty for each pair of incompatible oriented edges that both have positive probability under Q. In particular, if $\mathrm{supp}(Q)$ is compatible, then all the penalties vanish and $\Phi(Q) = H(Q)$. In the next lemma, we bound Φ in terms of the clique number of $\bar{G}$. This bound plays, for the privacy funnel, the same role that the bound on $g_{\bar{G}}$ in Lemma 12 plays for the information bottleneck.

Lemma 21. *For every probability distribution Q on $A(\bar{G})$, it holds that*

$$\Phi(Q) \leq \log\big(\omega(\bar{G})(\omega(\bar{G}) - 1)\big). \tag{94}$$

Proof. Let Q be a probability distribution on $A(\bar{G})$. We show that, whenever $\mathrm{supp}(Q)$ is not compatible, that is, it contains both elements of at least one pair $\{x, x'\}$ in $\mathcal{I}(\bar{G})$, we can move all the probability of x to x', or of x' to x, without decreasing the value of Φ. Reiterating this operation, we reach a probability distribution Q' with compatible support and $\Phi(Q') \geq \Phi(Q)$. Then, the claim follows from Lemma 20.

Assume that $Q(x) > 0$ and $Q(x') > 0$ for some $\{x, x'\} \in \mathcal{I}(\bar{G})$, and let $c = Q(x) + Q(x')$. For $a \in [0, c]$, let Q_a be the probability distribution that coincides with Q except that $Q_a(x) = a$ and

$Q_a(x') = c - a$. Note that $Q = Q_{Q(x)}$, and that Q_0 and Q_c are the probability distributions obtained from Q by moving all the probability of x to x', and of x' to x, respectively. We show that the function $a \mapsto \Phi(Q_a)$ is convex on $[0, c]$. To this end, we recall that

$$\Phi(Q_a) = H(Q_a) - \sum_{\{y,y'\} \in \mathcal{I}(\bar{G})} \big(Q_a(y) + Q_a(y')\big)\, h_2\Big(\frac{Q_a(y)}{Q_a(y) + Q_a(y')}\Big). \tag{95}$$

We now split the terms of (95) that depend on a into three groups.

First, since Q_a coincides with Q outside $\{x, x'\}$, the terms of $H(Q_a)$ that depend on a are

$$-a \log a - (c - a) \log(c - a). \tag{96}$$

Second, the term of the sum in (95) corresponding to the pair $\{y, y'\} = \{x, x'\}$ is equal to

$$c\, h_2\Big(\frac{a}{c}\Big) = -a \log a - (c - a) \log(c - a) + c \log c. \tag{97}$$

Hence, the difference between (96) and (97) gives the term $-c \log c$, which is constant and does not depend on a.

Third, the only other terms of the sum in (95) that depend on a are those corresponding to pairs $\{x, z\}$ or $\{x', z\}$ in $\mathcal{I}(\bar{G})$, with $z \notin \{x, x'\}$. Each of these terms has the form

$$f(b) = (b + w)\, h_2\Big(\frac{b}{b + w}\Big) = -b \log b - w \log w + (b + w) \log(b + w),$$

where $w = Q(z)$, which does not depend on a, and $b = a$ or $b = c - a$. Since $f''(b) = \frac{1}{\ln 2}\big(\frac{1}{b+w} - \frac{1}{b}\big) \leq 0$ for $b > 0$, the function f is concave, and consequently each of these terms is a concave function of a.

Putting the three groups together, we get that, as a function of $a \in [0, c]$, $\Phi(Q_a)$ is equal to a constant minus a sum of concave functions of a. Hence, $a \mapsto \Phi(Q_a)$ is convex on $[0, c]$, and, since a convex function on an interval attains its maximum at an endpoint, we get that

$$\Phi(Q) = \Phi(Q_{Q(x)}) \leq \max\{\Phi(Q_0), \Phi(Q_c)\}. \tag{98}$$

Since both Q_0 and Q_c have one element less in their support than Q, by repeating this operation, after at most $|\operatorname{supp}(Q)| - 1$ steps we obtain a probability distribution Q' with $\Phi(Q') \geq \Phi(Q)$ whose support is compatible. For such a distribution Q', all the terms of the sum in (93) vanish, so that $\Phi(Q') = H(Q')$. Hence, since $\operatorname{supp}(Q')$ is a compatible set, by Lemma 20 we finally obtain that

$$\Phi(Q) \leq \Phi(Q') = H(Q') \leq \log |\operatorname{supp}(Q')| \leq \log\big(\omega(\bar{G})(\omega(\bar{G}) - 1)\big),$$

which concludes the proof. □

B. The joint distribution

We now describe the joint distribution of the instance of PF-THRESHOLD. As in Section III, we let the random variable X be uniformly distributed on $\mathcal{X} = A(\bar{G})$, with $|A(\bar{G})| = M = Ns$, where $N = |E(G)|$ and $s = m(m-1)$. Moreover, we let

$$R_0 = \log N, \tag{99}$$

which, by Lemma 5, is the rate of the encoder T_0, that is, $I(X;T_0) = R_0$.

We now define the random variable Y. For each pair $\{x, x'\} \in \mathcal{I}(\bar{G})$, we introduce a symbol $y_{\{x,x'\}}$. For each $x \in A(\bar{G})$, let $d(x)$ be the number of oriented edges incompatible with x, and note that $d(x) \leq M-1$. We also introduce $M - d(x)$ symbols $(x,1), \ldots, (x, M-d(x))$. The alphabet $\mathcal{Y}$ of Y is the set of all these symbols. For every $x \in A(\bar{G})$, let

$$\mathcal{Y}_x = \big\{y_{\{x,z\}} : z \in A(\bar{G}),\ \{x,z\} \in \mathcal{I}(\bar{G})\big\} \cup \big\{(x,1), \ldots, (x, M-d(x))\big\} \tag{100}$$

be the set of values that Y can take when $X = x$, and note that $|\mathcal{Y}_x| = d(x) + (M - d(x)) = M$. Given $X = x$, we let Y be uniformly distributed on $\mathcal{Y}_x$, that is,

$$W(y \mid x) = P_{Y|X}(y \mid x) = \begin{cases} 1/M, & \text{if } y \in \mathcal{Y}_x, \\ 0, & \text{otherwise.} \end{cases} \tag{101}$$

Consequently, since P_X is uniform on $A(\bar{G})$, the joint distribution P_{XY} of (X,Y) is

$$P_{XY}(x,y) = P_X(x)\, W(y \mid x) = \begin{cases} \dfrac{1}{M^2}, & \text{if } y \in \mathcal{Y}_x, \\ 0, & \text{otherwise,} \end{cases} \qquad \forall (x,y) \in A(\bar{G}) \times \mathcal{Y}. \tag{102}$$

By (100) and (102), we notice that, after observing Y, we can draw the following conclusions about X. If $Y = (x,i)$, then, since the symbol (x,i) belongs to $\mathcal{Y}_x$ only, $X = x$ with certainty. If $Y = y_{\{x,x'\}}$, then, since the symbol $y_{\{x,x'\}}$ belongs to both $\mathcal{Y}_x$ and $\mathcal{Y}_{x'}$, and to no other set $\mathcal{Y}_z$, X is equal to x or to x', each with probability $1/2$. Consequently, Y determines X, unless Y is a symbol $y_{\{x,x'\}}$, in which case X is one of the two incompatible oriented edges x and x'. This property is crucial in our reduction. Indeed, if we also know that X belongs to a compatible set S, then at most one of x and x' belongs to S, and hence X is determined by Y. Let

$$h_0 = H(X \mid Y) \tag{103}$$

be the uncertainty about X that remains after observing Y. The next lemma provides some useful properties of the joint distribution P_{XY} in (102) and of the channel W in (101).

Lemma 22. *Let P_{XY} be the joint distribution defined in* (102)*, and let W be the channel from $\mathcal{X}$ to $\mathcal{Y}$ defined in* (101)*. Then, the following properties hold:*

1) $h_0 = 2|\mathcal{I}(\bar{G})|/M^2$ *and* $H(Y) = 2\log M - h_0$*;*
2) *for every probability distribution Q on $A(\bar{G})$,*

$$H(WQ) = \log M + \Big(1 - \frac{1}{M}\Big)H(Q) + \frac{\Phi(Q)}{M}; \tag{104}$$

3) *if every posterior of an encoder T of X is supported on a compatible set, then $I(Y;T) = I(X;T) - h_0$.*

Proof. We prove the three properties separately.

Property 1. We first compute h_0. By definition of the conditional entropy in (7), we have that

$$h_0 = H(X \mid Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, H(X \mid Y = y). \tag{105}$$

As observed before, if $y = (x,i)$, then, by (102), $P_{XY}(z,y) > 0$ only for $z = x$. Therefore, X is determined by Y, and $H(X \mid Y = y) = 0$. If, instead, $y = y_{\{x,x'\}}$, then, by (102), $P_{XY}(x, y_{\{x,x'\}}) = P_{XY}(x', y_{\{x,x'\}}) = 1/M^2$ and $P_{XY}(z, y_{\{x,x'\}}) = 0$ for every other z. Hence, $P_Y(y_{\{x,x'\}}) = 2/M^2$, and, given $Y = y_{\{x,x'\}}$, X is uniformly distributed on $\{x, x'\}$, so that $H(X \mid Y = y_{\{x,x'\}}) = 1$. Consequently, since there are $|\mathcal{I}(\bar{G})|$ symbols of the form $y_{\{x,x'\}}$, splitting the sum in (105) according to the two kinds of symbols in $\mathcal{Y}$, we obtain that

$$\begin{aligned} h_0 &= \sum_{x \in A(\bar{G})} \sum_{i=1}^{M-d(x)} P_Y\big((x,i)\big)\, H\big(X \mid Y = (x,i)\big) + \sum_{\{x,x'\} \in \mathcal{I}(\bar{G})} P_Y\big(y_{\{x,x'\}}\big)\, H\big(X \mid Y = y_{\{x,x'\}}\big) \\ &= \sum_{x \in A(\bar{G})} \sum_{i=1}^{M-d(x)} P_Y\big((x,i)\big) \cdot 0 + \sum_{\{x,x'\} \in \mathcal{I}(\bar{G})} \frac{2}{M^2} \cdot 1 = |\mathcal{I}(\bar{G})| \cdot \frac{2}{M^2} = \frac{2|\mathcal{I}(\bar{G})|}{M^2}. \end{aligned}$$

We now compute $H(Y)$. By the chain rule (9), $H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$, and hence

$$H(Y) = H(X) + H(Y \mid X) - H(X \mid Y). \tag{106}$$

Since X is uniform on $A(\bar{G})$, we have that $H(X) = \log M$, and, by (101), given $X = x$, Y is uniform on the M elements of $\mathcal{Y}_x$, so that $H(Y \mid X) = \log M$. Putting these results together and recalling that $H(X \mid Y) = h_0$, from (106) we obtain that $H(Y) = 2\log M - h_0$.

Property 2. Let Q be a probability distribution on $A(\bar{G})$. We compute $H(WQ)$ explicitly. Let $a_x = Q(x)/M$ for every $x \in A(\bar{G})$. By the definition of the channel W in (101), each symbol (x,i) has

probability $(WQ)\big((x,i)\big) = a_x$, and each symbol $y_{\{x,x'\}}$ has probability $(WQ)\big(y_{\{x,x'\}}\big) = a_x + a_{x'}$. Since there are $M - d(x)$ symbols of the form (x,i) for each $x \in A(\bar{G})$, we obtain that

$$H(WQ) = -\sum_{x\in A(\bar{G})} \big(M - d(x)\big)\, a_x \log a_x - \sum_{\{x,x'\}\in\mathcal{I}(\bar{G})} (a_x + a_{x'}) \log(a_x + a_{x'}). \tag{107}$$

To rewrite the first sum in (107), we observe that the term $a_x \log a_x$ of a fixed $x \in A(\bar{G})$ appears once in the sum $\sum_{\{x,x'\}\in\mathcal{I}(\bar{G})} \big(a_x \log a_x + a_{x'} \log a_{x'}\big)$ for each pair of $\mathcal{I}(\bar{G})$ containing x, that is, exactly $d(x)$ times. Hence, we have that

$$\sum_{x\in A(\bar{G})} d(x)\, a_x \log a_x = \sum_{\{x,x'\}\in\mathcal{I}(\bar{G})} \big(a_x \log a_x + a_{x'} \log a_{x'}\big). \tag{108}$$

Substituting (108) into (107), we get that

$$H(WQ) = -\sum_{x\in A(\bar{G})} M\, a_x \log a_x - \sum_{\{x,x'\}\in\mathcal{I}(\bar{G})} \Big((a_x + a_{x'}) \log(a_x + a_{x'}) - a_x \log a_x - a_{x'} \log a_{x'}\Big). \tag{109}$$

We now evaluate the two sums in (109) separately. For the first one, since $M a_x = Q(x)$, we get that

$$-\sum_{x\in A(\bar{G})} M\, a_x \log a_x = -\sum_{x\in A(\bar{G})} Q(x) \log \frac{Q(x)}{M} = H(Q) + \log M. \tag{110}$$

For the second one, we use the identity $(a+b)\log(a+b) - a\log a - b\log b = (a+b)\, h_2\big(\frac{a}{a+b}\big)$, valid for $a, b \geq 0$ with $a + b > 0$. Since $a_x + a_{x'} = \frac{Q(x)+Q(x')}{M}$ and $\frac{a_x}{a_x+a_{x'}} = \frac{Q(x)}{Q(x)+Q(x')}$, each term of the second sum equals $\frac{Q(x)+Q(x')}{M}\, h_2\big(\frac{Q(x)}{Q(x)+Q(x')}\big)$, and, by the definition of Φ in (93), the whole second sum equals $\frac{1}{M}\big(H(Q) - \Phi(Q)\big)$. Substituting this and (110) into (109), we obtain that

$$H(WQ) = \log M + H(Q) - \frac{1}{M}\big(H(Q) - \Phi(Q)\big) = \log M + \Big(1 - \frac{1}{M}\Big) H(Q) + \frac{\Phi(Q)}{M},$$

which proves (104).

Property 3. Let T be an encoder of X whose posteriors are all supported on compatible sets. As in the proof of Lemma 2, we have that

$$I(Y;T) = I(X;T) - I(X;T \mid Y) = I(X;T) - h_0 + H(X \mid Y,T). \tag{111}$$

Hence, it suffices to show that $H(X \mid Y,T) = 0$, that is, that X is determined by Y and T. By definition of conditional entropy (see (7)), it reduces to showing that $H(X \mid Y, T = t) = 0$ for every t with $P_T(t) > 0$. Fix such a t, and let $S = \operatorname{supp}(P_{X|T}(\cdot \mid t))$ be the support of $P_{X|T}(\cdot \mid t)$, which, by assumption, is compatible. Since Y and T are conditionally independent given X, we have that

$$P_{XY|T}(x, y \mid t) = P_{X|T}(x \mid t)\, W(y \mid x),$$

where the factor $P_{X|T}(x \mid t)$ is positive only if $x \in S$, since S contains the support of $P_{X|T}(\cdot \mid t)$, and $W(y \mid x)$ is positive only if $y \in \mathcal{Y}_x$, by (101). Hence, $P_{XY|T}(x, y \mid t) > 0$ if and only if $x \in S$ and

$y \in \mathcal{Y}_x$. We now show that, for every $y \in \mathcal{Y}$, there is at most one $x \in S$ with $y \in \mathcal{Y}_x$, which implies that, given $T = t$, the value of Y determines the value of X. We distinguish two cases. If $y = (x, i)$ for some $x \in A(\bar{G})$ and $1 \le i \le M - d(x)$, then y belongs only to $\mathcal{Y}_x$, and consequently x is the only oriented edge with $y \in \mathcal{Y}_x$. If $y = y_{\{x,x'\}}$ for some $\{x, x'\} \in \mathcal{I}(\bar{G})$, then y belongs only to $\mathcal{Y}_x$ and $\mathcal{Y}_{x'}$. Since x and x' are incompatible, they cannot both belong to the compatible set S. Thus, at most one of them belongs to S. In both cases, there is at most one $x \in S$ with $y \in \mathcal{Y}_x$. Hence, given $T = t$ and $Y = y$, the random variable X can take only one value, and therefore $H(X \mid Y, T = t) = 0$. Thus, by (111), we conclude that $I(Y;T) = I(X;T) - h_0$. □

C. The reduction

We now show that the yes- and no-instances of CLIQUE can be distinguished by comparing $\mathrm{PF}(P_{XY}, r)$ with the lower bound of Lemma 2. Indeed, by Lemma 2, every encoder T of X with $I(X;T) \ge r$ satisfies $I(Y;T) \ge I(X;T) - h_0 \ge r - h_0$, so that $\mathrm{PF}(P_{XY}, r) \ge r - h_0$. We show that, for a rate r slightly smaller than $R_0 = \log N$, this lower bound is not attained if G does not contain a k-clique, while it is attained if G contains a k-clique.

1) No-instances: G does not contain a k-clique: Assume that G has no k-clique. Then $\omega(\bar{G}) = m$ by property 3 of Lemma 4, and hence, by Lemma 21, $\Phi(Q) \le \log s = \log(m(m-1))$ for every probability distribution Q on $A(\bar{G})$. Using this, the next lemma shows that no encoder T of X with $I(X;T) < R_0$ attains the lower bound of Lemma 2.

Lemma 23. *Assume that $\omega(G) < k$. Then every encoder T of X satisfies*

$$I(Y;T) \ge I(X;T) - h_0 + \frac{R_0 - I(X;T)}{M}. \tag{112}$$

Proof. The idea of the proof is to bound the entropy of Y given each value of T, using Lemma 21, and then to average over T and express the result in terms of mutual information.

We first bound the output entropy of the channel W defined in (101). Since $\omega(G) < k$, we have that $\omega(G) \le m$, and by property 3 of Lemma 4 it follows that $\omega(\bar{G}) = m$. Hence, $\omega(\bar{G})(\omega(\bar{G}) - 1) = m(m-1) = s$, and Lemma 21 gives $\Phi(Q) \le \log s$ for every probability distribution Q on $A(\bar{G})$. Substituting this bound into (104) of Lemma 22, we obtain that

$$H(WQ) \le \log M + \Big(1 - \frac{1}{M}\Big) H(Q) + \frac{\log s}{M} \tag{113}$$

for every probability distribution Q on $A(\bar{G})$.

We now apply (113) to the posterior distributions of an encoder. Let T be an encoder of X, and let t be such that $P_T(t) > 0$. By Lemma 1, we have that $P_{Y|T}(\cdot \mid t) = W P_{X|T}(\cdot \mid t)$. Hence, applying (113) with $Q = P_{X|T}(\cdot \mid t)$, we get that

$$H\big(P_{Y|T}(\cdot \mid t)\big) \leq \log M + \Big(1 - \frac{1}{M}\Big) H\big(P_{X|T}(\cdot \mid t)\big) + \frac{\log s}{M}.$$

Multiplying by $P_T(t)$, summing over t, by the definition of conditional entropy, we obtain that

$$H(Y \mid T) \leq \log M + \Big(1 - \frac{1}{M}\Big) H(X \mid T) + \frac{\log s}{M}. \tag{114}$$

We now rewrite (114) in terms of mutual information. By property 1 of Lemma 22, we have that $H(Y \mid T) = H(Y) - I(Y;T) = 2\log M - h_0 - I(Y;T)$. Moreover, since X is uniform on $A(\bar{G})$, we have that $H(X \mid T) = H(X) - I(X;T) = \log M - I(X;T)$. Substituting these two expressions into the inequality (114), we get that

$$2\log M - h_0 - I(Y;T) \leq \log M + \Big(1 - \frac{1}{M}\Big)\big(\log M - I(X;T)\big) + \frac{\log s}{M}.$$

Solving for $I(Y;T)$ and expanding the product, we obtain that

$$\begin{aligned} I(Y;T) &\geq \log M - h_0 - \Big(1 - \frac{1}{M}\Big)\log M + \Big(1 - \frac{1}{M}\Big) I(X;T) - \frac{\log s}{M} \\ &= I(X;T) - h_0 + \frac{\log M - \log s - I(X;T)}{M}. \end{aligned} \tag{115}$$

Since $\log M - \log s = \log(M/s) = \log N = R_0$ by (99), one has that the inequality (115) proves (112), that is, the lemma. □

Lemma 23 states that, if G has no k-clique, every encoder T exceeds the lower bound (16) of Lemma 2 by at least $\big(R_0 - I(X;T)\big)/M$, which is positive whenever $I(X;T) < R_0$. Hence, without a k-clique, no encoder with rate smaller than R_0 attains equality in (16). On the other hand, the rate R_0 itself is attained with equality by the encoder T_0. Indeed, by Lemma 5, $I(X;T_0) = R_0$, and the posteriors of T_0 are the distributions U_{C_e}, which are supported on the sets $A(C_e)$. Since each C_e is a clique of $\bar{G}$, these sets are compatible by Lemma 20, and, by property 3 of Lemma 22, it holds that

$$I(X;T_0) = R_0 \qquad \text{and} \qquad I(Y;T_0) = R_0 - h_0. \tag{116}$$

2) Yes-instances: G contains a k-clique: We now assume that G contains a k-clique K, and we show that the encoder T_1 of Section III-D, with a suitable choice of p, attains equality in (16) at a rate strictly smaller than R_0. To this end, we choose

$$p = \frac{1}{2(m+1)}, \qquad \text{so that} \qquad \varepsilon = 2\mu p = \frac{1}{s(m+1)} \quad \text{and} \quad \lambda = \frac{B\varepsilon}{N} = \frac{m}{2M}, \tag{117}$$

where the last equality follows from $B = \frac{m(m+1)}{2}$ and $M = Ns$.

The next lemma shows that, with this choice of p, the rate of the encoder T_1 is smaller than R_0 by at least $\frac{3}{4M}$.

Lemma 24. *Assume that $\omega(G) \geq k$, and let T_1 be the encoder defined in* (32) *with p as in* (117)*. Then, T_1 satisfies*

$$I(Y;T_1) = I(X;T_1) - h_0 \qquad \textit{and} \qquad I(X;T_1) < R_0 - \frac{3}{4M}. \tag{118}$$

Proof. We first prove that the encoder T_1 attains equality in (16), that is, $I(Y;T_1) = I(X;T_1) - h_0$. By Lemma 6, the posteriors of T_1 are U_K, which is supported on $A(K)$, and the distributions V_e and U_{C_e}, which are supported on the sets $A(C_e)$ by (34). Since K and the C_e are cliques of $\bar{G}$, all these sets are compatible by Lemma 20. Hence, since the posteriors of T_1 are supported on compatible sets, the equality $I(Y;T_1) = I(X;T_1) - h_0$ follows from property 3 of Lemma 22.

We now bound the rate of T_1. By Lemma 8 and $R_0 = \log N$, one has that

$$I(X;T_1) = R_0 - \eta, \qquad \text{where} \qquad \eta = \lambda \log \frac{m+1}{m-1} - \frac{B(1-\varepsilon)}{N} D_V \tag{119}$$

is the increase in conditional entropy obtained by replacing T_0 with T_1. Hence, to prove the rightmost inequality in (118) it suffices to show that $\eta > \frac{3}{4M}$. To this end, we bound the two terms of η in (119) separately.

We start with the first term. Using $\ln(1+z) \geq \frac{z}{1+z}$ with $z = \frac{2}{m-1}$, for which $1 + z = \frac{m+1}{m-1}$ and $\frac{z}{1+z} = \frac{2}{m+1}$, we get that $\ln \frac{m+1}{m-1} \geq \frac{2}{m+1}$. Since $\log z = \frac{\ln z}{\ln 2}$, this gives

$$\lambda \log \frac{m+1}{m-1} \geq \frac{2\lambda}{(m+1)\ln 2}. \tag{120}$$

We now bound the second term of η in (119). By (38) in Lemma 7, rearranging the factors, we have that

$$\frac{B(1-\varepsilon)}{N} D_V \leq \frac{B(1-\varepsilon)}{N} \cdot \frac{\varepsilon^2}{2\mu(1-\varepsilon)^2 \ln 2} = \frac{B\varepsilon}{N} \cdot \frac{\varepsilon}{\mu} \cdot \frac{1}{2(1-\varepsilon)\ln 2}. \tag{121}$$

By the choice of p in (117), we have that $\frac{B\varepsilon}{N} = \lambda$ and $\frac{\varepsilon}{\mu} = \varepsilon s = \frac{1}{m+1}$. Moreover, since $m \geq 3$, we have that $s(m+1) \geq 24$, so that $\varepsilon = \frac{1}{s(m+1)} \leq \frac{1}{24}$ and $\frac{1}{1-\varepsilon} \leq \frac{24}{23}$. Substituting these facts into (121), we obtain that

$$\frac{B(1-\varepsilon)}{N} D_V \leq \frac{\lambda}{2(m+1)(1-\varepsilon)\ln 2} \leq \frac{12}{23} \cdot \frac{\lambda}{(m+1)\ln 2}. \tag{122}$$

We can now combine the two bounds. Substituting (120) and (122) into the expression of η in (119), we get that

$$\eta \geq \Big(2 - \frac{12}{23}\Big) \frac{\lambda}{(m+1)\ln 2} = \frac{34}{23} \cdot \frac{\lambda}{(m+1)\ln 2}. \tag{123}$$

Since $\lambda = \frac{m}{2M}$ by (117) and $\frac{m}{m+1} \geq \frac{3}{4}$ since $m \geq 3$, we have that $\frac{\lambda}{m+1} = \frac{m}{2(m+1)M} \geq \frac{3}{8M}$. Hence, using $\ln 2 < \frac{7}{10}$, it follows from (123) that

$$\eta > \frac{34}{23} \cdot \frac{3}{8M} \cdot \frac{10}{7} = \frac{255}{322\,M} > \frac{3}{4M},$$

where the last inequality follows from $4 \cdot 255 = 1020 > 966 = 3 \cdot 322$. Hence, by (119), $I(X;T_1) = R_0 - \eta < R_0 - \frac{3}{4M}$, which concludes the proof. □

We can now define the instance of PF-THRESHOLD associated with (G, k). We choose the rate r slightly smaller than R_0, namely such that

$$R_0 - \frac{3}{4M} \leq r \leq R_0 - \frac{5}{8M}. \tag{124}$$

With this choice, the two cases of CLIQUE lead to different values of $\mathrm{PF}(P_{XY}, r)$. Indeed, if G has no k-clique, then, by Lemma 23 and the rightmost inequality in (124), every encoder T of X with $I(X;T) \geq r$ satisfies $I(Y;T) \geq r - h_0 + \frac{R_0 - r}{M} \geq r - h_0 + \frac{5}{8M^2}$. If, instead, G contains a k-clique, then, by (116) and Lemma 24, the encoders T_0 and T_1 attain the lower bound (16) of Lemma 2 at rates R_0 and less than $R_0 - \frac{3}{4M}$, respectively. Since r lies between these two rates by (124), we show in the proof of Theorem 4 that, by combining T_0 and T_1, the lower bound is also attained at rate r, so that $\mathrm{PF}(P_{XY}, r) = r - h_0$. Since the two cases are separated by a gap of $\frac{5}{8M^2}$, we place the threshold in the middle of this gap, that is,

$$\vartheta = r - h_0 + \frac{5}{16M^2}. \tag{125}$$

Finally, since $R_0 = \log N$ is in general irrational, to obtain a rational r satisfying (124), we take a rational number $\tilde{R}_0$ with $|\tilde{R}_0 - R_0| \leq \frac{1}{16M}$, and we set

$$r = \tilde{R}_0 - \frac{11}{16M}. \tag{126}$$

The instance of PF-THRESHOLD associated with (G, k) is the triple (P_{XY}, r, ϑ), where P_{XY} is defined in (102), r in (126), and ϑ in (125).

D. Proof of Theorem 4

Proof. We start by showing that PF-THRESHOLD is NP-hard. To this end, we show that the construction of the instance (P_{XY}, r, ϑ) described above is a polynomial-time reduction from CLIQUE to PF-THRESHOLD. Given an instance (G, k) of CLIQUE satisfying (25), the reduction constructs the extended graph $\bar{G}$, the joint distribution P_{XY} defined in (102), and the rational numbers r and ϑ defined in (126) and (125), and outputs the instance (P_{XY}, r, ϑ) of PF-THRESHOLD.

We first show that the reduction runs in time polynomial in $n = |V(G)|$. The reduction has to compute $\bar{G}$, P_{XY}, r and ϑ. By property 4 of Lemma 4, $\bar{G}$ can be constructed in polynomial time, and $M = |A(\bar{G})| \leq n^4$. We now consider the joint distribution P_{XY} and show that it can be computed in polynomial time. To this end, by (100) and (102), to compute P_{XY} we need the set $\mathcal{I}(\bar{G})$ of incompatible pairs, which determines the alphabet $\mathcal{Y}$, and we need to bound $|\mathcal{Y}|$. We first compute $\mathcal{I}(\bar{G})$. There are at most M^2 pairs of oriented edges, and, for each pair $\{x, x'\}$, deciding whether x and x' are compatible requires checking whether the set of their endpoints, which has at most four vertices, is a clique of $\bar{G}$, that is, checking at most $\binom{4}{2} = 6$ adjacencies. Hence, the set $\mathcal{I}(\bar{G})$ and the numbers $d(x)$ can be computed in polynomial time. Moreover, by (100), $\mathcal{Y}$ contains one symbol for each pair in $\mathcal{I}(\bar{G})$ and $M - d(x)$ symbols (x, i) for each $x \in A(\bar{G})$. Since each pair of $\mathcal{I}(\bar{G})$ contains exactly two oriented edges, we have that $\sum_{x \in A(\bar{G})} d(x) = 2|\mathcal{I}(\bar{G})|$, and hence

$$|\mathcal{Y}| = |\mathcal{I}(\bar{G})| + \sum_{x \in A(\bar{G})} \big(M - d(x)\big) = |\mathcal{I}(\bar{G})| + M^2 - 2|\mathcal{I}(\bar{G})| = M^2 - |\mathcal{I}(\bar{G})| \leq M^2. \tag{127}$$

Therefore, P_{XY} is a table with at most M^3 entries, each equal to $1/M^2$ or to 0 by (102), which are rational numbers with $O(\log n)$ bits, and it can be computed in polynomial time. We now consider r and ϑ. By property 1 of Lemma 22, $h_0 = 2|\mathcal{I}(\bar{G})|/M^2$ is a rational number that can be computed exactly, and, by Lemma 25 in Appendix A, a rational $\tilde{R}_0$ with $|\tilde{R}_0 - R_0| \leq \frac{1}{16M}$ can be computed in polynomial time. Hence, by (126) and (125), r and ϑ are rational numbers computable in polynomial time. Therefore, the whole reduction runs in polynomial time.

It remains to check that r and ϑ are non-negative, as required in the definition of PF-THRESHOLD. By assumption (25) on the graph G, we have that $k \geq 4$ and $N \geq B = \binom{k}{2} \geq \binom{4}{2} = 6$, so that $R_0 = \log N > 2$. Hence, by the leftmost inequality in (124), we get that $r \geq R_0 - \frac{3}{4M} > 1$. Moreover, every pair in $\mathcal{I}(\bar{G})$ consists of two distinct oriented edges, so that $|\mathcal{I}(\bar{G})| \leq \binom{M}{2} < \frac{M^2}{2}$. Hence, by property 1 of Lemma 22, it follows that $h_0 = \frac{2|\mathcal{I}(\bar{G})|}{M^2} < 1$, and therefore, by (125), $\vartheta > r - h_0 > 1 - 1 = 0$.

We now show that the reduction is correct, that is, (P_{XY}, r, ϑ) is a yes-instance of PF-THRESHOLD if and only if (G, k) is a yes-instance of CLIQUE. We recall that (P_{XY}, r, ϑ) is a yes-instance of PF-THRESHOLD if there exists an encoder T of X such that $I(X;T) \geq r$ and $I(Y;T) \leq \vartheta$, while (G, k) is a yes-instance of CLIQUE if $\omega(G) \geq k$, that is, if G contains a clique of size k.

Assume first that (G, k) is a yes-instance of CLIQUE, that is, $\omega(G) \geq k$. By (116) and Lemma 24, the encoders T_0 and T_1 attain the lower bound in (16), at rates R_0 and less than $R_0 - \frac{3}{4M}$, respectively. We now show that, by combining the two encoders, we obtain an encoder that attains the lower bound in (16) at rate exactly r. For $\theta \in [0, 1]$, let T_θ be the encoder of Lemma 3 with $T_a = T_0$ and $T_b = T_1$, which applies T_0 with probability θ and T_1 with probability $1 - \theta$, and reveals which of the two has

been applied. By (18), the posteriors of T_θ are posteriors of T_0 or of T_1, and are therefore supported on compatible sets. Hence, by property 3 of Lemma 22, it follows that the encoder T_θ satisfies

$$I(Y;T_\theta) = I(X;T_\theta) - h_0 \qquad \text{for every } \theta \in [0,1]. \tag{128}$$

We now choose θ so that $I(X;T_\theta) = r$. By the expression of the rate of T_θ in (19), $I(X;T_\theta) = \theta\, I(X;T_0) + (1-\theta)\, I(X;T_1)$ is a continuous function of θ. Moreover, by Lemma 24, by the choice (124) of r, and by the rate of T_0 in (116), we have that

$$I(X;T_1) < R_0 - \frac{3}{4M} \le r \le R_0 - \frac{5}{8M} < R_0 = I(X;T_0). \tag{129}$$

Hence, there exists $\theta \in (0,1)$ such that $I(X;T_\theta) = r$, and, for this θ, (128) and the definition (125) of ϑ give

$$I(Y;T_\theta) = r - h_0 = \vartheta - \frac{5}{16M^2} < \vartheta. \tag{130}$$

Therefore, the encoder T_θ satisfies both $I(X;T_\theta) \ge r$ and $I(Y;T_\theta) \le \vartheta$, and (P_{XY}, r, ϑ) is a yes-instance of PF-THRESHOLD. Moreover, by (130), $\mathrm{PF}(P_{XY}, r) \le r - h_0$, while, by Lemma 2, $\mathrm{PF}(P_{XY}, r) \ge r - h_0$. Hence, it follows that

$$\mathrm{PF}(P_{XY}, r) = r - h_0 = \vartheta - \frac{5}{16M^2}. \tag{131}$$

Assume now that (G,k) is a no-instance of CLIQUE, that is, $\omega(G) < k$. Let T be an encoder of X with $I(X;T) \ge r$. We show that $I(Y;T) > \vartheta$. By Lemma 23, we have that

$$\begin{aligned} I(Y;T) &\ge I(X;T) - h_0 + \frac{R_0 - I(X;T)}{M} \\ &= \Big(1 - \frac{1}{M}\Big) I(X;T) + \frac{R_0}{M} - h_0. \end{aligned} \tag{132}$$

The right-hand side of (132) is increasing in $I(X;T)$. Hence, since $I(X;T) \ge r$, we can replace $I(X;T)$ with r, and, using the rightmost inequality in (124) and the definition of ϑ in (125), we get from (132) that

$$I(Y;T) \ge r - h_0 + \frac{R_0 - r}{M} \ge r - h_0 + \frac{5}{8M^2} = \vartheta + \frac{5}{16M^2} > \vartheta. \tag{133}$$

Since (133) holds for any arbitrary encoder T with $I(X;T) \ge r$, it follows that no encoder satisfies both $I(X;T) \ge r$ and $I(Y;T) \le \vartheta$, and hence (P_{XY}, r, ϑ) is a no-instance of PF-THRESHOLD. Moreover, taking the infimum over all encoders with $I(X;T) \ge r$ in (133) yields

$$\mathrm{PF}(P_{XY}, r) \ge \vartheta + \frac{5}{16M^2}. \tag{134}$$

This proves that the reduction is correct, and hence that PF-THRESHOLD is NP-hard.

It remains to prove the inapproximability result. By (131) and (134), $\mathrm{PF}(P_{XY}, r)$ lies below $\vartheta - \frac{5}{16M^2}$ if $\omega(G) \geq k$, and above $\vartheta + \frac{5}{16M^2}$ otherwise. Suppose that a polynomial-time algorithm outputs, given P_{XY} and r, a value v with $|v - \mathrm{PF}(P_{XY}, r)| \leq \frac{1}{4|\mathcal{X}|^2} = \frac{4}{16M^2}$. If $\omega(G) \geq k$, then, by (131), $v \leq \mathrm{PF}(P_{XY}, r) + \frac{4}{16M^2} = \vartheta - \frac{1}{16M^2} < \vartheta$, while, if $\omega(G) < k$, then, by (134), $v \geq \mathrm{PF}(P_{XY}, r) - \frac{4}{16M^2} \geq \vartheta + \frac{1}{16M^2} > \vartheta$. Thus, comparing v with ϑ would decide CLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$. □

E. The Lagrangian formulation

We now show that PF-LAGRANGIAN$_\beta$ is NP-hard, and hard to approximate, for every fixed rational $\beta \in (0, 1)$. The proof combines the constructions in the proofs of Theorem 4 and Proposition 2. Specifically, we first reduce CLIQUE to PF-LAGRANGIAN$_{1/2}$, starting from the same instance of Theorem 4, and then we reduce PF-LAGRANGIAN$_{1/2}$ to PF-LAGRANGIAN$_\beta$, as in Proposition 2.

Proposition 3. *Let $\beta \in \mathbb{Q} \cap (0, 1)$ be fixed, and let $\kappa_\beta = 2\min\{\beta, 1 - \beta\}$. Then* PF-LAGRANGIAN$_\beta$ *is NP-hard. Moreover, unless* $\mathrm{P} = \mathrm{NP}$*, there is no polynomial-time algorithm that, given a rational joint distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$, outputs a value v such that*

$$\big|v - \Lambda_\beta(P_{XY})\big| \leq \frac{\kappa_\beta}{8|\mathcal{X}|^2}.$$

Proof. We first prove that PF-LAGRANGIAN$_\beta$ is NP-hard for every $\beta \in \mathbb{Q} \cap (0, 1)$. To this end, we give a polynomial-time reduction from CLIQUE to PF-LAGRANGIAN$_{1/2}$, and then a polynomial-time reduction from PF-LAGRANGIAN$_{1/2}$ to PF-LAGRANGIAN$_\beta$.

We start with the reduction from CLIQUE to PF-LAGRANGIAN$_{1/2}$. Let (G, k) be an instance of CLIQUE satisfying (25), let P_{XY} be the joint distribution defined in (102), and let

$$\beta_0 = 1 - \frac{1}{M}, \qquad \ell_0 = \frac{R_0}{M} - h_0, \qquad a = \frac{1}{2\beta_0} = \frac{M}{2(M-1)}. \tag{135}$$

Note that $\frac{1}{2} < a \leq 1$, since $M \geq 2$. We define the joint distribution $P_{XY'}$ on $\mathcal{X} \times (\mathcal{Y} \cup \{\perp\})$, where $\perp$ is a new symbol, by

$$P_{XY'}(x, y) = a\, P_{XY}(x, y), \quad \forall y \in \mathcal{Y}, \qquad \text{and} \quad P_{XY'}(x, \perp) = (1 - a)\, P_X(x), \tag{136}$$

that is, Y' is equal to Y with probability a, and to $\perp$ otherwise, independently of (X, Y). Moreover, we let τ be a rational number such that

$$\left|\tau - \left(a\ell_0 - \frac{6}{32M^2}\right)\right| \leq \frac{1}{32M^2}. \tag{137}$$

The reduction maps (G, k) to the instance $(P_{XY'}, \tau)$ of PF-LAGRANGIAN$_{1/2}$. We first show that the instance $(P_{XY'}, \tau)$ can be computed in polynomial time. Indeed, the joint distribution P_{XY} can be

computed in polynomial time, as shown in the proof of Theorem 4, and, since a is rational, it follows by (136) that $P_{XY'}$ can also be computed in polynomial time. Moreover, $a\ell_0 = \frac{a}{M}\log N - ah_0$, where a and h_0 are rational. Hence, by Lemma 25, τ can be computed in polynomial time. Therefore, the whole instance $(P_{XY'}, \tau)$ can be computed in polynomial time.

We now show that the reduction is correct. By Lemmas 23 and 24, the Lagrangian of P_{XY} with multiplier $\beta_0 = 1 - \frac{1}{M}$ already separates the two cases of CLIQUE. However, β_0 depends on the instance, while the multiplier of PF-LAGRANGIAN$_{1/2}$ is fixed. For this reason, we relate the Lagrangian of $P_{XY'}$ with multiplier $\frac{1}{2}$ to the Lagrangian of P_{XY} with multiplier β_0. Since X is unchanged, the encoders of X for P_{XY} and for $P_{XY'}$ coincide. Let T be an encoder of X, and let E be the indicator of the event $\{Y' \neq \perp\}$, so that $E = 1$ with probability a. Since E is a function of Y', by the chain rule, we have that

$$I(Y';T) = I(Y', E; T) = I(E;T) + I(Y';T \mid E). \tag{138}$$

By construction, E is independent of (X, Y), and T depends only on X. Thus, E is independent of (X, Y, T), and it follows that $I(E;T) = 0$, and that, for each value of E, the distribution of (Y, T) is unchanged. Moreover, if $E = 1$, then $Y' = Y$, so that $I(Y';T \mid E = 1) = I(Y;T)$, while, if $E = 0$, then $Y' = \perp$ is constant, so that $I(Y';T \mid E = 0) = 0$. Substituting these facts into (138), we obtain that

$$I(Y';T) = a\, I(Y;T). \tag{139}$$

Finally, since $a\beta_0 = \frac{1}{2}$ by (135), subtracting $\frac{1}{2}I(X;T)$ from both sides of (139) gives

$$I(Y';T) - \tfrac{1}{2}I(X;T) = a\, I(Y;T) - a\beta_0\, I(X;T) = a\big[I(Y;T) - \beta_0\, I(X;T)\big]. \tag{140}$$

We now use the equivalence (140) to show that G contains a k-clique if and only if there exists an encoder T of X such that

$$I(Y';T) - \tfrac{1}{2}I(X;T) \leq \tau. \tag{141}$$

Assume first that (G, k) is a no-instance of CLIQUE, that is, $\omega(G) < k$. By Lemma 23, and since $\beta_0 = 1 - \frac{1}{M}$ and $\ell_0 = \frac{R_0}{M} - h_0$ by (135), every encoder T of X satisfies

$$I(Y;T) \geq I(X;T) - h_0 + \frac{R_0 - I(X;T)}{M} = \beta_0\, I(X;T) + \ell_0,$$

that is, $I(Y;T) - \beta_0 I(X;T) \geq \ell_0$. Hence, by the equivalence (140) and the definition of τ in (137), every encoder T of X satisfies

$$I(Y';T) - \tfrac{1}{2}I(X;T) \geq a\ell_0 \geq \tau + \frac{5}{32M^2}. \tag{142}$$

Thus, no encoder satisfies (141), and $(P_{XY'}, \tau)$ is a no-instance of PF-LAGRANGIAN$_{1/2}$.

Assume now that (G, k) is a yes-instance of CLIQUE, that is, $\omega(G) \geq k$. By Lemma 24, the encoder T_1 satisfies $I(Y;T_1) = I(X;T_1) - h_0$ and $I(X;T_1) < R_0 - \frac{3}{4M}$. Hence, substituting $I(Y;T_1) = I(X;T_1) - h_0$ and $\beta_0 = 1 - \frac{1}{M}$, we get that

$$I(Y;T_1) - \beta_0\, I(X;T_1) = \frac{I(X;T_1)}{M} - h_0 < \frac{R_0}{M} - \frac{3}{4M^2} - h_0 = \ell_0 - \frac{3}{4M^2}, \tag{143}$$

where the inequality follows from $I(X;T_1) < R_0 - \frac{3}{4M}$, and the last equality from the definition of ℓ_0 in (135). Multiplying both sides of (143) by $a > 0$, and using the identity (140) and $a > \frac{1}{2}$, we get that

$$I(Y';T_1) - \tfrac{1}{2} I(X;T_1) < a\ell_0 - \frac{3a}{4M^2} < a\ell_0 - \frac{3}{8M^2} = a\ell_0 - \frac{12}{32M^2}. \tag{144}$$

Moreover, by the definition (137) of τ, we have that $a\ell_0 \leq \tau + \frac{6}{32M^2} + \frac{1}{32M^2} = \tau + \frac{7}{32M^2}$. Substituting this into (144), we obtain that

$$I(Y';T_1) - \tfrac{1}{2} I(X;T_1) < \tau + \frac{7}{32M^2} - \frac{12}{32M^2} = \tau - \frac{5}{32M^2}, \tag{145}$$

Thus, the encoder T_1 satisfies (141), and $(P_{XY'}, \tau)$ is a yes-instance of PF-LAGRANGIAN$_{1/2}$. This proves that the reduction is correct, and hence that PF-LAGRANGIAN$_{1/2}$ is NP-hard.

We now prove that $\Lambda_{1/2}$ is hard to approximate. By (142) and (145), we have that $\Lambda_{1/2}(P_{XY'}) \geq \tau + \frac{5}{32M^2}$ if $\omega(G) < k$, and that $\Lambda_{1/2}(P_{XY'}) < \tau - \frac{5}{32M^2}$ if $\omega(G) \geq k$. Suppose that a polynomial-time algorithm outputs, given $P_{XY'}$, a value v with $|v - \Lambda_{1/2}(P_{XY'})| \leq \frac{1}{8|\mathcal{X}|^2}$. Since $P_{XY'}$ has the same input alphabet $\mathcal{X}$ as P_{XY}, we have that $|\mathcal{X}| = M$, and hence $\frac{1}{8|\mathcal{X}|^2} = \frac{4}{32M^2}$. Therefore, if $\omega(G) < k$, then $v \geq \Lambda_{1/2}(P_{XY'}) - \frac{4}{32M^2} \geq \tau + \frac{1}{32M^2} > \tau$, while, if $\omega(G) \geq k$, then $v \leq \Lambda_{1/2}(P_{XY'}) + \frac{4}{32M^2} < \tau - \frac{1}{32M^2} < \tau$. Thus, running the reduction, then the algorithm, and comparing v with τ would decide CLIQUE in polynomial time, which is impossible unless $\mathrm{P} = \mathrm{NP}$.

It remains to give the reduction from PF-LAGRANGIAN$_{1/2}$ to PF-LAGRANGIAN$_\beta$. Let $(P_{XY'}, \tau)$ be an instance of PF-LAGRANGIAN$_{1/2}$, and let $\tilde{P}_{XZ}$ be the distribution obtained from $P_{XY'}$ by the construction (80) in the proof of Proposition 2. The reduction maps $(P_{XY'}, \tau)$ to the instance $(\tilde{P}_{XZ}, \kappa_\beta \tau)$ of PF-LAGRANGIAN$_\beta$, which, as shown in that proof, can be computed in polynomial time. Moreover, by (82), every encoder T of X satisfies

$$I(Z;T) - \beta I(X;T) = \kappa_\beta \Big(I(Y';T) - \tfrac{1}{2} I(X;T) \Big). \tag{146}$$

Since $\kappa_\beta > 0$, by (146) an encoder satisfies $I(Y';T) - \frac{1}{2} I(X;T) \leq \tau$ if and only if it satisfies $I(Z;T) - \beta I(X;T) \leq \kappa_\beta \tau$. Hence, the reduction is correct, and PF-LAGRANGIAN$_\beta$ is NP-hard. Finally, taking the infimum over all encoders in (146), we get that $\Lambda_\beta(\tilde{P}_{XZ}) = \kappa_\beta \Lambda_{1/2}(P_{XY'})$. Since $\tilde{P}_{XZ}$ has the same input alphabet $\mathcal{X}$ as $P_{XY'}$, an approximation of Λ_β within $\frac{\kappa_\beta}{8|\mathcal{X}|^2}$, divided by κ_β, would give an approximation of $\Lambda_{1/2}$ within $\frac{1}{8|\mathcal{X}|^2}$, which, as shown above, is impossible unless $\mathrm{P} = \mathrm{NP}$. $\square$

We observe that the argument of Remark 1 for the information bottleneck applies also to the privacy funnel. Therefore, the result of Proposition 3 extends to every fixed polynomial-time computable real $\beta \in (0,1)$. Finally, note that, for $\beta \geq 1$, the problem is trivial since the optimum is attained by $T = X$. Indeed, by Lemma 2, every encoder T of X satisfies

$$I(Y;T) - \beta I(X;T) \geq (1-\beta)\, I(X;T) - H(X \mid Y) \geq (1-\beta)\, H(X) - H(X \mid Y) = I(X;Y) - \beta H(X),$$

where the second inequality follows from $\beta \geq 1$ and $I(X;T) \leq H(X)$, and equality holds for $T = X$. Hence, $\Lambda_\beta(P_{XY}) = I(X;Y) - \beta H(X)$ for every $\beta \geq 1$.

VI. Conclusions

We have established the NP-hardness of the threshold decision problems for both the constrained and the Lagrangian formulations of the information bottleneck. Theorems 1 and 2 also rule out polynomial-time approximation within additive errors of order $|\mathcal{X}|^{-3/2}$, unless P = NP, and, by padding the instances of CLIQUE with isolated edges, the exponent $3/2$ can be lowered to $1 + \alpha$, for every fixed $\alpha > 0$ (Theorem 3).

By adapting our hardness framework for the information bottleneck, we also showed that it is NP-hard both to compute the optimal privacy funnel [20] exactly and to approximate it within an additive error of order $|\mathcal{X}|^{-2}$ (Theorem 4). The same holds for its Lagrangian formulation, for every fixed polynomial-time computable multiplier $\beta \in (0,1)$ (Proposition 3), while, for $\beta \geq 1$, $\Lambda_\beta(P_{XY}) = I(X;Y) - \beta H(X)$.

An open question is whether approximating the information bottleneck and the privacy funnel within a larger additive error remains NP-hard. Thus, whether hardness holds within errors of order $|\mathcal{X}|^{-1}$, or even within a constant additive error, remains open.

Given the many applications of the information bottleneck framework, perhaps the most important open problem is the design of efficient approximation algorithms with provable guarantees. Another direction is to identify structural conditions on the joint distribution that permit efficient exact computation or approximation.

AI statement

The authors used OpenAI's ChatGPT Astra and Anthropic's Claude Opus 5.5 throughout the preparation of this manuscript for language editing, expository suggestions, mathematical consistency checks, and literature and reference checks. ChatGPT drew the authors' attention to the relevance of the Motzkin–Straus lemma [21] and the entropic form of Turán's theorem [6] to this work. The overall problem formulation and the overall structure of the paper are the authors' own. The authors reviewed and edited the AI-assisted material and take full responsibility for the content of the paper.

REFERENCES

[1] Y. S. Abu-Mostafa, Essential average mutual information, in: *Open Problems in Communication and Computation*, T. M. Cover and B. Gopinath (Eds.), Springer–Verlag, pp. 75, 1987.

[2] S. Agmon, The information bottleneck's ordinary differential equation: First-order root tracking for the information bottleneck, *Entropy*, vol. **25**, no. 10, 2023.

[3] S. Asoodeh and F. P. Calmon, Bottleneck problems: An information and estimation-theoretic view, *Entropy*, vol. **22**, no. 11, 2020.

[4] E. Benger, S. Asoodeh and J. Chen, The cardinality bound on the information bottleneck representations is tight, in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Taipei, Taiwan, pp. 1478–1483, 2023.

[5] R. P. Brent, Fast multiple-precision evaluation of elementary functions, *Journal of the ACM*, vol. **23**, pp. 242–251, 1976.

[6] T.-W. Chao and H.-H. H. Yu, When entropy meets Turán: new proofs and hypergraph Turán results, *J. London Math. Soc.*, vol. **113**, no. 3, 2026.

[7] N. Chayat and S. Shamai, Extension of an entropy property for binary input memoryless symmetric channels, *IEEE Trans. Inf. Theory*, vol. **35**, no. 5, pp. 1077–1079, 1989.

[8] G. Chechik, A. Globerson, N. Tishby and Y. Weiss, Information bottleneck for Gaussian variables, *Journal of Machine Learning Research*, vol. **6**, pp. 165–188, 2005.

[9] M. Dikshtein, N. Weinberger and S. Shamai, On information bottleneck for Gaussian processes, in *Proc. IEEE Information Theory Workshop (ITW)*, Mumbai, India, pp. 546–551, 2022.

[10] K. Etessami, A. Stewart and M. Yannakakis, A note on the complexity of comparing succinctly represented integers, with an application to maximum probability parsing, *ACM Trans. Comput. Theory*, vol. **6**, no. 2, 2014.

[11] N. Friedman, O. Mosenzon, N. Slonim and N. Tishby, Multivariate information bottleneck, in *Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI)*, Seattle, WA, USA, pp. 152–161, 2001.

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

[13] Z. Goldfeld and Y. Polyanskiy, The information bottleneck problem and its applications in machine learning, *IEEE J. Sel. Areas Inf. Theory*, vol. **1**, no. 1, pp. 19–38, 2020.

[14] S. Hu, Z. Lou, X. Yan and Y. Ye, A survey on information bottleneck, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. **46**, no. 8, pp. 5325–5344, 2024.

[15] T.-H. Huang and A. El Gamal, A provably convergent information bottleneck solution via ADMM, in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Melbourne, Australia, pp. 43—48, 2021.

[16] R. M. Karp, Reducibility among combinatorial problems, in *Complexity of Computer Computations* (R. E. Miller and J. W. Thatcher, eds.), Plenum Press, New York, pp. 85–103, 1972.

[17] A. Kolchinsky, B. D. Tracey and S. Van Kuyk, Caveats for information bottleneck in deterministic scenarios, in *Proc. International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, 2019.

[18] J. Lewandowsky and G. Bauch, Eds., *Theory and Application of the Information Bottleneck Method*. Basel, Switzerland: MDPI, 2024.

[19] H. Liu and Y. Guo, Survey of information bottleneck frameworks in multimodal representation learning, *Journal of Frontiers of Computer Science and Technology*, vol. **20**, no. 9, pp. 2454-–2468, 2026.

[20] A. Makhdoumi, S. Salamatian, N. Fawaz and M. Médard, From the information bottleneck to the privacy funnel, in *Proc. IEEE Information Theory Workshop (ITW)*, Hobart, Australia, pp. 501–505, 2014.

[21] T. S. Motzkin and E. G. Straus, Maxima for graphs and a new proof of a theorem of Turán, *Canad. J. Math.*, vol. **17**, pp. 533–540, 1965.

[22] P. Piantanida and L. Rey Vega, Information bottleneck and representation learning, in: *Information-Theoretic Methods in Data Science*, M. R. D. Rodrigues and Y. C. Eldar (Eds.), Cambridge University Press, pp. 330—358, 2021.

[23] I. Sason and S. Verdú, f-divergence inequalities, *IEEE Trans. Inf. Theory*, vol. **62**, no. 11, pp. 5973–6006, 2016.

[24] R. Shwartz-Ziv and Y. LeCun, To compress or not to compress—self-supervised learning and information theory: A review, *Entropy*, vol. **26**, no. 3, 2024.

[25] D. J. Strouse and D. J. Schwab, The deterministic information bottleneck, *Neural Computation*, vol. **29**, no. 6, pp. 1611–1630, 2017.

[26] I. Sutskover, S. Shamai and J. Ziv, Extremes of information combining, *IEEE Trans. Inf. Theory*, vol. **51**, no. 4, pp. 1313–1325, 2005.

[27] N. Tishby, F. C. Pereira and W. Bialek, The information bottleneck method, in *Proc. 37th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, pp. 368–377, 1999.

[28] H. S. Witsenhausen and A. D. Wyner, A conditional entropy bound for a pair of discrete random variables, *IEEE Trans. Inf. Theory*, vol. **21**, pp. 493–501, 1975.

[29] A. D. Wyner, A theorem on the entropy of certain binary sequences and applications: Part II, *IEEE Trans. Inf. Theory*, vol. **19**, pp. 772–777, 1973.

[30] A. D. Wyner and J. Ziv, A theorem on the entropy of certain binary sequences and applications: Part I, *IEEE Trans. Inf. Theory*, vol. **19**, pp. 769–772, 1973.

[31] A. Zaidi, I. Estella-Aguerri and S. Shamai (Shitz), On the information bottleneck problems: Models, connections, applications and information theoretic views, *Entropy*, vol. **22**, no. 2, 2020.

APPENDIX A
APPROXIMATION OF LOGARITHMS

The reductions of Theorems 1, 2 and 4, and of Proposition 3, must compute rational approximations of finitely many logarithmic expressions. Exact comparison of such expressions is a delicate matter (e.g., see [10]). However, the reductions of Theorems 1, 2 and 4, and of Proposition 3 only need the following standard approximation result, which follows from classical algorithms for the multiple-precision evaluation of the logarithm (see [5]).

Lemma 25. *There is an algorithm that, given rational numbers $\zeta_0, \ldots, \zeta_\ell$, positive rational numbers $z_1, \ldots, z_\ell$, and a rational number $\gamma > 0$, outputs a rational number $\tilde{F}$ such that*

$$\left|\tilde{F} - \left(\zeta_0 + \sum_{i=1}^{\ell} \zeta_i \log z_i\right)\right| \leq \gamma,$$

in time polynomial in the total bit length of the input.