



HAL
open science

On the Additivity of Optimal Rates for Independent Zero-Error Source and Channel Problems

Nicolas Charpenay, Maël Le Treust, Aline Roumy

► **To cite this version:**

Nicolas Charpenay, Maël Le Treust, Aline Roumy. On the Additivity of Optimal Rates for Independent Zero-Error Source and Channel Problems. *IEEE Transactions on Information Theory*, 2026, pp.1-1. <10.1109/TIT.2025.3639292>. <hal-05401169>

HAL Id: hal-05401169

<https://inria.hal.science/hal-05401169v1>

Submitted on 5 Dec 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

On the Additivity of Optimal Rates for Independent Zero-Error Source and Channel Problems

Nicolas Charpenay, *Student member, IEEE*, Maël Le Treust, *Member, IEEE*, and Aline Roumy, *Member, IEEE*

Abstract—Zero-error coding encompasses a variety of source and channel problems where the probability of error must be exactly zero. This condition is stricter than that of the vanishing error regime, where the error probability goes to zero as the code blocklength goes to infinity. In general, zero-error coding is an open combinatorial question. We investigate two unsolved zero-error problems: the source coding problem with side information and the channel coding problem. We focus our attention on families of independent problems for which the probability distribution decomposes into a product of probability distributions. A crucial step is the additivity property of the optimal rate, which does not always hold in the zero-error regime, unlike in the vanishing error regime. When the additivity holds, the concatenation of optimal codes is optimal. We derive a condition under which the additivity of the complementary graph entropy \bar{H} for the AND product of graphs and for the disjoint union of graphs are equivalent. Then we establish the connection with a recent result obtained by Wigderson and Zuydam and by Schrijver, for the zero-error capacity C_0 . As a consequence, we provide new single-letter characterizations of \bar{H} and C_0 , for example when the graph is a product of perfect graphs, which is not perfect in general, and for the class of graphs obtained by the product of a perfect graph G with the pentagon graph C_5 . By building on Hamers result for C_0 , we also show that the additivity of \bar{H} does not hold for the product of the Schläfli graph with its complementary graph.

I. INTRODUCTION

Transmitting information without any errors has been a concern for Shannon since the beginning of his work. In his seminal paper [3], Shannon proposed a construction for zero-error source coding, a problem soon solved by Huffman in [4]. Shortly after establishing the channel capacity in [3], Shannon turned his attention to channel coding with zero-error in [5], instead of vanishing error. This subtle difference radically changes the nature of the problem, essentially combinatorial rather than probabilistic. The single-letter characterization of the zero-error capacity is a notoriously difficult open problem. For example, the zero-error capacity of the noisy-typewriter channel with 7 letters is unknown [6], some lower and upper

The work of Nicolas Charpenay was conducted during his PhD at IRISA UMR 6074 and Centre Inria de l'Université de Rennes, funded by CDSN ENS Paris-Saclay. The work of Maël Le Treust is funded by the French government under the France 2030 ANR program "PEPR Networks of the Future" (ref. ANR-22-PEFT-0010).

This work was presented in part at the IEEE International Symposium on Information Theory (ISIT) 2023 [1], and in part at the IEEE Information Theory Workshop (ITW) 2023 [2].

Nicolas Charpenay is with Univ Rennes, CNRS, IRMAR UMR 6625, F-35000 Rennes, France (e-mail: nicolas.charpenay@univ-rennes.fr).

Maël Le Treust is with Univ. Rennes, CNRS, Inria, IRISA UMR 6074, F-35000 Rennes, France (e-mail: mael.le-treust@cnrs.fr).

Aline Roumy is with Inria Center at Rennes University, France (e-mail: aline.roumy@inria.fr).

bounds are stated in [7], [8], [9], [10], see also [11, pp. 122]. In the zero-error regime, the exact values of the non-zero channel transition probabilities are irrelevant, only the support of the channel distribution matters. The zero-error property is translated into the *characteristic graph* that encompasses the problem data in its structure: the vertices are the channel inputs \mathcal{X} , and two symbols x and x' are adjacent if they are "confusable", i.e. if they can produce the same channel output y with positive probability. For sequences of symbols x^n , the characteristic graph is obtained by taking iteratively the AND product (\wedge), where two sequences x^n and x'^n are adjacent if for all $t \in \{1, \dots, n\}$ either $x_t = x'_t$ or the symbols x_t and x'_t are adjacent. In order to prevent any decoding error, a zero-error codebook must be composed of non-adjacent codewords. Thus, the size of the optimal codebook is given by the size of the maximal independent set, called the *independence number*. In other words, the zero-error capacity C_0 is the asymptotic limit of the independence number of iterated AND product of the characteristic graph. Determining this independence number is an open question that has attracted a lot of attention in Information Theory [12, Chap. 11] and in Combinatorics and Graph Theory, see [13, Chap. 27].

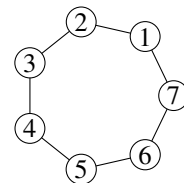


Fig. 1. The characteristic graph C_7 of the noisy-typewriter channel with 7 letters.

This problem inspired Berge's notion of perfect graphs [14, pp. 382], for which the zero-error capacity is given by the one-shot independence number [15, Theorem 4.18]. Graphs with odd cycles are also related to Berge's conjecture [16], later proved in [17] by Chudnovsky et al., namely "a graph G is perfect if and only if neither G nor its complementary graph \bar{G} , have odd cycles of length 5 or more." Since the zero-error capacity of the pentagon graph C_5 has been characterized by Lovász [18], as well as the zero-error capacity for perfect graphs, see [15, Theorem 4.18], the graph C_7 depicted in Fig. 1, is the minimal connected graph for which the zero-error capacity is an open problem.

In the source coding framework, Witsenhausen in [19] posed the question of the optimal compression rate when the decoder has side information. In this problem depicted in Fig. 2, the encoder shares information about a source X , exploiting

the side-information Y observed by the decoder but not by itself. In the vanishing error regime, Slepian and Wolf in [20] showed that the optimal rate is $H(X|Y)$. In the zero-error regime, no single-letter characterization is available. Similarly to the channel coding problem, the zero-error property is embedded into the same characteristic graph G , constructed with respect to the conditional distribution $P_{Y|X}$ of the “side-information channel”. The source distribution P_X assigns a probability value to each vertex of the characteristic graph G , this defines the probabilistic graph (G, P_X) . The main difference between zero-error source and zero-error channel coding problems is that the source distribution P_X is given, whereas in channel coding it is a degree of freedom allowing to maximize the communication rate. This is a property common to both zero-error and vanishing-error regimes. Witsenhausen showed in [19] that fixed-length source coding amounts to coloring the characteristic graph. In [21], Alon and Orlitsky studied variable-length coding and determined an asymptotic expression for the optimal rate based on the chromatic entropy. In [22], Koulgi et al. proved that the optimal rate coincides with the *complementary graph entropy* of the probabilistic graph $\bar{H}(G, P_X)$, introduced by Körner and Longo in [23] for the second step of the two-step source coding. These two expressions are asymptotic, as optimal source rates are determined by coloring an infinite product of graphs. Similarly to the zero-error capacity, single-letter characterizations are known only in a few cases, including the pentagon graph C_5 [22] and perfect graphs [24]. Another concept, called the *Körner graph entropy* [25], provides a single-letter expression for the unrestricted input setting as considered by Alon and Orlitsky in [21]. In this context, the zero-error constraint is met even outside the source’s support, providing an upper bound on the optimal rate.

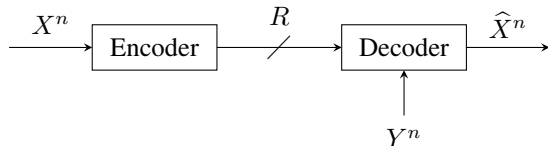


Fig. 2. The source coding problem with decoder side information, called *side-information problem*.

The difficulty in zero-error source and channel coding problems with side information lies in the asymmetry between the information known to the encoder and to the decoder. For example in the source problem, if the side information is known to both encoder and decoder, the problem can be solved by using conditional Huffman coding [4], with the same rate $H(X|Y)$ as in the vanishing error regime. Similarly, when the information known to the decoder is communicated to the encoder via a feedback channel, the zero-error capacity has a single-letter expression established by Shannon in [5].

A. Related Literature

Various questions have been addressed in point-to-point zero-error information theory. For instance, in [26], Sason characterized the zero-error capacity of two subclasses of

strongly regular graphs by leveraging the Lovász θ -function [18]. In [27], Lavi and Sason established sufficient conditions under which the zero-error capacity of a polynomial in graphs equals the corresponding polynomial of the individual capacities. The computability of zero-error capacity has been studied by Boche and Deppe in [28]. In [29], Guo and Watanabe examined a family of graphs where no finite-length code can achieve the zero-error capacity. In [30], Dalai et al. improved the upper bound on the zero-error list-decoding capacity for the $4/3$ channel. In [31], Noorzad et al. provided necessary and sufficient conditions for the randomly generated codebook to satisfy the zero-error property under list-decoding. In [32], Wang et al. analyse the zero-error performance of superposition codes with expurgation processes. In [33], Cao et al. studied the zero-error capacity of channels with memory which are represented by graphs with one edge.

Network zero-error information theory, have been studied, with significant recent advances. In [34], Devroye et al. investigated the zero-error capacity of the primitive relay channel, by proposing a one-shot relaying scheme, termed color-and-forward. The zero-error capacity with noisy channel feedback was studied in [35] and [36], where dynamic programming provides lower and upper bounds. Bracher and Lapidoth established, in [37], the zero-error feedback capacity of the state-dependent channel. In [38], Lapidoth and Yan characterized the zero-error helper capacity of modulo-additive noise channel in the presence of feedback, and derived lower and upper bounds when there is no feedback. In [39], Li and Zuyddam studied the asymptotic spectrum of graphs and introduced the quantum version of the zero-error capacity. In [40], Cao and Yeung determined the zero-error capacity regions of several noisy networks. In [41], Saberi et al. introduced the zero-error feedback capacity of causal discrete channels with memory, and they provided a tight condition for bounded stabilization of unstable noisy linear systems.

In [42], Wang and Shayevitz investigated the combination of zero-error source and channel coding schemes, by introducing the notion of “graph information ratio”, which refers to the relative Shannon capacity of two graphs of Körner and Marton in [43]. In [44], Alipour et al. defined a new quantity called the relative fractional independence number, which provides new upper and lower bounds on the zero-error capacity. In [45], Gu and Shayevitz provided outer and inner bounds on the zero-error capacity region for the two-way channel. In [46], Shayevitz investigated the zero-error broadcasting problem by introducing the ρ -capacity function, for which upper and lower bounds are derived. In [47], Ordentlich and Shayevitz investigated the zero-error capacity region of the binary adder multiple-access channel. They provided a new outer bound that strictly improves upon the bound obtained by Urbanke and Li, in [48]. In [49], Wiese et al. defined zero-error wiretap codes by requiring that every output at the eavesdropper can be generated by at least two inputs. They defined the zero-error secrecy capacity as the supremum of rates for which there exists a zero-error wiretap code. They showed that it is either zero or the zero-error capacity of the channel between the encoder and the legitimate receiver. In [50], Khooshemehr et al. characterized the zero-error capacity of the molecular

delay channel by assuming that the maximum number of the released molecules is fixed.

In [51], Orlitsky and Roche formulated the graph-based problem of coding for computing, which has also seen recent advances. For instance, in [52], Doshi et al. investigated the distributed functional compression problem and proposed a layered architecture that uses graph coloring and distributed source coding. In [53] and [54], Malak demonstrated that coding gains can be obtained by using fractional coloring instead of the traditional graph coloring. In [55], Deylam Salehi et al. determined achievability and converse bounds for the broadcast problem when receivers compute distinct functions.

B. Summary of the Contributions

In this paper, we investigate the additivity property of optimal rates for zero-error source and channel problems. First, we consider a source coding problem with side-information that decomposes into a family of *independent* subproblems $(X_a, Y_a)_{a \in \mathcal{A}}$ with probabilistic graphs $(G_a, P_{X_a})_{a \in \mathcal{A}}$. This problem has a characteristic graph with a specific structure given by the AND product $G = \wedge_{a \in \mathcal{A}} G_a$, and a product distribution $\otimes_{a \in \mathcal{A}} P_{X_a}$. In the vanishing error regime, the optimal rate is *additive*, meaning that it is equal to the sum of the optimal rates of each subproblem. Consequently, the concatenation of optimal codes is optimal. On the contrary, by building on Haemers results for the zero-error capacity of the Schläfli graph in [56], we show that independence alone doesn't ensure the additivity property in the zero-error source problem. This inspired us to investigate the conditions under which the *complementary graph entropy* is additive, i.e. $\overline{H}(G_1 \wedge G_2, P_{X_1} \otimes P_{X_2}) = \overline{H}(G_1, P_{X_1}) + \overline{H}(G_2, P_{X_2})$.

Our first contribution Theorem 3 in Sec. II-B shows that \overline{H} is additive for the AND product (\wedge) of graphs if and only if it is additive for the disjoint union of graphs (\sqcup), also called "sum of graphs" in [57]. The disjoint union is obtained by taking the union of the vertices, and two vertices are adjacent if they are adjacent in the original graph. Additivity for the disjoint union holds if $\overline{H}(G_1 \sqcup G_2, \alpha P_{X_1} + (1 - \alpha)P_{X_2}) = \alpha \overline{H}(G_1, P_{X_1}) + (1 - \alpha) \overline{H}(G_2, P_{X_2})$. We demonstrate an equality that ensures the equivalence between the additivity of the AND product and that of the disjoint union. More precisely, we show that all pairs of probabilistic graphs (G_1, P_{X_1}) and (G_2, P_{X_2}) satisfy

$$\begin{aligned} & \overline{H}\left(G_1 \sqcup G_2, \frac{1}{2}(P_{X_1} + P_{X_2})\right) \\ &= \frac{1}{2} \cdot \overline{H}(G_1 \wedge G_2, P_{X_1} \otimes P_{X_2}), \end{aligned} \quad (1)$$

where $P_{X_1} \otimes P_{X_2}$ is the product of the probability distributions, and where the probability distribution over the vertices of $G_1 \sqcup G_2$ is obtained by taking the uniform distribution $P_A = (\frac{1}{2}, \frac{1}{2})$ over the two probabilistic graphs (G_1, P_{X_1}) and (G_2, P_{X_2}) . It follows directly from (1), that the additivity of $\overline{H}(\wedge \cdot)$ is equivalent to that of $\overline{H}(\sqcup \cdot)$. Then, we extend (1) by showing that if the distribution P_A over the family of probabilistic graphs $(G_a, P_{X_a})_{a \in \mathcal{A}}$ is a type, see [12, Definition 2.1], then $\overline{H}(\sqcup \cdot)$ is a function of the complementary

graph entropy $\overline{H}(\wedge \cdot)$ of a specific product of graphs. Since the set of types is dense in the set of distributions, we show that the additivity of $\overline{H}(\wedge \cdot)$ and of $\overline{H}(\sqcup \cdot)$ are equivalent for all distributions P_A .

Second, we consider a channel coding problem that decomposes into a family of *independent* zero-error subproblems. Recently, Wigderson and Zuiddam in [58] and Schrijver in [59] showed the equivalence of the additivity properties of $C_0(\wedge \cdot)$ and $C_0(\sqcup \cdot)$, where C_0 denotes the zero-error channel capacity. In [59], the key ingredients in the proof are the superadditivity of C_0 , and the additivity property of the independence number $\alpha(G \sqcup G') = \alpha(G) + \alpha(G')$ with respect to the disjoint union of the graphs. Since the characteristic graph G is defined similarly, we explore the connections and differences between the additivity results for $C_0(G)$ and for $\overline{H}(G, P_X)$. For instance, the probability distribution P_X is given in the source problem, whereas it is *a priori* not specified in the channel problem.

In order to relate $C_0(G)$ and $\overline{H}(G, P_X)$, we use the zero-error capacity $C(G, P_X)$ of a graph G relative to a distribution P_X , introduced by Csiszár and Körner in [60] to bound the capacity of the arbitrarily varying channel with maximal error probability. The zero-error capacity $C_0(G)$ and $C(G, P_X)$ are related via the result of [12, equation (11.2)] where

$$C_0(G) = \max_{P_X} C(G, P_X). \quad (2)$$

Moreover, Marton showed in [61] that the complementary graph entropy satisfies

$$C(G, P_X) + \overline{H}(G, P_X) = H(X). \quad (3)$$

Equations (2) and (3) are the analogues of the channel capacity $C = \max_{P_X} I(X; Y)$ and the entropy property $I(X; Y) + H(X|Y) = H(X)$ in the vanishing error regime. Proposition 5 in Sec. III-C uses (3) to prove the equivalence between the additivity of $C(G, P_X)$ and that of $\overline{H}(G, P_X)$. In order to complete this study, it remains to relate the additivity of $C(G, P_X)$ to the one of $C_0(G)$. This requires careful analysis of the distribution that maximizes $C(G, P_X)$ and how it decomposes over the AND product and over the disjoint union of graphs. Suppose that for all subproblems the source distribution P_X maximizes $C(G, P_X)$, Theorem 9 in Sec. III-D and Theorem 10 in Sec. III-E conclude that the additivity properties of $C_0(G)$, $C(G, P_X)$ and $\overline{H}(G, P_X)$ for the AND product (\wedge) and for the disjoint union (\sqcup) are all equivalent.

A crucial notion is the set of *capacity-achieving distributions* that contains all the distributions P_X for which $C_0(G) = C(G, P_X)$. We show that the uniform distribution is *capacity-achieving* when the graph is vertex-transitive, i.e. when all vertices play the same role within the graph. Since the Schläfli graph S and its complementary graph \overline{S} are vertex-transitive, so is their product $S \wedge \overline{S}$, and the uniform distribution is capacity achieving for S , \overline{S} and $S \wedge \overline{S}$. Together with Haemers result [56], this shows a counterexample where the additivity property of $C_0(G)$, $C(G, P_X)$ and $\overline{H}(G, P_X)$ for the AND product and for the disjoint union of S and \overline{S} does not hold.

The additivity property enlarges the class of graphs for which C_0 and \overline{H} have a single-letter characterization. For

perfect graphs, we show that the additivity of C_0 and \overline{H} always hold for the AND product and for the disjoint union. This allows us to characterize new single-letter solutions for all products of perfect graphs, that are not perfect in general, and for the product of a perfect graph G with the pentagon graph C_5 , by building on the characterization of $\overline{H}(G \sqcup C_5)$ by Koulgi et al. in [57].

The path taken to demonstrate all these equivalences is shown in Fig. 3. In Sec. II, we study the additivity of the complementary graph entropy \overline{H} for the source problem with side information. The connection with the additivity of the zero-error capacity C_0 is investigated in Sec. III. New single-letter solutions for C_0 and \overline{H} are provided in Sec. IV, as well as the counter-example for the additivity based on Haemers result [56] for the Schläfli graph.

II. ZERO-ERROR SOURCE CODING WITH DECODER SIDE INFORMATION

A. Results from the Literature and Open Problems

We investigate the source coding problem depicted in Fig. 2, which we call the *side-information problem*. This situation arises in data compression where the decoder has side-information Y about the source X that must be retrieved. This problem has been solved in the vanishing error probability regime by Slepian and Wolf [20], but remains open in the zero-error regime. Asymptotic expressions have been derived by Witsenhausen [19] for fixed-length codes, and by Alon and Orlitsky [21] for variable-length codes.

More formally, we assume that a sequence of length $n \in \mathbb{N}^*$ of i.i.d. random variables (X^n, Y^n) is drawn according to the probability distribution $P_{X,Y} \in \Delta(\mathcal{X} \times \mathcal{Y})$ where \mathcal{X} and \mathcal{Y} are finite sets and $\Delta(\mathcal{X} \times \mathcal{Y})$ denotes the set of probability distributions over $\mathcal{X} \times \mathcal{Y}$. We assume that the marginal distributions $P_X \in \Delta(\mathcal{X})$ and $P_Y \in \Delta(\mathcal{Y})$ have full support. We denote by $\{0, 1\}^*$ the set of binary words. We consider variable-length source coding, which encompasses the special case of fixed-length source coding. An (n, ϕ_e, ϕ_d) *variable-length side-information source code* for X and Y consists of

- an encoder $\phi_e : \mathcal{X}^n \rightarrow \{0, 1\}^*$ that assigns to each x^n a binary word such that its image $\text{Im } \phi_e$ is prefix-free,
- a decoder $\phi_d : \mathcal{Y}^n \times \{0, 1\}^* \rightarrow \mathcal{X}^n$ that assigns an estimate \hat{x}^n to each pair $(y^n, \phi_e(x^n))$.

The rate of the (n, ϕ_e, ϕ_d) -code is the average length of the codeword per source symbol, i.e. $R \doteq \frac{1}{n} \mathbb{E}[\ell \circ \phi_e(X^n)]$ where ℓ denote the codeword length. We denote the probability of error by $P_e^{(n)} \doteq \mathbb{P}(\hat{X}^n \neq X^n)$.

Definition 1. *The optimal rate in the zero-error regime is the minimal rate among all coding schemes that satisfy the zero-error constraint:*

$$R_0^* \doteq \inf_{(n, \phi_e, \phi_d) : P_e^{(n)} = 0} \frac{1}{n} \mathbb{E}[\ell \circ \phi_e(X^n)]. \quad (4)$$

When the side-information Y is available at both encoder and decoder, the optimal rate in the zero-error regime coincides with the optimal rate $H(X|Y)$ in the vanishing error

regime. The zero-error coding construction relies on a conditional Huffman coding [4]. In the side-information problem, the encoder does not observe the side-information Y . Due to this information asymmetry, the optimal rates in the zero-error and vanishing-error regimes are distinct. The characterisation of the optimal rate R_0^* is a notoriously difficult open problem of combinatorial nature. The key features are captured by the “characteristic graph” of Witsenhausen [19], which is constructed based on the support of the side information channel $P_{Y|X}$.

Definition 2 (Characteristic graph). *Let \mathcal{X}, \mathcal{Y} be two finite sets and $P_{Y|X}$ be a conditional distribution. The characteristic graph $G = (\mathcal{X}, \mathcal{E})$ associated to $P_{Y|X}$ is defined by:*

- \mathcal{X} as set of vertices,
- $x \neq x' \in \mathcal{X}$ are adjacent $xx' \in \mathcal{E}$, if for some $y \in \mathcal{Y}$, $P_{Y|X}(y|x) \cdot P_{Y|X}(y|x') > 0$.

We denote by (G, P_X) the probabilistic graph induced by the characteristic graph G and the distribution P_X on its vertices.

When the side information y does not allow to distinguish exactly between the source realizations x and x' , then x and x' are adjacent, and must be mapped to different codewords. A zero-error encoding consists of a graph coloring where adjacent vertices are mapped to different colors.

Definition 3 (Coloring, chromatic number χ). *Let $G = (\mathcal{X}, \mathcal{E})$ be a graph. A mapping $c : \mathcal{X} \rightarrow \mathcal{C}$ is a coloring if for all adjacent vertices x, x' with $xx' \in \mathcal{E}$, we have $c(x) \neq c(x')$. The chromatic number $\chi(G)$ is the smallest $|\mathcal{C}|$ such that there exists a coloring $c : \mathcal{X} \rightarrow \mathcal{C}$ of G .*

For all graphs G_1 and G_2 , we have $\chi(G_1 \wedge G_2) \leq \chi(G_1) \cdot \chi(G_2)$. For sequences of symbols with underlying i.i.d. conditional distribution $P_{Y|X}^{\otimes n}(y^n|x^n) = \prod_{t=1}^n P_{Y|X}(y_t|x_t)$, two sequences of source symbols x^n, x'^n are adjacent in the graph if $P_{Y|X}^{\otimes n}(y^n|x^n) P_{Y|X}^{\otimes n}(y^n|x'^n) > 0$ for some sequence of channel outputs y^n , i.e. if and only if either $x_t = x'_t$ or $x_t x'_t \in \mathcal{E}$, for all $1 \leq t \leq n$. This implies that for sequences of symbols, the characteristic graph is built by using the AND product of graphs, denoted by \wedge , and also called “strong product” or “normal product” in [18], [61], and defined below.

Definition 4 (AND product \wedge). *Let $G_1 = (\mathcal{X}_1, \mathcal{E}_1)$, $G_2 = (\mathcal{X}_2, \mathcal{E}_2)$ be two graphs, their AND product $G_1 \wedge G_2$ is defined by*

- $\mathcal{X}_1 \times \mathcal{X}_2$ as set of vertices,
- $(x_1 x_2), (x'_1 x'_2)$ are adjacent if $x_1 x'_1 \in \mathcal{E}_1$ and $x_2 x'_2 \in \mathcal{E}_2$, with the convention of self-adjacency for all vertices.

We denote by $G_1^{\wedge n}$ the n -th AND power $G_1^{\wedge n} = G_1 \wedge \dots \wedge G_1$ n times.

When (G_1, P_{X_1}) and (G_2, P_{X_2}) are two probabilistic graphs, the distribution induced on the product of vertices $\mathcal{X}_1 \times \mathcal{X}_2$ is the product of the two distributions $P_{X_1} \otimes P_{X_2}$.

Alon and Orlitsky introduced an asymptotic expression in [21], for the optimal rate R_0^* in the “restricted inputs” setting. It relies on the notion of chromatic entropy $H_\chi(G^{\wedge n}, P_X^{\otimes n})$, which is the minimal entropy of a coloring of $G^{\wedge n}$ induced by the i.i.d. distribution $P_X^{\otimes n}$.

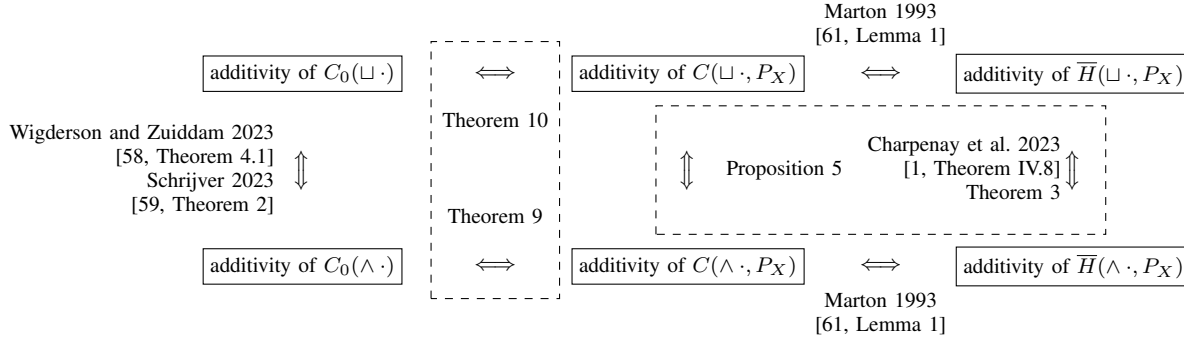


Fig. 3. Equivalences of additivity properties for the zero-error capacity C_0 , for the zero-error capacity relative to a distribution C , and for the complementary graph entropy \bar{H} . The equivalences of our Theorem 3, Proposition 5, Theorem 9 and Theorem 10 are valid under specific conditions on the probability distributions.

Theorem 1 (from [21, Lemma 6]). *Given a probabilistic graph (G, P_X) ,*

$$R_0^* = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n}, P_X^{\otimes n}), \quad (5)$$

where the chromatic entropy is defined by

$$H_\chi(G, P_X) \doteq \min \left\{ H(c(X)) \mid c \text{ is a coloring of } G \right\}. \quad (6)$$

The sequence $(H_\chi(G^{\wedge n}, P_X^{\otimes n}))_{n \in \mathbb{N}^*}$ is subadditive, Fekete's Lemma [62, pp. 103] ensures the existence of the limit in (5), which coincides with the infimum.

There is no single-letter expression for R_0^* . In [21], Alon and Orlitsky provided a single-letter upper bound of R_0^* which corresponds to the “unrestricted inputs” scenario where the source X has to be recovered with zero-error for all side information y , even when (X, y) is outside the support of $P_{X,Y}$.

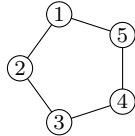


Fig. 4. The pentagon graphs C_5 with uniform distribution $P_X = \text{Unif}(\{1, \dots, 5\})$ over the vertices.

With high probability, the source sequence X^n is typical with respect to P_X . Let $G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]$ be the subgraph of $G^{\wedge n}$ induced by the set of *typical sequences* $\mathcal{T}_\varepsilon^n(P_X)$ with tolerance $\varepsilon > 0$, see [12, Definition 2.8]. We can construct a zero-error code by taking the minimal coloring of this induced subgraph $\chi(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)])$. The encoder sends the color index to the decoder if X^n is typical, otherwise it sends the index of the sequence X^n in \mathcal{X}^n . This coding strategy has a rate upper-bounded by

$$\frac{1}{n} + \mathbb{P}(X^n \notin \mathcal{T}_\varepsilon^n(P_X)) \log |\mathcal{X}| + \frac{1}{n} \log \chi(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]). \quad (7)$$

The zero-error property is satisfied since the decoder is able to retrieve X^n thanks to Y^n and the color symbol. Koulgi et al. have shown in [22, Theorem 1] that taking the limit when n goes to infinity and ε goes to 0 yields the best achievable

rate in the zero-error side-information problem. This quantity, introduced by Körner and Longo in [23] for the two-step source coding, is called the complementary graph entropy. Here, we adopt the formalism of [63, Definition 7] and [57, Definition 7].

Definition 5. *Given a probabilistic graph (G, P_X) , the complementary graph entropy $\bar{H}(G, P_X)$ is defined by:*

$$\bar{H}(G, P_X) \doteq \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \chi(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]). \quad (8)$$

The question of the subadditivity of the sequence $(\log \chi(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]))_{n \in \mathbb{N}^*}$ is open.

Theorem 2 (from [22, Theorem 1]). *The optimal rate in the zero-error regime is*

$$R_0^* = \bar{H}(G, P_X), \quad (9)$$

where (G, P_X) is the probabilistic graph formed of the characteristic graph associated to the distribution $P_{Y|X}$, with the underlying distribution P_X on its vertices.

Note that when restricting to zero-error fixed-length code, the optimal rate $H_0(G)$ introduced by Witsenhausen in [19] does not depend on the source distribution P_X , since

$$H_0(G) \doteq \lim_{n \rightarrow +\infty} \frac{1}{n} \log \chi(G^{\wedge n}). \quad (10)$$

The limit exists and coincides with the infimum thanks to Fekete's Lemma [62, pp. 103] for the subadditive sequence $(\log \chi(G^{\wedge n}))_{n \in \mathbb{N}^*}$. In fact, the *Witsenhausen rate* $H_0(G)$ is equal to the maximum of \bar{H} over the set of source distributions $Q_X \in \Delta(\mathcal{X})$.

Lemma 1 (from [63, Lemma 3]). *Given the graph $G = (\mathcal{X}, \mathcal{E})$, the Witsenhausen rate satisfies*

$$H_0(G) = \max_{Q_X} \bar{H}(G, Q_X). \quad (11)$$

The proof relies on the type counting Lemma, see [12, Lemma 2.2]. Interestingly (11) shows that the use of zero-error fixed-length codes is a strong restriction compared to the use of zero-error variable-length codes. Indeed, the optimal rate for variable-length code satisfies $R_0^* = \bar{H}(G, P_X) \leq \max_{Q_X} \bar{H}(G, Q_X) = H_0(G)$. Moreover, note that the rate

$H_0(G)$ corresponds to a universal source coding problem in which the characteristic graph is known, but the source distribution is unknown.

We now focus on variable-length coding. The optimal rate R_0^* has a trivial single-letter upper bound given by $H(X)$, where the zero-error coding construction relies on Huffman coding and the decoder ignores the side information Y . In fact, this upper bound is tight for a dense subset of distributions in $\Delta(\mathcal{X} \times \mathcal{Y})$.

Proposition 1 (Full support, from [19]). *If the distribution $P_{X,Y}$ has full support, then $R_0^* = H(X)$.*

Indeed, since the distribution $P_{X,Y}$ has full support, the characteristic graph G is complete, i.e. every pair of symbols $x \in \mathcal{X}$, $x' \in \mathcal{X}$ are adjacent in G , thus $H_X(G, P_X) = H(X)$.

There are a few other cases where the optimal zero-error rate is known such as perfect graphs [24, Corollary 12], or the pentagon C_5 with uniform distribution shown in Fig. 4 where $R_0^* = \frac{1}{2} \log_2(5)$, see [22, Example 1]. In general, the single-letter characterization of R_0^* remains a difficult open question.

B. Problem Statement for Independent Sources

In order to understand this difficulty, we examine a specific scenario where the source and the side information decompose into *independent* variables. In the vanishing error regime, independence is a key assumption that ensures the additivity of optimal rates, shedding light on practical coding techniques. In the zero-error regime, the independence hypothesis alone is not sufficient to ensure the additivity of optimal rates. We formulate a condition that implies additivity, enabling us to enlarge the set of problems for which the optimal rate has a known single-letter characterization.

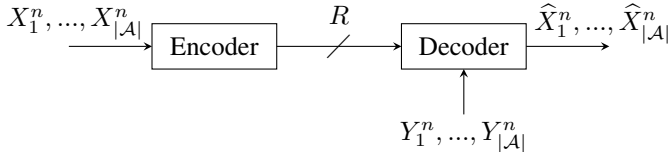


Fig. 5. Independent side-information problems

More formally, for a finite set \mathcal{A} , we assume a set of pairs $(X_a, Y_a)_{a \in \mathcal{A}}$, referred to as an *independent family*, that consists of $|\mathcal{A}|$ pairs with a joint distribution that decomposes into a product of distributions. This independent family generates sequences of i.i.d. random variables,

$$(X_1^n, Y_1^n, \dots, X_{|\mathcal{A}|}^n, Y_{|\mathcal{A}|}^n) \sim \left(P_{X_1, Y_1} \otimes \dots \otimes P_{X_{|\mathcal{A}|}, Y_{|\mathcal{A}|}} \right)^{\otimes n}.$$

Independent side-information problems correspond to a side-information problem in which the source and the side information form an independent family, as shown in Fig. 5. In the vanishing error regime, the optimal rate is additive,

$$R^* = H(X_1, \dots, X_{|\mathcal{A}|} | Y_1, \dots, Y_{|\mathcal{A}|}) \quad (12)$$

$$= \sum_{a \in \mathcal{A}} H(X_a | Y_a) = \sum_{a \in \mathcal{A}} R_a^*. \quad (13)$$

This property is fundamental to guarantee the optimality of the codebook constructed by concatenating the codewords of the optimal codebooks of each subproblem.

But does the *additivity* property also hold in the zero-error regime for independent side-information problems? To answer this question, we first derive an asymptotic expression for the optimal zero-error rate. This derivation follows from the fact that the independent family can be characterized by a product of graphs.

Proposition 2. *Let $(X_a, Y_a)_{a \in \mathcal{A}}$ be an independent family. The optimal rate for the independent zero-error side-information problem is*

$$R_0^* = \overline{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right), \quad (14)$$

where for all $a \in \mathcal{A}$, G_a is the characteristic graph associated to the conditional distribution $P_{Y_a | X_a}$, with the underlying probability distribution P_{X_a} on its vertices.

It is known that the complementary graph entropy \overline{H} is sublinear with respect to the AND product. Indeed, [57, Theorem 2] states that for all probabilistic graphs (G, P_X) and (G', P'_X)

$$\overline{H}(G \wedge G', P_X \otimes P'_X) \leq \overline{H}(G, P_X) + \overline{H}(G', P'_X). \quad (15)$$

However, \overline{H} is not additive in general. Inspired by Haemers result [56], we show in Theorem 16, that the inequality (15) is strict for the product of the Schläfli graph S with its complement \overline{S} . In the following, we study a condition that ensures the additivity of \overline{H} , i.e. where (15) holds with equality. To do this, we introduce the disjoint union of graphs, also called “sum of graphs” in [57].

Definition 6 (Disjoint union of probabilistic graphs \sqcup). *Let \mathcal{A} be a finite set, let $P_A \in \Delta(\mathcal{A})$, and for all $a \in \mathcal{A}$, let (G_a, P_{X_a}) be a probabilistic graph with $G_a = (\mathcal{X}_a, \mathcal{E}_a)$. The disjoint union with respect to P_A is a probabilistic graph denoted by $\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right)$, where the vertices \mathcal{X} and the edges \mathcal{E} are defined by:*

- $\mathcal{X} = \bigsqcup_{a \in \mathcal{A}} \mathcal{X}_a$ is the disjoint union of the sets $(\mathcal{X}_a)_{a \in \mathcal{A}}$,
- For all $x, x' \in \mathcal{X}$, $xx' \in \mathcal{E}$ if and only if they both belong to the same \mathcal{X}_a and $xx' \in \mathcal{E}_a$.

The distribution $P_X = \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ is constructed with respect to P_A and the family of distributions $(P_{X_a})_{a \in \mathcal{A}}$ that have disjoint support in \mathcal{X} .

In Fig. 6, we provide an example of an AND product and of a disjoint union of probabilistic graphs. As for the AND product, the complementary graph entropy \overline{H} is sublinear with respect to the disjoint union. Indeed, [57, Theorem 2] states that for all probabilistic graphs (G, P_X) and (G', P'_X) and the distribution $P_A = (s, 1 - s)$ with $s \in [0, 1]$,

$$\overline{H}(G \sqcup G', sP_X + (1 - s)P'_X) \leq s\overline{H}(G, P_X) + (1 - s)\overline{H}(G', P'_X). \quad (16)$$

We present our first contribution, which establishes the equivalence between the additivity of the AND product and the

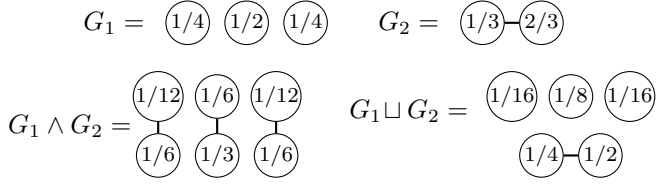


Fig. 6. An empty graph $(G_1, P_{X_1}) = (N_3, (\frac{1}{4}, \frac{1}{2}, \frac{1}{4}))$ and a complete graph $(G_2, P_{X_2}) = (K_2, (\frac{1}{3}, \frac{2}{3}))$, along with their AND product $(G_1 \wedge G_2, P_{X_1} \otimes P_{X_2})$ and their disjoint union $(G_1 \sqcup G_2, \frac{1}{4}P_{X_1} + \frac{3}{4}P_{X_2})$. The values on each vertex correspond to the underlying probability distributions $P_{X_1} \otimes P_{X_2}$ and $\frac{1}{4}P_{X_1} + \frac{3}{4}P_{X_2}$.

additivity of the the disjoint union. An important consequence when the additivity holds is that the concatenation of optimal codes of each subproblems is optimal.

Theorem 3 (Equivalence of the additivity for \wedge and for \sqcup , from [1, Theorem IV.8]). *Let \mathcal{A} be a finite set, P_A a distribution with full-support, and let $(G_a, P_{X_a})_{a \in \mathcal{A}}$ a family of probabilistic graphs. The following equivalence holds:*

$$\overline{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) = \sum_{a \in \mathcal{A}} \overline{H}(G_a, P_{X_a}), \quad (17)$$

$$\iff \overline{H} \left(\bigcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) = \sum_{a \in \mathcal{A}} P_A(a) \overline{H}(G_a, P_{X_a}). \quad (18)$$

We say that the additivity property holds for the AND product if (17) is satisfied. Similarly, the additivity holds for the disjoint union if (18) is satisfied.

This result allows to characterize new single-letter solutions, as discussed in Sec. IV. By using Theorem 3, we characterize the optimal rate of all products of perfect graphs, which are not perfect in general. Since the disjoint product of perfect graphs is perfect, the optimal rates add up.

Without loss of generality, we consider that \mathcal{A} is the support of P_A . We observe that (17) does not depend on the distribution P_A , therefore if (18) holds for one distribution P_A with full support, then it holds for all distributions with full support. This remark allows us to state the following corollary.

Corollary 1. *If the additivity properties (17) or (18) hold for a family of probabilistic graphs $(G_a, P_{X_a})_{a \in \mathcal{A}}$, then (17) and (18) also hold for any sub-family of probabilistic graphs $(G_{\tilde{a}}, P_{X_{\tilde{a}}})_{\tilde{a} \in \tilde{\mathcal{A}}}$ with $\tilde{\mathcal{A}} \subset \mathcal{A}$.*

C. Proof of Theorem 3

In this section, we present the three main ingredients for the proof of Theorem 3. In Lemma 2, we provide an asymptotic formulation that relies solely on the AND product of graphs.

Lemma 2. *Let $P_A \in \Delta(\mathcal{A})$ a probability distribution with full-support and let $(\bar{a}_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$ be any sequence such*

that its type $T_{\bar{a}^n} \rightarrow P_A$ when $n \rightarrow \infty$. Then we have

$$\begin{aligned} & \overline{H} \left(\bigcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} H_{\chi} \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)} \right). \end{aligned} \quad (19)$$

As a consequence, if P_A belongs to the set $\Delta_k(\mathcal{A})$ of types of sequences of length $k \in \mathbb{N}^*$, we have

$$\begin{aligned} & \overline{H} \left(\bigcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= \frac{1}{k} \overline{H} \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge k P_A(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes k P_A(a)} \right). \end{aligned} \quad (20)$$

Moreover, if P_A is the uniform distribution, we have

$$\overline{H} \left(\bigcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} \frac{1}{|\mathcal{A}|} P_{X_a} \right) = \frac{1}{|\mathcal{A}|} \overline{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right). \quad (21)$$

The proof of Lemma 2 is stated in App. A-A. It follows from the properties of the product of graphs induced by typical sequences $a^n \in \mathcal{T}_{\varepsilon}^n(P_A)$, such as the distributivity of \wedge with respect to \sqcup , the subadditivity of the sequence of chromatic entropy $(H_{\chi}(G^{\wedge n}, P_X^{\otimes n}))_{n \in \mathbb{N}^*}$, and the property $H_{\chi}(G_1 \sqcup G_2, P_A(a_1)P_{X_1} + P_A(a_2)P_{X_2}) = H_{\chi}(G_1, P_{X_1}) = H_{\chi}(G_2, P_{X_2})$, for all P_A when the probabilistic graphs (G_1, P_{X_1}) and (G_2, P_{X_2}) are isomorphic. Lemma 2 shows that the complementary graph entropy of the disjoint union of graphs $\overline{H}(\sqcup \cdot)$ is a function of the complementary graph entropy of the AND product of graphs $\overline{H}(\wedge \cdot)$, on a dense subset of distributions. In that case, the equivalence of the additivity with respect to \wedge and \sqcup is direct.

A consequence of (21) is that both additivity properties of (17) and (18) are equivalent when $P_A = \text{Unif}(\mathcal{A})$. In order to prove Theorem 3, we extend this property to all distributions $P_A \in \Delta(\mathcal{A})$ by using the convexity property of the function

$$\eta : P_A \mapsto \overline{H} \left(\bigcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right). \quad (22)$$

Lemma 3. *Given a family of graphs $(G_a)_{a \in \mathcal{A}}$, the function η is convex in P_A and $(\log \max_a |\mathcal{X}_a|)$ -Lipschitz.*

The proof of Lemma 3 is stated in App. A-F. Now, since η is convex, if it meets the linear interpolation of $(\eta(\mathbb{1}_a))_{a \in \mathcal{A}}$, where $(\mathbb{1}_a)_{a \in \mathcal{A}}$ are the extreme points of the set of distributions $\Delta(\mathcal{A})$, then η is linear. This statement is provided by Lemma 4 whose proof is given in App. A-H.

Lemma 4. *Let \mathcal{A} be a finite set, and $\gamma : \Delta(\mathcal{A}) \rightarrow \mathbb{R}$ be a convex function, and for all $a \in \mathcal{A}$, let $\mathbb{1}_a$ be the distribution that assigns 1 to the symbol a and 0 to the others. Then the following holds:*

$$\exists P_A \in \text{int}(\Delta(\mathcal{A})), \quad \gamma(P_A) = \sum_{a \in \mathcal{A}} P_A(a) \gamma(\mathbb{1}_a) \quad (23)$$

$$\iff \forall P_A \in \Delta(\mathcal{A}), \quad \gamma(P_A) = \sum_{a \in \mathcal{A}} P_A(a) \gamma(\mathbb{1}_a). \quad (24)$$

where $\text{int}(\Delta(\mathcal{A}))$ is the interior of $\Delta(\mathcal{A})$, i.e. the set of full-support distributions on \mathcal{A} .

Now let us prove Theorem 3.

(\implies) We assume that $\bar{H}(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}) = \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a})$. By taking $P_A = \text{Unif}(\mathcal{A})$, the equation (21) in Lemma 2 implies that

$$\begin{aligned} & \bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} \frac{1}{|\mathcal{A}|} P_{X_a}\right) \\ &= \frac{1}{|\mathcal{A}|} \bar{H}\left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}\right) \\ &= \sum_{a \in \mathcal{A}} \frac{1}{|\mathcal{A}|} \bar{H}(G_a, P_{X_a}). \end{aligned} \quad (25)$$

The function η is convex by Lemma 3, and satisfies (23) with the interior point $P_A = \text{Unif}(\mathcal{A})$, thus Lemma 4 implies that for all $P_A \in \Delta(\mathcal{A})$,

$$\bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right) = \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}). \quad (26)$$

(\Leftarrow) Conversely, we assume (26) and we consider $P_A = \text{Unif}(\mathcal{A})$. According to (21) in Lemma 2, we conclude that

$$\begin{aligned} & \bar{H}\left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}\right) \\ &= |\mathcal{A}| \cdot \bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} \frac{1}{|\mathcal{A}|} P_{X_a}\right) \\ &= \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a}). \end{aligned} \quad (27)$$

This concludes the proof of Theorem 3.

D. Problem Statement with Partial Side Information at the Encoder

In this section, we highlight the operational significance of the disjoint union of graphs. Recall that Theorem 3 uses the additivity for the disjoint union in order to establish the additivity for the AND product of graphs. This is generally simpler because the disjoint union has fewer vertices and edges than the AND product. Yet, the usefulness extends further as the disjoint union corresponds to the problem in Fig. 7, where the encoder has partial information about the decoder's side information, obtained through the deterministic function $g : \mathcal{Y} \rightarrow \mathcal{A}$. This setting in Fig. 7, is a specific case of the setting of Fig. 2 where the decoder must retrieve the source $(X, g(Y))$. We refer to this as the *partial-side-information problem*.

Proposition 3. *When the encoder has partial side information, the optimal rate is*

$$R_0^* = \bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right), \quad (28)$$

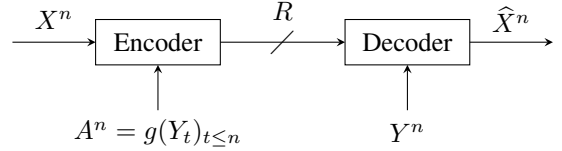


Fig. 7. The partial-side-information problem.

where for all $a \in \mathcal{A}$, G_a is the characteristic graph associated to the conditional distribution $P_{Y|X, A=a}$ with the underlying probability distribution $P_{X|A=a}$ on its vertices. These conditional distributions and P_A are obtained from the joint distribution $P_{XY} \mathbb{1}_{A=g(Y)}$ depending on the deterministic function $g : \mathcal{Y} \rightarrow \mathcal{A}$.

Indeed, for each realization of the encoder side information $a = g(y)$, we construct a characteristic graph G_a to model the sub-problem indexed by $a \in \mathcal{A}$. Since, both encoder and decoder have access to $a \in \mathcal{A}$, the characteristic graph consists in the disjoint union of the graphs $(G_a)_{a \in \mathcal{A}}$. Moreover, each G_a contains all realizations $x \in \mathcal{X}$, and there is an edge between two vertices x, x' if and only if $P_{Y|X, a}(y|x, a) P_{Y|X, a}(y|x', a) > 0$ for some $y \in g^{-1}(a)$.

If \bar{H} is additive for the disjoint union in the sense of (18), then the following coding scheme is optimal:

- For each symbol $a \in \mathcal{A}$, we select the indices $t \in \{1, \dots, n\}$ of the sequence a^n such that $a_t = a$. We denote by (X^{n_a}, Y^{n_a}) the corresponding subsequences of length n_a , extracted from (X^n, Y^n) .
- For each $a \in \mathcal{A}$, we use the optimal codebook for the independent sources $(X^{n_a}, Y^{n_a})_{a \in \mathcal{A}}$ with distinct length of sequences $(n_a)_{a \in \mathcal{A}}$ and we concatenate the codewords obtained.

With high probability, the sequence A^n belongs to the set of typical sequences $\mathcal{T}_\varepsilon^n(P_A)$, therefore the empirical distribution $(\frac{n_a}{n})_{a \in \mathcal{A}}$ converges to P_A in probability. The coding rate of the above scheme converges to

$$\sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}) = \bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right),$$

which is optimal for the partial-side-information problem when \bar{H} is additive.

III. ZERO-ERROR CHANNEL CODING PROBLEM

Recently, Wigderson and Zuiddam in [58] and Schrijver in [59] established the equivalence between the additivity of the zero-error channel capacity for the AND product $C_0(\wedge \cdot)$ and for the disjoint union of graphs $C_0(\sqcup \cdot)$. The main difference with the side information problem is that the channel input distribution P_X is *a priori* not specified in the zero-error channel coding problem.

In order to establish the equivalence between the additivity properties of \bar{H} and C_0 , we introduce the *zero-error capacity* $C(G, P_X)$ of a graph G relative to a distribution P_X due to Csiszár and Körner [60]. We show the equivalence of all additivity properties, i.e. those of [58, Theorem 4.1], of [59, Theorem 2], and of Theorem 3, provided that the source

distribution maximizes the zero-error capacity relative to a distribution of the AND product of graphs.

A. Open Problem of Zero-Error Channel Capacity

The channel coding problem in Fig. 8 is introduced in [3] in the vanishing error regime, and in [5] in the zero-error regime. We consider a *discrete memoryless channel* that consists of an input alphabet \mathcal{X} , a finite output alphabet \mathcal{Y} and a conditional distribution $P_{Y|X}$. A $(n, \mathcal{C}_n, \phi_d)$ -code consists of

- an encoder that selects uniformly a codeword x^n from the codebook $\mathcal{C}_n \subseteq \mathcal{X}^n$, and sends it over the channel,
- a decoder ϕ_d that assigns an estimate \hat{x}^n to each received sequence y^n .

The rate of the $(n, \mathcal{C}_n, \phi_d)$ -code is the average number of messages transmitted per channel use, i.e. $\frac{1}{n} \log |\mathcal{C}_n|$, and the probability of error is $P_e^{(n)} \doteq \mathbb{P}(\hat{X}^n \neq X^n)$.

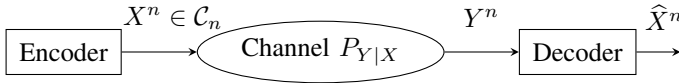


Fig. 8. The channel coding problem.

Definition 7. The zero-error channel capacity is the maximal rate among all coding schemes that satisfy the zero-error property:

$$C_0^* \doteq \sup_{(n, \mathcal{C}_n, \phi_d) : P_e^{(n)} = 0} \frac{1}{n} \log |\mathcal{C}_n|. \quad (29)$$

In the zero-error regime, the capacity depends on the independence number of the characteristic graph, defined below.

Definition 8 (Independent subset, independence number α). Let $G = (\mathcal{X}, \mathcal{E})$ be a graph. A subset $\mathcal{S} \subseteq \mathcal{X}$ is independent in G if $xx' \notin \mathcal{E}$ for all $x \neq x' \in \mathcal{S}$. The independence number is the maximal size of an independent set in G , and is denoted by $\alpha(G)$.

The zero-error channel capacity relies on the same characteristic graph, see Definition 2.

Theorem 4 (from [5]). Let G be the characteristic graph corresponding to the discrete memoryless channel $(\mathcal{X}, \mathcal{Y}, P_{Y|X})$. The zero-error channel capacity satisfies

$$C_0^* = C_0(G) \doteq \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\wedge n}). \quad (30)$$

Since $\alpha(G_1 \wedge G_2) \geq \alpha(G_1) \cdot \alpha(G_2)$ for all graphs G_1 and G_2 , the sequence $(\log \alpha(G^{\wedge n}))_{n \in \mathbb{N}^*}$ is superadditive, Fekete's Lemma [62, pp. 103] ensures the existence of the limit in (30), which coincides with the supremum.

Remark 1. Note that, by convention, we define the zero-error capacity with the logarithm. Another existing convention (for example in [18]) for the zero-error capacity is $\Theta(G) \doteq \lim_{n \rightarrow \infty} \sqrt[n]{\alpha(G^{\wedge n})}$, which is equivalent in the sense that $C_0 = \log \Theta$.

We present in Sec. IV-A some examples from the literature where $C_0(G)$ is known, in particular when G is a perfect

graph. The Lovász θ -function, introduced in [18], is an upper bound on the zero-error capacity. This function is used to show that $C_0(C_5) = \frac{1}{2} \log 5$, which makes C_5 the smallest non-perfect graph for which C_0 is known. Further observations on the θ -function are derived by Sason in [64], and new zero-error capacity results are characterized for two subclasses of strongly regular graphs in [26]. The zero-error capacity of C_7 is still unknown. Several existing lower bounds on $C_0(C_7)$ were found via computer programming, in particular in [7], [9] and [65].

B. Problem Statement for Independent Channels

To understand why no single-letter exists for the zero-error channel capacity, we study the case where the conditional distribution of the channel decomposes into a product $\bigotimes_{a \in \mathcal{A}} P_{Y_a|X_a}$, as depicted in Fig. 9. This is called the *independent channel coding problem*.

Proposition 4 (from [5]). The zero-error capacity of the independent channel $\bigotimes_{a \in \mathcal{A}} P_{Y_a|X_a}$ is given by

$$C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right), \quad (31)$$

where for all $a \in \mathcal{A}$, G_a is the characteristic graph associated to the conditional distribution $P_{Y_a|X_a}$.

In the vanishing error regime, the capacity of independent channels is additive since

$$\begin{aligned} C &= \max_{P_{X_1, \dots, X_{|\mathcal{A}|}}} I(X_1, \dots, X_{|\mathcal{A}|}; Y_1, \dots, Y_{|\mathcal{A}|}) \\ &= \sum_{a \in \mathcal{A}} \max_{P_{X_a}} I(X_a; Y_a) = \sum_{a \in \mathcal{A}} C_a. \end{aligned} \quad (32)$$

Therefore, it is optimal to concatenate the optimal codebooks designed for each channel $P_{Y_a|X_a}$.

In the zero-error regime, the capacity is super-linear as shown by Shannon in [5, Theorem 4],

$$C_0(G) + C_0(G') \leq C_0(G \wedge G'). \quad (33)$$

Haemers shows in [56] that the inequality (33) is strict for the product of the Schläfli graph S and its complementary graph \bar{S} , as stated in Theorem 5. An explicit construction of the Schläfli graph is provided in [66, Sec. 6.1]. Haemers's result relies on a bound on the zero-error capacity based on the rank of the adjacency matrix of the graph. Refinements of this bound are developed by Bukh and Cox in [67], and by Gao et al. in [68].

Theorem 5 (from [56]). Let S be the Schläfli graph and \bar{S} its complementary graph, then

$$C_0(S) + C_0(\bar{S}) < C_0(S \wedge \bar{S}). \quad (34)$$

Since the additivity does not hold in general, Wigderson and Zuydam in [58] and Schrijver in [59], established a condition under which the additivity holds for the zero-error capacity.

Theorem 6 (from [58, Theorem 4.1] and [59, Theorem 2]). *For all graphs G, G' ,*

$$C_0(G) + C_0(G') = C_0(G \wedge G'), \quad (35)$$

$$\iff \log \left(2^{C_0(G)} + 2^{C_0(G')} \right) = C_0(G \sqcup G'). \quad (36)$$

The zero-error capacity of a disjoint union of graphs $C_0(G \sqcup G')$ is the optimal coding rate when each input letter can be selected from any of the two channels. We establish the connection between the additivity properties of Theorem 3 and Theorem 6, by introducing the zero-error capacity $C(G, P_X)$ of a graph relative to a distribution P_X , defined by Csiszár and Körner in [60].

C. Zero-Error Capacity $C(G, P_X)$ of a Graph Relative to a Distribution P_X

Theorems 3 and 6 both establish the equivalence of additivity properties for the AND product and for the disjoint union of graphs but they do so for different underlying quantities: $\bar{H}(G, P_X)$ and $C_0(G)$. We study these two quantities and we establish the equivalence between additivity properties. To do this, we introduce the zero-error capacity $C(G, P_X)$ of a graph relative to a distribution $P_X \in \Delta(\mathcal{X})$.

Definition 9. *A sequence of codes $(\mathcal{C}_n)_n$ is said to be typical with respect to P_X , or in short P_X -typical, if*

$$\max_{x^n \in \mathcal{C}_n} \|T_{x^n} - P_X\|_\infty \xrightarrow[n \rightarrow \infty]{} 0. \quad (37)$$

The zero-error channel capacity relative to the input distribution P_X is the maximal rate among all P_X -typical sequence of codes that satisfy the zero-error property:

$$C^*(P_X) \doteq \sup_{(n, \mathcal{C}_n, \phi_d) : (\mathcal{C}_n)_n \text{ } P_X\text{-typical, } P_e^{(n)} = 0} \frac{1}{n} \log |\mathcal{C}_n|.$$

Csiszár and Körner [60, equation (3.2)] derived an upper bound on the channel capacity for arbitrarily varying channel with maximal error probability. This expression, denoted by $C(G, P_X)$, corresponds to the zero-error capacity relative to a distribution.

Lemma 5 (from [60]). *The zero-error capacity of the graph $G = (\mathcal{X}, \mathcal{E})$ relative to the distribution P_X is*

$$C^*(P_X) = C(G, P_X) \doteq \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \alpha(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]), \quad (38)$$

where $G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]$ is the subgraph of $G^{\wedge n}$ induced by the set of typical sequences $\mathcal{T}_\varepsilon^n(P_X)$ with tolerance $\varepsilon > 0$.

The superior limit in (38) can be replaced by a regular limit, thanks to the superadditivity of the sequence $(\log \alpha(G^{\wedge n}[\mathcal{T}_\varepsilon^n(P_X)]))_{n \in \mathbb{N}^*}$ and Fekete's Lemma [62, pp. 103]. In [61, Lemma 1], Marton established the relationship between the complementary graph entropy $\bar{H}(G, P_X)$ and $C(G, P_X)$.

Theorem 7 (from [61, Lemma 1]). *Given a graph $G = (\mathcal{X}, \mathcal{E})$ and a probability distribution P_X ,*

$$C(G, P_X) + \bar{H}(G, P_X) = H(X). \quad (39)$$

Equation (39) can be seen as an analog for zero-error regime of the formula $I(X; Y) + H(X|Y) = H(X)$.

Remark 2 (Zero-error source and channel duality at the code-level). *We interpret the formula in Theorem 7 in the following way. The quantities $n \cdot \bar{H}(G, P_X)$ and $n \cdot C(G, P_X)$ represent in bits, respectively, the minimum number of colors and the maximum size of an independent set. A color class, i.e. the set of vertices to which the same color is assigned, is an independent subset of vertices. In the case where all color classes have the same size, we would need $n \cdot \log \alpha(G)$ bits to describe the source sequence within its color class. Thus, $n \cdot C(G, P_X)$ can be seen as the average number of bits needed to specify the index of the source sequence in its color class. These two quantities sum up to $n \cdot H(X)$, which is the information needed to describe the source sequence with zero-error. However, this interpretation does not allow to conclude about the source and channel duality at the code level. In particular it does not guarantee that all sources are covered by a codeword, i.e. a color and an index within that color class, which is required in the source-coding problem.*

We establish below the connection between the additivity properties of $\bar{H}(G, P_X)$ and $C(G, P_X)$, where the equivalences in (41) and (43) follow from Marton's formula in Theorem 7, and the equivalence (42) from Theorem 3. The complete proof is in App. B.

Proposition 5. *Let \mathcal{A} be a finite set, P_A a distribution with full-support, and let $(G_a, P_{X_a})_{a \in \mathcal{A}}$ a family of probabilistic graphs. The following equivalences hold:*

$$\begin{aligned} C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \end{aligned} \quad (40)$$

$$\begin{aligned} \iff \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ = \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}) \end{aligned} \quad (41)$$

$$\iff \bar{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) = \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a}) \quad (42)$$

$$\iff C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}). \quad (43)$$

Proposition 5 is an important result, as it enables the passage from additivity equivalences for $\bar{H}(G, P_X)$ to those for $C(G, P_X)$, and vice versa.

D. Capacity Achieving Distributions and Equivalence of the Additivity of C_0 and \bar{H} for the AND Product

The key element of this section is the set of input distributions P_X that achieve the zero-error capacity in the sense that $C_0(G) = C(G, P_X)$. As in the vanishing error regime, it is optimal to consider codebooks composed of codewords that are typical with respect to the input distribution P_X that maximizes $C(G, P_X)$.

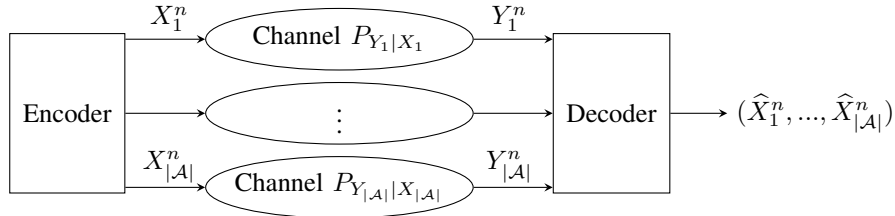


Fig. 9. The independent channel coding problem: the information is transmitted via $|\mathcal{A}|$ parallel channels $(P_{Y_a|X_a})_{a \in \mathcal{A}}$.

Lemma 6 (from [12, equation (11.2)]). For all graph $G = (\mathcal{X}, \mathcal{E})$,

$$C_0(G) = \max_{P_X} C(G, P_X). \quad (44)$$

The proof, stated in [69, Theorem 2], follows similar arguments to those in the proof of [63, Lemma 3]. As a consequence of Theorem 7 and Lemma 6, the zero-error capacity reformulates

$$C_0(G) = \max_{P_X} \left(H(X) - \bar{H}(G, P_X) \right). \quad (45)$$

In order to show the equivalence of the additivity of $C_0(G)$ and of $C(G, P_X)$, we define the set of capacity-achieving distributions.

Definition 10. Let $G = (\mathcal{X}, \mathcal{E})$ be a graph. The set of capacity-achieving distributions of G is the subset of $\Delta(\mathcal{X})$ defined by

$$\mathcal{P}^*(G) \doteq \operatorname{argmax}_{P_X} C(G, P_X). \quad (46)$$

Proposition 6. For all graphs G , the mapping $P_X \mapsto C(G, P_X)$ is concave and the set of capacity-achieving distributions $\mathcal{P}^*(G)$ is convex, nonempty.

The proof of Proposition 6 is stated in App. C, and relies on Lemma 6. The following Theorem is essential to demonstrate the equivalence of the additivity properties depicted in Fig. 3. It establishes that if a joint distribution achieves capacity, then the product of its marginals also achieves it.

Theorem 8. If $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$, then $\bigotimes_{a \in \mathcal{A}} P_{X_a} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$.

The proof of Theorem 8 is stated in App. D and relies on a *codebook shifting argument*: given a codebook composed of codewords (x_1^n, x_2^n) that are typical with respect to the joint distribution P_{X_1, X_2} , we construct a set of permuted codebooks by applying a cyclic permutation only to the first component x_1^n of each codeword. We concatenate all the permuted codebooks and we replicate them n times so that the codewords length is equal to $n' = n^3$. Then, we remove the codewords $(x_1^{n'}, x_2^{n'})$ that are not typical with respect to the product of marginal distributions $P_X \otimes P_{X'}$. We show that this construction has the same rate and preserves the zero-error property. However, it modifies the types of the codewords, which become the product of marginals as wished.

We can now establish the equivalence between the additivity of $C_0(G)$ and of $C(G, P_X)$ for the AND product.

Theorem 9. Let \mathcal{A} be a finite set, and $(G_a)_{a \in \mathcal{A}} = (\mathcal{X}_a, \mathcal{E}_a)_{a \in \mathcal{A}}$ be a family of graphs. The following equivalence holds:

$$C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right) = \sum_{a \in \mathcal{A}} C_0(G_a) \quad (47)$$

$$\iff \exists P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right),$$

$$C \left(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}} \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}). \quad (48)$$

Furthermore, any distribution $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$ that satisfies (48) also satisfies $P_{X_a} \in \mathcal{P}^*(G_a)$ for all $a \in \mathcal{A}$.

The proof of Theorem 9 is given in App. F. The main difficulty comes from the decomposition of the capacity-achieving distributions over the AND product of graphs. On the one hand, if C_0 is additive, then the product of capacity-achieving distributions $P_{X_a} \in \mathcal{P}^*(G_a)$ is optimal for the graph $\bigwedge_{a \in \mathcal{A}} G_a$ and thus, C is additive. On the other hand, if C is additive for the graph $\bigwedge_{a \in \mathcal{A}} G_a$ with the capacity-achieving distribution $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$, then C_0 is additive. These are the statements of Lemma 14 and Lemma 15 in App. F.

E. Additivity of the Sum of Independent Channels

The equivalences in Theorem 6 and Proposition 5 rely on the zero-error capacity of a disjoint union of graphs. Such graphs have an operational interpretation in terms of sum of channels which is depicted in Fig. 10, see also [5, pp. 13]. The codewords are intermingled, each letter in a codeword can be selected from any of the channels $(P_{Y_a|X_a})_{a \in \mathcal{A}}$. Since the output alphabets of each individual channel are disjoint, the channel output symbol uniquely identifies the channel that is used. This allows to embed additional information with rate $H(P_A)$.

In the vanishing error regime, the additivity of the capacity for the sum of channels holds since

$$C = \log \left(\sum_{a \in \mathcal{A}} 2^{C_a} \right), \quad (49)$$

where $C_a \doteq \max_{P_{X_a}} I(X_a; Y_a)$ is the capacity of the channel $P_{Y_a|X_a}$.

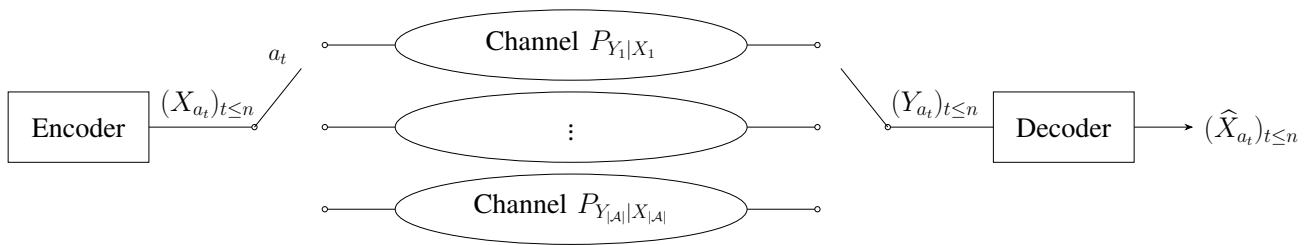


Fig. 10. Sum of the $|\mathcal{A}|$ channels $(P_{Y_a|X_a})_{a \in \mathcal{A}}$: only the channel $a_t \in \mathcal{A}$ is used at instant $t \in \{1, \dots, n\}$.

Proposition 7 (from [5]). *The zero-error capacity of the sum of channels is given by*

$$C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right). \quad (50)$$

In the zero-error regime, Shannon in [5, Theorem 4] shows that

$$\forall G, G', \quad \log \left(2^{C_0(G)} + 2^{C_0(G')} \right) \leq C_0(G \sqcup G'). \quad (51)$$

For the sum of channels, a natural coding scheme consists in using the optimal codebooks for each channel in a time sharing manner, with respect to the distribution P_A that maximizes $H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C_0(G_a)$. In other words, communicating over the sum channel is equivalent to sending 2 types of information: one related to identification of the chosen channel of rate $H(P_A)$, and the other to the information sent at rate $C_0(G_a)$ over each channel $a \in \mathcal{A}$.

Lemma 7. *The mapping $P_A \mapsto H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C_0(G_a)$ has a unique maximum*

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}, \quad (52)$$

which gives

$$H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C_0(G_a) = \log \left(\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})} \right). \quad (53)$$

The proof of Lemma 7 is given in App. E-B and relies on the fact that the function $(w_a)_{a \in \mathcal{A}} \mapsto \log \left(\sum_{a \in \mathcal{A}} 2^{w_a} \right)$ is the Legendre-Fenchel conjugate of the entropy function $P_A \mapsto H(P_A)$, see [70, Chap. 1].

We consider the time-sharing strategy between the optimal codebooks along with the distribution $P_A^* \in \Delta(\mathcal{A})$ defined in (52). If this strategy is optimal, then

$$C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right), \quad (54)$$

which means that the additivity holds for the disjoint union of graphs.

Remark 3. *Note that $P_A^* \in \Delta(\mathcal{A})$ in (52) is full-support. It can be observed that $P_A \mapsto H(P_A)$ has an infinite slope at the frontier of $\Delta(\mathcal{A})$, consequently the maximizer of $P_A \mapsto H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C_0(G_a)$ is always an interior point. In other words, the information carried by the channel index $H(P_A)$ offsets the loss in rate, if the channels with smaller*

capacities are not chosen too often. Therefore, in the sum of channels setting, always choosing the channel with highest capacity is suboptimal, and never choosing a channel is also suboptimal, even if this channel has a capacity equal to 0.

Similar to Theorem 9, we establish the equivalence between the additivity between $C_0(G)$ and $C(G, P_X)$ for the disjoint union of a family of graphs $(G_a)_{a \in \mathcal{A}}$.

Theorem 10. *The following equivalence holds*

$$\begin{aligned} C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) &= \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right) \\ \iff \exists P_X \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right), \\ C \left(\bigsqcup_{a \in \mathcal{A}} G_a, P_X \right) &= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}), \end{aligned} \quad (55)$$

where for all $a \in \mathcal{A}$, $P_A(a) = \sum_{x \in \mathcal{X}_a} P_X(x)$ and $P_{X_a} = P_{X|X \in \mathcal{X}_a} = P_X / P_A(a)$. Furthermore, any $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ that satisfies (56) also satisfies the following for all $a \in \mathcal{A}$:

$$P_A(a) = \frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}}, \text{ and } P_{X_a} \in \mathcal{P}^*(G_a). \quad (57)$$

The proof of Theorem 10 is given in App. G. The key ingredient is the decomposition of the capacity-achieving distributions over the disjoint union of graphs. The first result is Lemma 7 which proves that the distribution given by $P_A(a) = \frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}}$ maximizes $P_A \mapsto H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C_0(G_a)$. On the one hand, if C_0 is additive, then the average distribution $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ with $\forall a \in \mathcal{A}$, $P_{X_a} \in \mathcal{P}^*(G_a)$ is optimal for the graph $\bigsqcup_{a \in \mathcal{A}} G_a$ and thus, C is additive. On the other hand, if C is additive for the graph $\bigsqcup_{a \in \mathcal{A}} G_a$ and $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ is capacity-achieving, then C_0 is additive. These are the statements of Lemma 16 and Lemma 17 in App. G.

Remark 4. *One could try an alternative route to prove Theorem 10, by successively using the equivalences in Theorem 6¹,*

¹i.e. [58, Theorem 4.1] and [59, Theorem 2]

Theorem 9, and Proposition 5. However, this approach yields the following statement:

$$C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right) \quad (58)$$

$$\iff \begin{cases} \exists P_A \in \Delta(\mathcal{A}) \text{ full-support,} \\ \exists P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right), \\ C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}), \end{cases} \quad (59)$$

Nevertheless, this does not complete the proof, as we still need to relate the sets of capacity-achieving distributions $\mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right)$ required in Theorem 10 to $\mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right)$ appearing in (59).

Theorem 10 together with Theorem 9, Theorem 6 and Proposition 5, establish the equivalence of the additivity between $C_0(G)$, $C(G, P_X)$ and $\bar{H}(G, P_X)$ for the AND product \wedge , and for the disjoint union \sqcup of a family of graphs $(G_a)_{a \in \mathcal{A}}$, as depicted in Fig. 3.

IV. MAIN EXAMPLE AND COUNTER-EXAMPLES FOR THE ADDITIVITY OF OPTIMAL RATES

In this section, we exploit the equivalences in the additivity properties depicted in Fig. 3, in order to provide single-letter characterizations of \bar{H} and C_0 for several new classes of graphs.

A. Perfect Graphs

In this section, we show that perfect graphs allow for additivity of \bar{H} and C_0 with respect to both \sqcup and \wedge with any underlying probability distribution. Perfect graphs are one of the only known examples of graphs with a single-letter formula for \bar{H} and C_0 . Theorem 3 allows us to provide new single-letter characterizations for \bar{H} and C_0 for all products of perfect graphs, which are not perfect in general.

Definition 11 (Graph complement, clique number ω). For all $G = (\mathcal{X}, \mathcal{E})$, the complementary graph of G is defined by $\bar{G} \doteq (\mathcal{X}, \mathcal{E}^c)$. The clique number of G is defined by $\omega(G) \doteq \alpha(\bar{G})$, where the independence number α is stated in Definition 8.

Definition 12 (Perfect graph). A graph $G = (\mathcal{X}, \mathcal{E})$ is perfect if for all subset of vertices $S \subseteq \mathcal{X}$, $\chi(G[S]) = \omega(G[S])$. A probabilistic graph (G, P_X) is perfect if G is perfect.

A remarkable property of perfect graphs is their single-letter characterization for zero-error source and channel coding problems. For example, when G is a perfect graph, the optimal rate of the side information problem $\bar{H}(G, P_X)$ is equal to the Körner graph entropy introduced in [25], see also [21, equation (13)] and [51, Sec.II.B].

Definition 13 (Körner graph entropy H_κ). For all probabilistic graph (G, P_X) , let $\Gamma(G)$ be the collection of independent sets of vertices in G . The Körner graph entropy of (G, P_X) is defined by

$$H_\kappa(G, P_X) = \min_{X \in W \in \Gamma(G)} I(W; X), \quad (60)$$

where the minimum is taken over all distributions $P_{W|X}$ with the constraint that the random vertex X belongs to the random

independent set W with probability one, i.e. $X \in W \in \Gamma(G)$ in (60).

Theorem 11 (from [24, Corollary 12]). Let (G, P_X) be a perfect probabilistic graph, then

$$\bar{H}(G, P_X) = H_\kappa(G, P_X). \quad (61)$$

Similarly, a single-letter characterization exists for the zero-error capacity of perfect graphs, as stated below. This is a consequence of a more general result due to Shannon (see [5, Theorem 3]) that states that a graph G whose vertex set can be partitioned into $\alpha(G)$ cliques, i.e. complete induced subgraphs, satisfies $C_0(G) = \log \alpha(G)$. A perfect graph G satisfies this property as its complementary \bar{G} is also perfect, and satisfy $\chi(\bar{G}) = \omega(\bar{G}) = \alpha(G)$, where $\omega(\bar{G})$ is the clique number, see [14, pp. 382].

Theorem 12 (from [5, Theorem 3]). If G is a perfect graph, then $C_0(G) = \log \alpha(G)$.

We now provide single-letter characterizations for C_0 , C , and \bar{H} that were previously unknown. These characterizations are consequences of the additivity results of Wigderson and Zuydam [58] and Schrijver [59] for C_0 , and of Theorem 3 for \bar{H} .

More precisely, we consider some perfect graphs. Their disjoint union is perfect, as shown in Lemma 19, and C_0 is additive since $C_0(G \sqcup G') = \log \alpha(G \sqcup G') = \log(\alpha(G) + \alpha(G')) = \log(2^{C_0(G)} + 2^{C_0(G')})$ holds for all perfect graphs G, G' . According to the additivity result of [58] and [59], since C_0 is additive for the disjoint union, so does C_0 for the AND product $C_0(G \wedge G') = C_0(G) + C_0(G')$. This leads to the following proposition.

Proposition 8. Let G and G' be perfect graphs, then

$$C_0(G \sqcup G') = \log \left(2^{C_0(G)} + 2^{C_0(G')} \right) = \log(\alpha(G) + \alpha(G')), \quad (62)$$

$$C_0(G \wedge G') = C_0(G) + C_0(G') = \log \alpha(G) + \log \alpha(G'). \quad (63)$$

According to the previous proposition, $C_0(G \wedge G')$ can now be computed for any pairs of perfect graphs, as it is additive. This result was previously unknown, as the AND product of perfect graphs is not necessarily perfect. For instance, cycle graphs C_6 and C_8 are perfect, due to the strong perfect graph Theorem mentioned below, but their AND product is not. This is also due to the strong perfect graph Theorem since it contains an odd cycle of length 7, illustrated in Fig. 11.

Theorem 13 (Strong Perfect Graph Theorem, from [71, Theorem 1.2]). A graph G is perfect if and only if neither G nor \bar{G} have an induced odd cycle of length at least 5.

Similarly, we show that the additivity property of \bar{H} and C holds for perfect graphs and for all underlying probability distributions, and we provide new single-letter expression for \bar{H} and C in that case.

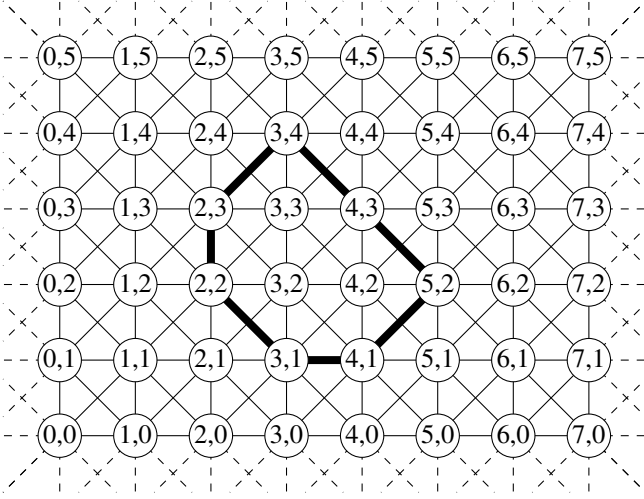


Fig. 11. A non-perfect AND product of perfect graphs: $C_6 \wedge C_8$ with an induced C_7 .

Theorem 14. When $(G_a, P_{X_a})_{a \in \mathcal{A}}$ is a family of perfect probabilistic graphs, we have the following single-letter characterizations

$$\begin{aligned} \overline{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) &= \sum_{a \in \mathcal{A}} \overline{H}(G_a, P_{X_a}) \\ &= \sum_{a \in \mathcal{A}} H_\kappa(G_a, P_{X_a}), \end{aligned} \quad (64)$$

$$\begin{aligned} \overline{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) &= \sum_{a \in \mathcal{A}} P_A(a) \overline{H}(G_a, P_{X_a}) \\ &= \sum_{a \in \mathcal{A}} P_A(a) H_\kappa(G_a, P_{X_a}), \end{aligned} \quad (65)$$

$$\begin{aligned} C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) &= \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) \\ &= \sum_{a \in \mathcal{A}} (H(X_a) - H_\kappa(G_a, P_{X_a})), \end{aligned} \quad (66)$$

$$\begin{aligned} C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) &= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \\ &= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) (H(X_a) - H_\kappa(G_a, P_{X_a})). \end{aligned} \quad (67)$$

The proof of Theorem 14 is given in App. H. As an example, we consider the AND product of the cycle graphs C_6 and C_8 .

Corollary 2. We consider the cycle graphs C_6 and C_8 and we denote by P_{X_6} and P_{X_8} the probability distributions over their vertices. We have

$$\overline{H}(C_6 \wedge C_8, P_{X_6} \otimes P_{X_8}) = H_\kappa(C_6, P_{X_6}) + H_\kappa(C_8, P_{X_8}), \quad (68)$$

$$\begin{aligned} C(C_6 \wedge C_8, P_{X_6} \otimes P_{X_8}) &= H(P_{X_6}) - H_\kappa(C_6, P_{X_6}) \\ &\quad + H(P_{X_8}) - H_\kappa(C_8, P_{X_8}). \end{aligned} \quad (69)$$

We now explore the combination of a perfect graph with a non-perfect graph. More specifically, we consider the graph $C_5 \sqcup G$ where G is perfect, for which the additivity of \overline{H} was studied by Tuncel et al. in [57]. The pentagon graph C_5 is not perfect, thereby making any disjoint union or AND product involving it non-perfect. However, by using Theorem 3, we provide a non-perfect example where the additivity property holds, offering a single-letter characterization of \overline{H} for the class of graphs $C_5 \wedge G$ when G is perfect.

Theorem 15 (from [57, Lemma 3]). Let $s \in [0, 1]$, let $(C_5, \text{Unif}(\{1, \dots, 5\}))$ and (G, P_X) be a perfect probabilistic graph, we have

$$\begin{aligned} \overline{H}(C_5 \sqcup G, s \text{Unif}(\{1, \dots, 5\}) + (1-s)P_X) \\ = s \overline{H}(C_5, \text{Unif}(\{1, \dots, 5\})) + (1-s) \overline{H}(G, P_X) \end{aligned} \quad (70)$$

$$= \frac{s}{2} \log 5 + (1-s) H_\kappa(G, P_X). \quad (71)$$

Corollary 3. For all perfect probabilistic graph (G, P_X) ,

$$\begin{aligned} \overline{H}(G \wedge C_5, P_X \otimes \text{Unif}(\{1, \dots, 5\})) \\ = \overline{H}(G, P_X) + \overline{H}(C_5, \text{Unif}(\{1, \dots, 5\})) \\ = H_\kappa(G, P_X) + \frac{1}{2} \log 5. \end{aligned} \quad (72)$$

B. Vertex Transitive Graphs

We study the class of vertex-transitive graphs, where all the vertices play the same ‘‘role’’. For these graphs, we show that the uniform distribution achieves the zero-error capacity.

Definition 14 (Vertex-transitive graph). An automorphism of a graph $G = (\mathcal{X}, \mathcal{E})$ is a bijection $\psi : \mathcal{X} \rightarrow \mathcal{X}$ such that for all $x, x' \in \mathcal{X}$, $xx' \in \mathcal{E}$ if and only if $\psi(x)\psi(x') \in \mathcal{E}$. The group of automorphisms of G is denoted by $\text{Aut}(G)$.

A graph $G = (\mathcal{X}, \mathcal{E})$ is vertex-transitive if $\text{Aut}(G)$ acts transitively on its vertices, i.e. for all $x, x' \in \mathcal{X}$, there exists $\psi \in \text{Aut}(G)$ such that $\psi(x) = x'$.

Proposition 9. If $G = (\mathcal{X}, \mathcal{E})$ is vertex-transitive, then

$$\text{Unif}(\mathcal{X}) \in \mathcal{P}^*(G). \quad (73)$$

The proof of Proposition 9 is given in App. E-A.

Corollary 4. Let $(G_a)_{a \in \mathcal{A}} = (\mathcal{X}_a, \mathcal{E}_a)_{a \in \mathcal{A}}$ be vertex-transitive graphs, their product is also vertex-transitive and

$$\text{Unif} \left(\prod_{a \in \mathcal{A}} \mathcal{X}_a \right) = \bigotimes_{a \in \mathcal{A}} \text{Unif}(\mathcal{X}_a) \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right). \quad (74)$$

C. The Schläfli Graph

We now study the important case of the Schläfli graph S as it offers a counterexample for the additivity of \overline{H} , C and C_0 . In [56], Haemers showed that the additivity property does not hold for the zero-error capacity of the product of the Schläfli graph S with its complement \overline{S} ,

$$C_0(S) + C_0(\overline{S}) < C_0(S \wedge \overline{S}). \quad (75)$$

More specifically, Haemers showed that $C_0(S) = \log_2(3)$, $C_0(\overline{S}) \leq \log_2(7)$ and $\log_2(27) \leq C_0(S \wedge \overline{S})$. In this section,

we show that a similar conclusion holds for $C(G, P_X)$ and for $\overline{H}(G, P_X)$.

According to [72, Lemma 3.7], the Schläfli graph S and its complement \overline{S} are vertex transitive, as well as their product $S \wedge \overline{S}$. By Proposition 9, the uniform distribution is capacity-achieving for S , \overline{S} , and $S \wedge \overline{S}$.

Corollary 5. *We denote by \mathcal{X}_S and $\mathcal{X}_{\overline{S}}$ the sets of vertices of the Schläfli graph S and its complement \overline{S} . Then,*

$$C(S, \text{Unif}(\mathcal{X}_S)) = C_0(S), \quad C(\overline{S}, \text{Unif}(\mathcal{X}_{\overline{S}})) = C_0(\overline{S}), \quad (76)$$

$$C(S \wedge \overline{S}, \text{Unif}(\mathcal{X}_S) \otimes \text{Unif}(\mathcal{X}_{\overline{S}})) = C_0(S \wedge \overline{S}). \quad (77)$$

In Theorem 16, we extend Haemers's results of [56] and we show that the additivity of \overline{H} and C does not hold when the distribution P_X is uniform. By using (21) in Lemma 2, we also equalize \overline{H} , and similarly C_0 , for the AND product $S \wedge \overline{S}$ and for the disjoint union $S \sqcup \overline{S}$, up to a certain constant.

Theorem 16. *Let $s \in (0, 1)$, let S be the Schläfli graph and \overline{S} its complementary, with uniform distributions on their vertices. Then,*

$$\begin{aligned} & C(S \wedge \overline{S}, \text{Unif}(\mathcal{X}_S) \otimes \text{Unif}(\mathcal{X}_{\overline{S}})) \\ & > C(S, \text{Unif}(\mathcal{X}_S)) + C(\overline{S}, \text{Unif}(\mathcal{X}_{\overline{S}})), \end{aligned} \quad (78)$$

$$\begin{aligned} & C(S \sqcup \overline{S}, s \text{Unif}(\mathcal{X}_S) + (1-s) \text{Unif}(\mathcal{X}_{\overline{S}})) \\ & > h_b(s) + sC(S, \text{Unif}(\mathcal{X}_S)) + (1-s)C(\overline{S}, \text{Unif}(\mathcal{X}_{\overline{S}})), \end{aligned} \quad (79)$$

$$\begin{aligned} & \overline{H}(S \wedge \overline{S}, \text{Unif}(\mathcal{X}_S) \otimes \text{Unif}(\mathcal{X}_{\overline{S}})) \\ & < \overline{H}(S, \text{Unif}(\mathcal{X}_S)) + \overline{H}(\overline{S}, \text{Unif}(\mathcal{X}_{\overline{S}})), \end{aligned} \quad (80)$$

$$\begin{aligned} & \overline{H}(S \sqcup \overline{S}, s \text{Unif}(\mathcal{X}_S) + (1-s) \text{Unif}(\mathcal{X}_{\overline{S}})) \\ & < s\overline{H}(S, \text{Unif}(\mathcal{X}_S)) + (1-s)\overline{H}(\overline{S}, \text{Unif}(\mathcal{X}_{\overline{S}})); \end{aligned} \quad (81)$$

where h_b is the binary entropy.

We obtain (78) from Theorem 9 and Corollary 5, equations (79) and (80) come from Proposition 5, and equation (81) comes from Theorem 3.

Remark 5. *In [73], Alon has built infinite families of graphs that satisfy $C_0(G \sqcup G') > \log(2^{C_0(G)} + 2^{C_0(G')})$. Similar results as in Theorem 16 can be derived for these graphs, by using their respective capacity-achieving distributions.*

V. CONCLUSION

We have shown the equivalences of additivity properties between C_0 , C , and \overline{H} for the AND product (\wedge) and for the disjoint union (\sqcup) of graphs, as depicted in Fig. 3. For zero-error problems with same characteristic graphs and capacity-achieving distributions, we show that it is optimal to code separately for independent channels if and only if it is optimal to code separately for independent sources with side-information. We also state the following open questions:

- As pointed out in Theorem 8, for all capacity-achieving distributions of a product graph, the product of its marginals is also capacity-achieving. Are these marginals capacity-achieving for the respective graphs in the product, and conversely, if we consider the product of

capacity-achieving distributions of graphs, is this distribution capacity-achieving for the product of graphs? In other words,

$$\mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right) \cap \bigotimes_{a \in \mathcal{A}} \Delta(\mathcal{X}_a) \stackrel{?}{=} \bigotimes_{a \in \mathcal{A}} \mathcal{P}^*(G_a). \quad (82)$$

Theorem 9 partially answer this question in the sense that inclusion holds when the additivity holds for the AND product.

- Theorem 9 and Theorem 10 show that the additivity of C_0 holds if and only if the additivity of C holds when P_X is capacity-achieving. Can we find graphs G such that the additivity of $C(G, P_X)$ holds when P_X is capacity-achieving, but does not hold for some P_X that is not capacity-achieving? A negative answer would imply that the additivity of C_0 is equivalent to the additivity of $C(G, P_X)$ and $\overline{H}(G, P_X)$ for all P_X , similarly to perfect graphs.
- In the vanishing error regime, Wyner showed in [74, Sec. VI.C] that a linear channel code can be used to construct a source code with side information, where the encoder transmits the syndrome. Similarly in Theorem 7, the index of a source sequence can be decomposed into a color class and an index within the color class. Does such a source-channel duality hold at the code level in the zero-error regime? Can we show that the decomposition into the color classes is optimal for both zero-error problems?
- Finally, we have seen in Corollary 3 that $\overline{H}(G \wedge C_5, P_X \otimes P_{X_5})$ is additive when G perfect. Is the non-additivity of \overline{H} tied to specific non-perfect induced subgraphs in each graph in the product? And if so, can we find a minimal sub-family for these graphs?

APPENDIX A

RESULTS FOR THE PROOF OF THEOREM 3

A. Proof of Lemma 2

We need several lemmas for this result. Lemma 8 establishes the distributivity of \wedge with respect to \sqcup for probabilistic graphs, similarly as in [75] for graphs without underlying distribution. Lemma 9 states that \overline{H} can be computed with subgraphs induced by sets that have an asymptotic probability one, in particular we will use it with typical sets of vertices. Lemma 10 gives the chromatic entropy of a disjoint union of isomorphic probabilistic graphs. The proofs of Lemma 8, Lemma 9 and Lemma 10 are respectively given in App. A-B, App. A-C, and Appendix A-D.

Lemma 8. *Let \mathcal{A}, \mathcal{B} be finite sets, let $P_A \in \Delta(\mathcal{A})$ and $P_B \in \Delta(\mathcal{B})$. For all $a \in \mathcal{A}$ and $b \in \mathcal{B}$, let (G_a, P_{X_a}) and (G_b, P_{X_b}) be probabilistic graphs. Then*

$$\left(\bigsqcup_{a \in \mathcal{A}} G_a \right) \wedge \left(\bigsqcup_{b \in \mathcal{B}} G_b \right) = \bigsqcup_{(a,b) \in \mathcal{A} \times \mathcal{B}} G_a \wedge G_b. \quad (83)$$

For each $(a, b) \in \mathcal{A} \times \mathcal{B}$ and $(x_a, x_b) \in \mathcal{X}_a \times \mathcal{X}_b$, the induced probability distribution takes value $P_A(a)P_B(b)P_{X_a}(x_a)P_{X_b}(x_b)$.

Lemma 9. Let a probabilistic graph (G, P_X) , and let $(\mathcal{S}_n)_{n \in \mathbb{N}^*}$ be a sequence of sets such that for all $n \in \mathbb{N}^*$, $\mathcal{S}_n \subseteq \mathcal{X}^n$, and $P_X^{\otimes n}(\mathcal{S}_n) \rightarrow 1$ when $n \rightarrow \infty$. Then $\overline{H}(G, P_X) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n}[\mathcal{S}_n], P'_{X^n})$ where $P'_{X^n}(x^n) = P_X^{\otimes n}(x^n)/P_X^{\otimes n}(\mathcal{S}_n)$ for all $x^n \in \mathcal{S}_n$, and $P'_{X^n}(x^n) = 0$ otherwise.

Definition 15 (Isomorphic probabilistic graphs). Let (G_1, P_{X_1}) and (G_2, P_{X_2}) be two probabilistic graphs with $G_1 = (\mathcal{X}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{X}_2, \mathcal{E}_2)$. We say that (G_1, P_{X_1}) is isomorphic to (G_2, P_{X_2}) (denoted by $(G_1, P_{X_1}) \simeq (G_2, P_{X_2})$) if there exists an isomorphism between them, i.e. a bijection $\psi : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ such that: label = -

- For all $x_1, x'_1 \in \mathcal{X}_1$, $x_1 x'_1 \in \mathcal{E}_1 \iff \psi(x_1) \psi(x'_1) \in \mathcal{E}_2$,
- For all $x_1 \in \mathcal{X}_1$, $P_{X_1}(x_1) = P_{X_2}(\psi(x_1))$.

Lemma 10. Let \mathcal{B} be a finite set, let $P_B \in \Delta(\mathcal{B})$ and let $(G_b, P_{X_b})_{b \in \mathcal{B}}$ be a family of isomorphic probabilistic graphs, then $H_\chi(\bigsqcup_{b' \in \mathcal{B}} G_{b'}, \sum_{b' \in \mathcal{B}} P_B(b') P_{X_{b'}}) = H_\chi(G_b, P_{X_b})$ for all $b \in \mathcal{B}$.

Now let us prove Lemma 2. Let $P_A \in \Delta(\mathcal{A})$, and let $(G, P_X) = (\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a})$, by definition we have

$$\begin{aligned} & \overline{H}(G, P_X) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi \left(\left(\bigsqcup_{a \in \mathcal{A}} G_a \right)^{\wedge n}, \left(\sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right)^n \right) \quad (84) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi \left(\bigsqcup_{a^n \in \mathcal{A}^n} \bigwedge_{t \leq n} G_{a_t}, \sum_{a^n \in \mathcal{A}^n} P_A^{\otimes n}(a^n) \bigotimes_{t \leq n} P_{X_{a_t}} \right), \quad (85) \end{aligned}$$

where (85) is a consequence of Lemma 8.

We focus our attention on typical sequences, we denote by T_{a^n} the type of the sequence a^n . Let $\varepsilon > 0$, and for all $n \in \mathbb{N}^*$, we define

$$\begin{aligned} \mathcal{T}_\varepsilon^n(P_A) &\doteq \{a^n \in \mathcal{A}^n \mid \|T_{a^n} - P_A\|_\infty \leq \varepsilon\}, \quad (86) \\ \mathcal{S}_{n,\varepsilon} &\doteq \bigsqcup_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} \prod_{t \leq n} \mathcal{X}_{a_t}. \end{aligned}$$

We introduce the conditional distribution of A^n given $A^n \in \mathcal{T}_\varepsilon^n(P_A)$ by

$$P'_{A^n}(a^n) \doteq \begin{cases} \frac{P_A^{\otimes n}(a^n)}{P_A^{\otimes n}(\mathcal{T}_\varepsilon^n(P_A))}, & \text{if } a^n \in \mathcal{T}_\varepsilon^n(P_A), \\ 0 & \text{otherwise,} \end{cases} \quad (87)$$

and the distribution P'_{X^n} induced over the subgraph $G^{\wedge n}[\mathcal{S}_{n,\varepsilon}]$, defined by

$$P'_{X^n} = \sum_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} P'_{A^n}(a^n) \bigotimes_{t \leq n} P_{X_{a_t}}. \quad (88)$$

Since for all $\varepsilon > 0$, $P_X^{\otimes n}(\mathcal{S}_{n,\varepsilon}) \rightarrow 1$ when $n \rightarrow \infty$, we have by Lemma 9

$$\overline{H}(G, P_X) = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n}[\mathcal{S}_{n,\varepsilon}], P'_{X^n}). \quad (89)$$

Let us study the limit in (89). Let $(\bar{a}_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$ be a sequence such that $\bar{a}^n \in \mathcal{T}_\varepsilon^n(P_A)$ as $T_{\bar{a}^n} \rightarrow P_A$. Therefore, for all $a^n \in \mathcal{T}_\varepsilon^n(P_A)$, $a \in \mathcal{A}$, and n large enough, we have

$$|T_{\bar{a}^n}(a) - T_{a^n}(a)| \leq 2\varepsilon. \quad (90)$$

We have on one hand

$$\begin{aligned} & H_\chi(G^{\wedge n}[\mathcal{S}_{n,\varepsilon}], P'_{X^n}) \\ &= H_\chi \left(\left(\bigsqcup_{a^n \in \mathcal{A}^n} \bigwedge_{t \leq n} G_{a_t} \right) [\mathcal{S}_{n,\varepsilon}], \right. \\ & \quad \left. \sum_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} P'_{A^n}(a^n) \bigotimes_{t \leq n} P_{X_{a_t}} \right) \quad (91) \end{aligned}$$

$$\begin{aligned} &= H_\chi \left(\bigsqcup_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} \bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{a^n}(a)}, \right. \\ & \quad \left. \sum_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} P'_{A^n}(a^n) \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{a^n}(a)} \right) \quad (92) \end{aligned}$$

$$\begin{aligned} &\leq H_\chi \left(\bigsqcup_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} \bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a) + \lceil 2n\varepsilon \rceil}, \right. \\ & \quad \left. \sum_{a^n \in \mathcal{T}_\varepsilon^n(P_A)} P'_{A^n}(a^n) \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a) + \lceil 2n\varepsilon \rceil} \right) \quad (93) \end{aligned}$$

$$= H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a) + \lceil 2n\varepsilon \rceil}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a) + \lceil 2n\varepsilon \rceil} \right) \quad (94)$$

$$\begin{aligned} &\leq H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)} \right) \\ & \quad + H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge \lceil 2n\varepsilon \rceil}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes \lceil 2n\varepsilon \rceil} \right) \quad (95) \end{aligned}$$

$$\leq H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)} \right) + \lceil 2n\varepsilon \rceil |\mathcal{A}| \log |\mathcal{X}|, \quad (96)$$

where (91) comes from Lemma 8; (92) comes from the definition of $\mathcal{S}_{n,\varepsilon}$ in (86) and a rearrangement of the terms inside the product; (93) comes from (90) and the fact that $H_\chi(G_1, P_{X_1}) \leq H_\chi(G_1 \wedge G_2, P_{X_1} \otimes P_{X_2})$ for all pair of probabilistic graphs (G_1, P_{X_1}) and (G_2, P_{X_2}) ; (94) follows from Lemma 10, the graphs $(\bigwedge_{a' \in \mathcal{A}} G_{a'}^{\wedge n T_{\bar{a}^n}(a) + \lceil 2n\varepsilon \rceil})_{a^n \in \mathcal{T}_\varepsilon^n(P_A)}$ are isomorphic as they do not depend on a^n ; (95) follows from the subadditivity of H_χ ; and (96) is the upper bound on H_χ given by the highest entropy of a coloring.

On the other hand, we obtain by using similar arguments

$$\begin{aligned}
& H_\chi(G^{\wedge n}[\mathcal{S}_{n,\varepsilon}], P'_{X_n}) \\
& \geq H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a) - \lfloor 2n\varepsilon \rfloor}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a) - \lfloor 2n\varepsilon \rfloor}\right) \quad (97) \\
& \geq H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)}\right) \\
& \quad - H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge \lfloor 2n\varepsilon \rfloor}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes \lfloor 2n\varepsilon \rfloor}\right) \quad (98) \\
& \geq H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)}\right) - \lfloor 2n\varepsilon \rfloor |\mathcal{A}| \log |\mathcal{X}|. \quad (99)
\end{aligned}$$

Note that (98) also comes from the subadditivity of H_χ , as $H_\chi(G_2, P_{X_2}) \geq H_\chi(G_1 \wedge G_2, P_{X_1} \otimes P_{X_2}) - H_\chi(G_1, P_{X_1})$ for all (G_1, P_{X_1}) and (G_2, P_{X_2}) .

By combining (96) and (99) we obtain

$$\begin{aligned}
& \left| \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi(G^{\wedge n}[\mathcal{S}_{n,\varepsilon}], P'_{X_n}) - \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes n T_{\bar{a}^n}(a)}\right) \right| \\
& \leq 2\varepsilon |\mathcal{A}| \log |\mathcal{X}|. \quad (100)
\end{aligned}$$

As this holds for all $\varepsilon > 0$, combining (89) and (100) yields the result of (19).

Now let us prove (20). We assume that P_A is a type. Let $(\bar{a}_n)_{n \in \mathbb{N}^+}$ be a k -periodic sequence such that $T_{\bar{a}^k} = P_A$, then $T_{\bar{a}^{nk}} = T_{\bar{a}^k}$ for all $n \in \mathbb{N}^*$, and $T_{\bar{a}^n} \xrightarrow{n \rightarrow \infty} P_A$. We can use Lemma 2 and consider every k -th term in the limit:

$$\begin{aligned}
& \bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, P_X\right) \\
& = \lim_{n \rightarrow \infty} \frac{1}{kn} H_\chi\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge kn T_{\bar{a}^{kn}}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes kn T_{\bar{a}^{kn}}(a)}\right) \quad (101)
\end{aligned}$$

$$= \lim_{n \rightarrow \infty} \frac{1}{kn} H_\chi\left(\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge k T_{\bar{a}^k}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes k T_{\bar{a}^k}(a)}\right)^{\wedge n}\right) \quad (102)$$

$$= \frac{1}{k} \bar{H}\left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge k P_A(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\otimes k T_{\bar{a}^k}(a)}\right). \quad (103)$$

B. Proof of Lemma 8

The probabilistic graphs in both sides of (83) have

$$\left(\bigsqcup_{a \in \mathcal{A}} \mathcal{X}_a\right) \times \left(\bigsqcup_{b \in \mathcal{B}} \mathcal{X}_b\right) = \bigsqcup_{(a,b) \in \mathcal{A} \times \mathcal{B}} \mathcal{X}_a \times \mathcal{X}_b \quad (104)$$

as set of vertices, with underlying distribution

$$\begin{aligned}
& \left(\sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right) \left(\sum_{b \in \mathcal{B}} P_B(b) P_{X_b}\right) \\
& = \sum_{(a,b) \in \mathcal{A} \times \mathcal{B}} P_A(a) P_B(b) P_{X_a} P_{X_b}. \quad (105)
\end{aligned}$$

Now let us show that these two graphs have the same edges. Let $(x_A, x_B), (x'_A, x'_B) \in \left(\bigsqcup_{a \in \mathcal{A}} \mathcal{X}_a\right) \times \left(\bigsqcup_{b \in \mathcal{B}} \mathcal{X}_b\right)$, let $a_*, a'_* \in \mathcal{A}$ and $b_*, b'_* \in \mathcal{B}$ be the unique indexes such that

$$(x_A, x_B) \in \mathcal{X}_{a_*} \times \mathcal{X}_{b_*} \quad \text{and} \quad (x'_A, x'_B) \in \mathcal{X}_{a'_*} \times \mathcal{X}_{b'_*}. \quad (106)$$

We have:

$$(x_A, x_B), (x'_A, x'_B) \text{ adjacent in } \left(\bigsqcup_{a \in \mathcal{A}} G_a\right) \wedge \left(\bigsqcup_{b \in \mathcal{B}} G_b\right) \quad (107)$$

$$\iff x_A, x'_A \text{ adjacent in } \bigsqcup_{a \in \mathcal{A}} G_a \quad \text{and} \quad x_B, x'_B \text{ adjacent in } \bigsqcup_{b \in \mathcal{B}} G_b \quad (108)$$

$$\iff a_* = a'_* \text{ and } x_A x'_A \in \mathcal{E}_{a_*} \text{ and } b_* = b'_* \text{ and } x_B x'_B \in \mathcal{E}_{b_*} \quad (109)$$

$$\iff (a_*, b_*) = (a'_*, b'_*) \text{ and } (x_A, x_B), (x'_A, x'_B) \text{ are adjacent in } G_{a_*} \wedge G_{b_*} \quad (110)$$

$$\iff (x_A, x_B), (x'_A, x'_B) \text{ are adjacent in } \bigsqcup_{(a,b) \in \mathcal{A} \times \mathcal{B}} G_a \wedge G_b. \quad (111)$$

C. Proof of Lemma 9

In order to prove Lemma 9, we need Lemma 11 which gives upper and lower bounds on the chromatic entropy of an induced subgraph $G[\mathcal{S}]$, using the chromatic entropy of the whole graph G and the probability $P_X(\mathcal{S})$. The core idea is that if $P_X(\mathcal{S})$ is close to 1 and $H_\chi(G)$ is large, then $H_\chi(G[\mathcal{S}])$ is close to $H_\chi(G)$. The proof of Lemma 11 is given in App. A-E

Lemma 11. *Let a probabilistic graph (G, P_X) and $\mathcal{S} \subseteq \mathcal{X}$, then*

$$\begin{aligned}
& H_\chi(G, P_X) - 1 - (1 - P_X(\mathcal{S})) \log |\mathcal{X}| \\
& \leq H_\chi(G[\mathcal{S}], P_X/P_X(\mathcal{S})) \leq \frac{H_\chi(G, P_X)}{P_X(\mathcal{S})}. \quad (112)
\end{aligned}$$

Remark 6. $H_\chi(G[\mathcal{S}], P_X/P_X(\mathcal{S}))$ can be greater than $H_\chi(G, P_X)$, even if $G[\mathcal{S}]$ has less vertices and inherits the structure of G . This stems from the normalized distribution $P_X/P_X(\mathcal{S})$ on the vertices of $G[\mathcal{S}]$ which gives more weight to the vertices in \mathcal{S} . For example, consider

$$G = N_5 \sqcup K_5,$$

where K_n (resp. N_n) is the complete (resp. empty) graph with n vertices, i.e. there is an edge (resp. no edge) between any pair of distinct vertices, and with \mathcal{S} being the vertices in the connected component K_5 in G . Then $H_\chi(G, \text{Unif}(\{1, \dots, 5\})) = \varepsilon \log 5$ and $H_\chi(G[\mathcal{S}], \text{Unif}(\{1, \dots, 5\})/\varepsilon) = \log 5$.

Now let us prove Lemma 9. By Lemma 11, we have for all $n \in \mathbb{N}^*$:

$$\begin{aligned}
& H_\chi(G^{\wedge n}, P_X^n) - 1 - (1 - P_X^{\otimes n}(\mathcal{S}_n)) \log |\mathcal{X}| \\
& \leq H_\chi(G^{\wedge n}[\mathcal{S}^n], P_X^n/P_X^{\otimes n}(\mathcal{S}_n)) \leq \frac{H_\chi(G^{\wedge n}, P_X^n)}{P_X^{\otimes n}(\mathcal{S}_n)}. \quad (113)
\end{aligned}$$

Since $P_X^{\otimes n}(\mathcal{S}_n) \rightarrow 1$, and $H_\chi(G^{\wedge n}, P_X^n) = n \bar{H}(G, P_X) + o(n)$ when $n \rightarrow \infty$, the desired results follows immediately by normalization and limit.

D. Proof of Lemma 10

Let $(\tilde{G}_i, P_i)_{i \leq N}$ be isomorphic probabilistic graphs and G such that $G = \bigsqcup_i \tilde{G}_i$. Let $c_1^* : \mathcal{X}_1 \rightarrow \mathcal{C}$ be the coloring of \tilde{G}_1

with minimal entropy, and let c^* be the coloring of G defined by

$$c^* : \mathcal{X} \rightarrow \mathcal{C} \quad (114)$$

$$x \mapsto c_1^* \circ \psi_{i_x \rightarrow 1}(x), \quad (115)$$

where i_x is the unique integer such that $x \in \mathcal{X}_{i_x}$, and $\psi_{i_x \rightarrow 1} : \mathcal{X}_{i_x} \rightarrow \mathcal{X}_1$ is an isomorphism between \tilde{G}_{i_x} and \tilde{G}_1 . In other words c^* applies the same coloring pattern c_1^* on each connected component of G . We have

$$H_X(G, P_X) \leq H(c^*(X)) \quad (116)$$

$$= h\left(\sum_{j \leq N} P_{i_X}(j) P_{c^*(X_j)}\right) \quad (117)$$

$$= h\left(\sum_{j \leq N} P_{i_X}(j) P_{c_1^*(X_1)}\right) \quad (118)$$

$$= H(c_1^*(X_1)) \quad (119)$$

$$= H_X(\tilde{G}_1), \quad (120)$$

where h denotes the entropy of a distribution; (118) comes from the definition of c^* ; and (120) comes from the definition of c_1^* .

Now let us prove the upper bound on $H_X(\tilde{G}_1)$. Let c be a coloring of G , and let $i^* \doteq \operatorname{argmin}_i H(c(X_i))$ (i.e. i^* is the index of the connected component for which the entropy of the coloring induced by c is minimal). We have

$$H(c(X)) = h\left(\sum_{j \leq N} P_{i_X}(j) P_{c(X_j)}\right) \quad (121)$$

$$\geq \sum_{j \leq N} P_{i_X}(j) h(P_{c(X_j)}) \quad (122)$$

$$\geq \sum_{j \leq N} P_{i_X}(j) H(c(X_{i^*})) \quad (123)$$

$$\geq H_X(\tilde{G}_{i^*}), \quad (124)$$

$$= H_X(\tilde{G}_1, P_{X_1}), \quad (125)$$

where (122) follows from the concavity of h ; (123) follows from the definition of i^* ; (124) comes from the fact that c induces a coloring of \tilde{G}_{i^*} ; (125) comes from the fact that \tilde{G}_1 and \tilde{G}_{i^*} are isomorphic. Now, we can combine the bounds (120) and (125): for all coloring c of G we have

$$H_X(G, P_X) \leq H_X(\tilde{G}_1, P_{X_1}) \leq H(c(X)), \quad (126)$$

which yields the desired equality when taking the infimum over c .

E. Proof of Lemma 11

Let $c^* : \mathcal{X} \rightarrow \mathcal{C}$ and $c_S^* : \mathcal{S} \rightarrow \mathcal{C}$ be the optimal colorings of G and $G[\mathcal{S}]$, respectively. Consider the coloring $c : \mathcal{X} \rightarrow \mathcal{C} \sqcup \mathcal{X}$ of G defined by $c(x) = c_S^*$ if $x \in \mathcal{S}$, $c(x) = x$ otherwise.

(Lower bound) On one hand, we have

$$\begin{aligned} & H_X(G, P_X) \\ & \leq H(c(X), \mathbb{1}_{X \in \mathcal{S}}) \end{aligned} \quad (127)$$

$$\begin{aligned} & = H(\mathbb{1}_{X \in \mathcal{S}}) + P_X(\mathcal{S}) H(c(X) | X \in \mathcal{S}) \\ & \quad + (1 - P_X(\mathcal{S})) H(c(X) | X \notin \mathcal{S}) \end{aligned} \quad (128)$$

$$\leq 1 + H(c_S^*(X) | X \in \mathcal{S}) + (1 - P_X(\mathcal{S})) \log |\mathcal{X}| \quad (129)$$

$$= H_X(G[\mathcal{S}], P_X/P_X(\mathcal{S})) + 1 + (1 - P_X(\mathcal{S})) \log |\mathcal{X}|; \quad (130)$$

where (127) comes from the fact that c is a coloring of G ; (128) is a decomposition using conditional entropies; (129) comes from the construction of c : $c|_{\mathcal{S}} = c_S^*$; (130) follows from the optimality of c_S^* as a coloring of $G[\mathcal{S}]$.

(Upper bound) On the other hand,

$$\begin{aligned} & H_X(G[\mathcal{S}], P_X/P_X(\mathcal{S})) \\ & \leq H(c^*(X) | X \in \mathcal{S}) \end{aligned} \quad (131)$$

$$\begin{aligned} & = \frac{1}{P_X(\mathcal{S})} \left(H(c^*(X) | \mathbb{1}_{X \in \mathcal{S}}) \right. \\ & \quad \left. - (1 - P_X(\mathcal{S})) H(c^*(X) | X \notin \mathcal{S}) \right) \end{aligned} \quad (132)$$

$$\leq \frac{H(c^*(X))}{P_X(\mathcal{S})} = \frac{H_X(G, P_X)}{P_X(\mathcal{S})} \quad (133)$$

where (131) comes from the fact that c^* induces a coloring of $G[\mathcal{S}]$; (132) is a decomposition using conditional entropies; (133) results from the elimination of negative terms and the optimality of c^* .

F. Proof of Lemma 3

In order to prove Lemma 3, we need Lemma 2 which can be found in App. A-A; and Lemma 12, which is a generalization for infinite sequences of the following observation: if $T_{\bar{a}^n} = P_A \in \Delta_n(\mathcal{A})$ satisfies $P_A = \frac{i}{n} P'_A + \frac{n-i}{n} P''_A$ with $P'_A \in \Delta_i(\mathcal{A})$ and $P''_A \in \Delta_{n-i}(\mathcal{A})$, then \bar{a}^n can be separated into two subsequences a'^i and a''^{n-i} such that $T_{a'^i} = P'_A$ and $T_{a''^{n-i}} = P''_A$.

Lemma 12 (Type-splitting lemma). *Let $(\bar{a}_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$ be a sequence such that $T_{\bar{a}^n} \rightarrow P_A \in \Delta(\mathcal{A})$ when $n \rightarrow \infty$, let $\beta \in (0, 1)$ and $P'_A, P''_A \in \Delta(\mathcal{A})$ such that*

$$P_A = \beta P'_A + (1 - \beta) P''_A. \quad (134)$$

Then there exists a sequence $(b_n)_{n \in \mathbb{N}^} \in \{0, 1\}^{\mathbb{N}^*}$ such that the two extracted sequences $a' \doteq (\bar{a}_n)_{\substack{n \in \mathbb{N}^* \\ b_n = 0}}$ and $a'' \doteq (\bar{a}_n)_{\substack{n \in \mathbb{N}^* \\ b_n = 1}}$, satisfy*

$$T_{b_n} \xrightarrow{n \rightarrow \infty} (\beta, 1 - \beta), \quad (135)$$

$$T_{a'^n} \xrightarrow{n \rightarrow \infty} P'_A, \quad T_{a''^n} \xrightarrow{n \rightarrow \infty} P''_A. \quad (136)$$

The proof of Lemma 12 is given in App. A-G. Now let us prove Lemma 3. We recall the definition of the function

$$\eta : P_A \mapsto \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right). \quad (137)$$

(η Lipschitz) Let us first prove that the function η is Lipschitz. For all $P_A, P'_A \in \Delta(\mathcal{A})$ we need to bound the quantity $|\eta(P_A) - \eta(P'_A)|$; by Lemma 2 this is equivalent to bounding

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \left| H_X \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}^n}(a)} \right) \right. \\ & \quad \left. - H_X \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}'^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}'^n}(a)} \right) \right| \end{aligned} \quad (138)$$

where $(T_{\bar{a}^n}, T_{\bar{a}'^n}) \rightarrow (P_A, P'_A)$ when $n \rightarrow \infty$.

Fix $n \in \mathbb{N}^*$, we assume that the quantity inside $|\cdot|$ in (138) is positive; the other case can be treated with the same arguments by symmetry of the roles. We have

$$\begin{aligned} & H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}^n}(a)} \right) \\ & - H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}'^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}'^n}(a)} \right) \end{aligned} \quad (139)$$

$$\begin{aligned} & \leq H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}^n}(a)} \right) \\ & - H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n \min(T_{\bar{a}^n}(a), T_{\bar{a}'^n}(a))}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n \min(T_{\bar{a}^n}(a), T_{\bar{a}'^n}(a))} \right) \end{aligned} \quad (140)$$

$$\begin{aligned} & = H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n \min(T_{\bar{a}^n}(a), T_{\bar{a}'^n}(a))} \right. \\ & \quad \left. \bigwedge_{a \in \mathcal{A}} G_a^{\wedge n |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|_+}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}^n}(a)} \right) \\ & - H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n \min(T_{\bar{a}^n}(a), T_{\bar{a}'^n}(a))}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n \min(T_{\bar{a}^n}(a), T_{\bar{a}'^n}(a))} \right) \end{aligned} \quad (141)$$

$$\begin{aligned} & \leq H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|_+}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|_+} \right) \end{aligned} \quad (142)$$

$$\leq \log \left(\max_a |\mathcal{X}_a| \right) \sum_{a \in \mathcal{A}} n |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|_+ \quad (143)$$

$$\leq n \log \left(\max_a |\mathcal{X}_a| \right) \|T_{\bar{a}^n} - T_{\bar{a}'^n}\|_1, \quad (144)$$

where $|\cdot|_+ = \max(\cdot, 0)$ and $\|T_{\bar{a}^n} - T_{\bar{a}'^n}\|_1 = \sum_{a \in \mathcal{A}} |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|$; (140) follows from the removal of terms in the second product, as $H_\chi(G \wedge G', P_X \otimes P'_X) \geq H_\chi(G, P_X)$ for all probabilistic graphs (G, P_X) and (G', P'_X) ; (141) is an arrangement of the terms in the first product, as $\min(s, t) + \max(s - t, 0) = s$ for all real numbers s, t ; (142) comes from the subadditivity of H_χ ; (143) follows from $H_\chi(G_a, P_{X_a}) \leq \log \max_{a'} |\mathcal{X}_{a'}|$ for all $a \in \mathcal{A}$; (144) results from $|T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|_+ \leq |T_{\bar{a}^n}(a) - T_{\bar{a}'^n}(a)|$ for all $a \in \mathcal{A}$.

By normalization and limit, it follows that

$$|\eta(P_A) - \eta(P'_A)| \leq \lim_{n \rightarrow \infty} \log \left(\max_a |\mathcal{X}_a| \right) \cdot \|T_{\bar{a}^n} - T_{\bar{a}'^n}\|_1 \quad (145)$$

$$= \log \left(\max_a |\mathcal{X}_a| \right) \cdot \|P_A - P'_A\|_1. \quad (146)$$

Hence η is $(\log \max_a |\mathcal{X}_a|)$ -Lipschitz.

(η convex) Let us now prove that η is convex. Let $P'_A, P''_A \in \Delta(\mathcal{A})$, and $\beta \in (0, 1)$, we have by Lemma 2

$$\begin{aligned} & \eta(\beta P'_A + (1 - \beta)P''_A) \\ & = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge n T_{\bar{a}^n}(a)}, \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n T_{\bar{a}^n}(a)} \right), \end{aligned} \quad (147)$$

where $T_{\bar{a}^n} \rightarrow \beta P'_A + (1 - \beta)P''_A$ when $n \rightarrow \infty$. By Lemma 12, there exists $(b_n)_{n \in \mathbb{N}^*} \in \{0, 1\}^{\mathbb{N}^*}$ such that the decomposition of $(\bar{a}_n)_{n \in \mathbb{N}^*}$ into two subsequences $a' \doteq (\bar{a}_n)_{n \in \mathbb{N}^*, b_n=0}$, and $a'' \doteq (\bar{a}_n)_{n \in \mathbb{N}^*, b_n=1}$, satisfies

$$T_{b^n} \xrightarrow{n \rightarrow \infty} (\beta, 1 - \beta), \quad (148)$$

$$T_{a'^n} \xrightarrow{n \rightarrow \infty} P'_A, \quad T_{a''^n} \xrightarrow{n \rightarrow \infty} P''_A. \quad (149)$$

For all $n \in \mathbb{N}^*$, let $\Xi(n) \doteq n T_{b^n}(0)$, we have

$$\begin{aligned} & \eta(\beta P'_A + (1 - \beta)P''_A) \\ & = \lim_{n \rightarrow \infty} \frac{1}{n} H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge \Xi(n) T_{a'^{\Xi(n)}}(a) + (n - \Xi(n)) T_{a''^{n - \Xi(n)}}(a)}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{n \Xi(n) T_{a'^{\Xi(n)}}(a) + (n - \Xi(n)) T_{a''^{n - \Xi(n)}}(a)} \right) \end{aligned} \quad (150)$$

$$\begin{aligned} & \leq \lim_{n \rightarrow \infty} \frac{\Xi(n)}{n} \frac{1}{\Xi(n)} H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge \Xi(n) T_{a'^{\Xi(n)}}(a)}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{\Xi(n) T_{a'^{\Xi(n)}}(a)} \right) \end{aligned} \quad (151)$$

$$\begin{aligned} & + \frac{n - \Xi(n)}{n} \frac{1}{n - \Xi(n)} H_\chi \left(\bigwedge_{a \in \mathcal{A}} G_a^{\wedge (n - \Xi(n)) T_{a''^{n - \Xi(n)}}(a)}, \right. \\ & \quad \left. \bigotimes_{a \in \mathcal{A}} P_{X_a}^{(n - \Xi(n)) T_{a''^{n - \Xi(n)}}(a)} \right) \end{aligned} \quad (152)$$

$$= \beta \eta(P'_A) + (1 - \beta) \eta(P''_A); \quad (153)$$

where (151) comes from (147); (153) follows from the subadditivity of H_χ ; (154) comes from (148), (149) and Lemma 2. Since (154) holds for all $P'_A, P''_A \in \Delta(\mathcal{A})$ and $\beta \in (0, 1)$, we have that η is convex.

G. Proof of Lemma 12

Let $(\bar{a}_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$ be a sequence such that $T_{\bar{a}^n} \rightarrow P_A = \beta P'_A + (1 - \beta)P''_A$ when $n \rightarrow \infty$.

Consider a sequence $(B_n)_{n \in \mathbb{N}^*}$ of independent Bernoulli random variables such that for all $n \in \mathbb{N}^*$,

$$\mathbb{P}(B_n = 0) = \frac{\beta P'_A(\bar{a}_n)}{P_A(\bar{a}_n)}. \quad (155)$$

By the strong law of large numbers,

$$\mathbb{P} \left(T_{B^n, \bar{a}^n} \xrightarrow{n \rightarrow \infty} (\beta P'_A, (1 - \beta)P''_A) \right) = 1. \quad (156)$$

Therefore, there exists at least one realization $(b_n)_{n \in \mathbb{N}^*}$ of $(B_n)_{n \in \mathbb{N}^*}$ such that T_{b^n, \bar{a}^n} converges to $(\beta P'_A, (1 - \beta)P''_A)$. The convergences of marginal and conditional types yield

$$T_{b^n} \xrightarrow{n \rightarrow \infty} (\beta, 1 - \beta), \quad (157)$$

$$T_{a'^n} \xrightarrow{n \rightarrow \infty} P'_A, \quad T_{a''^n} \xrightarrow{n \rightarrow \infty} P''_A, \quad (158)$$

where $a' \doteq (\bar{a}_n)_{n \in \mathbb{N}^*, b_n=0}$, and $a'' \doteq (\bar{a}_n)_{n \in \mathbb{N}^*, b_n=1}$, are the extracted sequences.

H. Proof of Lemma 4

It can be easily observed that

$$\exists P_A \in \text{int}(\Delta(\mathcal{A})), \gamma(P_A) = \sum_{a \in \mathcal{A}} P_A(a) \gamma(\mathbb{1}_a) \quad (159)$$

$$\iff \forall P_A \in \Delta(\mathcal{A}), \gamma(P_A) = \sum_{a \in \mathcal{A}} P_A(a) \gamma(\mathbb{1}_a). \quad (160)$$

Now let us prove (159) \Rightarrow (160). Let $P_A^* \in \text{int} \Delta(\mathcal{A})$ such that $\gamma(P_A^*) = \sum_{a \in \mathcal{A}} P_A^*(a) \gamma(\mathbb{1}_a)$. Let $m : \Delta(\mathcal{A}) \rightarrow \mathbb{R}$ linear such that $m(P_A^*) = \gamma(P_A^*)$ and $\forall P_A \in \Delta(\mathcal{A}), m(P_A) \leq \gamma(P_A)$. We have

$$0 = \gamma(P_A^*) - m(P_A^*) = \sum_{a \in \mathcal{A}} P_A^*(a) (\gamma(\mathbb{1}_a) - m(\mathbb{1}_a)); \quad (161)$$

and therefore $\gamma(\mathbb{1}_a) = m(\mathbb{1}_a)$ for all $a \in \mathcal{A}$, as $\gamma - m \geq 0$ and $P_A^*(a) > 0$ for all $a \in \mathcal{A}$. For all $P_A \in \Delta(\mathcal{A})$, we have

$$f(P_A) \leq \sum_{a \in \mathcal{A}} P_A(a) \gamma(\mathbb{1}_a) \quad (162)$$

$$= \sum_{a \in \mathcal{A}} P_A(a) m(\mathbb{1}_a) = m(P_A), \quad (163)$$

hence $\gamma = m$ and γ is linear.

APPENDIX B PROOF OF PROPOSITION 5

In order to prove Proposition 5, we need Lemma 13, which is a consequence of Marton's formula in Theorem 7 applied to a disjoint union. The proof of Lemma 13 can be found in App. B-A.

Lemma 13. *Let $P_A \in \Delta(\mathcal{A})$, then*

$$\begin{aligned} & \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & + C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) H(P_{X_a}). \end{aligned} \quad (164)$$

Let us prove Proposition 5. We have on one hand:

$$H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \quad (165)$$

$$\begin{aligned} & = H(P_A) - \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}) \\ & + \sum_{a \in \mathcal{A}} P_A(a) H(P_{X_a}) \end{aligned} \quad (166)$$

$$\begin{aligned} & \leq H(P_A) - \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & + \sum_{a \in \mathcal{A}} P_A(a) H(P_{X_a}) \end{aligned} \quad (167)$$

$$= C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right); \quad (168)$$

where (166) comes from Theorem 7; (167) follows from (16), see [57, Theorem 2]; and (168) follows from Lemma 13. Therefore,

$$\begin{aligned} & C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \end{aligned} \quad (169)$$

$$\begin{aligned} & \iff \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & = \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}). \end{aligned} \quad (170)$$

On the other hand:

$$\begin{aligned} & \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) \\ & = - \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a}) + H(P_{X_a}) \end{aligned} \quad (171)$$

$$\leq - \bar{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) + \sum_{a \in \mathcal{A}} H(P_{X_a}) \quad (172)$$

$$= - \bar{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) + H \left(\bigotimes_{a \in \mathcal{A}} P_{X_a} \right) \quad (173)$$

$$= C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right); \quad (174)$$

where (171) comes from Theorem 7; (172) follows from (15), see [57, Theorem 2]; and (174) also follows from Theorem 7. Therefore,

$$\sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) = C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) \quad (175)$$

$$\iff \bar{H} \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a} \right) = \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a}). \quad (176)$$

A. Proof of Lemma 13

The graph $\bigsqcup_{a \in \mathcal{A}} G_a$ has $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ as underlying distribution. Let A, X be two random variables such that A is drawn with P_A , and X is drawn with $P_{X|A}(\cdot|a) \doteq P_{X_a}$, so that

$$P_X = \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}. \quad (177)$$

We have

$$\begin{aligned} & \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ & + C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) = H(X) \end{aligned} \quad (178)$$

$$= H(A, X) = H(A) + H(X|A) \quad (179)$$

$$= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) H(P_{X_a}); \quad (180)$$

where (178) comes from Theorem 7 and (177); and (179) comes from the fact that A can be written as a function of X : by definition, the vertex set of $\bigsqcup_{a \in \mathcal{A}} G_a$ is given by $\mathcal{X} = \bigsqcup_{a \in \mathcal{A}} \mathcal{X}_a$ and $\text{supp } P_{X_a} \subseteq \mathcal{X}_a$, therefore A is the unique index such that $X \in \mathcal{X}_A$.

APPENDIX C PROOF OF PROPOSITION 6

Let us show that for all graph $G = (\mathcal{X}, \mathcal{E})$, the function $P_X \mapsto C(G, P_X)$ is concave. Let $P_X, P'_X \in \Delta(\mathcal{X})$ and $\beta \in [0, 1]$. Let $(b_n)_{n \in \mathbb{N}}$ be a sequence of integers such that $\frac{b_n}{n} \xrightarrow{n \rightarrow \infty} \beta$.

By Lemma 5, there exists two sequences $(\mathcal{C}_n)_{n \in \mathbb{N}}$ and $(\mathcal{C}'_n)_{n \in \mathbb{N}}$ that satisfy the following:

$$\forall n \in \mathbb{N}^*, \mathcal{C}_n \subseteq \mathcal{X}^n \text{ and } \mathcal{C}'_n \subseteq \mathcal{X}^n \text{ are independent in } G^{\wedge n}; \quad (181)$$

and

$$\frac{\log |\mathcal{C}_n|}{n} \xrightarrow{n \rightarrow \infty} C(G, P_X), \quad \frac{\log |\mathcal{C}'_n|}{n} \xrightarrow{n \rightarrow \infty} C(G, P'_X), \quad (182)$$

$$\max_{x^n \in \mathcal{C}_n} \|T_{x^n} - P_X\|_\infty \xrightarrow{n \rightarrow \infty} 0, \quad \max_{x^n \in \mathcal{C}'_n} \|T_{x^n} - P_X\|_\infty \xrightarrow{n \rightarrow \infty} 0. \quad (183)$$

Let us build a sequence of codebooks $(\mathcal{C}_n'')_{n \in \mathbb{N}^*}$ adapted to the distribution $\beta P_X + (1 - \beta)P_X'$ by using a time-sharing between $(\mathcal{C}_n)_{n \in \mathbb{N}^*}$ and $(\mathcal{C}_n')_{n \in \mathbb{N}^*}$. For all $n \in \mathbb{N}^*$, let

$$\mathcal{C}_n'' \doteq \mathcal{C}_n^{b_n} \times \mathcal{C}_n'^{n-b_n}. \quad (184)$$

For all $n \in \mathbb{N}^*$, $\mathcal{C}_n'' \subseteq \mathcal{X}^{n^2}$ is independent in $G^{\wedge n^2}$ as a product of independent sets.

The rate associated to \mathcal{C}_n'' is given by

$$\frac{\log |\mathcal{C}_n''|}{n^2} = \frac{b_n \log |\mathcal{C}_n| + (n - b_n) \log |\mathcal{C}_n'|}{n^2} \quad (185)$$

$$= \frac{b_n}{n} \frac{\log |\mathcal{C}_n|}{n} + \frac{n - b_n}{n} \frac{\log |\mathcal{C}_n'|}{n} \quad (186)$$

$$\xrightarrow{n \rightarrow \infty} \beta C(G, P_X) + (1 - \beta)C(G, P_X'); \quad (187)$$

and the types of the codewords in \mathcal{C}_n'' satisfy

$$\max_{x^{n^2} \in \mathcal{C}_n''} \left\| T_{x^{n^2}} - \beta P_X - (1 - \beta)P_X' \right\|_{\infty} \quad (188)$$

$$= \max_{x^{nb_n} \in \mathcal{C}_n} \max_{x'^{n(n-b_n)} \in \mathcal{C}_n'} \left\| \frac{nb_n}{n^2} T_{x^{nb_n}} + \frac{n(n-b_n)}{n^2} T_{x'^{n(n-b_n)}} - \beta P_X - (1 - \beta)P_X' \right\|_{\infty} \quad (189)$$

$$\leq \max_{x^{nb_n} \in \mathcal{C}_n} \left\| \frac{b_n}{n} T_{x^{nb_n}} - \beta P_X \right\|_{\infty} + \max_{x'^{n(n-b_n)} \in \mathcal{C}_n'} \left\| \frac{n-b_n}{n} T_{x'^{n(n-b_n)}} - (1 - \beta)P_X' \right\|_{\infty} \quad (190)$$

$$= \beta \max_{x^{nb_n} \in \mathcal{C}_n} \left\| T_{x^{nb_n}} - P_X + o(1)T_{x^{nb_n}} \right\|_{\infty} + (1 - \beta) \times \max_{x'^{n(n-b_n)} \in \mathcal{C}_n'} \left\| T_{x'^{n(n-b_n)}} - P_X' + o(1)T_{x'^{n(n-b_n)}} \right\|_{\infty} \quad (191)$$

$$\leq \beta \max_{x^{nb_n} \in \mathcal{C}_n} \left\| T_{x^{nb_n}} - P_X \right\|_{\infty} + o(1) \left\| T_{x^{nb_n}} \right\|_{\infty} + (1 - \beta) \max_{x'^{n(n-b_n)} \in \mathcal{C}_n'} \left\| T_{x'^{n(n-b_n)}} - P_X' \right\|_{\infty} + o(1) \left\| T_{x'^{n(n-b_n)}} \right\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (192)$$

By Lemma 5, $\lim_{n \rightarrow \infty} \frac{\log |\mathcal{C}_n''|}{n^2} \leq C(G, \beta P_X + (1 - \beta)P_X')$, thus

$$\beta C(G, P_X) + (1 - \beta)C(G, P_X') \leq C(G, \beta P_X + (1 - \beta)P_X'). \quad (193)$$

The function $P_X \mapsto C(G, P_X)$ is concave on the convex compact set $\Delta(\mathcal{X})$, therefore its set of maximizers $\mathcal{P}^*(G) = \operatorname{argmax}_{P_X \in \Delta(\mathcal{X})} C(G, P_X)$ is convex. Furthermore, by Lemma 6, the set $\mathcal{P}^*(G)$ is nonempty and satisfies

$$\forall P_X \in \mathcal{P}^*(G), C(G, P_X) = C_0(G). \quad (194)$$

APPENDIX D PROOF OF THEOREM 8

Let us start by showing that Theorem 8 is true when \mathcal{A} has two elements. Let $G = (\mathcal{X}, \mathcal{E})$, and $G' = (\mathcal{X}', \mathcal{E}')$ be

two graphs, and let $P_{X, X'} \in \mathcal{P}^*(G \wedge G')$. We will prove that $P_X \otimes P_{X'}$ is also capacity-achieving by building an adequate sequence of codebooks.

For all $n \in \mathbb{N}^*$, let $\mathcal{C}_n \subseteq (\mathcal{X} \times \mathcal{X}')^n$ such that \mathcal{C}_n is an independent set in $(G \wedge G')^{\wedge n}$, and

$$\frac{1}{n} \log |\mathcal{C}_n| \xrightarrow{n \rightarrow \infty} C_0(G \wedge G'), \quad (195)$$

$$\max_{(x^n, x'^n) \in \mathcal{C}_n} \|T_{x^n, x'^n} - P_{X, X'}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (196)$$

The existence of such a sequence is given by Lemma 5, and Proposition 6. Let

$$Q_{X, X'}^{(n)} \doteq \frac{1}{|\mathcal{C}_n|} \sum_{(x^n, x'^n) \in \mathcal{C}_n} T_{x^n, x'^n}. \quad (197)$$

An immediate observation is that

$$Q_{X, X'}^{(n)} \xrightarrow{n \rightarrow \infty} P_{X, X'} \quad (198)$$

as a consequence of (196).

Let us build a sequence of codebooks with asymptotic rate $C_0(G \wedge G')$, such that the type of their codewords converge uniformly to $P_X \otimes P_{X'}$:

$$\mathcal{C}_{n^3}^* \doteq \mathcal{T}_{\varepsilon_n}^{n^3}(P_X \otimes P_{X'}) \cap \left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n; \quad (199)$$

where

$$\varepsilon_n \doteq \|Q_X^{(n)} \otimes Q_{X'}^{(n)} - P_X \otimes P_{X'}\|_{\infty} + \frac{1}{\sqrt[4]{n}}; \quad (200)$$

and where for all $t \leq n$, the shifted codebook $\mathcal{C}_n^{(t)}$ is defined by

$$\mathcal{C}_n^{(t)} \doteq \left\{ \left((x_t, x_{t+1}, \dots, x_n, x_1, \dots, x_{t-1}), x'^n \right) \mid (x^n, x'^n) \in \mathcal{C}_n \right\}. \quad (201)$$

By construction, $\mathcal{C}_{n^3}^* \subseteq \mathcal{T}_{\varepsilon_n}^{n^3}(P_X \otimes P_{X'})$ thanks to (199), and $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$ thanks to (200) and (198); therefore we have

$$\max_{x^{n^3} \in \mathcal{C}_{n^3}^*} \|T_{x^{n^3}} - P_X \otimes P_{X'}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (202)$$

Furthermore, $\mathcal{C}_{n^3}^*$ is an independent set in $(G \wedge G')^{\wedge n^3}$, as it is contained in the product independent set $\left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n$; note that this holds because the shifted codebook $\mathcal{C}_n^{(t)}$ is an independent set in $(G \wedge G')^{\wedge n}$ for all $t \leq n$.

Now let us prove that $\frac{\log |\mathcal{C}_{n^3}^*|}{n^3} \xrightarrow{n \rightarrow \infty} C_0(G \wedge G')$. Let us draw a codeword uniformly from $\left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n$:

$$C^{n^3} \doteq (C_1^{n^2}, \dots, C_n^{n^2}), \quad (203)$$

where for all $t \leq n$, $C_t^{n^2}$ is a random $n \times n$ -sequence drawn uniformly from $\prod_{t \leq n} \mathcal{C}_n^{(t)}$. We want to prove that $C^{n^3} \in \mathcal{T}_{\varepsilon_n}^{n^3}(P_X \otimes P_{X'})$ with high probability.

On one hand we have to determine the average type of the random variables $(C_t^{n^2})_{t \leq n}$ which are iid copies of $C^{n^2} =$

(C_1^n, \dots, C_n^n) ; where each C_t^n is drawn uniformly from $\mathcal{C}_n^{(t)}$, and the $(C_t^n)_{t \leq n}$ are mutually independent.

$$\mathbb{E} [T_{C_t^{n^2}}] = \frac{1}{n} \sum_{t \leq n} \mathbb{E} [T_{C_t^n}] \quad (204)$$

$$= \frac{1}{n} \sum_{t \leq n} \frac{1}{|\mathcal{C}_n^{(t)}|} \sum_{(x^n, x'^n) \in \mathcal{C}_n^{(t)}} T_{x^n, x'^n} \quad (205)$$

$$= \frac{1}{n} \sum_{t \leq n} \frac{1}{|\mathcal{C}_n|} \sum_{(x^n, x'^n) \in \mathcal{C}_n} T_{\sigma_t(x^n), x'^n} \quad (206)$$

$$= \frac{1}{|\mathcal{C}_n|} \sum_{(x^n, x'^n) \in \mathcal{C}_n} \frac{1}{n} \sum_{t \leq n} T_{\sigma_t(x^n), x'^n} \quad (207)$$

$$= \frac{1}{|\mathcal{C}_n|} \sum_{(x^n, x'^n) \in \mathcal{C}_n} T_{x^n} \otimes T_{x'^n} \quad (208)$$

$$= Q_X^{(n)} \otimes Q_{X'}^{(n)}, \quad (209)$$

where $\sigma_t(x^n) = (x_t, x_{t+1}, \dots, x_n, x_1, \dots, x_{n-1})$; (206) comes from the construction of $\mathcal{C}_n^{(t)}$ in (201); and (208) comes from the following observation:

$$\sum_{t \leq n} T_{\sigma_t(x^n), x'^n} = \sum_{t \leq n} \sum_{s \leq n} T_{x_{s+t}, x'_s} = \sum_{s \leq n} \sum_{t \leq n} T_{x_{s+t}, x'_s} \quad (210)$$

$$= \sum_{s \leq n} T_{x^n, (x'_s, \dots, x'_s)} = \sum_{s \leq n} T_{x^n} \otimes T_{x'_s} = T_{x^n} \otimes T_{x'^n}, \quad (211)$$

where the index $s + t$ is taken modulo n .

On the other hand we have

$$\frac{|\mathcal{C}_{n^3}^*|}{\left| \left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n \right|} \quad (212)$$

$$= \frac{|\mathcal{T}_{\varepsilon_n}^{n^3}(P_X \otimes P_{X'}) \cap \left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n|}{\left| \left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n \right|} \quad (213)$$

$$= \mathbb{P} \left(C^{n^3} \in \mathcal{T}_{\varepsilon_n}^{n^3}(P_X \otimes P_{X'}) \right) \quad (214)$$

$$= \mathbb{P} \left(\left\| \frac{1}{n} \sum_{t \leq n} T_{C_t^{n^2}} - P_X \otimes P_{X'} \right\|_{\infty} \leq \varepsilon_n \right) \quad (215)$$

$$\geq \mathbb{P} \left(\left\| \frac{1}{n} \sum_{t \leq n} T_{C_t^{n^2}} - Q_X^{(n)} \otimes Q_{X'}^{(n)} \right\|_{\infty} + \left\| Q_X^{(n)} \otimes Q_{X'}^{(n)} - P_X \otimes P_{X'} \right\|_{\infty} \leq \varepsilon_n \right) \quad (216)$$

$$= \mathbb{P} \left(\left\| \sum_{t \leq n} T_{C_t^{n^2}} - n Q_X^{(n)} \otimes Q_{X'}^{(n)} \right\|_{\infty} \leq n^{3/4} \right) \quad (217)$$

$$\geq 1 - \sum_{(x, x') \in \mathcal{X} \times \mathcal{X}'} \mathbb{P} \left(\left| \sum_{t \leq n} T_{C_t^{n^2}}(x, x') - n Q_X^{(n)} \otimes Q_{X'}^{(n)}(x, x') \right| > n^{3/4} \right) \quad (218)$$

$$\geq 1 - \sum_{(x, x') \in \mathcal{X} \times \mathcal{X}'} \frac{1}{n^{3/2}} \mathbb{X} \left[\sum_{t \leq n} T_{C_t^{n^2}}(x, x') \right] \quad (219)$$

$$\geq 1 - \frac{|\mathcal{X}| |\mathcal{X}'|}{n^{1/2}} \xrightarrow{n \rightarrow \infty} 1; \quad (220)$$

where (214) and (215) come from the construction of C^{n^3} ; (217) comes from the construction of ε_n ; (218) follows from the union bound; (219) comes from Chebyshev's inequality and (209); and (220) comes from the fact that

$\mathbb{V} \left[\sum_{t \leq n} T_{C_t^{n^2}}(x, x') \right] = \sum_{t \leq n} \mathbb{V} [T_{C_t^{n^2}}(x, x')] \leq n$, as the random variables $T_{C_t^{n^2}}(x, x')$ are iid and takes values in $[0, 1]$. Hence

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\log |\mathcal{C}_{n^3}^*|}{n^3} &= \lim_{n \rightarrow \infty} \frac{\log \left| \left(\prod_{t \leq n} \mathcal{C}_n^{(t)} \right)^n \right|}{n^3} \\ &= \lim_{n \rightarrow \infty} \frac{\log |\mathcal{C}_n|}{n} = C_0(G \wedge G'); \end{aligned} \quad (221)$$

where the second equality holds as the shifted codebooks $(\mathcal{C}_n^{(t)})_{t \leq n}$ all have cardinality $|\mathcal{C}_n|$.

Thus, by combining (221), Lemma 5, and Proposition 6 we obtain

$$\begin{aligned} C_0(G \wedge G') &= \lim_{n \rightarrow \infty} \frac{\log |\mathcal{C}_{n^3}^*|}{n^3} \leq C(G \wedge G', P_X \otimes P_{X'}) \\ &\leq C_0(G \wedge G'), \end{aligned} \quad (222)$$

hence $P_X \otimes P_{X'} \in \mathcal{P}^*(G \wedge G')$.

Therefore, Theorem 8 is proved when \mathcal{A} has two elements:

$$P_{X, X'} \in \mathcal{P}^*(G \wedge G') \implies P_X \otimes P_{X'} \in \mathcal{P}^*(G \wedge G'). \quad (223)$$

Now let us consider the case where \mathcal{A} has a cardinality greater than 2. Let $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$. By considering the product graphs

$$\bigwedge_{a \in \mathcal{A}} G_a = \left(\bigwedge_{1 \leq i < i^*} G_i \right) \wedge \left(\bigwedge_{i^* \leq i \leq |\mathcal{A}|} G_i \right); \quad (224)$$

for all $i^* \leq |\mathcal{A}|$, and applying (223) successively, we obtain

$$\begin{aligned} P_{X_1, \dots, X_{|\mathcal{A}|}} &\in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a) \\ \implies P_{X_1} \otimes P_{X_2, \dots, X_{|\mathcal{A}|}} &\in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a) \end{aligned} \quad (225)$$

$$\implies (P_{X_1} \otimes P_{X_2}) \otimes P_{X_3, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a) \quad (226)$$

$$\implies \bigotimes_{a \in \mathcal{A}} P_{X_a} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a). \quad (227)$$

APPENDIX E

RESULTS ON CAPACITY-ACHIEVING DISTRIBUTIONS

A. Proof of Proposition 9

Let G be a vertex-transitive graph, and let $P_X \in \mathcal{P}^*(G)$. Let $\psi \in \text{Aut}(G)$, we first prove that $P_{\psi(X)} \in \mathcal{P}^*(G)$, then we will conclude by using the convexity of $\mathcal{P}^*(G)$.

Let $(\mathcal{C}_n)_{n \in \mathbb{N}^*}$ be a sequence such that

$$\forall n \in \mathbb{N}^*, \mathcal{C}_n \subseteq \mathcal{X}^n \text{ is an independent set in } G^{\wedge n}, \quad (228)$$

$$\max_{x^n \in \mathcal{C}_n} \|T_{x^n} - P_X\|_{\infty} \xrightarrow{n \rightarrow \infty} 0, \quad (229)$$

$$\frac{\log |\mathcal{C}_n|}{n} \xrightarrow{n \rightarrow \infty} C(G, P_X) = C_0(G). \quad (230)$$

The existence of such a sequence is given by Lemma 5. Note that the last equality in (230) comes from the assumption $P_X \in \mathcal{P}^*(G)$.

Now, for all $n \in \mathbb{N}^*$ the codebook

$$\psi(\mathcal{C}_n) \doteq \{(\psi(x_1), \dots, \psi(x_n)) \mid x^n \in \mathcal{C}_n\} \quad (231)$$

is also independent in $G^{\wedge n}$, as ψ is a graph automorphism and therefore preserves adjacencies. We have by construction

$$\max_{x^n \in \psi(\mathcal{C}_n)} \|T_{x^n} - P_{\psi(X)}\|_{\infty} \xrightarrow{n \rightarrow \infty} 0. \quad (232)$$

Furthermore, since ψ is a bijection we have $|\psi(\mathcal{C}_n)| = |\mathcal{C}_n|$ for all $n \in \mathbb{N}^*$, thus

$$\frac{\log |\psi(\mathcal{C}_n)|}{n} = \frac{\log |\mathcal{C}_n|}{n} \xrightarrow{n \rightarrow \infty} C_0(G). \quad (233)$$

Hence

$$P_{\psi(X)} \in \mathcal{P}^*(G). \quad (234)$$

Now, for all $x \neq x' \in \mathcal{X}$, denote by $\mathcal{S}_{x' \rightarrow x} \subseteq \text{Aut}(G)$ the set of automorphisms that map x' to x ; note that this set is nonempty thanks to the vertex-transitivity of G . We have for all $x \in \mathcal{X}$

$$\text{Aut}(G) = \bigsqcup_{x' \in \mathcal{X}} \mathcal{S}_{x' \rightarrow x}. \quad (235)$$

Furthermore, for all $x \in \mathcal{X}$, all the sets $(\mathcal{S}_{x' \rightarrow x})_{x' \rightarrow x}$ have the same cardinality: for all $x', x'' \in \mathcal{X}$,

$$\mathcal{S}_{x'' \rightarrow x} \circ \psi_1 \subseteq \mathcal{S}_{x' \rightarrow x}, \quad (236)$$

where $\psi_1 \in \mathcal{S}_{x' \rightarrow x''}$. It follows that for all $x, x' \in \mathcal{X}$,

$$|\mathcal{S}_{x' \rightarrow x}| = \frac{|\text{Aut}(G)|}{|\mathcal{X}|}. \quad (237)$$

Therefore, for all $x \in \mathcal{X}$ we have

$$\frac{1}{|\text{Aut}(G)|} \sum_{\psi \in \text{Aut}(G)} P_{\psi(X)} \in \mathcal{P}^*(G) \quad (238)$$

$$= \left(\frac{1}{|\text{Aut}(G)|} \sum_{\psi \in \text{Aut}(G)} P_X(\psi^{-1}(x)) \right)_{x \in \mathcal{X}} \quad (239)$$

$$= \left(\frac{1}{|\text{Aut}(G)|} \sum_{x' \in \mathcal{X}} |\mathcal{S}_{x' \rightarrow x}| P_X(x') \right)_{x \in \mathcal{X}} \quad (240)$$

$$= \left(\frac{1}{|\text{Aut}(G)|} \sum_{x' \in \mathcal{X}} \frac{|\text{Aut}(G)|}{|\mathcal{X}|} P_X(x') \right)_{x \in \mathcal{X}} \quad (241)$$

$$= \text{Unif}(\mathcal{X}); \quad (242)$$

where (238) comes from the convexity of $\mathcal{P}^*(G)$ given by Proposition 6 and (234); (240) comes from (235); and (241) comes from (237).

B. Proof of Lemma 7

Let $(w_a)_{a \in \mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|}$, and maximize

$$\zeta : P_A \mapsto H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) w_a. \quad (243)$$

It can be easily observed that ζ is strictly concave, hence the existence and uniqueness of the maximum. We have

$$\nabla \zeta(P_A) = \left(-\log P_A(a) - \frac{1}{\ln 2} + w_a \right)_{a \in \mathcal{A}}, \quad (244)$$

hence

$$\nabla \zeta(P_A) \perp \Delta(\mathcal{A}) \iff \exists C \in \mathbb{R}, \nabla \zeta(P_A) = (C, \dots, C) \quad (245)$$

$$\iff \exists C' \in \mathbb{R}, (-\log P_A(a) + w_a)_{a \in \mathcal{A}} = (C', \dots, C') \quad (246)$$

$$\iff \exists C' \in \mathbb{R}, P_A = 2^{-C'} (2^{w_a})_{a \in \mathcal{A}} \quad (247)$$

The value of C' can be deduced from the fact that P_A is a probability distribution: $2^{C'}$ is the normalization constant $\sum_{a' \in \mathcal{A}} 2^{w_{a'}}$. Hence the maximum of ζ is given by

$$P_A^* = \left(\frac{2^{w_a}}{\sum_{a' \in \mathcal{A}} 2^{w_{a'}}} \right)_{a \in \mathcal{A}}; \quad (248)$$

and we have

$$\zeta(P_A^*) = \sum_{a \in \mathcal{A}} P_A^*(a) \left(\log \left(\frac{\sum_{a' \in \mathcal{A}} 2^{w_{a'}}}{2^{w_a}} \right) + w_a \right) \quad (249)$$

$$= \log \left(\sum_{a' \in \mathcal{Z}} 2^{w_{a'}} \right). \quad (250)$$

APPENDIX F PROOF OF THEOREM 9

We prove Theorem 9 in two steps, which are Lemma 14 and Lemma 15. The proofs are respectively given in App. F-A and F-B.

Lemma 14.

$$C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right) = \sum_{a \in \mathcal{A}} C_0(G_a) \quad (251)$$

$$\implies \forall (P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a),$$

$$\left\{ \begin{array}{l} \bigotimes_{a \in \mathcal{A}} P_{X_a}^* \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right) \text{ and} \\ C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^* \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}^*). \end{array} \right. \quad (252)$$

Lemma 15. For all $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right)$, the following holds

$$C \left(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}} \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) \quad (253)$$

$$\implies C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right) = \sum_{a \in \mathcal{A}} C_0(G_a) \quad (254)$$

and $\forall a \in \mathcal{A}, P_{X_a} \in \mathcal{P}^*(G_a)$.

Let us prove Theorem 9. We consider a family of distributions $P_{X_a} \in \mathcal{P}^*(G_a)$ with $a \in \mathcal{A}$. By Lemma 14, we have

$$C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right) = \sum_{a \in \mathcal{A}} C_0(G_a) \quad (255)$$

$$\implies \bigotimes_{a \in \mathcal{A}} P_{X_a}^* \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right) \text{ and}$$

$$C \left(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^* \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}^*) \quad (256)$$

$$\implies \exists P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right),$$

$$C \left(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}} \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}). \quad (257)$$

Conversely, by Lemma 15 we have

$$\exists P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^* \left(\bigwedge_{a \in \mathcal{A}} G_a \right),$$

$$C \left(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}} \right) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) \quad (258)$$

$$\implies C_0 \left(\bigwedge_{a \in \mathcal{A}} G_a \right) = \sum_{a \in \mathcal{A}} C_0(G_a). \quad (259)$$

Moreover, all distribution $P_{X_1, \dots, X_{|\mathcal{A}|}}$ that satisfies (258), also satisfies $P_{X_a} \in \mathcal{P}^*(G_a)$, for all $a \in \mathcal{A}$.

A. Proof of Lemma 14

For all family of graphs $(G_a)_{a \in \mathcal{A}}$, and family of distributions $P_{X_a} \in \mathcal{P}^*(G_a)$ with $a \in \mathcal{A}$, we have

$$C_0(\bigwedge_{a \in \mathcal{A}} G_a) = \max_{P_{X_1, \dots, X_{|\mathcal{A}|}}} C(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}}) \quad (260)$$

$$\geq C(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^*) \quad (261)$$

$$\geq \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}^*) \quad (262)$$

$$= \sum_{a \in \mathcal{A}} C_0(G_a); \quad (263)$$

where (260) comes from Lemma 6; (262) comes from Proposition 5; and (263) follows from the hypothesis $P_{X_a} \in \mathcal{P}^*(G_a)$ with $a \in \mathcal{A}$.

Now assume that $\sum_{a \in \mathcal{A}} C_0(G_a) = C_0(\bigwedge_{a \in \mathcal{A}} G_a)$, then equality holds between the left-hand side of (260) and the term in (263). Therefore, we have

$$\begin{aligned} C_0(\bigwedge_{a \in \mathcal{A}} G_a) &= C(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^*), \\ \implies \bigotimes_{a \in \mathcal{A}} P_{X_a}^* &\in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a), \\ C(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^*) &= \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}^*). \end{aligned} \quad (264)$$

B. Proof of Lemma 15

Let $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$, and let $P_{X_a}^* \in \mathcal{P}^*(G_a)$, for all $a \in \mathcal{A}$. The following holds

$$C_0(\bigwedge_{a \in \mathcal{A}} G_a) = C(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}}) \quad (265)$$

$$\geq C(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}^*) \quad (266)$$

$$\geq \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}^*) \quad (267)$$

$$\geq \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}), \quad (268)$$

where (266) comes from the hypothesis $P_{X_1, \dots, X_{|\mathcal{A}|}} \in \mathcal{P}^*(\bigwedge_{a \in \mathcal{A}} G_a)$; (267) comes from Proposition 5; and (268) comes from the hypothesis $P_{X_a}^* \in \mathcal{P}^*(G_a)$, for all $a \in \mathcal{A}$.

Now assume that

$$C(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}}) = \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}). \quad (269)$$

Then equality holds in between the right-hand side of (265) and the term in (268). In particular, we have for all $a \in \mathcal{A}$

$$C(G_a, P_{X_a}) = C(G_a, P_{X_a}^*), \quad (270)$$

which implies that P_{X_a} also maximizes $C(G_a, \cdot)$ for all $a \in \mathcal{A}$:

$$\forall a \in \mathcal{A}, P_{X_a} \in \mathcal{P}^*(G_a). \quad (271)$$

Furthermore,

$$C_0(\bigwedge_{a \in \mathcal{A}} G_a) = C(\bigwedge_{a \in \mathcal{A}} G_a, P_{X_1, \dots, X_{|\mathcal{A}|}}) \quad (272)$$

$$= \sum_{a \in \mathcal{A}} C(G_a, P_{X_a}) \quad (273)$$

$$= \sum_{a \in \mathcal{A}} C_0(G_a); \quad (274)$$

where (273) is a consequence of the equality in equations (265)-(268), and (274) comes from (271).

APPENDIX G PROOF OF THEOREM 10

The techniques used in this proof are the same as in the proof of Theorem 9. We prove Theorem 10 in two steps, which are Lemma 16 and Lemma 17; their proofs are respectively given in App. G-A and G-B.

Lemma 16. *Let*

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \quad (275)$$

We have

$$\begin{aligned} C_0\left(\bigsqcup_{a \in \mathcal{A}} G_a\right) &= \log\left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)}\right) \\ \implies \forall (P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \\ \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^* &\in \mathcal{P}^*\left(\bigsqcup_{a \in \mathcal{A}} G_a\right) \text{ and} \\ C\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^*\right) \\ &= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{X_a}^*), \end{aligned} \quad (276)$$

Lemma 17. *Let*

$$P_A \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \quad (277)$$

For all $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \in \mathcal{P}^(\bigsqcup_{a \in \mathcal{A}} G_a)$ the following holds*

$$\begin{aligned} C\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right) \\ = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \\ \implies C_0\left(\bigsqcup_{a \in \mathcal{A}} G_a\right) = \log\left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)}\right), \\ (P_{X_a})_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \text{ and } P_A = P_A^*. \end{aligned} \quad (278)$$

Now let us prove Theorem 10. Let $(P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a)$, we have by Lemma 16

$$\begin{aligned} C_0\left(\bigsqcup_{a \in \mathcal{A}} G_a\right) &= \log\left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)}\right) \quad (279) \\ \implies \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^* &\in \mathcal{P}^*\left(\bigsqcup_{a \in \mathcal{A}} G_a\right) \text{ and} \\ C\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^*\right) \\ &= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{X_a}^*), \quad (280) \\ \implies \exists P_X \in \mathcal{P}^*\left(\bigsqcup_{a \in \mathcal{A}} G_a\right), \\ C\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^*\right) \\ &= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{X_a}^*), \end{aligned}$$

where $P_{X_a}^* = P_{X|X \in \mathcal{X}_a}$ and $P_A^*(a) = P_X(\mathcal{X}_a)$ for all $a \in \mathcal{A}$.

Conversely, by Lemma 17 we have

$$\exists P_X \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right), \quad (281)$$

$$\begin{aligned} & C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \\ \implies & C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right), \end{aligned} \quad (282)$$

and any $P_X = \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}$ that satisfies (281) also satisfies

$$\begin{aligned} (P_{X_a})_{a \in \mathcal{A}} &\in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \\ \text{and } P_A &= \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \end{aligned} \quad (283)$$

A. Proof of Lemma 16

Now let us prove Lemma 16. Let

$$\sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right), \quad (284)$$

$$(P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \quad (285)$$

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \quad (286)$$

We have

$$C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \quad (287)$$

$$\geq C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^* \right) \quad (288)$$

$$\geq H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{X_a}^*) \quad (289)$$

$$= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C_0(G_a) \quad (290)$$

$$= \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right); \quad (291)$$

where (287) and (288) come from (284) and Proposition 6; (289) comes from Proposition 5; (290) comes from (285) and Proposition 6; and (291) comes from (286) and Lemma 7.

Assume that $C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right)$, then equality holds in (287) to (291), therefore the following holds:

$$\begin{aligned} & C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right) \\ \implies & \forall (P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \\ & \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^* \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) \text{ and} \\ & C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A^*(a) P_{X_a}^* \right) \\ &= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C(G_a, P_{X_a}^*). \end{aligned} \quad (292)$$

B. Proof of Lemma 17

Let

$$\sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right), \quad (293)$$

$$(P_{X_a}^*)_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \quad (294)$$

$$P_A^* \doteq \left(\frac{2^{C_0(G_a)}}{\sum_{a' \in \mathcal{A}} 2^{C_0(G_{a'})}} \right)_{a \in \mathcal{A}}. \quad (295)$$

We have

$$\begin{aligned} & C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) \end{aligned} \quad (296)$$

$$\geq \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right) \quad (297)$$

$$= H(P_A^*) + \sum_{a \in \mathcal{A}} P_A^*(a) C_0(G_a) \quad (298)$$

$$\geq H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C_0(G_a) \quad (299)$$

$$= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}^*) \quad (300)$$

$$\geq H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}); \quad (301)$$

where (296) comes from (293) and Proposition 6; (297) comes from (33), see [5, Theorem 4]; (298) and (299) come from (295) and Lemma 7, which can be found in App. G-A; (300) and (301) come from (294) and Proposition 6.

Assume that $C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) = H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a})$, then equality holds in (296) to (301). In particular $P_A = P_A^*$ as a consequence of the equality between (298) and (299); and $(P_{X_a})_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a)$ as a consequence of the equality between (300) and (301). Thus, for all $\sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \in \mathcal{P}^* \left(\bigsqcup_{a \in \mathcal{A}} G_a \right)$ the following holds:

$$\begin{aligned} & C \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= H(P_A) + \sum_{a \in \mathcal{A}} P_A(a) C(G_a, P_{X_a}) \\ \implies & C_0 \left(\bigsqcup_{a \in \mathcal{A}} G_a \right) = \log \left(\sum_{a \in \mathcal{A}} 2^{C_0(G_a)} \right), \\ & (P_{X_a})_{a \in \mathcal{A}} \in \prod_{a \in \mathcal{A}} \mathcal{P}^*(G_a), \text{ and } P_A = P_A^*. \end{aligned} \quad (302)$$

APPENDIX H

PROOF OF THEOREM 14

Lemma 18 comes from [76, Corollary 1], and states that the function $P_A \mapsto H_\kappa \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right)$, defined analogously to $P_A \mapsto \bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right)$, is always linear. The proof of Lemma 19 is given in App. H-A.

Lemma 18 (from [77, Corollary 3.4]). *For all probabilistic graphs $(G_a, P_{X_a})_{a \in \mathcal{A}}$ and $P_A \in \Delta(\mathcal{A})$, we have*

$$H_\kappa \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) = \sum_{a \in \mathcal{A}} P_A(a) H_\kappa(G_a, P_{X_a}). \quad (303)$$

Lemma 19. *The graph $\bigsqcup_{a \in \mathcal{A}} G_a$ is perfect if and only if G_a is perfect for all $a \in \mathcal{A}$.*

Now let us prove Theorem 14.

For all $a \in \mathcal{A}$, let (G_a, P_{X_a}) a probabilistic graph where $G_a = (\mathcal{X}_a, \mathcal{E}_a)$ is a perfect graph. By Lemma 19, $\bigsqcup_{a \in \mathcal{A}} G_a$ is also perfect; and we have

$$\bar{H} \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) = H_\kappa \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \quad (304)$$

by Theorem 11. We also have

$$\begin{aligned} & H_\kappa \left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a} \right) \\ &= \sum_{a \in \mathcal{A}} P_A(a) H_\kappa(G_a, P_{X_a}) = \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}) \end{aligned} \quad (305)$$

by Lemma 18 and Theorem 11 used on the perfect graphs $(G_a)_{a \in \mathcal{A}}$. Thus

$$\bar{H}\left(\bigsqcup_{a \in \mathcal{A}} G_a, \sum_{a \in \mathcal{A}} P_A(a) P_{X_a}\right) = \sum_{a \in \mathcal{A}} P_A(a) \bar{H}(G_a, P_{X_a}). \quad (306)$$

By Theorem 3, it follows that $\bar{H}(\bigwedge_{a \in \mathcal{A}} G_a, \bigotimes_{a \in \mathcal{A}} P_{X_a}) = \sum_{a \in \mathcal{A}} \bar{H}(G_a, P_{X_a}) = \sum_{a \in \mathcal{A}} H_\kappa(G_a, P_{X_a})$, where the last equality comes from Theorem 11.

A. Proof of Lemma 19

(\implies) Let $G = \bigsqcup_{a \in \mathcal{A}} G_a$ be a perfect graph. Let $a' \in \mathcal{A}$ and $\mathcal{S}_{a'} \subset \mathcal{X}_{a'}$. We have $\chi((\bigsqcup_{a \in \mathcal{A}} G_a)[\mathcal{S}_{a'}]) = \omega((\bigsqcup_{a \in \mathcal{A}} G_a)[\mathcal{S}_{a'}])$ since G is perfect, and therefore $\chi(G_{a'}[\mathcal{S}_{a'}]) = \omega(G_{a'}[\mathcal{S}_{a'}])$, as $(\bigsqcup_{a \in \mathcal{A}} G_a)[\mathcal{S}_{a'}] = G_{a'}[\mathcal{S}_{a'}]$. Thus all the graphs $(G_a)_{a \in \mathcal{A}}$ are perfect.

(\impliedby) Conversely, assume that for all $a \in \mathcal{A}$, $G_a = (\mathcal{X}_a, \mathcal{E}_a)$ is perfect. Then for all $\mathcal{S} \subset \bigsqcup_{a \in \mathcal{A}} \mathcal{X}_a$, \mathcal{S} can be written as $\bigsqcup_{a \in \mathcal{A}} \mathcal{S}_a$ where $\mathcal{S}_a \subset \mathcal{X}_a$ for all $a \in \mathcal{A}$, and we have for all $P_A \in \Delta(\mathcal{A})$:

$$\begin{aligned} \chi((\bigsqcup_{a \in \mathcal{A}} G_a)[\mathcal{S}]) &= \chi(\bigsqcup_{a \in \mathcal{A}} G_a[\mathcal{S}_a]) \\ &= \max_{a \in \mathcal{A}} \chi(G_a[\mathcal{S}_a]) = \max_{a \in \mathcal{A}} \omega(G_a[\mathcal{S}_a]), \end{aligned} \quad (307)$$

and similarly, $\omega((\bigsqcup_{a \in \mathcal{A}} G_a)[\mathcal{S}]) = \max_{a \in \mathcal{A}} \omega(G_a[\mathcal{S}_a])$. Hence $\bigsqcup_{a \in \mathcal{A}} G_a$ is also perfect.

REFERENCES

- [1] N. Charpenay, M. Le Treust, and A. Roumy, "Complementary graph entropy, and product, and disjoint union of graphs," in *IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 2446–2451.
- [2] —, "Optimal zero-error coding for computing under pairwise shared side information," in *IEEE Information Theory Workshop (ITW)*, 2023, pp. 2446–2451.
- [3] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [4] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.
- [5] C. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, 1956.
- [6] D. de Boer, P. Buys, and J. Zuiddam, "The asymptotic spectrum distance, graph limits, and the Shannon capacity," *Proceedings of the 13th European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB'25)*, Budapest, 2025.
- [7] A. Vesel and J. Žerovnik, "Improved lower bound on the Shannon capacity of C_7 ," *Information processing letters*, vol. 81, no. 5, pp. 277–282, 2002.
- [8] B. Codenotti, I. Gerace, and G. Resta, "Some remarks on the Shannon capacity of odd cycles," *Ars Combinatoria*, vol. 66, pp. 243–258, 2003.
- [9] S. C. Polak and A. Schrijver, "New lower bound on the Shannon capacity of C_7 from circular graphs," *Information Processing Letters*, vol. 143, pp. 37–40, 2019.
- [10] P. Buys, S. Polak, and J. Zuiddam, "A limit theorem for Shannon capacity from lattice packings," *Proceedings of the 13th European Conference on Combinatorics, Graph Theory and Applications (EUROCOMB'25)*, Budapest, 2025.
- [11] M. Jurkiewicz, "A survey on known values and bounds on the Shannon capacity," *Selected Topics in Modern Mathematics-Edition*, pp. 115–128, 2014.
- [12] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [13] R. Hammack, W. Imrich, and S. Klavzar, "Handbook of graph products," *Discrete Mathematics and its Applications*, CRC Press, 2011.
- [14] C. Berge, *Graphs and Hypergraphs*, ser. North-Holland mathematical library. Amsterdam, 1973.
- [15] M. Grötschel, L. Lovász, and A. Schrijver, "Polynomial algorithms for perfect graphs," *Ann. Discrete Math.*, vol. 21, pp. 325–356, 1984.
- [16] C. Berge, "Färbung von graphen, deren sämtliche bzw. deren ungerade kreise starr sind," *Wissenschaftliche Zeitschrift*, 1961.
- [17] M. Chudnovsky, N. Robertson, P. D. Seymour, and R. Thomas, "Progress on perfect graphs," *Mathematical Programming*, vol. 97, no. 1-2, pp. 405–422, 2003.
- [18] L. Lovász, "On the Shannon capacity of a graph," *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, 1979.
- [19] H. Witsenhausen, "The zero-error side information problem and chromatic numbers (corresp.)," *IEEE Transactions on Information Theory*, vol. 22, no. 5, pp. 592–593, 1976.
- [20] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on information Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [21] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Transactions on Information Theory*, vol. 42, no. 5, pp. 1329–1339, 1996.
- [22] P. Koulgi, E. Tuncel, S. L. Regunathan, and K. Rose, "On zero-error source coding with decoder side information," *IEEE Transactions on Information Theory*, vol. 49, no. 1, pp. 99–111, 2003.
- [23] J. Körner and G. Longo, "Two-step encoding for finite sources," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 778–782, 1973.
- [24] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, "Entropy splitting for antiblocking corners and perfect graphs," *Combinatorica*, vol. 10, no. 1, pp. 27–40, 1990.
- [25] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," *Transactions of the 6th Prague Conference on Information Theory (1971)*, pp. 411–425, 1973.
- [26] I. Sason, "Observations on graph invariants with the Lovász ϑ -function," *AIMS Mathematics*, vol. 9, no. 6, pp. 15 385–15 468, 2024.
- [27] N. Lavi and I. Sason, "Advances in the Shannon capacity of graphs," 2025. [Online]. Available: <https://arxiv.org/abs/2509.24600>
- [28] H. Boche and C. Deppe, "Computability of the zero-error capacity of noisy channels," in *2021 IEEE Information Theory Workshop (ITW)*. IEEE, 2021, pp. 1–6.
- [29] F. Guo and Y. Watanabe, "On graphs in which the Shannon capacity is unachievable by finite product," *IEEE Transactions on Information Theory*, vol. 36, no. 3, pp. 622–623, May 1990.
- [30] M. Dalai, V. Guruswami, and J. Radhakrishnan, "An improved bound on the zero-error list-decoding capacity of the 4/3 channel," *IEEE Transactions on Information Theory*, vol. 66, no. 2, pp. 749–756, 2020.
- [31] P. Noorzad, M. Effros, M. Langberg, and V. Kostina, "The birthday problem and zero-error list codes," *IEEE Transactions on Information Theory*, vol. 67, no. 9, pp. 5791–5803, 2021.
- [32] C. Wang, A. Irfan, Y. Zhang, M. Langberg, and S. Jaggi, "Zero-error superposition codes," in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025, pp. 1–6.
- [33] Q. Cao, Q. Chen, and B. Bai, "On zero-error capacity of graphs with one edge," *IEEE Transactions on Information Theory*, vol. 71, no. 5, pp. 3350–3359, 2025.
- [34] Y. Chen and N. Devroye, "Zero-error relaying for primitive relay channels," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7708–7715, Dec 2017.
- [35] M. Asadi and N. Devroye, "On the zero-error capacity of channels with noisy feedback," in *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct 2017, pp. 642–649.
- [36] L. Zhao and H. H. Permuter, "Zero-error feedback capacity of channels with state information via dynamic programming," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2640–2650, June 2010.
- [37] A. Bracher and A. Lapidoth, "The zero-error feedback capacity of state-dependent channels," *IEEE Transactions on Information Theory*, vol. 64, no. 5, pp. 3538–3578, 2018.
- [38] A. Lapidoth and Y. Yan, "On the zero-error capacity of the modulo-additive noise channel with help," *IEEE Transactions on Information Theory*, vol. 70, no. 7, pp. 4721–4730, 2024.
- [39] Y. Li and J. Zuiddam, "Quantum asymptotic spectra of graphs and non-commutative graphs, and quantum Shannon capacities," *IEEE Transactions on Information Theory*, vol. 67, no. 1, pp. 416–432, 2021.
- [40] Q. Cao and R. W. Yeung, "Zero-error capacity regions of noisy networks," *IEEE Transactions on Information Theory*, vol. 68, no. 7, pp. 4201–4223, 2022.
- [41] A. Saberi, F. Farokhi, and G. N. Nair, "Zero-error feedback capacity for bounded stabilization and finite-state additive noise channels," *IEEE Transactions on Information Theory*, vol. 68, no. 10, pp. 6335–6355, 2022.

- [42] L. Wang and O. Shayevitz, “Graph information ratio,” *SIAM Journal on Discrete Mathematics*, vol. 31, no. 4, pp. 2703–2734, 2017.
- [43] J. Körner and K. Marton, “Relative capacity and dimension of graphs,” *Discrete Mathematics*, vol. 235, no. 1, pp. 307 – 315, 2001.
- [44] S. Alipour, A. Gohari, and M. Taziki, “Relative fractional independence number,” in *2024 IEEE Information Theory Workshop (ITW)*, 2024, pp. 777–782.
- [45] Y. Gu and O. Shayevitz, “On the non-adaptive zero-error capacity of the discrete memoryless two-way channel,” *Entropy*, vol. 23, no. 11, p. 1518, 2021.
- [46] S. Hu and O. Shayevitz, “The ρ -capacity of a graph,” *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2241–2253, April 2017.
- [47] O. Ordentlich and O. Shayevitz, “A VC-dimension-based outer bound on the zero-error capacity of the binary adder channel,” in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2366–2370.
- [48] R. Urbanke and Quinn Li, “The zero-error capacity region of the 2-user synchronous BAC is strictly smaller than its Shannon capacity region,” in *Information Theory Workshop (ITW)*, June 1998.
- [49] M. Wiese, T. J. Oechtering, K. H. Johansson, P. Papadimitratos, H. Sandberg, and M. Skoglund, “Secure estimation and zero-error secrecy capacity,” *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1047–1062, March 2019.
- [50] N. A. Khooshemehr, A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, “On zero-error molecular communication with multiple molecule types,” *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4311–4324, 2020.
- [51] A. Orłitsky and J. Roche, “Coding for computing,” *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 903–917, 2001.
- [52] V. Doshi, D. Shah, M. Médard, and M. Effros, “Functional compression through graph coloring,” *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3901–3917, 2010.
- [53] D. Malak, “Fractional graph coloring for functional compression with side information,” in *2022 IEEE Information Theory Workshop (ITW)*, 2022, pp. 750–755.
- [54] —, “Weighted graph coloring for quantized computing,” in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 2290–2295.
- [55] M. R. Deylam Salehi, V. K. Kizhakke Purakkal, and D. Malak, “Non-linear function computation broadcast,” in *2025 IEEE International Symposium on Information Theory (ISIT)*, 2025.
- [56] W. Haemers, “On some problems of Lovász concerning the Shannon capacity of a graph,” *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 231–232, 1979.
- [57] E. Tuncel, J. Nayak, P. Koulgi, and K. Rose, “On complementary graph entropy,” *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2537–2546, 2009.
- [58] A. Wigderson and J. Zuydam, *Asymptotic spectra: Theory, applications and extensions*. manuscript, 2023.
- [59] A. Schrijver, “On the Shannon capacity of sums and products of graphs,” *Indagationes Mathematicae*, vol. 34, no. 1, pp. 37–41, 2023.
- [60] I. Csiszár and J. Körner, “On the capacity of the arbitrarily varying channel for maximum probability of error,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 57, no. 1, pp. 87–101, 1981.
- [61] K. Marton, “On the Shannon capacity of probabilistic graphs,” *Journal of Combinatorial Theory, Series B*, vol. 57, no. 2, pp. 183–195, 1993.
- [62] J. H. Van Lint and R. M. Wilson, *A course in combinatorics*. Cambridge university press, 2001.
- [63] G. Simonyi, “On Witsenhausen’s zero-error rate for multiple sources,” *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3258–3260, 2003.
- [64] I. Sason, “Observations on the Lovász θ -function, graph capacity, eigenvalues, and strong products,” *Entropy*, vol. 25, no. 1, p. 104, 2023.
- [65] K. A. Mathew and P. R. Östergård, “New lower bounds for the Shannon capacity of odd cycles,” *Designs, Codes and Cryptography*, vol. 84, pp. 13–22, 2017.
- [66] M. Chudnovsky and P. D. Seymour, “The structure of claw-free graphs,” in *BCC*, 2005, pp. 153–171.
- [67] B. Bukh and C. Cox, “On a fractional version of Haemers’ bound,” *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3340–3348, 2018.
- [68] L. Gao, S. Gribling, and Y. Li, “On a fractional version of Haemers bound,” *IEEE Transactions on Information Theory*, 2022.
- [69] L. Gargano, J. Körner, and U. Vaccaro, “Capacities: from information theory to extremal set theory,” *Journal of Combinatorial Theory, Series A*, vol. 68, no. 2, pp. 296–316, 1994.
- [70] J.-B. Hiriart-Urruty and C. Lemaréchal, “Convex analysis and minimization algorithms ii,” *Grundlehren der mathematischen Wissenschaften*, 1993.
- [71] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, “The strong perfect graph theorem,” *Annals of mathematics*, pp. 51–229, 2006.
- [72] P. J. Cameron, “6-transitive graphs,” *Journal of Combinatorial Theory, Series B*, vol. 28, no. 2, pp. 168–179, 1980.
- [73] N. Alon, “The Shannon capacity of a union,” *Combinatorica*, vol. 18, no. 3, pp. 301–310, 1998.
- [74] A. Wyner, “Recent results in the shannon theory,” *IEEE Transactions on Information Theory*, vol. 20, no. 1, pp. 2–10, 1974.
- [75] J. Zuydam, *Algebraic complexity, asymptotic spectra and entanglement polytopes*. Institute for Logic, Language and Computation, University of Amsterdam, 2018.
- [76] G. Simonyi, “Perfect graphs and graph entropy. An updated survey,” in *Perfect Graphs*, B. A. Reed and J. L. R. Alfonsín, Eds. John Wiley & Sons, 2001, ch. 13, pp. 293–328.
- [77] —, “Graph entropy: a survey,” *Combinatorial Optimization*, vol. 20, pp. 399–441, 1995.

Nicolas Charpenay received the M.Sc. degree in Optimization from the ENS Paris-Saclay and the University of Paris-Saclay in 2019. In 2019–2020 he was visiting Bernard Ries at University of Fribourg, Switzerland. He received the PhD degree from the University of Rennes in 2023 under the supervision of Aline Roumy and Maël Le Treust at IRISA UMR 6074 and Inria Center of Rennes University.

Maël Le Treust is researcher at CNRS since Oct. 2013, working at IRISA UMR 6074 in Rennes, France, since Sept. 2022. He serves as an associate editor for the IEEE Transactions on Information Theory since 2021. Since 2023, he co-organizes the weekly Game Theory Seminar at Institut Henri Poincaré in Paris, and the mentorship program at IRISA in Rennes. Since 2022, he serves in the technical program committees of the annual IEEE International Symposium on Information Theory (ISIT). In 2025 he co-organizes the workshop “30 Years of Game Theory at Institut Henri Poincaré”. In 2024 and in 2019, he was elected as a member of the scientific council of CNRS Informatics, of which he was secretary from 2019 to 2023. From Oct. 2013 to Aug. 2022, he was researcher at ETIS laboratory UMR 8051, CY Cergy Paris Université, ENSEA, CNRS, where he received the HDR in Dec. 2022. He received the M.Sc. degree in Optimization, Game Theory and Economics (OJME) from the Université de Paris VI in 2008 and the Ph.D. degree in Information Theory from the Université de Paris Sud XI, L2S UMR 8506 Gif-sur-Yvette, in 2011. In 2012–2013, he was a post-doctoral researcher at LIGM UMR 8049 in Marne-la-Vallée, France, and at Université INRS and McGill University in Montréal, Canada. His research interests include Information Theory, Game Theory, Optimization, Physical Layer Security and Wireless Communications.

Aline Roumy is a Research Director at the Inria Center at Rennes University, France, where she leads a research group on data compression. She received her Ph.D. degree in 2000 from the University of Cergy-Pontoise, France. From 2000 to 2001, she was a research associate at Princeton University, Princeton, NJ, and she joined Inria in 2001. She has held visiting positions at Eurecom and the University of California, Berkeley. Her research interests include statistical signal and image processing, coding theory, and information theory. From 2018 to 2023, she served as an Associate Editor for the IEEE Transactions on Image Processing and received the Outstanding Editorial Board Member Award in 2023. Since 2023, she has been a Senior Associate Editor for the IEEE Transactions on Image Processing.