Graph-Entropic Characterization of Optimal Zero-Error Coding with Side Information

Prashant Koulgi, Ertem Tuncel, Shankar Regunathan and Kenneth Rose
{prashant, ertem, otto, rose}@ece.ucsb.edu
Department of ECE, University of California, Santa Barbara, CA 93106

I. Introduction

The problem of zero-error source coding when the decoder has side-information unknown to the encoder is considered. With the advent of networks such as the Internet, distributed storage and retrieval of very large databsaes is seen as a promising possibility. Recently, this has renewed interest in multi-terminal source coding frameworks such as distributed source codes. The scenario of the *side-information problem* - where the encoder tries to exploit side information about the source available to the decoder but not to itself - is important both as a canonical distributed source coding system, and as a fundamental building block of more intricate real-world systems. The zero-error version of this problem, apart from its significance in practical applications, has also been studied due to its connections with basic graph-theoretic quantities.

We characterize the minimum asymptotic zero-error rate for variable-length side-information codes as the *complementary graph entropy* of an associated graph. We then briefly discuss some new properties of the complementary graph entropy revealed by this result. After formulating the problem in section 2, we summarize and discuss our results in section 3. In section 4 we prove the main result of the paper.

II. Preliminaries

Let (X, Y) be pair of random variables distributed over a finite product set $V \times U$ according to a probability distribution $P(x, y)$. A sender $\mathcal{P}_x$ knows X while a receiver $\mathcal{P}_y$ knows Y and wants to learn X *without error*. We assume: 1) communication is permitted only from $\mathcal{P}_x$ to $\mathcal{P}_y$; 2) there are no transmission errors; 3) $\mathcal{P}_y$ must be able to tell when $\mathcal{P}_x$'s codeword ends; 4) both communicators use an agreed-upon code designed with knowledge of the underlying probability distribution P. We call this the *side-information problem*.

Distinct $x, x' \in V$ are *confusable* if there is a $y \in U$ such that $P(x, y) > 0$ and $P(x', y) > 0$. Two confusable letters may not be assigned the same codeword in any valid code. Thus confusability defines a binary symmetric relation on the letters of V. Witsenhausen, [1], captured this confusability relation of the source pair (X, Y) in the *characteristic graph* G. $G = (V, E)$ is defined on the vertex set V, and distinct $x, x' \in V$ are connected by an edge if they are confusable. The pair (G, P) denotes the probabilistic graph consisting of $G = (V, E)$ together with the distribution P over its vertices. (Here we denote also by P the marginal distribution on V derived from $P(x, y)$.)

Variable-length codes for the side-information problem were introduced by Alon and Orlitsky in [2]. A valid code for (G, P) is a mapping $\phi : V \to \{0, 1\}^*$ such that if $(x, x') \in E$ then $\phi(x)$ is not a prefix of $\phi(x')$. The rate of a code ϕ is the expected number of bits transmitted:

$$\bar{l}(\phi) = \sum_{x \in V} P(x)|\phi(x)|. \tag{1}$$

To define variable-length block codes, we extend the notion of confusability to vectors. Thus, distinct $x^n = (x_1, x_2, \cdots, x_n), x'^n = (x'_1, x'_2, \cdots, x'_n) \in V^n$ are confusable if every distinct pair $(x_i, x'_i), i = 1, 2, \cdots, n$ is confusable. The characteristic graph for (X^n, Y^n) is then the so-called n-fold *normal* product of G with itself, denoted G^n. $G^n = (V^n, E_n)$, with $(x^n, x'^n) \in E_n$ if $(x_i, x'_i) \in E$ for all distinct pairs $x_i, x'_i \in V, i = 1, 2, \cdots, n$. We denote by P^n the product distribution induced on V^n by P: $P^n((x_1, x_2, \cdots, x_n)) = \Pi_{i=1}^n P(x_i)$. The previous definitions of valid codes for (G, P) may now be extended to block codes for (G^n, P^n).

We shall briefly summarize some standard notations and concepts from graph theory, which we will use extensively in the sequel (see, for example, [10]). We assume that all graphs are undirected and have no loops or multiple edges. For our purposes, these assumptions do not entail loss of generality. Two distinct nodes are connected in $\bar{G}$ - the complement of $G = (V, E)$ - if they are not connected in G. A subset of the vertex set V is an independent set of G if it induces an edge-free subgraph in G. Let $\alpha(G)$ - the *independence number* of G - be the maximum size of an independent set of G, and let $\chi(G)$ - the *chromatic number* of G - be the minimum cardinality of a coloring of G, i.e., a partition of V into independent sets. It is clear that $\alpha(G) \leq \chi(\bar{G})$, and $\alpha(G)\chi(G) \geq |V|$. The subgraph $G' = (V', E')$ induced in G by a subset $V' \subseteq V$ is called an induced subgraph. G is a *perfect* graph if $\chi(G') = \alpha(\bar{G}')$ for every induced subgraph G' of G. For the extensive literature on perfect graphs, see [10], and the references therein.

Finally, note that all logarithms are to base two.

This work was supported in part by the NSF under grant no. MIP-9707764, EIA 9986057, the UC MICRO program, Conexant Systems, Inc., Dolby Labs, Inc., Lucent Technologies, Inc., Medio Stream, Inc., and Qualcomm, Inc.

III. Summary of Results

A. Characterization of Minimum Asymptotic Rate

Let $\bar{L}(G^n, P^n)$ denote the minimum rate of a valid variable-length block code for (G^n, P^n). The minimum asymptotic rate per source letter required for the side information problem is

$$R^* = \lim_{n\to\infty} \frac{\bar{L}(G^n, P^n)}{n}. \tag{2}$$

The characterization of R^* was first considered by Alon and Orlitsky in [2]. They defined the *chromatic entropy* of a probabilistic graph, $H_\chi(G, P)$, as the minimum entropy of its colorings. They then showed that

$$R^* = \lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n}, \tag{3}$$

but a single-letter characterization of R^* remained elusive.

In Section 4, we build on the results of Alon and Orlitsky to characterize the minimum asymptotic rate as the *complementary graph entropy*, $\bar{H}(G, P)$, of the characteristic graph (G, P). In particular, we prove that

$$\lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} = \bar{H}(G, P). \tag{4}$$

Motivated by a two-step source coding problem, [3], Körner and Longo defined two information-theoretic functionals on probabilistic graphs: the *graph entropy*, $H(G, P)$, and the *complementary graph entropy*, $\bar{H}(G, P)$ (this is also referred to as the co-entropy or the π-entropy in the literatture). They then showed that these quantities characterize the minimum asymptotic rates for the coding problems they considered. While Körner derived a formula for $H(G, P)$ in [4], no formula is currently known for $\bar{H}(G, P)$. In [5], Marton revealed the close connection between the complementary graph entropy and the Shannon (zero-error) capacity, [6]. Thus, a formula for the complementary graph entropy of an arbitrary probabilistic graph would imply, via her results, a formula for the Shannon capacity of the corresponding graph. This, in turn, would resolve a major unsolved problem of information theory and graph theory.

Upper and lower bounds for $\bar{H}(G, P)$ have been studied by Csiszár, Körner, Marton and others. In [3], Körner and Longo established bounds for $\bar{H}(G, P)$ in terms of $H(G, P)$ and $H(\bar{G}, P)$:

$$H(P) - H(\bar{G}, P) \le \bar{H}(G, P) \le H(G, P), \tag{5}$$

where $H(P)$ is the Shannon entropy of P.

Csiszár et al., in [7], showed that both the bounds above are tight for all distributions P if the graph G is perfect. Other bounds on $\bar{H}(G, P)$ include Marton's bounds in [5], in terms of a generalization of the Lovász θ-functional, [8], to probabilistic graphs.

B. Further Results

The result (4) may be viewed as a new characterization of the complementary graph entropy in terms of the chromatic entropy. Exploring this viewpoint, and using the properties of $H_\chi(G, P)$, we continued the investigation of $\bar{H}(G, P)$. We will only briefly summarize our results here; due to lack of space, we will not be able to supply detailed proofs.

An interesting re-formulation of (5) is the following inequality, for arbitrary (G, P):

$$\bar{H}(G, P) + H(\bar{G}, P) \ge H(P). \tag{6}$$

We prove a generalization of (6) in the following inequality, valid for arbitrary (G_1, P) and (G_2, P):

$$\bar{H}(G_1, P) + H(G_2, P) \ge \bar{H}(G_1 \cup G_2, P). \tag{7}$$

Here, $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are defined on the same vertex set V. $G_1 \cup G_2 = (V, E_1 \cup E_2)$, is the graph union of G_1 and G_2.

Graph entropy is subadditive w.r.t graph union. Körner posed the question: Is the complementary graph entropy [5] also subadditive? It was previously known that

$$\bar{H}(G_1, P) + \bar{H}(G_2, P) \ge \bar{H}(G_1 \cup G_2, P) \tag{8}$$

when (i) $G_1 \cup G_2$ is perfect [5] (ii) G_1 and G_2 are both perfect [7]. It follows from (7) that it is in fact sufficient for *either* G_1 *or* G_2 to be perfect.

In some applications, fixed-rate channels - which rule out the use of buffers - may force the use of fixed-length codes. The minimum rate for fixed-length coding is the number

$$R(G) = \lim_{n\to\infty} \frac{1}{n} \log \chi(G^n). \tag{9}$$

This quantity is also of interest in graph theory, as characterizing the growth of chromatic numbers of the normal products of a graph. We prove the following equation, relating $R(G)$ to $\bar{H}(G, P)$:

$$R(G) = \max_P \bar{H}(G, P). \tag{10}$$

This relation has a simple coding interpretation. If the underlying graph G is fixed, but the source distribution P is not known, the encoder, using variable-length codes designed for the worst case, transmits at a rate $\max_P \bar{H}(G, P)$. Alternately, the encoder uses fixed-length codes, in which case the rate required is $R(G)$. The result (10) ensures that this alternate strategy is, in fact, optimal.

The zero-error capacity of a graph $G = (V, E)$ is [6]

$$C(G) = \lim \frac{1}{n} \log \alpha(G^n). \tag{11}$$

It follows, from $\chi(G^n)\alpha(G^n) \ge |V|^n$, that $R(G) + C(G) \ge \log |V|$. We show that equality holds, i.e.,

$$R(G) + C(G) = \log |V|, \tag{12}$$

when G is vertex-transitive. (A permutation of V is an automorphism if it preserves adjacency of the vertices. If for each pair of vertices $i, j \in V$ there exists an automorphism mapping i onto j, then G is said to be vertex-transitive.)

IV. Minimum Asymptotic Rate and the Complementary Graph Entropy

The chromatic entropy of a probabilistic graph (G, P) (where $G = (V, E)$), $H_\chi(G, P)$, was defined in [2]. If c is a function defined over V, then $c(V)$ is a random variable with entropy

$$H(c(V)) = \sum_{\gamma \in c(V)} P[c^{-1}(\gamma)] \log \frac{1}{P[c^{-1}(\gamma)]},$$

where c^{-1} is the inverse of c.

Definition 1: The chromatic entropy of (G, P) is the lowest entropy of any coloring of G:

$$H_\chi(G, P) = \min\{H(c(V)) : c \text{ is a coloring of } G\}. \quad (13)$$

Let R_n be the minimum rate of a *uniquely decodable* (not necessarily instantaneous) code for (G^n, P^n). The following lemma bounds R_n in terms of $H_\chi(G^n, P^n)$:

Lemma 1:

$$H_\chi(G^n, P^n) - \log\{H_\chi(G^n, P^n) + 1\} - \log e \le R_n;$$
$$R_n \le H_\chi(G^n, P^n) + 1. \quad (14)$$

Proof: Let $\phi : V^n \to \{0, 1\}^*$ be a code for (G, P). If distinct $x^n, x'^n \in V^n$ are confusable and, further, if $\phi(x^n) = \phi(x'^n)$, then the decoder cannot distinguish between x^n and x'^n, and ϕ is not uniquely decodable. In other words, if ϕ is uniquely decodable, $\phi(x^n) = \phi(x'^n)$ for distinct x^n, x'^n implies that x^n and x'^n are not connected in G^n. Thus ϕ may be written as the composition of a coloring of G^n and a one-to-one encoding of the colors. Now, (14) follows from known upper and lower bounds on the rates of one-to-one codes [2]. ■

Identical bounds as in (14) were proved in [2] for the restricted class of instantaneous codes, and were then used to calculate the minimum asymptotic rate of such codes. We can therefore parallel these calculations, to determine the minimum asymptotic rate for uniquely decodable codes.

Lemma 2:

$$\lim_{n\to\infty} \frac{R_n}{n} = \lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} \quad (15)$$

Proof: The proof is identical to that of Lemma 6 in [2]. ■

Since the same asymptotic rate as in (15) is achievable with instantaneous codes, Lemma 2 shows that the possibly larger class of uniquely decodable codes offers no *asymptotic* advantage. While this situation is identical to that obtained in regular lossless source coding, we are unable to answer whether uniquely decodable codes also offer no advantage in the case of finite block lengths.

The complementary graph entropy was introduced as an information-theoretic functional in [3].

Definition 2: The complementary graph entropy of (G, P) is the normalized logarithm of the "essential chromatic number of G^n with respect to P," i.e., the number

$$\bar{H}(G, P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \min_{P^n(A)\ge 1-\epsilon} \{\chi(G^n(A))\}, \quad (16)$$

where $G^n(A)$ is the subgraph induced in G^n by $A \subseteq V^n$.

Thus G^n has a high-probability induced subgrapph which can be colored with approximately $2^{n\bar{H}(G,P)}$ colors. Körner and Longo used this fact to show that the complementary graph entropy is the rate required for the following two-step source coding problem: Consider a memoryless source emitting symbols from a finite alphabet V according to a distribution P. Assume that some pairs of elements of the alphabet are distinguishable, while some others are not, and let G be the graph on V where connectedness means distinguishability. We want to encode the n-length source vector X^n in two steps. In the first step, an encoding function f on V^n is used, and it is required that, on the basis of X^n, the decoder be able to determine a sequence $\hat{x}^n$ that is, with high probability, indistinguishable from X^n in every co-ordinate. Call an encoder f achieving this goal "G-faithful." In the second step we want to encode X^n by an encoding function g such that the following holds: the encoded source $g(X^n)$, together with an *arbitrary* G-faithful encoding of X^n, determines X^n with high probability. It was shown in [3] that the minimum asymptotic rate needed for such a "complementary encoding" in the second step is $\bar{H}(G, P)$.

We will also need the generalization of the Shannon capacity, [6], to probabilistic graphs. This quantity was introduced by Csiszár and Körner in [9] to study the capacity of an arbitrarily varying channel with maximum probability of error.

Definition 3: Let $\mathcal{T}^n(P, \epsilon)$ be the set of "(P, ϵ)-typical" sequences in V^n, i.e., the set of sequences $x^n \in V^n$ for which the frequency $\pi(i|x^n)$ of each element $i \in V$ satisfies

$$|\pi(i|x^n) - P(i)| < \epsilon.$$

Let $G^n(P, \epsilon)$ be the subgraph of G^n induced by $\mathcal{T}^n(P, \epsilon)$.

Definition 4: The capacity of the graph G relative to P is

$$C(G, P) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log \alpha(G^n(P, \epsilon)). \quad (17)$$

We will need the following relation between $\bar{H}(G, P)$ and $C(G, P)$ established by Marton in [5]:

$$\bar{H}(G, P) + C(G, P) = H(P). \quad (18)$$

Consider a fixed-length encoding function $f : V^n \to \{1, 2, \cdots, 2^{nR'}\}$ for $G^n = (V^n, E_n)$ of which we require the following property: if $(x^n, x'^n) \in E_n$ then, with high probability, $f(x^n) \neq f(x'^n)$. It follows from (16) that the minimum rate required is $\bar{H}(G, P)$. In the following theorem, we show that $\bar{H}(G, P)$ is also the minimum rate required if f is allowed to be a variable-length encoding function, but where $(x^n, x'^n) \in E_n \Rightarrow f(x^n) \neq f(x'^n)$, is always required.

Theorem 1:

$$R^* = \bar{H}(G, P), \quad (19)$$

where R^* is defined in (2).

Proof: We will show that

$$\lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} = \bar{H}(G, P).$$

We begin by proving $\lim_{n\to\infty} \frac{H_\chi(G^n,P^n)}{n} \le \bar{H}(G, P)$. Fix $\epsilon > 0$. Let

$$\bar{H}_\epsilon(G, P) = \limsup_{n\to\infty} \frac{1}{n} \log \left[\min_{A\subseteq V^n, P^n(A)\ge 1-\epsilon} \chi(G^n(A)) \right].$$

Then, for fixed $\delta > 0$, for each $n > n_0(\delta)$ there is a subset $A \subseteq V^n$ with $P^n(A) \ge 1 - \epsilon$, and a coloring c of G satisfying:

$$|c(G^n(A))| \le 2^{n\{\bar{H}_\epsilon(G,P)+\delta\}}. \tag{20}$$

For $x^n \in V^n$, define the function $\Phi : V^n \to \{0, 1\}$ as

$$\Phi(x^n) = \begin{cases} 1 & \text{if } x^n \in A \\ 0 & \text{else.} \end{cases}$$

Thus Φ is the indicator function of A. Estimating the entropy of the coloring c,

$$\begin{aligned} &H(c(V^n)) \le H(\Phi) + H(c(V^n)|\Phi), \\ &\le H(\Phi) + H(c(V^n)|V^n \in A) + \epsilon H(c(V^n)|V^n \notin A), \\ &\le 1 + n\{\bar{H}_\epsilon(G, P) + \delta + \epsilon \log |V|\}, \end{aligned}$$

where we used (20) in the last step. But, by the definition of the chromatic entropy,

$$H_\chi(G^n, P^n) \le H(c(V^n)).$$

Normalizing by n and taking limits, the inequality follows.

Next, we show that $\lim_{n\to\infty} \frac{H_\chi(G^n,P^n)}{n} \ge \bar{H}(G, P)$. We lower bound $H_\chi(G^n, P^n)$ in terms of the maximum size of an independent set induced by $\mathcal{T}^n(P, \epsilon)$ in G^n. But this size is related to the capacity $C(G, P)$, and the inequality will then follow from (18). Let us fill in the details. Fix $\epsilon > 0$. Define Φ as the indicator function of $\mathcal{T}^n(P, \epsilon)$: for $x^n \in V^n$,

$$\Phi(x^n) = \begin{cases} 1 & \text{if } x^n \in \mathcal{T}^n(P, \epsilon) \\ 0 & \text{else.} \end{cases}$$

Let the coloring function c on G^n achieve $H_\chi(G^n, P^n)$, so that

$$H_\chi(G^n, P^n) = H(c(V^n)).$$

To lower bound $H(c(V^n))$, we use the following elementary lower bound for the entropy function: if Q is a probability distribution over the set $\mathcal{Q}$, and $S \subseteq \mathcal{Q}$, then

$$H(Q) \ge -\{\sum_{j\in S} Q(j)\} \log \max_{j\in S} Q(j).$$

Thus we have the following estimate for $H_\chi(G^n, P^n)$:

$$H(c(V^n)) \ge -P(\mathcal{T}^n(P, \epsilon)) \log \max_{x^n\in\mathcal{T}^n(P,\epsilon)} P(c(x^n)). \tag{21}$$

But the set of (P, ϵ)-typical sequences $\mathcal{T}^n(P, \epsilon)$ captures most of the probability [11]:

$$P(\mathcal{T}^n(P, \epsilon)) \ge 1 - \frac{|V|}{4n\epsilon^2}. \tag{22}$$

Further, in any coloring of G^n, the maximum cardinality of a single-colored subset of $\mathcal{T}^n(P, \epsilon)$ cannot exceed $\alpha(G^n(P, \epsilon))$, the size of the largest independent set induced by $\mathcal{T}^n(P, \epsilon)$ in G^n. Thus,

$$\begin{aligned} &\max_{x^n\in\mathcal{T}^n(P,\epsilon)} P(c(x^n)) \le \alpha(G^n(P, \epsilon)) \max_{x^n\in\mathcal{T}^n(P,\epsilon)} P(x^n), \\ &\le \alpha(G^n(P, \epsilon)) 2^{-n \min\{H(P')+D(P'||P):|P'(i)-P(i)|<\epsilon \forall i\in V\}}, \\ &\le \alpha(G^n(P, \epsilon)) 2^{-n \min\{H(P)+\epsilon|V|\log\epsilon\}}, \end{aligned} \tag{23}$$

where we use a known formula for the probability of a typical sequence, and the uniform continuity of entropy, [11].

Substituting (22) and (23) in (21), we obtain

$$\frac{H_\chi(G^n, P^n)}{n} \ge \left(1 - \frac{|V|}{4n\epsilon^2}\right) \left\{ H(P) - \frac{1}{n} \log \alpha(G^n(P, \epsilon)) + \epsilon|V| \log \epsilon \right\},$$

and taking the limit,

$$\lim_{n\to\infty} \frac{H_\chi(G^n, P^n)}{n} \ge H(P) - C_\epsilon(G, P) + \epsilon|V| \log \epsilon,$$

where $C_\epsilon(G, P)$ is defined as (cf. (17)):

$$C_\epsilon(G, P) = \limsup_{n\to\infty} \frac{1}{n} \log \alpha(G^n(P, \epsilon)).$$

Now, letting $\epsilon \to 0$ and using (18), the result follows. ■

References

[1] H. S. Witsenhausen, "The zero-error side information problem and chromatic numbers," *IEEE Trans. on Inform. Theory*, vol.IT-22, (no.5), pp.592-3, Sept. 1976.

[2] N. Alon and A. Orlitsky, "Source coding and graph entropies," *IEEE Trans. on Inform. Theory*, vol.IT-42, (no.5), pp.1329-39, Sept. 1996.

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

[4] J. Körner, "Coding of an information source having ambiguous alphabet and the entropy of graphs," in: *Trans. of the 6th Prague Conf. on Inform. Theory, etc.,* Academia, Prague, 1973, 411-425.

[5] K. Marton, "On the Shannon capacity of probabilistic graphs," *J. of Comb. Theory*, Series B, vol. 57, no. 2, March 1993.

[6] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. IT-2, (no.3), pp. 8-19, Sept. 1956.

[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] L. Lovász, "On the Shannon Capacity of a Graph," *IEEE Trans. on Inform. Theory*, vol. IT-25, (no.1), pp. 1-7, Jan. 1979.

[9] I. Csiszár and J.Körner, "On the capacity of the arbitrarily varying channel for maximum probability of error," *Z. Wahrsch. Verw. Gebiete* 57 (1981), 87-101.

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

[11] I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless sources*. Academic Press, New York, 1982.