On Complementary Graph Entropy

Ertem Tuncel, *Member, IEEE*, Jayanth Nayak, Prashant Koulgi, *Student Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*—It has been recently discovered that complementary graph entropy characterizes (and offers new insights into) the minimum asymptotic rate for zero-error source coding with decoder side information. This paper presents new results that build on and complement this discovery. Specifically, i) previously unknown subadditivity properties of complementary graph entropy are derived, and ii) zero-error coding rates are characterized in terms of complementary graph entropy for two multiterminal coding scenarios. For both scenarios, the rate characterization implies no rate loss relative to point-to-point source coding.**

***Index Terms*—Chromatic entropy, complementary graph entropy, graph entropy, side information, zero-error capacity, zero-error source coding.**

I. Introduction

COMPLEMENTARY graph entropy was introduced by Körner and Longo [14] for the characterization of achievable coding rates in the following two-step source coding problem.

- **Scenario A:** For a discrete memoryless source (DMS) with finite alphabet $\mathcal{X}$ and underlying distribution P, let the graph $G=(\mathcal{X},\mathcal{E})$, with vertex set $\mathcal{X}$, and edge set $\mathcal{E}$, represent the *distinguishability* relationship between source symbols. That is, two vertices are connected by an edge if and only if they are distinguishable. Also assume that two vectors x_1^n and x_2^n are considered distinguishable if and only if they are distinguishable in at least one component. The task of the first stage encoder is to describe any source vector $X^n \in \mathcal{X}^n$ so that the first stage reconstruction $\hat{X}_1^n$ and X^n are indistinguishable with high probability. The second stage encoder works in a complementary fashion. Given that the decoder has access to *some* reconstruction that is with high probability indistinguishable from X^n, the second stage encoder ensures, also with high probability, perfect reconstruction of the source, i.e., $\hat{X}_2^n = X^n$.

An early result due to Körner [13] states that the minimum achievable rate at the first stage is given by $H(G,P)$, the *graph entropy* of G under the vertex distribution P. In [14], Körner and Longo defined *complementary* graph entropy for any graph F with a distribution Q on its vertices, denoted $\bar{H}(F,Q)$, and argued that there exists a second-stage scheme with a rate of $\bar{H}(\bar{G},P)$ ($\bar{G}$ denotes the complement of G) that works uniformly well on top of any valid first stage scheme.

One of the key results in [13] was a single-letter expression for graph entropy. This alternative expression makes it amenable to analysis and the properties of graph entropy have been extensively studied (cf. [20] and the references therein). However, no single-letter expression for complementary graph entropy is known and it is considerably less understood.

Three decades after its introduction, complementary graph entropy was shown to be the minimum coding rate in an unrelated scenario.

- **Scenario B:** Let $\{X^n,Y^n\}$ be a sequence of independent drawings of a pair of random variables X and Y jointly distributed according to P_{XY}. The side information sequence Y^n is available only at the decoder, and the encoder is required to convey the source sequence X^n to the decoder without error. In contrast with the Slepian–Wolf problem [22] which only requires asymptotically vanishing error $\Pr[\hat{X}^n \neq X^n] \to 0$, it is required here that $\Pr[\hat{X}^n \neq X^n] = 0$ for all $n>0$.

The zero-error requirement of this scenario was first addressed by Witsenhausen [24]. He defined the *confusability* graph $G=(\mathcal{X},\mathcal{E})$ for given P_{XY}, and observed that there is a one-to-one relationship between colorings of confusability graphs and valid fixed-length codes. Later on, Alon and Orlitsky [2] gave a characterization of the minimum achievable rate for variable-length coding in terms of a third type of graph entropy, the *chromatic entropy* H_χ, which they defined as the minimum $H(c(X))$ that can be achieved using a valid coloring $c(\cdot)$. That characterization is given by

$$R_{\min}(G,P_X)=\lim_{n\to\infty}\frac{1}{n}H_\chi(G^n,P_X^n) \quad (1)$$

where P_X is the marginal distribution induced by P_{XY} and G^n is the AND power of G, which is the confusability graph for blocks of length n. More recently, Koulgi *et al.* [12] proved that in fact

$$\lim_{n\to\infty}\frac{1}{n}H_\chi(G^n,P_X^n)=\bar{H}(G,P_X) \quad (2)$$

Manuscript received February 23, 2006; revised February 08, 2009. Current version published May 20, 2009. This work was supported in part by the National Science Foundation under CAREER Grant CCF-0643695, Grants IIS-0329267 and CCF-0728986 , the University of California MICRO Program, Applied Signal Technology Inc., Dolby Laboratories Inc., Qualcomm Inc., and Sony-Ericsson Inc. The material in this paper was presented in part at the IEEE International Symposium on Information Theory (ISIT), Yokohama, Japan, July 2003.

E. Tuncel is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: ertem@ee.ucr.edu).

J. Nayak was with the Department of Electrical Engineering, University of California, Riverside. He is now with Mayachitra, Inc, Santa Barbara, CA 93111 USA (e-mail: nayak@mayachitra.com).

P. Koulgi was with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93506 USA (e-mail: prashant_koulgi@yahoo.co.in).

K. Rose is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93506 USA (e-mail: rose@ece.ucsb.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2009.2018182

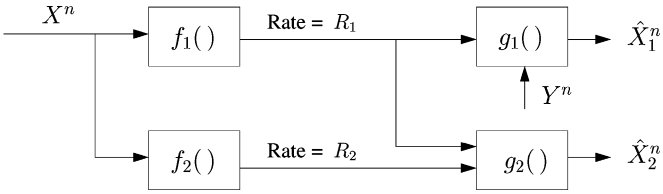

Fig. 1. Scenario C: Two-stage source coding when side information may be absent.

thereby also proving that $R_{\min}(G, P_X) = \bar{H}(G, P_X)$. While it should be noted that there is no known single-letter formula for either side of (2), $\bar{H}(G, P_X)$ offers additional useful insight on the minimum achievable rate for various reasons.

- **Insight 1:** The single-letter bounds on $\bar{H}(G, P_X)$ [14] as well as the conditions for their tightness [7], [15], [14] are directly applicable to the minimum achievable coding rate.
- **Insight 2:** $R_{\min}(G, P_X) = \bar{H}(G, P_X)$ reveals that the following simple variable-length coding scheme is optimal: Encode all the vertices in a high-probability subset of G^n using roughly $n\bar{H}(G, P_X)$ bits and the rest of the vertices with roughly $n \log |\mathcal{X}|$ bits.
- **Insight 3:** A new link between the problem of zero-error coding with decoder side information and the zero-error capacity problem [19] is revealed by applying Marton's equality [18]

$$\bar{H}(G, P) + C(G, P) = H(P) \tag{3}$$

where $C(G, P)$ is the zero-error capacity when the codewords are constrained to have an empirical distribution (or type) P [8]. Thus, any single-letter formula for the minimum achievable rate in the side-information problem would immediately yield a single-letter formula for the long-standing problem of zero-error capacity.

In this paper, we explore further benefits of the discovery of (2). Specifically, based on (2), we i) prove new subadditivity properties of complementary graph entropy, and using those properties derive a single-letter formula for $\bar{H}(G, P)$ for a family of (G, P) pairs, and ii) derive achievable minimum rates for two multiterminal source coding scenarios. We defer the discussion of the new properties of $\bar{H}(G, P_X)$ to Section III, and discuss below the two source coding scenarios we introduce and their relation to the complementary graph entropy.

- **Scenario C:** Consider the two-stage coding scheme shown in Fig. 1, which is an extension of Scenario B and is somewhat reminiscent of Scenario A. In the first stage, expending rate R_1, the encoder f_1 describes X^n in sufficient detail that would ensure reconstruction by decoder g_1 without error, i.e., $\Pr[\hat{X}_1^n = X^n] = 1$, in the presence of the side information Y^n. The complementary second stage description of rate R_2, supplied by the encoder f_2, allows the decoder g_2 to reconstruct X^n without error, i.e., so that $\Pr[\hat{X}_2^n = X^n] = 1$, even in the absence of Y^n. Unlike in Scenario A, the codes at the two stages are not to be chosen independently of each other.

This communication scenario can be useful in a variety of situations. For example, suppose that Y^n is itself transmitted in parallel to the receiver from another source through an unreliable communication link, and thus, can sometimes be absent. It is certainly inefficient to disregard the potential existence of side information, and send X^n expending rate $H(P_X)$. On the other hand, expending a rate of $\bar{H}(G, P_X)$ would be insufficient whenever Y^n is absent. It is therefore appropriate to encode X^n in a scalable fashion. If the transmission of Y^n proves unsuccessful (perhaps after some number of trials), the receiver requests the second-stage information from the main (reliable) source.

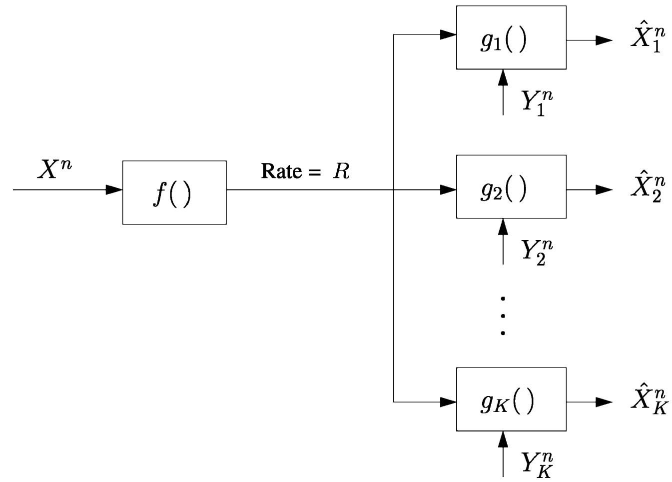

Fig. 2. Scenario D: Source coding with side information at multiple receivers.

Another example is a multicast scenario, where the decoders g_1 and g_2 in Fig. 1 may represent distinct receivers with and without access to the side information, respectively. In this case, it is more efficient to use scalable coding rather than transmit two descriptions independently.

It should also be noted that there is some similarity between Scenario C and what has been referred to in the literature as "successive refinement for the Wyner–Ziv problem" [23], and "rate–distortion when side information may be absent" [10], [11]. However, first of all, both problems are concerned with *lossy* coding. Moreover, in [23], the terminal that receives both layers also has better side information, while in Scenario C, this terminal has no side information. Finally, in [10], [11], there is only one direct description of the source and the side information is itself used as an "enhancement" description for reducing the distortion.

It is evident from (1) and (2) that the first and second stage rates must satisfy the trivial bounds

$$R_1 \geq \bar{H}(G, P_X) \tag{4}$$

$$R_1 + R_2 \geq H(P_X). \tag{5}$$

Using Insight 2 together with (3), we show that there always exists a two-stage variable-length instantaneously decodable scheme which achieves both (4) and (5) with equality.

- **Scenario D:** Suppose that it is desired to broadcast X^n by encoder f to several receivers $g_1, g_2, \ldots, g_K$, each equipped with a different side information vector Y_i^n, as depicted in Fig. 2. What is the asymptotic minimum rate R that a variable-length coder must expend to ensure $\Pr[\hat{X}_i^n = X^n] = 1$ for all i?

An immediate observation is that

$$R \geq \max_i R_{\min}(G_i, P_X) \tag{6}$$

where G_i is the confusability graph for the pair (X, Y_i). Can any zero-error coding scheme in fact achieve the trivial lower bound (6) with equality? Recently, Simonyi considered the corresponding question for the fixed-length version of the problem and proved that the lower bound is indeed achievable [21]. By extending (1) and (2) to multiple receivers, we prove that there exists a variable-length coding scheme that achieves (6) with equality, thereby generalizing Simonyi's result.

Note that both these scenarios become trivial if we only require perfect decoding with an asymptotically vanishing error probability. Again, simple lower bounds can be shown to be achievable using the results of Slepian and Wolf [22]

$$R_1 = H(X \,|\, Y)$$
$$R_1 + R_2 = H(X)$$

for Scenario C and

$$R = \max_i H(X \,|\, Y_i)$$

for Scenario D. However, handling the case where zero-error decoding is required at all block lengths, which we consider in this paper, is considerably more involved.

The organization of the paper is as follows. In the next section, we provide preliminaries. In Section III, we employ (2) to prove new subadditivity properties of complementary graph entropy. We then derive in Sections IV and V the achievable rate regions for Scenarios C and D, respectively. Section VI provides summary and conclusions.

II. Preliminaries

A. Graph-Theoretic Preliminaries

We denote by $G = (\mathcal{X}, \mathcal{E})$ an undirected graph with a finite vertex set $\mathcal{X}$ and edge set $\mathcal{E} \subset \mathcal{P}_2(\mathcal{X})$, where $\mathcal{P}_2(\mathcal{X})$ consists of all distinct pairs of vertices. If $\{x, x'\} \in \mathcal{E}$, we say that x and x' are connected or adjacent in G, and use the notation $x \overset{G}{\text{---}} x'$. We also use $x \overset{G}{\sim} x'$ to indicate that either $x = x'$ or $x \overset{G}{\text{---}} x'$. A graph $G = (\mathcal{X}, \mathcal{E})$ is called *empty* (denoted $E_{\mathcal{X}}$) or *complete* (denoted $K_{\mathcal{X}}$) if $\mathcal{E} = \emptyset$ or $\mathcal{E} = \mathcal{P}_2(\mathcal{X})$, respectively. A complete bipartite graph $K_{\mathcal{X}_1, \mathcal{X}_2}$ is a graph whose vertex set is the union of two disjoint sets $\mathcal{X}_1$ and $\mathcal{X}_2$ and $x \overset{K_{\mathcal{X}_1,\mathcal{X}_2}}{\text{---}} x'$ if and only if $x \in \mathcal{X}_1, x' \in \mathcal{X}_2$. The complement of $G = (\mathcal{X}, \mathcal{E})$ is $\bar{G} = (\mathcal{X}, \mathcal{E}^c)$, where $\mathcal{E}^c = \mathcal{P}_2(\mathcal{X}) \setminus \mathcal{E}$.

We say that $G(\mathcal{X}')$ is a subgraph of G *induced* by $\mathcal{X}' \subset \mathcal{X}$ if its vertex set is $\mathcal{X}'$ and its edge set is $\mathcal{E} \cap \mathcal{P}_2(\mathcal{X}')$. For $G_1 = (\mathcal{X}, \mathcal{E}_1)$ and $G_2 = (\mathcal{X}, \mathcal{E}_2)$, we use the notation $G_1 \subset G_2$ to indicate that $\mathcal{E}_1 \subset \mathcal{E}_2$. A set $\mathcal{X}'$ is called *independent* in G if it induces an empty subgraph. Similarly, a set $\mathcal{X}'$ is called a clique in G if it induces a complete subgraph. The *stability number* $\alpha(G)$ is the cardinality of the largest independent set in G. A coloring is a mapping $c : \mathcal{X} \longrightarrow \mathcal{I} = \{1, 2, \ldots\}$, such that connected vertices receive distinct indices. Coloring corresponds to a partition of the vertex set into independent sets, referred to as color classes. The *chromatic number* of G, $\chi(G)$, is the minimum number of colors that are required for coloring G. The stability and chromatic numbers are closely interrelated. For example, $\chi(G) \geq \alpha(\bar{G})$, and $\alpha(G)\chi(G) \geq |\mathcal{X}|$.

Definition 1: A graph $G = (\mathcal{X}, \mathcal{E})$ is called *perfect* if for all $\mathcal{X}' \subset \mathcal{X}$, $\chi(G(\mathcal{X}')) = \alpha(\bar{G}(\mathcal{X}'))$.

Perfect graphs are an important class of graphs. They include empty and complete graphs. A property of perfect graphs relevant to the purposes of this paper is that if G is perfect, so is its complement $\bar{G}$ [16].

We make extensive use of the AND-products and OR-products which are next defined.

Definition 2: The AND-product of $G_1 = (\mathcal{X}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{X}_2, \mathcal{E}_2)$, denoted by $G_1 \times G_2$ has the vertex set $\mathcal{X}_1 \times \mathcal{X}_2$, and for distinct (x_1, x_2) and (x_1', x_2'), $(x_1, x_2) \overset{G_1 \times G_2}{\text{---}} (x_1', x_2')$ if both $x_1 \overset{G_1}{\sim} x_1'$ and $x_2 \overset{G_2}{\sim} x_2'$.

Definition 3: The OR-product of $G_1 = (\mathcal{X}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{X}_2, \mathcal{E}_2)$, denoted by $G_1 \cdot G_2$ has the vertex set $\mathcal{X}_1 \times \mathcal{X}_2$, and $(x_1, x_2) \overset{G_1 \cdot G_2}{\text{---}} (x_1', x_2')$ if either $x_1 \overset{G_1}{\text{---}} x_1'$ or $x_2 \overset{G_2}{\text{---}} x_2'$.

We denote by $G^n = (\mathcal{X}^n, \mathcal{E}_n)$ and $G^{(n)} = (\mathcal{X}^n, \mathcal{E}_{(n)})$ the n-fold AND- and OR-products of G with itself, respectively. Note that $\overline{G_1 \cdot G_2} = \bar{G}_1 \times \bar{G}_2$, and more specifically, $\overline{G^{(n)}} = \bar{G}^n$.

Next, we define union and disjoint sum of two graphs.

Definition 4: For $G_1 = \{\mathcal{X}, \mathcal{E}_1\}$ and $G_2 = \{\mathcal{X}, \mathcal{E}_2\}$,

$$G_1 \cup G_2 = (\mathcal{X}, \mathcal{E}_1 \cup \mathcal{E}_2).$$

Definition 5: For $G_1 = \{\mathcal{X}_1, \mathcal{E}_1\}$ and $G_2 = \{\mathcal{X}_2, \mathcal{E}_2\}$ with $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$,

$$G_1 + G_2 = (\mathcal{X}_1 \cup \mathcal{X}_2, \mathcal{E}_1 \cup \mathcal{E}_2).$$

B. Types and Typical Sequences

Let $N(a \,|\, x^n)$ denote the number of occurrences of $a \in \mathcal{X}$ in x^n. The *type* of x^n, denoted P_{x^n} is then defined as

$$P_{x^n}(a) = \frac{1}{n} N(a \,|\, x^n), \quad \forall a \in \mathcal{X}.$$

The vector x^n is said to be ϵ-typical with P_X if $|P_{x^n}(a) - P_X(a)| \leq \epsilon$ for all $a \in \mathcal{X}$, and if $P_{x^n}(a) = 0$ whenever $P_X(a) = 0$. We denote by $\mathcal{T}^n_{P_X,\epsilon}$ the set of all x^n that are ϵ-typical with P_X. Many useful properties of types and typical sequences are widely known, for a thorough discussion of which we refer the reader to [5]. For the sake of completeness, we state below the most crucial properties for the purposes of this paper.

Lemma 1: [5, Lemma 1.2.12] If X^n is independent and identically distributed (i.i.d.) $\sim P_X$, then for all $n > n_0(\epsilon)$

$$P_X^n\left(\mathcal{T}^n_{P_X,\epsilon}\right) = \sum_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} P_X^n(x^n) > 1 - \epsilon. \quad (7)$$

Lemma 2: [5, Lemma 1.2.6] For any distribution Q

$$Q^n(x^n) = 2^{-n[H(P_{x^n}) + D(P_{x^n} \,||\, Q)]} \quad (8)$$

where $D(P\,||\,Q)$ denotes the Kullback–Leibler divergence measure given by

$$D(P\,||\,Q) = \sum_{x\in\mathcal{X}} P(x)\log_2\frac{P(x)}{Q(x)}.$$

C. Probabilistic Graphs and Graph Entropies

A probabilistic graph, denoted (G,P), attaches the probability measure P to the vertices of G. We first define the relevant entropies for probabilistic graphs.

Definition 6: Given a probabilistic graph (G,P), its graph entropy is [13]

$$H(G,P) = \lim_{n\to\infty}\frac{1}{n}\log_2\left\{\min_{\mathcal{A}:P^n(\mathcal{A})>1-\epsilon}\chi(G^{(n)}(\mathcal{A}))\right\}. \quad (9)$$

Definition 7: The complementary graph entropy of (G,P) is given by

$$\bar{H}(G,P) = \lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log_2\left\{\chi\left(G^n\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right)\right\}. \quad (10)$$

Unfortunately, it remains unknown whether the outer limit is necessary, or whether the inner $\limsup$ can be replaced by a regular limit.

Graph entropy and complementary graph entropy for probabilistic graphs were defined in [13], [14] motivated by Scenario A and its single-stage variant, where in the first and the second stages, coloring of high-probability subgraphs of $G^{(n)}$ and $\bar{G}^n$ satisfies the coding requirements, respectively. Since $H(G,P)$ and $\bar{H}(\bar{G},P)$ are the minimum achievable rates in the first and second stages, it is immediate from coding arguments that

$$H(G,P) + \bar{H}(\bar{G},P) \geq H(P), \quad (11)$$

as a two-stage scheme cannot outperform the optimal single-stage coding scheme. In [14], Körner and Longo also considered the question of conditions for equality in (11), i.e., conditions for no rate loss due to the two-stage scheme, and derived partial results. Specifically, observing that

$$H(\bar{G},P) \geq \bar{H}(\bar{G},P),$$

they proposed

$$H(G,P) + H(\bar{G},P) = H(P) \quad (12)$$

as a sufficient condition for equality in (11), and showed that (12) holds for *some* nowhere-vanishing P (with $P(x) > 0$ for all $x\in\mathcal{X}$) if G is *normal* (cf. [14]). Subsequent work on conditions for validity of (12) revealed several deeper connections between information theory and graph theory. It was shown in [15] that normality is necessary as well as sufficient for (12) to hold for some nowhere-vanishing P, and in [7] that (12) holds for *all* P if and only if G is *perfect* (cf. [3, Ch. 16]). A corollary of the result in [7] is that if a graph G is perfect, then we also have

$$H(G,P) = \bar{H}(G,P)$$

for all P.

A third graph entropic quantity, the *chromatic entropy*, was defined in [2].

Definition 8: The chromatic entropy of a probabilistic graph is

$$H_\chi(G,P) = \min_{\substack{X\sim P,\\ c(\cdot)\text{ is a coloring of }G}} H(c(X)) \quad (13)$$

These three graph entropies are useful in characterizing achievable rates in Scenario B. Given a source side-information pair (X,Y) distributed according to P_{XY}, the confusability graph [24] is the graph $G = (\mathcal{X},\mathcal{E})$ where $x \overset{G}{\text{—}} x' \Leftrightarrow \exists y\in\mathcal{Y} : P_{XY}(x,y)P_{XY}(x',y) > 0$. Instantaneous variable-length codes for Scenario B are those that satisfy [2]

$$x^n \overset{G^n}{\text{—}} x'^n \Longrightarrow \phi_n(x^n) \neq_P \phi_n(x'^n) \quad (14)$$

where $\neq_P$ indicates "prefix-free." The bit rate for any code ϕ_n is determined by

$$R(\phi_n) = \frac{1}{n}\sum_{x\in\mathcal{X}^n} P^n_X(x^n)|\phi_n(x^n)|$$

and therefore, once G is built, there is no further dependence of the coding algorithm on $P_{Y\,|\,X}$.

It was shown in [2] that the minimum achievable coding rate in Scenario B is given by

$$R_{\min}(G,P_X) = \lim_{n\to\infty}\frac{1}{n}H_\chi\left(G^n,P^n_X\right).$$

Later, it was shown in [12] that $H_\chi(G,P)$ and $\bar{H}(G,P)$ are in fact related via

$$\lim_{n\to\infty}\frac{1}{n}H_\chi\left(G^n,P^n_X\right) = \bar{H}(G,P_X) \quad (15)$$

thereby also showing that

$$R_{\min}(G,P_X) = \bar{H}(G,P_X). \quad (16)$$

Previously, it was known that the normalized limit of the chromatic entropy of the OR-products also yields a familiar expression [2]. Specifically

$$\lim_{n\to\infty}\frac{1}{n}H_\chi\left(G^{(n)},P^n_X\right) = H(G,P_X). \quad (17)$$

Alon and Orlitsky also gave a source coding interpretation of $H(G,P_X)$ for a variant of Scenario B in [2].

Finally, we recall the generalization by Csiszár and Körner [8] of the concept of zero-error capacity defined by Shannon [19].

Definition 9: The capacity of a graph G within type P is

$$C(G,P) = \lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{1}{n}\log_2\left\{\alpha\left(G^n\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right)\right\}.$$

In [18], Marton proved (3), which is a key result for our purposes in this work.

III. New Subadditivity Properties of Complementary Graph Entropy

The condition for the subadditivity of complementary graph entropy with respect to graph union, i.e.,

$$\bar{H}(G_1,P) + \bar{H}(G_2,P) \geq \bar{H}(G_1\cup G_2,P) \quad (18)$$

was previously studied, and it was shown that (18) holds when i) both G_1 and G_2 are perfect [7], and ii) $G_1 \cup G_2$ is perfect [18]. In the next theorem, we prove that a relaxed version of (18) *always* holds.

Theorem 1:

$$\bar{H}(G_1, P) + H(G_2, P) \geq \bar{H}(G_1 \cup G_2, P). \quad (19)$$

Remark 1: The inequality (11) is a special case of this result with $G_2 = G$ and $G_1 = \bar{G}$.

Proof: Let c_1 and c_2 be colorings of graphs F_1 and F_2 that achieve $H_\chi(F_1, P')$ and $H_\chi(F_2, P')$ for some fixed P', respectively. Also, let X' denote the random variable governed by P'. Since $c = (c_1, c_2)$ yields a coloring of $F_1 \cup F_2$, we have

$$\begin{aligned} H_\chi(F_1 \cup F_2, P') &\leq H(c(X')) \\ &= H(c_1(X'), c_2(X')) \\ &\leq H(c_1(X')) + H(c_2(X')) \\ &= H_\chi(F_1, P') + H_\chi(F_2, P') \end{aligned} \quad (20)$$

i.e., H_χ is subadditive with respect to graph union. We take $F_1 = G_1^n$, $F_2 = G_2^{(n)}$, and $P' = P^n$, and also observe that $(G_1 \cup G_2)^n \subset G_1^n \cup G_2^{(n)}$. To prove the latter, consider a pair $\{x^n, x'^n\}$ which is connected in $(G_1 \cup G_2)^n$ and let $\mathcal{J} = \{i : 1 \leq i \leq n, x_i \neq x'_i\}$. One can see that either $x_i \overset{G_1}{\text{---}} x'_i$ or $x_i \overset{G_2}{\text{---}} x'_i$ for all $i \in \mathcal{J}$. Suppose no $i \in \mathcal{J}$ satisfies $x_i \overset{G_2}{\text{---}} x'_i$. Then $x_i \overset{G_1}{\text{---}} x'_i$ must be satisfied for all $i \in \mathcal{J}$, which implies $x^n \overset{G_1^n}{\text{---}} x'^n$. On the other hand, if there is at least one $i \in \mathcal{J}$ with $x_i \overset{G_2}{\text{---}} x'_i$, then $x^n \overset{G_2^{(n)}}{\text{---}} x'^n$.

Now, $(G_1 \cup G_2)^n \subset G_1^n \cup G_2^{(n)}$ implies

$$H_\chi((G_1 \cup G_2)^n, P^n) \leq H_\chi\left(G_1^n \cup G_2^{(n)}, P^n\right) \quad (21)$$

because the coloring achieving the right-hand side is also a (not necessarily optimal) coloring for $(G_1 \cup G_2)^n$. Combining (21) with (20), normalizing by n, and letting $n \to \infty$ yields the desired result due to (15) and (17). □

Corollary 1: The original subadditivity property (18) holds when *either* G_1 or G_2 is perfect.

Proof: Since $H(G, P) = \bar{H}(G, P)$ for any perfect G, (19) and (18) are equivalent when G_2 is perfect. On the other hand, if G_1 is perfect, one can use the same argument on the inequality

$$H(G_1, P) + \bar{H}(G_2, P) \geq \bar{H}(G_1 \cup G_2, P)$$

which, in turn, follows by swapping G_1 and G_2 in (19). □

Now, for Scenario B, let $(G_1, P_X^{(1)})$ and $(G_2, P_X^{(2)})$ be the probabilistic graphs corresponding to two source–side information pairs distributed according to $P_{XY}^{(1)}$ and $P_{XY}^{(2)}$, with alphabets $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_2, \mathcal{Y}_2)$, respectively, such that $\mathcal{X}_1 \cap \mathcal{X}_2 = \emptyset$ and $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$. Suppose that at each time instant, the observed source–side information pair is determined by a Bernoulli trial with parameter α. The corresponding probabilistic graph for this scenario is $(G_1 + G_2, P_X^{(\alpha)})$, where

$$P_X^{(\alpha)}(x) = \begin{cases} \alpha P_X^{(1)}(x), & x \in \mathcal{X}_1 \\ (1-\alpha) P_X^{(2)}(x), & x \in \mathcal{X}_2. \end{cases}$$

Alternatively, if at each time instant, one sample from each pair is observed, $(G_1 \times G_2, P_X^{(1)} P_X^{(2)})$ is the governing probabilistic graph. The following theorem states that complementary graph entropy is subadditive with respect to both of these scenarios, the proof of which is purely based on (16).

Theorem 2:

$$\begin{aligned} &\bar{H}\left(G_1, P_X^{(1)}\right) + \bar{H}\left(G_2, P_X^{(2)}\right) \\ &\quad \geq \bar{H}\left(G_1 \times G_2, P_X^{(1)} P_X^{(2)}\right) \end{aligned} \quad (22)$$

and

$$\begin{aligned} &\alpha \bar{H}\left(G_1, P_X^{(1)}\right) + (1-\alpha) \bar{H}\left(G_2, P_X^{(2)}\right) \\ &\quad \geq \bar{H}\left(G_1 + G_2, P_X^{(\alpha)}\right). \end{aligned} \quad (23)$$

Proof: The concatenation of asymptotically optimal zero-error variable-length codes for $(G_1, P_X^{(1)})$ and $(G_2, P_X^{(2)})$ yields a valid code for $(G_1 \times G_2, P_X^{(1)} P_X^{(2)})$ with rate $\bar{H}(G_1, P_X^{(1)}) + \bar{H}(G_2, P_X^{(2)})$. Since the asymptotically optimal code for $(G_1 \times G_2, P_X^{(1)} P_X^{(2)})$ can only perform better, (22) is proven.

Similarly, the right-hand side of (23) corresponds to the asymptotic rate of the optimal coding strategy. The left-hand side, on the other hand, is the asymptotic rate achieved by separately encoding the two "pure" subsequences of X^n consisting of the $\mathcal{X}_1$-symbols and the $\mathcal{X}_2$-symbols. Note that since $\mathcal{Y}_1 \cap \mathcal{Y}_2 = \emptyset$, the decoder knows how to reconstruct X^n from the decoded subsequences. Therefore, (23) is proven. □

A. An Application of Results on Subadditivity of $\bar{H}(G, P)$

We now derive the complementary graph entropy $\bar{H}(G, P)$ for a class of probabilistic graphs. To date, the only known cases where this can be accomplished is i) when G is perfect, in which case $\bar{H}(G, P) = H(G, P)$, and thus $\bar{H}(G, P)$ has a single-letter characterization, and ii) when $G = G_5$, i.e., the five-cycle or pentagon graph with $\mathcal{X} = \mathcal{X}_5 = \{0, 1, 2, 3, 4\}$ and $x \overset{G_5}{\text{---}} x' \Leftrightarrow x = x' \pm 1 \bmod 5$, and $P = P_5$, the uniform distribution on $\mathcal{X}_5$. In the latter case

$$\bar{H}(G_5, P_5) = \frac{1}{2} \log 5$$

as was shown in [12].

Let $G = (\mathcal{X}, \mathcal{E})$ be an arbitrary perfect graph and P a distribution on its vertices. For all $0 < \alpha < 1$, we characterize the complementary graph entropy of $G_5 + G$ and its complement under the mixture probability distribution $P^{(\alpha)}$ in terms of $\bar{H}(G_5, P_5), \bar{H}(G, P)$, and α using the subadditivity properties derived earlier.

Lemma 3:

$$\bar{H}\left(G_5 + G, P^{(\alpha)}\right) = \frac{\alpha}{2} \log 5 + (1-\alpha) H(G, P)$$

and

$$\bar{H}\left(\overline{G_5 + G}, P^{(\alpha)}\right) = H_b(\alpha) + \frac{\alpha}{2} \log 5 + (1-\alpha) H(\bar{G}, P)$$

where $H_b(p) \triangleq -p \log p - (1-p) \log(1-p)$ is the binary entropy.

Proof: From (23), we have

$$\bar{H}\left(G_5+G,P^{(\alpha)}\right)\leq\alpha\bar{H}(G_5,P_5)+(1-\alpha)\bar{H}(G,P). \quad (24)$$

Turning to $\overline{G_5+G}$, we can decompose it into the union of three graphs $\overline{G_5}+E_{\mathcal{X}}, \bar{G}+E_{\mathcal{X}_5}$, and $K_{\mathcal{X}_5,\mathcal{X}}$. We then have

$$\begin{aligned}
&\bar{H}\left(\overline{G_5+G},P^{(\alpha)}\right)\\
&\overset{(a)}{\leq} H((\bar{G}+E_{\mathcal{X}_5})\cup K_{\mathcal{X}_5,\mathcal{X}},P^{(\alpha)})\\
&\quad+\bar{H}\left(\overline{G_5}+E_{\mathcal{X}},P^{(\alpha)}\right)\\
&\overset{(b)}{\leq} H\left(\bar{G}+E_{\mathcal{X}_5},P^{(\alpha)}\right)+H\left(K_{\mathcal{X}_5,\mathcal{X}},P^{(\alpha)}\right)\\
&\quad+\bar{H}\left(\overline{G_5}+E_{\mathcal{X}},P^{(\alpha)}\right)\\
&\overset{(c)}{\leq} H\left(\bar{G}+E_{\mathcal{X}_5},P^{(\alpha)}\right)\\
&\quad+H\left(K_{\mathcal{X}_5,\mathcal{X}},P^{(\alpha)}\right)+\alpha\bar{H}(\overline{G_5},P_5)\\
&\overset{(d)}{=}(1-\alpha)H(\bar{G},P)+H\left(K_{\mathcal{X}_5,\mathcal{X}},P^{(\alpha)}\right)\\
&\quad+\alpha\bar{H}(\overline{G_5},P_5)\\
&\overset{(e)}{=}(1-\alpha)H(\bar{G},P)+H_b(\alpha)+\alpha\bar{H}(\overline{G_5},P_5) \quad (25)
\end{aligned}$$

where (a) follows from Theorem 1, (b) from subadditivity of graph entropy under graph union [20, Lemma 3.2], (c) from (23) and using $\bar{H}(E_{\mathcal{X}},P)=0, (d)$ from additivity of graph entropy under graph sums [20, Corollary 3.4], and finally, (e) from the fact that the graph entropy of $K_{\mathcal{X}_5,\mathcal{X}}$ is simply given by the entropy of the indicator function $1_{\mathcal{X}}$ [20, Proposition 3.6].

Now, since $(G_5+G)\cup\overline{G_5+G}=K_{\mathcal{X}_5\cup\mathcal{X}}$, which is perfect, Marton's subadditivity result [18] implies that

$$\bar{H}\left(G_5+G,P^{(\alpha)}\right)+\bar{H}(\overline{G_5+G},P^{(\alpha)})\geq H(P^{(\alpha)}). \quad (26)$$

On the other hand, combining (24) and (25), we obtain

$$\begin{aligned}
&\bar{H}\left(G_5+G,P^{(\alpha)}\right)+\bar{H}\left(\overline{G_5+G},P^{(\alpha)}\right)\\
&\leq H_b(\alpha)+\alpha\left[\bar{H}(G_5,P_5)+\bar{H}(\overline{G_5},P_5)\right]\\
&\quad+(1-\alpha)\left[H(G,P)+H(\bar{G},P)\right]\\
&=H_b(\alpha)+\alpha\log 5\\
&\quad+(1-\alpha)\left[H(G,P)+H(\bar{G},P)\right] \quad (27)\\
&=H_b(\alpha)+\alpha\log 5+(1-\alpha)H(P). \quad (28)
\end{aligned}$$

To see (27), recall that $\bar{H}(G_5,P_5)=\frac{1}{2}\log 5$ is known [12] and $\bar{H}(G_5,P_5)=\bar{H}(\overline{G_5},P_5)$ because $\overline{G_5}$ itself is a pentagon. Equation (28) follows since G is perfect. But

$$H(P^{(\alpha)})=H_b(\alpha)+\alpha\log 5+(1-\alpha)H(P)$$

and hence the upper and lower bounds on $\bar{H}(G_5+G,P^{(\alpha)})+\bar{H}(\overline{G_5+G},P^{(\alpha)})$ are the same, implying that the upper bounds in both (24) and (25) are tight and the lemma is proved. □

By virtue of Insight 3, this result directly implies the following corollary about the capacity within a type.

Corollary 2:

$$C(G_5+G,P^{(\alpha)})=H_b(\alpha)+\frac{\alpha}{2}\log 5+(1-\alpha)H(\bar{G},P)$$

and

$$C(\overline{G_5+G},P^{(\alpha)})=\frac{\alpha}{2}\log 5+(1-\alpha)H(G,P).$$

IV. Achievable Rate Region For Scenario C

Let $\phi_n^{(1)}:\mathcal{X}^n\to\{0,1\}^*$ and $\phi_n^{(2)}:\mathcal{X}^n\to\{0,1\}^*$ be the first- and the second-stage encoders, respectively. Also denote by $\varphi_n^{(1)}:\{0,1\}^*\times\mathcal{Y}^n\to\mathcal{X}^n$ and $\varphi_n^{(2)}:\{0,1\}^*\times\{0,1\}^*\to\mathcal{X}^n$ the decoders at the two stages. The corresponding rates are

$$R_i\left(\phi_n^{(i)}\right)=\frac{1}{n}\sum_{x^n\in\mathcal{X}^n}P_X^n(x^n)\left|\phi_n^{(i)}(x^n)\right|$$

for $i=1,2$. Our goal is to characterize the region of rates achieved by instantaneous codes that ensure both

$$\Pr\left[X^n\neq\varphi_n^{(1)}\left(\phi_n^{(1)}(X^n),Y^n\right)\right]=0$$

and

$$\Pr\left[X^n\neq\varphi_n^{(2)}\left(\phi_n^{(1)}(X^n),\phi_n^{(2)}(X^n)\right)\right]=0.$$

We will focus on instantaneous coding schemes where the first-stage-encoded bitstream can be uniquely parsed even without the help of the side information, to which the second-stage decoder does not have access. Note that in contrast with the single-stage instantaneous coding mechanism of Scenario B, i.e., (14), this prohibits possible use of codewords which prefix each other. Instead, instantaneous coding of a coloring of G^n, i.e., the decomposition $\phi_n^{(1)}=\psi_n\circ c_n$, where ψ_n is a prefix-free code, is adopted. Both $\varphi_n^{(1)}$ and $\varphi_n^{(2)}$ can then instantaneously decode the color, and must utilize it together with Y^n and $\phi_n^{(2)}(X^n)$, respectively, to successfully decode X^n. This in particular implies that $\phi_n^{(2)}(X^n)$ must be instantaneously decodable when the color is known.

We now state the achievable rate region.

Theorem 3: There exists an instantaneous two-step coding strategy $(\phi_n^{(1)},\phi_n^{(2)})$ that achieves

$$\lim_{n\to\infty}R_1\left(\phi_n^{(1)}\right)=\bar{H}(G,P_X) \quad (29)$$

$$\lim_{n\to\infty}\left[R_1\left(\phi_n^{(1)}\right)+R_2\left(\phi_n^{(2)}\right)\right]=H(P_X) \quad (30)$$

where G is the confusability graph for P_{XY}.

Proof: Define

$$\bar{H}_\epsilon(G,P_X)=\limsup_{n\to\infty}\frac{1}{n}\log_2\chi\left(G^n\left(\mathcal{T}_{P_X,\epsilon}^n\right)\right)$$

and

$$C_\epsilon(G,P_X)=\limsup_{n\to\infty}\frac{1}{n}\log_2\alpha\left(G^n\left(\mathcal{T}_{P_X,\epsilon}^n\right)\right)$$

so that

$$\bar{H}(G,P_X)=\lim_{\epsilon\to 0}\bar{H}_\epsilon(G,P_X).$$

and

$$C(G,P_X)=\lim_{\epsilon\to 0}C_\epsilon(G,P_X).$$

These definitions imply, for any $\delta > 0$ and $n > n_1(\delta)$, the existence of a coloring $c_{\delta,n}(\cdot)$ of G^n satisfying

$$|c_{\delta,n}(\mathcal{T}^n_{P_X,\epsilon})| \le 2^{n[\bar{H}_\epsilon(G,P_X)+\delta]}$$

and

$$|c^{-1}_{\delta,n}(i)| \le 2^{n[C_\epsilon(G,P_X)+\delta]}$$

for all $i \in c_{\delta,n}(\mathcal{T}^n_{P_X,\epsilon})$. The latter follows from the fact that each color class is, by definition, an independent set. We can then use the following coding strategy: Fix $\epsilon, \delta > 0$ and $n > \max\{n_0(\epsilon), n_1(\delta)\}$. Denote by $\beta(\cdot|\mathcal{A}) : \mathcal{A} \to \{0,1\}^{\lceil \log_2 |\mathcal{A}| \rceil}$ the fixed-length canonical representation of elements in $\mathcal{A}$. Set the first stage encoder to

$$\phi^{(1)}_n(x^n) = \begin{cases} 0 \cdot \beta(x^n|\mathcal{X}^n), & x^n \notin \mathcal{T}^n_{P_X,\epsilon} \\ 1 \cdot \beta\left(c_{\delta,n}(x^n)|c_{\delta,n}\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right), & x^n \in \mathcal{T}^n_{P_X,\epsilon} \end{cases}$$

where $\cdot$ denotes concatenation. The decoder $\varphi^{(1)}_n$ can instantaneously recover all $x^n \notin \mathcal{T}^n_{P_X,\epsilon}$, and $i = c_{\delta,n}(x^n)$ whenever $x^n \in \mathcal{T}^n_{P_X,\epsilon}$. In the latter case, x^n is also easily recovered since it is the unique vertex which is simultaneously in both $c^{-1}_{\delta,n}(i)$ and the clique in G^n induced by y^n. The resultant first stage rate is bounded by

$$\begin{aligned} R_1\left(\phi^{(1)}_n\right) &\le \frac{1}{n} + P^n_X\left(\mathcal{T}^n_{P_X,\epsilon}\right)\left[\bar{H}_\epsilon(G,P_X)+\delta\right] \\ &\quad + \left[1 - P^n_X\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right]\log_2|\mathcal{X}| \\ &\le \bar{H}_\epsilon(G,P_X) + 2\delta + \epsilon\log_2|\mathcal{X}| \end{aligned}$$

for sufficiently large n. Setting the second-stage encoder to

$$\phi^{(2)}_n(x^n) = \begin{cases} \lambda, & x^n \notin \mathcal{T}^n_{P_X,\epsilon} \\ \beta\left(x^n|c^{-1}_{\delta,n}(i)\right), & x^n \in \mathcal{T}^n_{P_X,\epsilon}, c_{\delta,n}(x^n) = i \end{cases}$$

where λ is the null codeword, we ensure that x^n is instantaneously recovered in the absence of the side information. Therefore, the second-stage rate is bounded by

$$\begin{aligned} R_2\left(\phi^{(2)}_n\right) &\le P^n_X(\mathcal{T}^n_{P_X,\epsilon})[C_\epsilon(G,P_X)+\delta] \\ &\le C_\epsilon(G,P_X) + \delta. \end{aligned}$$

Letting $n \to \infty$ and then $\delta, \epsilon \to 0$ yields an asymptotic rate pair not larger than $\left(\bar{H}(G,P_X), C(G,P_X)\right)$. The proof is complete observing (3), (4), and (5). □

V. Achievable Rates for Scenario D

We will utilize $\mathcal{G} = \{G_1, G_2, \ldots, G_K\}$, the family of confusability graphs for pairs $(X, Y_i), i = 1, \ldots, K$, in Scenario D. For fixed-length coding, it has been observed in [21] that valid block codes have a one-to-one correspondence to colorings of

$$\mathcal{G}^n \triangleq \bigcup_{i=1}^{K} G^n_i.$$

That is because the encoder cannot send the same message for two different outcomes of X^n if they correspond to adjacent vertices in *any* G^n_i. On the other hand, a coloring of $\cup_i G^n_i$ is automatically a coloring for all G^n_i, and hence, each receiver can uniquely decode X^n upon knowing its color in $\cup_i G^n_i$ and the side information Y^n_i.

Turning to variable-length coding, an instantaneous code $\phi_n : \mathcal{X}^n \longrightarrow \{0,1\}^*$ must satisfy

$$x^n \overset{\mathcal{G}^n}{\text{---}} x'^n \Longrightarrow \phi_n(x^n) \neq_P \phi_n(x'^n).$$

Therefore, similar to Scenario B, the minimum achievable bit rate is purely characterized by $(\mathcal{G}, P_X)$, and there is no further dependence on $P_{X,Y_1,\ldots,Y_K}$. Denote by $R_{\min}(\mathcal{G}, P_X)$ the minimum asymptotically achievable rate. We now characterize this rate using chromatic entropy, thereby extending (1) to more than one receiver.

Lemma 4:

$$R_{\min}(\mathcal{G}, P_X) = \lim_{n\to\infty} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P^n_X\right). \tag{31}$$

Proof: As in Scenario B, the simplest instantaneous variable-length codes are constructed as $\phi_n = \psi_n \circ c_n$, where c_n is a coloring of $\mathcal{G}^n$ and ψ_n is a prefix-free coding of the colors. The rate achieved by this scheme is upper-bounded as

$$R(\phi_n) \le \frac{H_\chi\left(\mathcal{G}^n, P^n_X\right) + 1}{n}.$$

On the other hand, it is possible to decompose any instantaneous code as $\phi_n = \theta_n \circ c_n$, where θ_n is a one-to-one coding of the colors. Therefore

$$nR(\phi_n) \ge H_\chi\left(\mathcal{G}^n, P^n_X\right) - \log_2\left\{H_\chi\left(\mathcal{G}^n, P^n_X\right) + 1\right\} - \log_2 e. \tag{32}$$

To simplify this lower bound, we first prove subadditivity of chromatic entropy with respect to AND-power for families of graphs, i.e.,

$$H_\chi\left(\mathcal{G}^{m+n}, P^{m+n}_X\right) \le H_\chi\left(\mathcal{G}^m, P^m_X\right) + H_\chi\left(\mathcal{G}^n, P^n_X\right). \tag{33}$$

It is clear that the Cartesian product of colorings $c_m(\cdot)$ and $c_n(\cdot)$ for $\mathcal{G}^m$ and $\mathcal{G}^n$, respectively, constitutes a coloring for $\mathcal{G}^m \times \mathcal{G}^n$. Also observe that

$$\cup_i(G_i \times H_i) \subset \cup_i G_i \times \cup_i H_i$$

which implies

$$\cup_i G^{m+n}_i \subset \cup_i G^m_i \times \cup_i G^n_i.$$

Hence, any coloring for $\mathcal{G}^m \times \mathcal{G}^n$ is also a coloring for $\mathcal{G}^{m+n}$. In particular, if c_m and c_n above achieve $H_\chi(\mathcal{G}^m, P^m_X)$ and $H_\chi(\mathcal{G}^n, P^n_X)$, respectively, then $c_m \times c_n$ is a (potentially suboptimal) coloring for $\mathcal{G}^{m+n}$, proving (33).

Since subadditivity implies $H_\chi(\mathcal{G}^n, P^n_X) \le nH_\chi(\mathcal{G}, P_X)$, (32) becomes

$$\begin{aligned} nR(\phi_n) \ge H_\chi(\mathcal{G}^n, P^n_X) - \log_2\{H_\chi(\mathcal{G}, P_X) + 1\} \\ - \log_2 n - \log_2 e. \end{aligned}$$

The lemma follows by combining the upper and lower bounds on $nR(\phi_n)$ and again using the subadditivity to infer the existence of the limit in (31). □

In [4], the capacity of a probabilistic graph was generalized to a family of graphs, i.e., $K > 1$ as follows.

Definition 10: The capacity of $(\mathcal{G}, P)$ is given by

$$C(\mathcal{G}, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log_2 \alpha\left(\mathcal{G}^n\left(\mathcal{T}^n_{P,\epsilon}\right)\right). \tag{34}$$

Similarly, Simonyi [21] generalized the definition of complementary graph entropy to $K > 1$.

Definition 11: The complementary graph entropy $\bar{H}(\mathcal{G}, P)$ of a family of graphs $\mathcal{G}$ is the value

$$\bar{H}(\mathcal{G}, P) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log_2 \chi\left(\mathcal{G}^n\left(\mathcal{T}^n_{P,\epsilon}\right)\right).$$

Simonyi [21, Lemma 1] also generalized (3) to

$$\bar{H}(\mathcal{G}, P) = H(P) - C(\mathcal{G}, P). \tag{35}$$

Using this result, we now generalize (2) to a family of graphs.

Lemma 5:

$$\lim_{n \to \infty} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P_X\right) = \bar{H}(\mathcal{G}, P_X).$$

Proof: The proof follows generally the same lines as in the Proof of Theorem 1 in [12]. Let

$$\bar{H}_\epsilon(\mathcal{G}, P_X) = \limsup_{n \to \infty} \frac{1}{n} \log_2 \chi\left(\mathcal{G}^n\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right)$$

which implies that for any fixed $\delta > 0$ and for each $n \geq n_2(\delta)$ there exists a coloring $c(\cdot)$ of $\mathcal{G}^n$ satisfying

$$\left|c\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right| \leq 2^{n[\bar{H}_\epsilon(\mathcal{G}, P_X)+\delta]}.$$

Let $\Phi : \mathcal{X}^n \longrightarrow \{0, 1\}$ be the indicator function for the set $\mathcal{T}^n_{P_X,\epsilon}$. Then, for large enough n

$$\begin{aligned} H_\chi(\mathcal{G}^n, P^n_X) &\leq H(c(X^n)) \\ &\leq H(\Phi) + H(c(X^n)|\Phi) \\ &\leq 1 + H(c(X^n)|\Phi = 1) \\ &\quad + \Pr[\Phi = 0] H(c(X^n)|\Phi = 0) \\ &\leq 1 + n[\bar{H}_\epsilon(\mathcal{G}, P_X) + \delta + \epsilon \log_2 |\mathcal{X}|] \end{aligned}$$

where in the last inequality, we used (7). Therefore, normalizing both sides by n and taking limits $n \to \infty$ and $\epsilon, \delta \to 0$, in that order, yields

$$\lim_{n \to \infty} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P^n_X\right) \leq \bar{H}(\mathcal{G}, P_X).$$

For the reversed inequality, consider the coloring $c(\cdot)$ of $\mathcal{G}^n$ achieving $H_\chi(\mathcal{G}^n, P^n_X)$. An elementary lower bound for the entropy of an arbitrary distribution Q over the set $\mathcal{Q}$ is

$$H(Q) \geq -Q(\mathcal{S}) \log_2 \max_{j \in \mathcal{S}} Q(j)$$

for any $\mathcal{S} \subset \mathcal{Q}$. Using this inequality, we have

$$H_\chi\left(\mathcal{G}^n, P^n_X\right) \geq -P^n_X\left(\mathcal{T}^n_{P_X,\epsilon}\right) \log_2 \max_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} P^n_X(c(x^n)).$$

The maximum cardinality of a "color class" within $\mathcal{T}^n_{P_X,\epsilon}$ cannot exceed the maximum independent set size $\alpha(\mathcal{G}^n(\mathcal{T}^n_{P_X,\epsilon}))$. Thus

$$\begin{aligned} H_\chi\left(\mathcal{G}^n, P^n_X\right) \geq -P^n_X\left(\mathcal{T}^n_{P_X,\epsilon}\right) \Bigg\{ &\log_2 \alpha\left(\mathcal{G}^n\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right) \\ &+ \log_2 \max_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} P^n_X(x^n) \Bigg\}. \end{aligned} \tag{36}$$

From (8), we observe

$$\begin{aligned} &\log_2 \max_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} P^n_X(x^n) \\ &\quad = -n \min_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} \{H(P_{x^n}) + \mathcal{D}(P_{x^n}||P_X)\} \\ &\quad \leq -n \min_{x^n \in \mathcal{T}^n_{P_X,\epsilon}} H(P_{x^n}) \\ &\quad \leq -n\{H(P_X) + \epsilon|\mathcal{X}| \log_2 \epsilon\} \end{aligned} \tag{37}$$

where (37) follows from the definition of the typical set and the uniform continuity of the entropy [5, Lemma 1.2.7]. Substituting this into (36), we obtain

$$\begin{aligned} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P^n_X\right) \geq (1-\epsilon) \Bigg\{ &H(P_X) + \epsilon|\mathcal{X}| \log_2 \epsilon \\ &- \frac{1}{n} \log_2 \alpha\left(\mathcal{G}^n\left(\mathcal{T}^n_{P_X,\epsilon}\right)\right) \Bigg\}. \end{aligned}$$

Taking the $\liminf_{n \to \infty}$ of both sides and then letting $\epsilon \to 0$, we get (cf. (34) and (35))

$$\lim_{n \to \infty} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P^n_X\right) \geq H(P_X) - C(\mathcal{G}, P_X) = \bar{H}(\mathcal{G}, P_X).$$

□

Thus far, we have proved that

$$R_{\min}(\mathcal{G}, P_X) = \lim_{n \to \infty} \frac{1}{n} H_\chi\left(\mathcal{G}^n, P^n_X\right) = \bar{H}(\mathcal{G}, P_X)$$

where none of the formulas are single-letter. However, the latter characterization proves necessary in deriving the main result of this section, which we present below.

Theorem 4:

$$R_{\min}(\mathcal{G}, P_X) = \max_i R_{\min}(G_i, P_X)$$

where $R_{\min}(G_i, P_X)$ is the minimum achievable rate in Scenario B for the pair (X, Y_i).

Proof: The proof is trivially similar to that of Theorem 1 in [21]. However, we provide it here for completeness. We need the key result of Gargano, Körner, Vaccaro [9] stating that for any family of graphs $\mathcal{G}$ and probability distribution P

$$C(\mathcal{G}, P) = \min_i C(G_i, P).$$

Then, by Lemma 5

$$\begin{aligned} R_{\min}(\mathcal{G}, P_X) &= \bar{H}(\mathcal{G}, P_X) \\ &= H(P_X) - C(\mathcal{G}, P_X) \end{aligned}$$

$$= H(P_X) - \min_i C(G_i, P_X)$$
$$= \max_i [H(P_X) - C(G_i, P_X)]$$
$$= \max_i \bar{H}(G_i, P_X)$$
$$= \max_i R_{\min}(G_i, P_X). \qquad \square$$

Thus, asymptotically, no rate loss is incurred because of the need to transmit information to more than one receiver. In other words, conveying X^n to multiple receivers is no harder than conveying it to the most needy receiver.

VI. Conclusion

In light of the fact that complementary graph entropy characterizes the minimum asymptotic zero-error rate in the coding scenario where the decoder has access to side information unknown to the encoder, we derived new results regarding i) subadditivity properties of complementary graph entropy and ii) zero-error coding rates for two novel multiterminal coding scenarios.

The subadditivity properties turned out to be useful in deriving a single-letter formula for $\bar{H}(G, P)$ for a family of (G, P) pairs. Prior to this work, the only known cases where this could be accomplished was i) when G was perfect, and ii) when G was the pentagon graph and P was the uniform distribution on the vertices.

In the first coding scenario, considering that side information may sometimes be absent, we proposed a two-layer scheme where only the first-layer decoder has access to side information. We then proved that for all source–side information pairs characterized by (G, P), one can achieve a first-layer rate of $\bar{H}(G, P)$ and a total rate of $H(P)$. Thus, there is no rate loss in either layer with respect to point-to-point source coding.

In the second scenario, there are multiple decoders each having access to a different side information, and the encoder broadcasts the source information. It turns out that there exist coding schemes that asymptotically attain a trivial lower bound: the largest of the individual rates that would be necessary to convey the source to each decoder in a point-to-point fashion.

It is noteworthy that, in the less restrictive world of "vanishingly small error," the counterparts of the above coding results are relatively straightforward to show, whereas the zero-error setting requires quite involved proofs. Specifically, we found it highly beneficial to make use of several graph-theoretic results that had been derived previously in different contexts.

References

[1] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1670–1672, Sep. 1994.

[2] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1329–1339, Sep. 1996.

[3] C. Berge, *Graphs and Hypergraphs*. Amsterdam, The Netherlands: North-Holland, 1973.

[4] G. Cohen, J. Körner, and G. Simonyi, "Zero-error capacities and very different sequences," in *Sequences: Combinatorics, Compression, Security and Transmission*, R. M. Capocelli, Ed. Berlin, Germany: Springer-Verlag, pp. 144–155.

[5] I. Csiszár and J. Körner, *Information Theory, Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1982.

[6] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channel for maximum probability of error," *Z. Wahrsch. Verw. Gebiete*, vol. 57, pp. 87–101, 1981.

[7] I. Csiszár, J. Körner, L. Lovász, K. Marton, and G. Simonyi, "Entropy splitting for antiblocking corners and perfect graphs," *Combinatorica*, vol. 10, no. 1, 1990.

[8] I. Csiszár and J. Körner, "On the capacity of the arbitrarily varying channel for maximum probability of error," *Z. Wahrsch. Verw. Gebiete*, vol. 57, pp. 87–101, 1981.

[9] L. Gargano, J. Körner, and U. Vaccaro, "Capacities: From information theory to extremal set theory," *J. Comb. Theory Ser. A*, vol. 68, no. 2, pp. 296–316, Nov. 1994.

[10] C. Heegard and T. Berger, "Rate-distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 727–734, Nov. 1985.

[11] A. H. Kaspi, "Rate-distortion function when side information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[12] P. Koulgi, E. Tuncel, S. Regunathan, and K. Rose, "On zero-error source coding with decoder side information," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 99–111, Jan. 2003.

[13] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in *Trans. 6th Prague Conf. Inform. Theory, etc.*, Prague, Czechoslovakia, 1973, pp. 411–425.

[14] J. Körner and G. Longo, "Two-step encoding of finite memoryless sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 778–782, Nov. 1973.

[15] J. Körner and K. Marton, "Graphs that split entropies," *SIAM J. Discr. Math.*, vol. 1, pp. 71–79, 1988.

[16] L. Lovász, "Normal hypergraphs and the perfect graph conjecture," *Discr. Math.*, vol. 2, pp. 253–267, 1972.

[17] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 1, pp. 1–7, Jan. 1979.

[18] K. Marton, "On the Shannon capacity of probabilistic graphs," *J. Comb. Theory Ser. B*, vol. 57, no. 2, pp. 183–195, Mar. 1993.

[19] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. IT-2, no. 3, pp. 8–19, Sep. 1956.

[20] G. Simonyi, "Graph entropy: A survey," *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, vol. 20, pp. 399–441, 1995.

[21] G. Simonyi, "On Witsenhausen's zero-error rate for multiple sources," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3258–3261, Dec. 2003.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[23] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner–Ziv problem," *IEEE Trans. Inf. Theory*, vol. IT-50, no. 8, pp. 1636–1654, Aug. 2004.

[24] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 5, pp. 592–593, Sep. 1976.

Ertem Tuncel (S'99–M'04) received the Ph.D. degree in electrical and computer engineering from University of California, Santa Barbara, in 2002.

In 2003, he joined the Department of Electrical Engineering, University of California, Riverside, as an Assistant Professor. His research interests include rate–distortion theory, multiterminal source coding, joint source–channel coding, zero-error information theory, and content-based retrieval in high dimensional databases.

Prof. Tuncel received the 2007 National Science Foundation CAREER Award.

Jayanth Nayak received the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2005.

He is currently a Research Staff Member at Mayachitra, Inc., in Santa Barbara. His research interests include information theory, pattern recognition, and computer vision.

Prashant Koulgi (S'99) received the Ph.D. degree from the University of California, Santa Barbara, in 2002.

Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree in 1991 from the California Institute of Technology, Pasadena.

He then joined the Department of Electrical and Computer Engineering, University of California, Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate–distortion theory, source and source–channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Prof. Rose was corecipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.