On Network Functional Compression

Soheil Feizi, Student, IEEE, and Muriel Médard, Fellow, IEEE

Abstract—In this paper, we consider different aspects of the problem of compressing for function computation across a network, which we call network functional compression. In network functional compression, computation of a function (or, some functions) of sources located at certain nodes in a network is desired at receiver(s). The rate region of this problem has been considered in the literature under certain restrictive assumptions, particularly in terms of the network topology, the functions, and the characteristics of the sources. In this paper, we present results that significantly relax these assumptions. For a one-stage tree network, we characterize a rate region by introducing a necessary and sufficient condition for any achievable coloring-based coding scheme called coloring connectivity condition. We also propose a modularized coding scheme based on graph colorings to perform arbitrarily closely to rate lower bounds. For a general tree network, we provide a rate lower bound based on graph entropies and show that this bound is tight in the case of having independent sources. In particular, we show that, in a general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform computations. However, for a family of functions and random variables, which we call chain-rule proper sets, it is sufficient to have no computations at intermediate nodes to perform arbitrarily closely to the rate lower bound. In addition, we consider practical issues of coloring-based coding schemes and propose an efficient algorithm to compute a minimum entropy coloring of a characteristic graph under some conditions on source distributions and/or the desired function. Finally, extensions of these results for cases of having feedback and lossy function computations are discussed.

Index Terms—Functional compression, distributed computation, Slepian-Wolf compression, graph entropy, graph coloring, feedback, distortion.

I. INTRODUCTION

In this paper, we consider different aspects of the functional compression problem over networks from an information theoretic point of view. In the functional compression problem, we would like to compress source random variables for the purpose of computing a deterministic function (or some deterministic functions) at the receiver(s), when these sources and receivers are nodes in a network. Traditional data compression schemes are special cases of functional compression, where their desired function is the identity function. However, if the receiver is interested in computing a function (or some functions) of sources, further compression is possible.

There have been several approaches to investigate different aspects of this problem. One class of works considered the functional computation problem for specific functions. For example, reference [1] investigated computation of symmetric Boolean functions in tree networks, and references [2] and [3] studied the sum-network with three sources and three terminals. Some other references investigated the asymptotic analysis of the transmission rate in noisy broadcast networks [4], and also in random geometric graph models ([5] and [6]). Also, reference [7] investigated information-theoretic bounds for multiround function computation in collocated networks. Network flow techniques (also known as multi-commodity methods) have been used to study multiple unicast problems (references [8] and [9]). By some modifications, reference [10] used this framework for function computation considering communication constraints.

A major body of work on in-network computation investigates information-theoretic rate bounds, when a function of sources is desired to be computed at the receiver. These works can be categorized into the study of lossless functional compression and that of functional compression with distortion. By lossless computation, we mean asymptotically lossless computation of a function: the error probability goes to zero as the block length goes to the infinity. However, there are several works investigating zero-error computation of functions (references [11] and [12]).

Shannon was the first one who considered the function computation problem in reference [13] for a special case when \( f(X_1, X_2) = (X_1, X_2) \) (the identity function) and for the network topology depicted in Figure 1-a (the side information problem). For a general function, Orlitsky and Roche provided a single-letter characterization of the rate-region in [14]. In [15], Doshi et al. proposed an optimal coding scheme for this problem.

For the network topology depicted in Figure 1-b, and for the case that the desired function at the receiver is the identity function (i.e., \( f(X_1, X_2) = (X_1, X_2) \)), Slepian and Wolf provided a characterization of the rate region and an optimal achievable coding scheme in [16]. Some other practical but suboptimal coding schemes have been proposed by Pradhan and Ramachandran in [17]. Also, a rate-splitting technique for this problem is developed in [18] and [19]. Special cases when \( f(X_1, X_2) = X_1 \) and \( f(X_1, X_2) = (X_1 + X_2) \mod 2 \) have been investigated by Ahlswede and Körner in [20],
and Körner and Marton in [21], respectively. Under some special conditions on source distributions, Doshi et al. in [15] investigated this problem for a general function and proposed some achievable coding schemes.

There have been several prior works that studied lossy functional compression where the function at the receiver is desired to be computed within a distortion level. Wyner and Ziv [22] considered the side information problem for computing the identity function at the receiver within some distortion. Yamamoto solved this problem for a general function \( f(X_1, X_2) \) in [23]. Doshi et al. gave another characterization of the rate distortion function given by Yamamoto in [15]. Feng et al. [24] considered the side information problem for a general function at the receiver in the case the encoder and decoder have some noisy information. For the distributed function computation problem and for a general function, the rate-distortion region has been unknown, but some bounds have been given by Berger and Yeung [25], D insistive function, the rate-distortion region has been unknown, but some bounds have been given by Berger and Yeung [25], Doshi et al. [15]. By using C.C.C., we gave another characterization of the rate-distortion function given by Yamamoto [23].

In this paper, we present results that significantly relax previously considered restrictive assumptions particularly in terms of the network topology, the functions and the characteristics of the sources. For a one-stage tree network, we introduce a necessary and sufficient condition for any achievable coloring-based coding scheme called Coloring Connectivity Condition (C.C.C.), thus relaxing the previous sufficient zig-zag condition of Doshi et al. [15]. By using C.C.C., we characterize a rate region for distributed functional compression and propose a modularized coding scheme based on graph colorings to perform arbitrarily closely to rate lower bounds. These results are presented in Section III-A.

In Section III-B, we consider a general tree network and provide a rate lower bound based on graph entropies. We show that this bound is tight in the case of having independent sources. In particular, we show that, to achieve the rate lower bound, intermediate nodes should perform computations. However, for a family of functions and random variables, which we call \textit{chain-rule proper sets}, it is sufficient to have intermediate nodes act like relays (i.e., no computations at intermediate nodes) to perform arbitrarily closely to the rate lower bound.

In Section III-C, we discuss practical issues of coloring-based coding schemes and propose an efficient algorithm to compute a minimum entropy coloring of a characteristic graph under some conditions on source distributions and/or the desired function. Finally, extensions of proposed results for cases of having feedback and lossy function computations are discussed in Section IV. Particularly, we show that, in functional compression, unlike Slepian-Wolf, by having feedback, one may outperform rate bounds of the case without feedback. These results extend those of Bakshi \textit{et al.} We also present a practical coding scheme for the distributed lossy functional compression problem with a non-trivial performance guarantee.

II. PROBLEM SETUP AND PRIOR WORK

In this section, we setup the functional compression problem and review some prior work.

A. Problem Setup

Consider \( k \) discrete memoryless random processes, \( \{X^i_n\}_{i=1}^n, \{X^k_l\}_{l=1}^k \), as source processes. Memorylessness is not necessary, and one can approximate a source by a memoryless one with an arbitrary precision [28]. Suppose these sources are drawn from finite sets \( X^i = \{x^i_1, x^i_2, \ldots, x^i_{|X^i|} \} \), \( X^k = \{x^k_1, x^k_2, \ldots, x^k_{|X^k|} \} \). These sources have a joint probability distribution \( p(x^1_1, \ldots, x^k_k) \). Without loss of generality, we assume \( l = 1 \), and to simplify notation, \( n \) will be implied by the context if no confusion arises. We refer to the \( l \)-th element of \( X_j \) as \( x_{ji} \). We use \( x^1_j, x^2_j, \ldots, x^n_j \) as different \( n \)-sequences of \( X_j \). We shall omit the superscript when no confusion arises. Since the sequence \( (x^1_1, \ldots, x^k_k) \) is drawn i.i.d. according to \( p(x^1_1, \ldots, x^k_k) \), one can write \( p(x^1_1, \ldots, x^k_k) = \prod_{i=1}^k p(x^i_{1l}, \ldots, x^i_{kl}) \).

Consider a tree network with \( k \) source nodes in its leaves and a receiver in its root. Other nodes of this tree are referred as intermediate nodes. Source node \( j \) has an input random process \( \{X^i_n\}_{i=1}^n \). The receiver wishes to compute a deterministic function \( f : X^1_k \times \cdots \times X^k_k \rightarrow Z \), or \( f : X^1_n \times \cdots \times X^k_n \rightarrow Z^n \), its vector extension.

In this paper, we consider tree network topologies. Note that sources can be at any nodes of the network. However, without loss of generality, we can modify the network by adding some fake leaves to source nodes which are not located in leaves of the network. So, in the achieved network, sources are always located in leaves. Also, by adding some auxiliary nodes, one can make sources to be in the same distance from the receiver. Nodes of this tree are labeled as \( i \) for different \( i \)'s, where source nodes are denoted by \( \{1, \ldots, k\} \) and the outgoing link of node \( i \) is denoted by \( e_i \). Node \( i \) sends \( M_i \) over its outgoing edge \( e_i \) with a rate \( R_i \) (it maps length \( n \) blocks of \( M_i \), referred to as \( M_i \), to \( \{1, 2, \ldots, 2^{nR_i}\} \)).

For a source node, \( M_i = \text{en}_{X_i}(X_i) \), where \( \text{en}_{X_i} \) is the encoding function of the source node \( i \). For an intermediate node \( i, i \notin \{1, \ldots, k\} \), with incoming edges \( e_j, \ldots, e_q \), \( M_i = g_i(M_{f_1}, \ldots, M_{f_q}) \), where \( g_i(\cdot) \) is a function to be computed in that node.

The receiver maps incoming messages \( \{M_i, \ldots, M_j\} \) to \( Z^n \) by using a function \( r(\cdot) \); i.e., \( r : \prod_{l=1}^k \{1, \ldots, 2^{nR_l}\} \rightarrow Z^n \). Thus, the receiver computes \( r(M_i, \ldots, M_j) = r'(\text{en}_{X_i}(X_i), \ldots, \text{en}_{X_j}(X_j)) \). We refer to this encoding/decoding scheme as an \( n \)-distributed functional code. Intermediate nodes are
allowed to compute functions, but have no demand of their own. The desired function \( f(X_1, \ldots, X_k) \) at the receiver is the only demand in the network. For any encoding/decoding scheme, the probability of error is defined as

\[
P^e_n = Pr \{ (x_1, \ldots, x_k) : f(x_1, \ldots, x_k) \neq f'(en_{X_1}(x_1), \ldots, en_{X_k}(x_k)) \}.
\]  

(1)

A rate tuple of the network is the set of rates of its edges (i.e., \([R_i]\) for valid \(i\)’s). We say a rate tuple is achievable iff there exists a coding scheme operating at these rates so that \( P^e_n \to 0 \) as \( n \to \infty \). The achievable rate region is the set closure of the set of all achievable rates.

### B. Definitions and Prior Results

In this part, we present some definitions and review prior results.

**Definition 1:** The characteristic graph \( G_{X_1} = (V_{X_1}, E_{X_1}) \) of \( X_1 \) with respect to \( X_2 \), \( p(x_1, x_2) \), and function \( f(X_1, X_2) \) is defined as: \( V_{X_1} = X_1 \) and an edge \((x^1_1, x^1_2) \in \mathcal{X}^2_1 \) is in \( E_{X_1} \) iff there exists a \( x^2_j \in X_2 \) such that \( p(x^1_1, x^1_2) = f(x^1_1, x^2_j) \).

In other words, in order to avoid confusion about the function \( f(X_1, X_2) \) at the receiver, if \((x^1_1, x^1_2) \in E_{X_1} \), descriptions of \( x^1_1 \) and \( x^1_2 \) must be different. Shannon first defined this when studying the zero error capacity of noisy channels [13]. Witsenhausen [29] used this concept to study a simplified version of functional compression problem where one encodes \( X_1 \) to compute \( f(X_1) = X_1 \) with zero distortion and showed the chromatic number of the strong graph-product characterizes the rate. The characteristic graph of \( X_2 \) with respect to \( X_1 \), \( p(x_1, x_2) \), and \( f(X_1, X_2) \) is defined analogously and denoted by \( G_{X_2} \). One can extend the definition of the characteristic graph to the case of having more than two random variables as follows. Suppose \( X_1, \ldots, X_k \) are \( k \) random variables defined in Section II-A.

**Definition 2:** The characteristic graph \( G_{X_1} = (V_{X_1}, E_{X_1}) \) of \( X_1 \) with respect to random variables \( X_2, \ldots, X_k \), \( p(x_1, \ldots, x_k) \), and \( f(X_1, \ldots, X_k) \) is defined as: \( V_{X_1} = X_1 \) and an edge \((x^1_1, x^1_2) \in \mathcal{X}^2_1 \) is in \( E_{X_1} \) iff there exists a \( x^2_j \in X_2 \) such that \( p(x^1_1, x^1_2) = f(x^1_1, x^2_j) \).

**Example 3:** Consider the scenario described in Example 3.

The notation \( X_1 \in W_1 \in \Gamma(G_{X_1}) \) means that we are minimizing over all distributions \( p(w_1, x_1) \) such that \( p(w_1, x_1) > 0 \) implies \( x_1 \in w_1 \), where \( w_1 \) is a maximal independent set of the graph \( G_{X_1} \).

**Example 4:** Consider the scenario described in Example 3.

For the characteristic graph of \( X_1 \) shown in Figure 2-a, the set of maximal independent sets is \( W_1 = \{\{0, 2\}, \{1, 3\}\} \). To minimize \( I(X_1; W_1) = H(X_1) - H(X_1|W_1) = \log(2) - H(X_1|W_1) \), one should maximize \( H(X_1|W_1) \). Because of the symmetry of the problem, to maximize \( H(X_1|W_1) \), \( p(w_1) \) must be uniform over two possible maximal independent sets of \( G_{X_1} \). Since each maximal independent set \( w_1 \in W_1 \) has two \( X_1 \) values, thus, \( H(X_1|w_1) = \log(2) \) bit, and since \( p(w_1) \) is uniform, \( H(X_1|W_1) = \log(2) \) bit. Therefore, \( H_{G_{X_1}}(X_1) = \log(4) - \log(2) = 1 \) bit. One can see if we want to encode \( X_1 \) ignoring the effect of the function \( f \), we need \( H(X_1) = \log(4) = 2 \) bits. We will show that, for this example, functional compression saves us 1 bit in every 2 bits compared to the traditional data compression.

Witsenhausen [29] showed that the chromatic number of the strong graph-product characterizes the minimum rate at which a single source can be encoded so that the identity function of that source can be computed with zero distortion. Orlitsky and Roche [14] defined an extension of Körner’s graph entropy, the conditional graph entropy.

**Definition 6:** The conditional graph entropy is

\[
H_{G_{X_1}}(X_1|X_2) = \min_{X_1 \in W_1 \in \Gamma(G_{X_1})} I(W_1; X_1|X_2). \tag{3}
\]

Notation \( W_1 - X_1 - X_2 \) indicates a Markov chain. If \( X_1 \) and \( X_2 \) are independent, \( H_{G_{X_1}}(X_1|X_2) = H_{G_X}(X_1) \). To illustrate this concept, let us consider an example borrowed from [14].

**Example 7:** When \( f(X_1, X_2) = X_1 \), \( H_{G_{X_1}}(X_1|X_2) = H(X_1|X_2) \).

To show this, consider the characteristic graph of \( X_1 \), denoted as \( G_{X_1} \). Since \( f(X_1, X_2) = X_1 \), then for every \( x_1^2 \in X_2 \), the set \( \{ x_1^1 : p(x_1^1, x_1^2) > 0 \} \) of possible \( x_1^1 \) are connected to each other (i.e., this set is a clique of \( G_{X_1} \)). Since the intersection of a clique and a maximal independent set is a singleton, \( X_2 \) and the maximal independent set \( W_1 \)

\[ X_1^1 = 0 \]
\[ X_1^2 = 1 \]

\[ X_2^1 = 2 \]

\[ X_2^2 = 3 \]

\[ x_1^1 = 0 \]
\[ x_1^2 = 1 \]
\[ x_2^1 = 2 \]
\[ x_2^2 = 3 \]

Fig. 2. This figures shows characteristic graphs (a) \( G_{X_1} \), and (b) \( G_{X_2} \), for the setup of Example 3 (Different letters written over graph vertices indicate different colors.)
containing $X_1$ determine $X_1$. So,
\[
H_{G_{X_1}}(X_1|X_2) = \min_{X_1 \in \mathcal{X}_1} I(W_1; X_1|X_2) \\
= H(X_1|X_2) - \max_{X_1 \in \mathcal{X}_1} H(X_1|W_1, X_2) \\
= H(X_1|X_2). \tag{4}
\]

Definition 8: A vertex coloring of a graph is a function $c_{G_{X_1}}(X_1) : V_{X_1} \to \mathbb{N}$ of a graph $G_{X_1} = (V_{X_1}, E_{X_1})$ such that $(x_1^1, x_1^2) \in E_{X_1}$ implies $c_{G_{X_1}}(x_1^1) \neq c_{G_{X_1}}(x_1^2)$. The entropy of a coloring is the entropy of the induced distribution on colors. Here, $p(c_{G_{X_1}}(x_1^1) = p(c_{G_{X_1}}^{-1}(c_{G_{X_1}}(x_1^1))))$, where $c_{G_{X_1}}^{-1}(c_{G_{X_1}}(x_1^j)) = (x_1^j : c_{G_{X_1}}(x_1^j) = c_{G_{X_1}}(x_1^j))$ for all valid $j$. This subset of vertices with the same color is called a color class. We refer to a coloring which minimizes the entropy as a minimum entropy coloring. We use $C_{G_{X_1}}$ as the set of all valid colorings of a graph $G_{X_1}$.

Example 9: Consider again the random variable $X_1$ described in Example 3, whose characteristic graph $G_{X_1}$ and its valid coloring are shown in Figure 2-a. One can see that, in this coloring, two connected vertices are assigned to different colors. Specifically, $c_{G_{X_1}}(X_1) = [r, b]$. Therefore, $p(c_{G_{X_1}}(x_1^1) = r) = p(x_1^1 = 0) + p(x_1^1 = 2)$, and $p(c_{G_{X_1}}(x_1^1) = b) = p(x_1^1 = 1) + p(x_1^1 = 3)$.

We define a power graph of a characteristic graph as its co-normal products:

Definition 10: The $n$-th power of a graph $G_{X_1}$ is a graph $G_{X_1}^n = (V_{X_1}^n, E_{X_1}^n)$ such that $V_{X_1}^n = X_1^n$ and $(x_1^1, x_1^2) \in E_{X_1}^n$ when there exists at least one $i$ such that $(x_1^1, x_1^2) \in E_{X_1}$. We denote a valid coloring of $G_{X_1}^n$ by $c_{G_{X_1}^n}(X_1)$.

One may ignore atypical sequence in a sufficiently large power graph of a characteristic graph and then, color that graph. This coloring is called an $\epsilon$-coloring of a graph and is defined as follows:

Definition 11: Given a non-empty set $A \subseteq X_1 \times X_2$, define $\hat{p}(x_1, x_2) = p(x_1, x_2)/p(A)$ when $(x_1, x_2) \in A$, and $\hat{p}(x, y) = 0$ otherwise. $\hat{p}$ is the distribution over $(x_1, x_2)$ conditioned on $(x_1, x_2) \in A$. Denote the characteristic graph of $X_1$ with respect to $X_2$, $\hat{p}(x_1, x_2)$, and $(f_1, x_2)$ as $G_{X_1} = (V_{X_1}, E_{X_1})$ and the characteristic graph of $X_2$ with respect to $X_1$, $\hat{p}(x_1, x_2)$, and $(f_1, x_2)$ as $G_{X_2} = (V_{X_2}, E_{X_2})$. Note that $\hat{E}_{X_1} \subseteq E_{X_1}$ and $\hat{E}_{X_2} \subseteq E_{X_2}$. Suppose $p(A) \geq 1 - \epsilon$. We say that $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2)$ are $\epsilon$-colorings of $G_{X_1}$ and $G_{X_2}$ if they are valid colorings of $\hat{G}_{X_1}$ and $\hat{G}_{X_2}$.

In [31], the Chromatic entropy of a graph $G_{X_1}$ is defined as Definition 12:

\[
H_{G_{X_1}}^\epsilon(X_1) = \min_{c_{G_{X_1}} \text{isan}\epsilon\text{-coloringof } G_{X_1}} H(c_{G_{X_1}}(X_1)).
\]

The chromatic entropy is a representation of the chromatic number of high probability subgraphs of the characteristic graph. In [15], the conditional chromatic entropy is defined as Definition 13:

\[
H_{G_{X_1}}^\epsilon(X_1|X_2) = \min_{c_{G_{X_1}} \text{isan}\epsilon\text{-coloringof } G_{X_1}} H(c_{G_{X_1}}(X_1)|X_2).
\]

Regardless of $\epsilon$, the above optimizations are minima, rather than infima, because there are finitely many subgraphs of any fixed graph $G_{X_1}$, and therefore there are only finitely many $\epsilon$-colorings, regardless of $\epsilon$.

In general, these optimizations are NP-hard ([32]). But, depending on the desired function $f$, there are some interesting cases that optimal solutions can be computed efficiently. We discuss these cases in Section III-C.

Körner showed in [30] that, in the limit of large $n$, there is a relation between the chromatic entropy and the graph entropy.

Theorem 14:
\[
\lim_{n \to \infty} \frac{1}{n} H_{G_{X_1}}^\epsilon(X_1) = H_{G_{X_1}}(X_1). \tag{5}
\]

This theorem implies that the receiver can asymptotically compute a deterministic function of a discrete memoryless source, by first coloring a sufficiently large power of the characteristic graph of the source random variable with respect to the function, and then, encoding achieved colors using any encoding scheme which achieves the entropy bound of the coloring RV. In the previous approach, to achieve the encoding rate close to graph entropy of $X_1$, one should find the optimal distribution over the set of maximal independent sets of $G_{X_1}$. But, this theorem allows us to find the optimal coloring of $G_{X_1}^\epsilon$, instead of the optimal distribution on maximal independent sets. One can see that this approach modularizes the encoding scheme into two parts, a graph coloring module, followed by a Slepian-Wolf compression module.

The conditional version of the above theorem is proven in [15].

Theorem 15:
\[
\lim_{n \to \infty} \frac{1}{n} H_{G_{X_1}}^\epsilon(X_1|X_2) = H_{G_{X_1}}(X_1|X_2). \tag{6}
\]

This theorem implies a practical encoding scheme for the problem of functional compression with side information where the receiver wishes to compute $f(X_1, X_2)$, when $X_2$ is available at the receiver as the side information. Orlitsky and Roche showed in [14] that $H_{G_{X_1}}(X_1|X_2)$ is the minimum achievable rate for this problem. Their proof uses random coding arguments and shows the existence of an optimal coding scheme. This theorem presents a modularized encoding scheme where one first finds the minimum entropy coloring of $G_{X_1}^\epsilon$ for large enough $n$, and then uses a compression scheme on the coloring random variable (such as Slepian-Wolf [16]) to achieve a rate arbitrarily closely to $H(c_{G_{X_1}^\epsilon}(X_1)|X_2)$. This encoding scheme guarantees computation of the function at the receiver with a vanishing probability of error.

All these results considered only functional compression with side information at the receiver (Figure 1-a). In general, the rate region of the distributed functional compression problem (Figure 1-b) has not been determined. However, [15] characterized a rate-region of this network when source random variables satisfy a condition called the zigzag condition, defined below.

We refer to the $\epsilon$-joint-typical set of sequences of random variables $X_1, \ldots, X_k$ as $T^n_\epsilon$. $k$ is implied in this notation for simplicity. $T^n_\epsilon$ can be considered as a strong or weak typical set ([28]).
Definition 16: A discrete memoryless source \( \{ (X_i^1, X_i^2) \}_{i \in \mathbb{N}} \) with a distribution \( p(x_1, x_2) \) satisfies the zigzag condition if for any \( \epsilon \) and some \( n \), \( (x_1^1, x_2^1) \), \( (x_1^2, x_2^2) \) \( \in \mathcal{T}_n \), there exists some \( (x_1^3, x_2^3) \) \( \in \mathcal{T}_n \) such that \( (x_1^i, x_2^i) \) \( \in \mathcal{T}_n \) for each \( i \in \{1, 2\} \), and \( (x_1^j, x_2^j) = (x_1^{3-i}, x_2^{3-i}) \) for some \( i \in \{1, 2\} \) for each \( j \).

In fact, the zigzag condition forces many source sequences to be typical. Reference [15] shows that, if the source random variables satisfy the zigzag condition, an achievable rate region for this network is the set of all rates that can be achieved through graph colorings. The zigzag condition is a restrictive condition which does not depend on the desired function at the receiver. This condition is not necessary, but sufficient. In the next section, we relax this condition by introducing a necessary and sufficient condition for any achievable coloring-based coding scheme and characterize a rate region for distributed functional compression problem.

III. MAIN RESULTS
In this section, we present the main results of the paper.

A. A Rate Region for One-Stage Tree Networks
In this section, we compute a rate region for a general one-stage tree network without having any restrictive conditions such as the zigzag condition.

Consider a one-stage tree network with \( k \) sources.

Definition 17: A path with length \( m \) between two points \( Z_1 = (x_1^1, x_2^1, \ldots, x_1^i) \), and \( Z_m = (x_1^2, x_2^2, \ldots, x_1^i) \) is determined by \( m-1 \) points \( Z_i, 1 \leq i \leq m \) such that,

i) \( Pr(Z_i) > 0 \) for all \( 1 \leq i \leq m \).

ii) \( Z_i \) and \( Z_{i+1} \) only differ in one of their coordinates.

Definition 18: A path with length \( m \) between two points \( Z_1 = (x_1^1, x_2^1, \ldots, x_1^i) \) \( \in \mathcal{T}_n \), and \( Z_m = (x_1^2, x_2^2, \ldots, x_1^i) \) \( \in \mathcal{T}_n \) is determined by \( m-1 \) points \( Z_i, 1 \leq i \leq m \) such that,

i) \( Z_i \in \mathcal{T}_n \), for all \( 1 \leq i \leq m \).

ii) \( Z_i \) and \( Z_{i+1} \) only differ in one of their coordinates.

Note that, each coordinate of \( Z_i \) is a vector with length \( n \).

Definition 19: A joint-coloring family \( \mathcal{J}_C \) for random variables \( X_1, \ldots, X_k \) with characteristic graphs \( G_{X_1}, \ldots, G_{X_k} \), and any valid colorings \( c_{G_{X_1}}, \ldots, c_{G_{X_k}} \), respectively is defined as \( \mathcal{J}_C = \{ j^C_k \} \) where \( j^C_k \) is the collection of points \( (x_1^1, x_2^1, \ldots, x_k^i) \) whose coordinates have the same color (i.e., \( j^C_k = \{(x_1^{i_1}, x_2^{i_2}, \ldots, x_k^{i_k}), (x_1^{i_1}, x_2^{i_2}, \ldots, x_k^{i_k}) : c_{G_{X_1}}(x_1^{i_1}) = c_{G_{X_2}}(x_1^{i_2}), \ldots, c_{G_{X_k}}(x_k^{i_k}) = c_{G_{X_k}}(x_k^{i_k})\} \), for any valid \( i_1, \ldots, i_k \) and \( i_1, \ldots, i_k \). Each \( j^C_k \) is called a joint coloring class.

We say a joint coloring class \( j^C_k \) is connected if between any two points in \( j^C_k \), there exists a path that lies in \( j^C_k \). Otherwise, it is disconnected. Definition 19 can be expressed for random vectors \( X_1, \ldots, X_k \) with characteristic graphs \( G_{X_1}, \ldots, G_{X_k} \), and any valid \( \epsilon \)-colorings \( c_{G_{X_1}}, \ldots, c_{G_{X_k}} \), respectively.

In the following, we present the Coloring Connectivity Condition (C.C.C.) which is a necessary and sufficient condition for any coloring-based coding scheme.
Lemma 23: Consider random variables $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$ and any valid colorings $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ with joint coloring class $J_C = \{j_i^c : i \}$. For any two points $(x_1^1, \ldots, x_k^1)$ and $(x_1^2, \ldots, x_k^2)$ in $j_i^c$, $f(x_1^1, \ldots, x_k^1) = f(x_1^2, \ldots, x_k^2)$ if and only if $j_i^c$ satisfies C.C.C.

Proof of this lemma is presented in Section VI-B.

Lemma 24: Consider random variables $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$ and any valid $\epsilon$-colorings $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ with the joint coloring class $J_C = \{j_i^c : i\}$. For any two points $(x_1^1, \ldots, x_k^1)$ and $(x_1^2, \ldots, x_k^2)$ in $j_i^c$, $f(x_1^1, \ldots, x_k^1) = f(x_1^2, \ldots, x_k^2)$ if and only if $j_i^c$ satisfies C.C.C.

Proof of this lemma is presented in Section VI-C.

Next, we show that, if $X_1$ and $X_2$ satisfy the zigzag condition given in Definition 16, any valid colorings of their characteristic graphs satisfy C.C.C., but not vice versa. In other words, we show that the zigzag condition used in [15] is sufficient but not necessary.

Lemma 25: If two random variables $X_1$ and $X_2$ with characteristic graphs $G_{X_1}$ and $G_{X_2}$ satisfy the zigzag condition, any valid colorings $c_{G_{X_1}}$ and $c_{G_{X_2}}$ of $G_{X_1}$ and $G_{X_2}$ satisfy C.C.C., but not vice versa.

Proof of this lemma is presented in Section VI-D.

We use C.C.C. to characterize a rate region of functional compression for a one-stage tree network:

Definition 26: For random variables $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$, the joint graph entropy is defined as follows:

$$H_{G_{X_1}, \ldots, G_{X_k}}(X_1, \ldots, X_k)$$

$$= \lim_{n \to \infty} \min_{c_{G_{X_1}}, \ldots, c_{G_{X_k}}} \frac{1}{n} H(c_{G_{X_1}}(X_1), \ldots, c_{G_{X_k}}(X_k))$$

where the minimization is over $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ satisfying C.C.C.

Similarly, we can define the conditional graph entropy.

Definition 27: For random variables $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$, the conditional graph entropy can be defined as follows:

$$H_{G_{X_1}, \ldots, G_{X_k}}(X_1, \ldots, X_i | X_{i+1}, \ldots, X_k)$$

$$= \lim_{n \to \infty} \min_{c_{G_{X_1}}, \ldots, c_{G_{X_k}}} \frac{1}{n} H(c_{G_{X_1}}(X_1), \ldots, c_{G_{X_k}}(X_i)$$

$$| c_{G_{X_{i+1}}} (X_{i+1}), \ldots, c_{G_{X_k}} (X_k))$$

where the minimization is over $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ satisfying C.C.C.

Lemma 28: For $k = 2$, Definitions 6 and 27 are the same.

Proof of this lemma is presented in Section VI-E.

Note that, by this definition, the graph entropy does not satisfy the chain rule.

Suppose $S(k)$ denotes the power set of the set $\{1, 2, \ldots, k\}$ excluding the empty subset. Then, for any $S \in S(k)$,

$$X_S \triangleq \{X_i : i \in S\}.$$ 

Let $S^c$ denote the complement of $S$ in $S(k)$. For $S = \{1, 2, \ldots, k\}$, denote $S^c$ as the empty set. To simplify notation, we refer to a subset of sources by $X_S$. For instance, $S(2) = \{(1), (2), (1, 2)\}$, and for $S = \{1, 2\}$, we write $H_{G_{X_1},G_{X_2}}(X_S)$ instead of $H_{G_{X_1},G_{X_2}}(X_1, X_2)$.

Theorem 29: A rate region of a one-stage tree network is characterized by these conditions:

$$\forall S \in S(k) \Rightarrow \sum_{i \in S} R_i \geq H_{G_{X_1},G_{X_2}}(X_S|X_S^c).$$

Proof of this theorem is presented in Section VI-F.

If we have two transmitters ($k = 2$), Theorem 29 can be simplified as follows.

Corollary 30: A rate region of the network shown in Figure 1-b is determined by these three conditions:

$$R_{11} \geq H_{G_{X_1}}(X_1|X_2)$$

$$R_{12} \geq H_{G_{X_1}}(X_2|X_1)$$

$$R_{11} + R_{12} \geq H_{G_{X_1},G_{X_2}}(X_1, X_2).$$

Algorithm 31: The following algorithm proposes a modularized coding scheme which performs arbitrarily closely to rate bounds of Theorem 29:

- Source nodes compute $\epsilon$-colorings of sufficiently large power of their characteristic graphs satisfying C.C.C., followed by Slepian-Wolf compression.
- The receiver first uses a Slepian-Wolf decoder to decode transmitted coloring variables. Then, it uses a look-up table to compute the function values.

The achievability proof of this algorithm directly follows from the proof of Theorem 29.

B. A Rate Lower Bound for a General Tree Network

In this section, we consider a general tree structure with intermediate nodes that are allowed to perform computations. However, to simplify notations, we limit the arguments to the tree structure of Figure 4. Note that, all discussions can be extended to a general tree network.
The problem of function computations for a general tree network has been considered in [12] and [33]. Reference [12] derives a necessary and sufficient condition for the encoders on each edge of the tree for a zero-error computation of the desired function. Reference [33] shows that for a tree network with independent sources, a min-cut rate is a tight upper bound. Here, we consider an asymptotically lossless functional compression problem. For a general tree network with correlated sources, we derive rate bounds using graph entropies. We show that, these rates are achievable for the case of having independent sources and propose a modularized coding scheme based on graph colorings that performs arbitrarily closely to rate bounds. We also show that for a family of functions and random variables, which we call chain-rule proper sets, it is sufficient to have no computations at intermediate nodes to perform arbitrarily closely to the rate lower bound.

In the tree network depicted in Figure 4, nodes [1, . . . , 4] represent source nodes, nodes [5, 6] are intermediate nodes, and node 7 is the receiver. The receiver wishes to compute a deterministic function of source random variables. Intermediate nodes have no demand of their own, but they are allowed to perform computation. Computing the desired function \( f \) at the receiver is the only demand of the network. For this network, we compute a rate lower bound and show that, this bound is tight in the case of having independent sources. We also propose a modularized coding scheme to perform arbitrarily closely to derived rate lower bounds in this case.

Sources transmit variables \( M_1, . . . , M_4 \) through links \( e_1, . . . , e_4 \), respectively. Intermediate nodes transmit variables \( M_5 \) and \( M_6 \) over \( e_5 \) and \( e_6 \), respectively, where \( M_5 = g_5(M_1, M_2) \) and \( M_6 = g_6(M_3, M_4) \).

Let \( S(4) \) and \( S(5, 6) \) be the power sets of the set \{1, . . . , 4\} and the set \{5, 6\} except the empty set, respectively.

**Theorem 32:** A rate lower bound for the tree network of Figure 4 can be characterized as follows,

\[
\forall S \in S(4) \quad \sum_{i \in S} R_i \geq H_{G_{X_i}, k \in S}(X_S|X_{S'})
\]

\[
\forall S \in S(5, 6) \quad \sum_{i \in S} R_i \geq H_{G_{X_i}, k \in S}(X_S|X_{S'})
\]

Proof of this theorem is presented in Section VI-G. Note that, the result of Theorem 32 can be extended to a general tree network topology.

In the following, we show that for independent source variables, rate bounds of Theorem 32 are tight and we propose a coding scheme that performs arbitrarily closely to these bounds.

1) **Tightness of the Rate Lower Bound for Independent Sources:** Suppose random variables \( X_1, . . . , X_4 \) with characteristic graphs \( G^k_{X_1}, . . . , G^k_{X_4} \) are independent. Assume \( c_{G^k_{X_1}}, . . . , c_{G^k_{X_4}} \) are valid \( \epsilon \)-colorings of these characteristic graphs satisfying C.C.C. The following coding scheme performs arbitrarily closely to rate bounds of Theorem 32:

Source nodes first compute colorings of high probability subgraphs of their characteristic graphs satisfying C.C.C., and then, perform source coding on these coloring random variables. Intermediate nodes first compute their parents’ coloring random variables, and then, by using a look-up table, find corresponding source values of their received colorings. Then, they compute \( \epsilon \)-colorings of their own characteristic graphs. The corresponding source values of their received colorings form an independent set in the graph. If all are assigned to a single color in the minimum entropy coloring, intermediate nodes send this coloring random variable followed by a source coding. But, if vertices of this independent set are assigned to different colors, intermediate nodes send the coloring with the lowest entropy followed by source coding (Slepian-Wolf). The receiver first performs a minimum entropy decoding ([28]) on its received information and achieves coloring random variables. Then, it uses a look-up table to compute its desired function by using achieved colorings.

In the following, we summarize this proposed algorithm:

**Algorithm 33:** The following algorithm proposes a modularized coding scheme which performs arbitrarily closely to rate bounds of Theorem 32 when sources are independent:

- Source nodes compute \( \epsilon \)-colorings of sufficiently large power of their characteristic graphs satisfying C.C.C., followed by Slepian-Wolf compression.
- Intermediate nodes compute \( \epsilon \)-colorings of sufficiently large power of their characteristic graphs by using their parents colorings.
- The receiver first uses a Slepian-Wolf decoder to decode transmitted coloring variables. Then, it uses a look-up table to compute the function values.

The achievability proof of this algorithm is presented in Section VI-H. Also, in Section III-C, we show that, minimum entropy colorings of independent random variables can be computed efficiently.

2) **A Case When Intermediate Nodes Do not Need to Compute:** Though the proposed coding scheme in Algorithm 33 can perform arbitrarily closely to the rate lower bound, it may require computation at intermediate nodes. Here, we show that, for a family of functions and random variables, intermediate nodes do not need to perform computations.

**Definition 34:** Suppose \( f(X_1, . . . , X_k) \) is a deterministic function of random variables \( X_1, . . . , X_k \). \((f, X_1, . . . , X_k)\) is called a chain-rule proper set when for any \( s \in S(k) \),

\[ H_{G_{X_k}, k \in S}(X_s) = H_{G_{X_s}}(X_s). \]

**Theorem 35:** In a general tree network, if sources \( X_1, . . . , X_k \) are independent random variables and \((f, X_1, . . . , X_k)\) is a chain-rule proper set, it is sufficient to have intermediate nodes as relays (with no computations) to perform arbitrarily closely to the rate lower bound mentioned in Theorem 32.

Proof of this theorem is presented in Section VI-I.

In the following Lemma, we provide a sufficient condition to guarantee that a set is a chain-rule proper set.

**Lemma 36:** Suppose \( X_1 \) and \( X_2 \) are independent and \( f(X_1, X_2) \) is a deterministic function. If for any \( x_i^1 \) and \( x_i^2 \) in \( X_2 \) we have \( f(x_i^1, x_i^2) \neq f(x_j^1, x_j^2) \) for any possible \( i \) and \( j \), then, \((f, X_1, X_2)\) is a chain-rule proper set.

Proof of this lemma is presented in Section VI-J.
C. Polynomial Time Cases for Finding the Minimum Entropy Coloring of a Characteristic Graph

In proposed coding schemes of Sections III-A and III-B, one needs to compute a minimum entropy coloring (a coloring random variable which minimizes the entropy) of a characteristic graph. In general, finding this coloring is an NP-hard problem (as shown by Cardinal et al. [32]). However, in this section, we show that depending on the characteristic graph’s structure, there are some interesting cases where finding the minimum entropy coloring is not NP-hard, but tractable and practical. In one of these cases, we show that, having a non-zero joint probability condition on random variables’ distributions, for any desired function \( f \), makes characteristic graphs to be formed of some non-overlapping fully-connected maximal independent sets. We will show that, in this case, the minimum entropy coloring can be computed in polynomial time, according to Algorithm 41. In another case, we show that, if the function we seek to compute is a type of quantization function, this problem is also tractable.

For simplicity, we consider functions with two input random variables, but one can extend all discussions to functions with more input random variables than two. Suppose \( c_{min}^{G_{X_1}} \) represents a minimum entropy coloring of a characteristic graph \( G_{X_1} \).

1) Non-Zero Joint Probability Distribution Condition: Consider the network shown in Figure 1-b. Source random variables have a joint probability distribution \( p(x_1, x_2) \), and the receiver wishes to compute a deterministic function of sources (i.e., \( f(X_1, X_2) \)). In this section, we consider the effect of the probability distribution of sources in computations of minimum entropy colorings.

Theorem 37: Suppose for all \( (x_1, x_2) \in X_1 \times X_2 \), \( p(x_1, x_2) > 0 \). Then, maximal independent sets of the characteristic graph \( G_{X_1} \) (and, its \( n \)-th power \( G_{X_1}^n \), for any \( n \)) are some non-overlapping fully-connected sets. Under this condition, the minimum entropy coloring can be achieved by assigning different colors to its different maximal independent sets as described in Algorithm 41.

Here are some remarks about Theorem 37:

- The condition \( p(x_1, x_2) > 0 \), for all \( (x_1, x_2) \in X_1 \times X_2 \), is a necessary condition for Theorem 37. In order to illustrate this, consider Figure 5. In this example, \( x_1^1 \) and \( x_1^3 \) are in \( X_1 \), and \( x_2^1, x_2^2 \) and \( x_3^3 \) are in \( X_2 \). Suppose \( p(x_1^1, x_2^1) = 0 \). By considering the value of function \( f \) at these points depicted in the figure, one can see that, in \( G_{X_1} \), \( x_1^1 \) is not connected to \( x_1^2 \) and \( x_1^3 \). However, \( x_1^1 \) and \( x_1^3 \) are connected to each other. Thus, Theorem 37 does not hold here.

- The condition used in Theorem 37 only restricts the probability distribution and it does not depend on the function \( f \). Thus, for any function \( f \) at the receiver, if we have a non-zero joint probability distribution of source random variables (for example, when source random variables are independent), finding the minimum-entropy coloring is easy and tractable.

Reference [14] showed that, for the side information problem, having non-zero joint probability condition yields to a simplified graph entropy calculation. Here, we show that, under this condition, characteristic graphs have certain structures so that optimal coloring based coding schemes that perform arbitrarily closely to rate lower bounds can be designed efficiently.

2) Quantization Functions: In this section, we consider some special functions which leads to practical minimum entropy coloring computation.

An interesting function is a quantization function. A natural quantization function is a function which separates the \( X_1 - X_2 \) plane into some rectangles so that each rectangle corresponds to a different value of that function. Sides of these rectangles are parallel to the plane axes. Figure 6-a depicts such a quantization function.

Given a quantization function, one can extend different sides of each rectangle in the \( X_1 - X_2 \) plane. This may make some new rectangles. We call each of them a function region. Each function region can be determined by two subsets of \( X_1 \) and \( X_2 \). For example, in Figure 6-b, one of the function regions is distinguished by the shaded area.

Definition 38: Consider two function regions \( X_1^1 \times X_2^1 \) and \( X_1^2 \times X_2^2 \). If for any \( x_1^1 \in X_1^1 \) and \( x_2^1 \in X_2^1 \), there exist \( x_2^1 \) such that \( p(x_1^1, x_2^1) p(x_1^1, x_2^1) > 0 \) and \( f(x_1^1, x_2^1) \neq f(x_1^1, x_2^2) \), we say these two function regions are pairwise \( X_1 \)-proper.

Theorem 39: Consider a quantization function \( f \) such that its function regions are pairwise \( X_1 \)-proper. Then, \( G_{X_1} \) (and \( G_{X_1}^n \), for any \( n \)) is formed of some non-overlapping fully-connected maximal independent sets, and its minimum entropy coloring can be achieved by assigning different colors to different maximal independent sets.
A. Feedback in Functional Compression

If the function at the receiver is the identity function, the functional compression problem is Slepian-Wolf compression with feedback. For this case, having feedback does not improve rate bounds. For example, see reference [11] which considers both zero-error and asymptotically zero-error Slepian-Wolf compression with feedback. However, here by an example we show that, for a general desired function at the receiver, having feedback can improve rate bounds of the case without feedback.

Example 42: Consider a distributed functional compression problem with two sources and a receiver as depicted in Figure 7-a. Suppose each source has one byte (8 bits) to transmit to the receiver. Bits are sorted from MSB (most significant bit) to LSB (least significant bit). Bits can be 0 or 1 with the same probability. The desired function at the receiver is $f(X_1, X_2) = \max(X_1, X_2)$.

In the case of not having feedback, the characteristic graphs of sources are trivially complete graphs. Therefore, each source should transmit all bits to the receiver. The un-scaled rates are $R_1 = 8$ and $R_2 = 8$.

Now, suppose the receiver can broadcast some feedback bits to sources. In the following, we propose a communication scheme that has a reduced sum transmission rate compared to the case without feedback:

First, each source transmits its MSB. The receiver compares two received bits. If they are the same, the receiver broadcasts 0 to sources, otherwise it broadcasts 1. If sources receive 1 from feedback, they stop transmitting. Otherwise, they transmit their next significant bits. For this communication scheme, the un-scaled sum rate of the forward links can be calculated as follows:

$$ R_{1f} + R_{2f} = \frac{1}{2}(2 \times 1) + \frac{1}{4}(2 \times 2) + \ldots + \frac{1}{2n}(2 \times n) $$

where $n$ is the block length (in this example, $n = 8$), and $R_{1f}$ and $R_{2f}$ are transmission rates of sources $X_1$ and $X_2$, respectively. Sources stop transmitting after the first bit transmission if these bits are not equal. The probability of this event is $1/2$ (term (a) in equation (12)). Similarly, forward transmissions stop in the second round with probability $1/4$ (term (b) in equation (12)) and so on. Equality (c) follows from a closed-form solution for series $\sum_{i=1}^{\infty} \frac{1}{2i}$. For $n = 8$, this rate is around 3.92 which is less than the sum-rate without feedback. With similar calculations, the feedback rate is around 1.96. Hence, the total forward and feedback transmission rates is around 5.88 which is less than the one of without feedback.

B. A Practical Rate-Distortion Scheme for Distributed Functional Compression

In this section, we consider the problem of distributed functional compression with distortion. The objective is to compress correlated discrete sources so that an arbitrary
deterministic function of sources can be computed up to a distortion level at the receiver. Here, we present a practical coding scheme for this problem with a non-trivial performance guarantee. All discussions can be extended to more general networks similar to results of Section III.

Consider two sources as described in Section II-A. Here, we assume that, the receiver wants to compute a deterministic function \( f : X_1 \times X_2 \rightarrow Z \) or \( f : X_1^n \times X_2^n \rightarrow Z^n \), its vector extension up to distortion \( D \) with respect to a given distortion function \( d : Z \times Z \rightarrow [0, \infty) \). A vector extension of the distortion function is defined as follows:

\[
d(z_1, z_2) = \frac{1}{n} \sum_{i=1}^{n} d(z_{1i}, z_{2i}),
\]

where \( z_1, z_2 \in Z^n \). We assume that \( d(z_1, z_2) = 0 \) if and only if \( z_1 = z_2 \). This assumption causes vector extension to satisfy the same property (i.e., \( d(z_1, z_2) = 0 \) if and only if \( z_1 = z_2 \)).

The probability of error in this case is

\[
P^*_e = P(\{ (x_1, x_2) : d(f(x_1, x_2), r(en_{X_1}(x_1), en_{X_2}(x_2))) \geq D \}).
\]

We say a rate pair \((R_1, R_2)\) is achievable up to distortion \( D \) if there exist \( en_{X_1}, en_{X_2} \) and \( r \) such that \( P^*_e \rightarrow 0 \) as \( n \rightarrow \infty \).

Yamamoto gives a characterization of a rate-distortion function for the side information functional compression problem (i.e., \( X_2 \) is available at the receiver) in [23]. The rate-distortion function proposed in [23] is a generalization of the Wyner-Ziv side-information rate-distortion function [22]. Another multi-letter characterization of the rate distortion function for the side information problem given by Yamamoto was discussed in [15]. The multi-letter characterization of [15] can be extended naturally to a distributed functional compression case by using results of Section III.

Here, we present a practical coding scheme with a non-trivial performance guarantee for a given distributed lossy functional compression setup.

Define the \( D \)-characteristic graph of \( X_1 \) with respect to \( X_2 \), \( p(x_1, x_2) \), and \( f(X_1, X_2) \), as having vertices \( V = X_1 \) and the pair \((x_1^1, x_2^1)\) is an edge if there exists some \( x_2^1 \in X_2 \) such that \( p(x_1^1, x_2^1)p(x_2^1, x_2^1) > 0 \) and \( d(f(x_1^1, x_2^1), f(x_1^1, x_2^1)) > D \) as in [15]. Denote this graph as \( G_{X_1}(D) \). Similarly, we define \( G_{X_2}(D) \).

**Theorem 43:** For the network depicted in Figure 1-b with independent sources, if the distortion function is a metric, then the following rate pairs \((R_1, R_2)\) is achievable for the distributed lossy functional compression problem with distortion \( D \):

\[
R_1 \geq H_{G_{X_1}(D/2)}(X_1)
\]

\[
R_2 \geq H_{G_{X_2}(D/2)}(X_2)
\]

\[
R_1 + R_2 \geq H_{G_{X_1}(D/2), G_{X_2}(D/2)}(X_1, X_2). \tag{14}
\]

Proof of this theorem is presented in Section VI-N. A modularized coding scheme similar to Algorithm 31 can be used on \( D \)-characteristic graphs to perform arbitrarily closely to rate bounds of Theorem 43.

**C. Future Work**

Throughout this paper, we considered the case of having only one desired function at the receiver. However, all results can be extended naturally to the case of having several desired functions at the receiver by considering a vector extension of functions, and computing characteristic graphs of variables with respect to that vector. In fact, one can show that, the characteristic graph of a random variable with respect to several functions is equal to the union of individual characteristic graphs (union of two graphs with the same vertices is a graph with the same vertex set whose edges are the union of edges of individual graphs.

For possible future work, one may consider a general network topology rather than tree networks. For instance, one can consider a general multi-source multicast network in which receivers desire to have a deterministic function of source random variables. For the case of having the identity function at the receivers, this problem is well-studied in [34]–[36] under the name of network coding for multi-source multicast networks. Reference [36] shows that, random linear network coding can perform arbitrarily closely to min-cut max-flow bounds. To have an achievable scheme for the functional version of this problem, one may perform random network coding on coloring random variables satisfying C.C.C. If receivers desire different functions, one can use colorings of multi-functional characteristic graphs satisfying C.C.C., and then use random network coding for these coloring random variables. This achievable scheme can be extended to disjoint multicast and disjoint multicast plus multicast cases described in [35]. This scheme is an achievable scheme; however it is not optimal in general. If sources are independent, one may use encoding/decoding functions derived for tree networks at intermediate nodes, along with network coding.

Throughout this paper, we considered asymptotically lossless or lossy computation of a function. For possible future work, one may consider this problem for the zero-error computation of a function which leads to a communication complexity problem. One can use tools and schemes we have developed in this paper to attain some achievable schemes in the zero error computation case as well.

**V. Conclusions**

In this paper, we considered different aspects of the functional compression problem where computing a function (or, some functions) of sources is desired at the receiver(s). The rate region of this problem has been considered in the literature under certain restrictive assumptions, particularly in terms of the network topology, the functions and the characteristics of the sources. In this paper, we significantly relaxed these assumptions.

For a one-stage tree network, we introduced Coloring Connectivity Condition (C.C.C.), a necessary and sufficient condition for any achievable coloring-based coding scheme, relaxing the previous sufficient zig-zag condition of Doshi et al. By using C.C.C., we characterized a rate region for this problem and proposed a modularized coding scheme based on graph colorings to perform arbitrarily closely to rate lower bounds.
For a general tree network, we provided a rate lower bound based on graph entropies and showed that, this bound is tight in the case of having independent sources. In particular, we showed that in a general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform computations. However, for a family of functions and random variables, which we call chain-rule proper sets, it is sufficient to have intermediate nodes act like relays (with no computations) to perform arbitrarily closely to the rate lower bound. We discussed practical issues of coloring-based coding schemes and proposed an efficient algorithm to compute minimum entropy colorings under some conditions on source distributions and/or the desired function.

Finally, extensions of these results for cases of having feedback and lossy function computations were discussed. Particularly, we showed that, in functional compression, unlike Slepian-Wolf, by having feedback, one may outperform rate bounds of the case without feedback. These results extend those of Bakshi et al. For lossy distributed functional compression, we presented a practical coding scheme with a non-trivial performance guarantee.

VI. PROOFS

A. Proof of Lemma 22

Proof: First, we know that any random variable $X_2$ by itself is a trivial coloring of $G_{X_2}$ such that each vertex of $G_{X_2}$ is assigned to a different color. So, $J_C$ for $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ can be written as $J_C = \{j_1^X, \ldots, j_n^X\}$ such that $j_i^X = (x_i^1, x_i^2, \ldots, x_i^n)$, where $x_i^1$ is a generic color. Any two points in $j_i^X$ are connected to each other with a path with length one. So, $j_i^X$ satisfies C.C.C. This arguments hold for any $j_i^X$ for any valid $i$. Thus, all joint coloring classes and therefore, $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ satisfy C.C.C.

The argument for $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ is similar. ■

B. Proof of Lemma 23

Proof: We first show that if $j_i^j$ satisfies C.C.C., then, for any two points $(x_1^1, x_1^2, \ldots, x_1^k)$ and $(x_2^1, x_2^2, \ldots, x_2^k)$ in $j_i^j$, $f(x_1^1, x_1^2, \ldots, x_1^k) = f(x_2^1, x_2^2, \ldots, x_2^k)$. Since $j_i^j$ satisfies C.C.C., by definition, either $f(x_1^1, x_1^2, \ldots, x_1^k) = f(x_2^1, x_2^2, \ldots, x_2^k)$, or there exists a path with length $m - 1$ between these two points $Z_1 = (x_1^1, \ldots, x_1^k)$ and $Z_m = (x_2^1, \ldots, x_2^k)$, for some $m$, where two consecutive points $Z_j$ and $Z_{j+1}$ in this path, differ in exactly one of their coordinates. Without loss of generality, suppose $Z_j$ and $Z_{j+1}$ differ in their first coordinate, i.e., $Z_j = (x_1^1, x_2^2, \ldots, x_k)$ and $Z_{j+1} = (x_j^1, x_2^2, \ldots, x_k)$. Since these two points belong to $j_i^j$, $c_{G_{X_1}}(x_j^1) = c_{G_{X_1}}(x_j^0)$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $x_j^1$ and $x_j^0$ in $G_{X_1}$ and they could not have the same color. Therefore, $f(Z_j) = f(Z_{j+1})$. By applying the same argument inductively for all two consecutive points in the path between $Z_1$ and $Z_m$, we have $f(Z_1) = f(Z_2) = \cdots = f(Z_m)$.

If $j_i^j$ does not satisfy C.C.C., by definition there exists at least two points $Z_1$ and $Z_2$ in $j_i^j$ with different function values.

C. Proof of Lemma 24

Proof: The proof is similar to Lemma 23. The only difference is to use the definition of C.C.C. for $c_{G_{X_1}}$ and $c_{G_{X_2}}$. Since $j_i^j$ satisfies C.C.C., either $f(x_1^1, \ldots, x_1^k) = f(x_2^1, \ldots, x_2^k)$, or there exists a path with length $m - 1$ between any two points $Z_1 = (x_1^1, \ldots, x_1^k) \in T^e$ and $Z_m = (x_2^1, \ldots, x_2^k) \in T^e$ in $j_i^j$, for some $m$. Consider two consecutive points $Z_j$ and $Z_{j+1}$ in this path. They differ in one of their coordinates (suppose they differ in their first coordinate). In other words, suppose $Z_j = (x_j^1, x_2^2, \ldots, x_k^k)$ and $Z_{j+1} = (x_j^0, x_2^2, \ldots, x_k^k)$ in $T^e$. Since these two points belong to $j_i^j$, $c_{G_{X_1}}(x_i^1) = c_{G_{X_1}}(x_i^0)$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $x_j^1$ and $x_j^0$ in $G_{X_1}$, and they could not get the same color. Thus, $f(Z_j) = f(Z_{j+1})$. By applying the same argument for all two consecutive points in the path between $Z_1$ and $Z_m$, one can get $f(Z_1) = f(Z_2) = \cdots = f(Z_m)$. The converse part is similar to Lemma 23.

D. Proof of Lemma 25

Proof: Suppose $X_1$ and $X_2$ satisfy the zigzag condition, and $c_{G_{X_1}}$ and $c_{G_{X_2}}$ are two valid colorings of $G_{X_1}$ and $G_{X_2}$, respectively. We want to show that these colorings satisfy C.C.C. To do this, consider two points $(x_1^1, x_2^2)$ and $(x_1^2, x_2^2)$ in a joint coloring class $j_i^j$. The definition of the zigzag condition guarantees the existence of a path with length two between these two point. Thus, $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C.

The second part of this Lemma says that the converse part is not true. To have an example, one can see that in a special case considered in Lemma 22, those colorings always satisfy C.C.C. without having any condition such as the zigzag condition.

E. Proof of Lemma 28

Proof: By using the data processing inequality, we have

$$H_{G_{X_1}}(X_1|X_2) = \lim_{n \to \infty} \min_{c_{G_{X_1}}(X_1), c_{G_{X_2}}(X_2)} \frac{1}{n} H(c_{G_{X_1}}(X_1)|c_{G_{X_2}}(X_2))$$

$$= \lim_{n \to \infty} \min_{c_{G_{X_1}}(X_1)} \frac{1}{n} H(c_{G_{X_1}}(X_1)|X_2).$$

Then, Lemma 22 implies that $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ satisfy C.C.C. A direct application of Theorem 15 completes the proof.

F. Proof of Theorem 29

Proof: We first show the achievability of this rate region. We also propose a modularized encoding/decoding scheme in this part. Then, for the converse, we show that no encoding/decoding scheme can outperform this rate region.

1) Achievability:

**Lemma 44:** Consider random variables $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$, and any valid $e$-colorings $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ satisfying C.C.C. over typical points $T^n$, for sufficiently large $n$. There exists

$$\hat{f} : c_{G_{X_1}}(X_1) \times \cdots \times c_{G_{X_k}}(X_k) \to \mathbb{Z}^n \quad (15)$$
such that \( \hat{f}(c_{G^n_{X_1}}(x_1), \ldots, c_{G^n_{X_k}}(x_k)) = f(x_1, \ldots, x_k) \), for all \((x_1, \ldots, x_k) \in T^n_p\).

Proof: Suppose the joint coloring family for these colorings is \( J_C = \{ j_i \} \cdot \) We proceed by constructing \( f \).
Assume \((x_1^1, \ldots, x_1^n) \in j_i^1 \) and \((x_2^1, \ldots, x_2^n) \in j_i^2 \).
Define \( \hat{f}(\sigma_1, \ldots, \sigma_k) = f(x_1^1, \ldots, x_1^n) \).
To show this function is well-defined on elements in its support, we should show that for any two points \((x_1^1, \ldots, x_1^n) \) and \((x_2^1, \ldots, x_2^n) \) in \( T^n_p \), if \( c_{G^n_{X_1}}(x_1^1) = c_{G^n_{X_1}}(x_2^1) \), \ldots, \( c_{G^n_{X_k}}(x_1^n) = c_{G^n_{X_k}}(x_2^n) \), then \( f(x_1^1, \ldots, x_1^n) = f(x_2^1, \ldots, x_2^n) \).
Since \( c_{G^n_{X_1}}(x_1^1) = c_{G^n_{X_1}}(x_2^1) \), \ldots, \( c_{G^n_{X_k}}(x_1^n) = c_{G^n_{X_k}}(x_2^n) \), these two points belong to a joint coloring class such as \( j_i^1 \).
Since \( c_{G^n_{X_1}}, \ldots, c_{G^n_{X_k}} \) satisfy C.C.C., by using Lemma 24, \( f(x_1^1, \ldots, x_1^n) = f(x_2^1, \ldots, x_2^n) \). Therefore, our function \( \hat{f} \) is well-defined and has the desired property.

Lemma 44 implies that, given \( \epsilon \)-colorings of characteristic graphs of random variables satisfying C.C.C. at the receiver, we can successfully compute the desired function \( f \) with a vanishing probability of error as \( n \) goes to infinity.
Thus, if the decoder at the receiver is given colors, it can look up \( f \) based on its table of \( \hat{f} \).
The question remains of which rates encoders can transmit these colors to the receiver faithfully (with a probability of error less than \( \epsilon \)).

Lemma 45 (Slepian-Wolf Theorem): A rate-region of a one-stage tree network with the desired identity function at the receiver is characterized by these conditions:

\[ \forall S \in S(k) \implies \sum_{i \in S} R_i \geq H(X_S | X_{S^c}). \]  
(16)

Proof: See [16].

We now use the Slepian-Wolf (SW) encoding/decoding scheme on achieved coloring random variables. Suppose the probability of error in each decoder of SW is less than \( \frac{\epsilon}{2} \). Then, the total error in the decoding of colorings at the receiver is less than \( \epsilon \). Therefore, the total error in the coding scheme of first coloring \( G^n_{X_1}, \ldots, G^n_{X_k} \), and then encoding those colors by using SW encoding/decoding scheme is upper bounded by the sum of errors in each stage. By using Lemmas 44 and 45, the total error is less than \( \epsilon \), and goes to zero as \( n \) goes to infinity. By applying Lemma 45 on achieved coloring random variables, we have,

\[ \forall S \in S(k) \implies \sum_{i \in S} R_i \geq \frac{1}{n} H(c_{G^n_{X_1}} | c_{G^n_{X_k}}). \]  
(17)

where \( c_{G^n_{X_1}}, \ldots, c_{G^n_{X_k}} \) are \( \epsilon \)-colorings of characteristic graphs satisfying C.C.C. Thus, using Definition 27 completes the achievability part.

2) Converse: Here, we show that any distributed functional source coding scheme with a small probability of error induces \( \epsilon \)-colorings on characteristic graphs of random variables satisfying C.C.C. Suppose \( \epsilon > 0 \). Define \( F^n_\epsilon \) for all \((n, \epsilon)\) as follows,

\[ F^n_\epsilon = \{ \hat{f} : Pr[\hat{f}(X_1, \ldots, X_k) \neq f(X_1, \ldots, X_k)] < \epsilon \}. \]  
(18)

In other words, \( F^n_\epsilon \) is the set of all functions that differ from \( f \) with probability \( \epsilon \). Suppose \( \hat{f} \) is an achievable code with vanishing error probability where

\[ \hat{f}(x_1, \ldots, x_k) = \hat{r}(\epsilon_{X_1,a}(x_1), \ldots, \epsilon_{X_k,a}(x_k)). \]  
(19)

where \( n \) is the block length. Then there exists \( n_0 \) such that for all \( n > n_0 \), \( Pr(\hat{f} \neq f) < \epsilon \). In other words, \( \hat{f} \in F^n_\epsilon \). We can also use these codes \( \epsilon \)-error functional codes.

Lemma 46: Consider some function \( f : X_1 \times \cdots \times X_k \rightarrow Z \). Any distributed functional code which reconstructs this function with zero error probability induces colorings on \( G_{X_1}, \ldots, G_{X_k} \) satisfying C.C.C. with respect to this function.

Proof: Say we have a zero-error distributed functional code represented by encoders \( \epsilon_{X_1}, \ldots, \epsilon_{X_k} \) and a decoder \( r \). For any two points \((x_1^1, \ldots, x_1^n) \) and \((x_2^1, \ldots, x_2^n) \) with positive probabilities, if their encoded values are the same (i.e., \( \epsilon_{X_1}(x_1^1) = \epsilon_{X_1}(x_2^1), \ldots, \epsilon_{X_k}(x_1^n) = \epsilon_{X_k}(x_2^n) \)), their function values will be the same as well since it is an error free scheme.

\[ f(x_1^1, \ldots, x_1^n) = f(x_2^1, \ldots, x_2^n). \]  
(20)

We show that \( \epsilon_{X_1}, \ldots, \epsilon_{X_k} \) are in fact some valid colorings of \( G_{X_1}, \ldots, G_{X_k} \) satisfying C.C.C. We demonstrate this argument for \( X_1 \). The argument for other random variables is analogous. First, we show that \( \epsilon_{X_1} \) induces a valid coloring on \( G_{X_1} \), and then, we show that this coloring satisfies C.C.C.

Let us proceed by contradiction. If \( \epsilon_{X_1} \) did not induce a coloring on \( G_{X_1} \), there must be some edge in \( G_{X_1} \) connecting two vertices with the same color. Let us call these vertices \( x_1^1 \) and \( x_1^2 \). Since these vertices are connected in \( G_{X_1} \), there must exist a \((x_1^1, \ldots, x_k^n) \) such that \( p(x_1^1, x_1^2, \ldots, x_k^n) > 0, \epsilon_{X_1}(x_1^1) = \epsilon_{X_1}(x_1^2), \) and \( f(x_1^1, x_1^2, \ldots, x_k^n) \neq f(x_1^1, x_1^2, \ldots, x_k^n) \). Taking \( x_1^1 = x_1^2, \ldots, x_1^n = x_1^n \) as in equation (20) leads to a contradiction. Therefore, the contradiction assumption is wrong and \( \epsilon_{X_1} \) induces a valid coloring on \( G_{X_1} \).

Now, we show that these induced colorings satisfy C.C.C. If it was not true, it would mean that there must exist two point \((x_1^1, \ldots, x_1^n) \) and \((x_1^1, \ldots, x_1^n) \) in a joint coloring class \( j_i^1 \) so that there is no path between them in \( j_i^1 \). So, Lemma 23 says that the function \( f \) can get different values in these two points. In other words, it is possible to have \( f(x_1^1, \ldots, x_1^n) \neq f(x_1^1, \ldots, x_1^n) \), where \( c_{G_{X_1}}(x_1^1) = c_{G_{X_1}}(x_1^2), \ldots, c_{G_{X_1}}(x_1^n) = c_{G_{X_1}}(x_1^n) \), which is in contradiction with equation (20). Thus, these colorings satisfy C.C.C.

In the last step, we should show that any achievable functional code represented by \( F^n_\epsilon \) induces \( \epsilon \)-colorings on characteristic graphs satisfying C.C.C.

Lemma 47: Consider random variables \( X_1, \ldots, X_k \). All \( \epsilon \)-error functional codes of these random variables induce \( \epsilon \)-colorings on characteristic graphs satisfying C.C.C.

Proof: Suppose \( \hat{f}(x_1, \ldots, x_k) = r(\epsilon_{X_1}(x_1), \ldots, \epsilon_{X_k}(x_k)) \in F^n_\epsilon \) is such a code. If the desired function to compute was \( \hat{f} \), according to Lemma 46, a zero-error reconstruction of \( \hat{f} \) induces colorings on characteristic graphs satisfying C.C.C., with respect to \( \hat{f} \). Suppose the set of all points \((x_1, \ldots, x_k) \) such that \( \hat{f}(x_1, \ldots, x_k) \neq f(x_1, \ldots, x_k) \) be denoted by \( C \). Since \( Pr(\hat{f} \neq f, Pr[C] < \epsilon \). Therefore, functions \( \epsilon_{X_1}, \ldots, \epsilon_{X_k} \) restricted to \( C \) are \( \epsilon \)-colorings
of characteristic graphs satisfying C.C.C. with respect to $f$.

According to Lemmas 46 and 47, any distributed functional source codes with vanishing error probability induces $\epsilon$-colorings on characteristic graphs of source variables satisfying C.C.C. with respect to the desired function $f$.

Then, according to Slepian-Wolf Theorem 45, we have

$$\forall S \in S(k) \implies \sum_{i \in S} R_i \geq \frac{1}{n} \mathbb{H}(c_{G_{X_1}}^u | c_{G_{X_2}}^u),$$

where $c_{G_{X_1}}^u$ and $c_{G_{X_2}}^u$ are $\epsilon$-colorings of characteristic graphs satisfying C.C.C. with respect to $f$. Using Definition 27 completes the converse part.

G. Proof of Theorem 32

Proof: Here, we show that no coding scheme can outperform this rate region. Suppose source nodes $\{1, \ldots, 4\}$ are directly connected to the receiver. By direct application of Theorem 29, the first set of conditions of Theorem 32 can be derived. Repeating the argument for intermediate nodes $\{5, 6\}$ completes the proof.

H. Achievability Proof of Algorithm 33

Proof: To show the achievability, we show that, if nodes of each stage were directly connected to the receiver, the receiver could compute its desired function. For nodes $\{1, \ldots, 4\}$ in the first stage, the argument directly follows from Theorem 29. Now, we show that, the argument holds for intermediate nodes $\{5, 6\}$ as well. Consider node 5 in the second stage of the network. Since the corresponding source values of its received colorings form an independent set on its characteristic graph and since this node computes the minimum entropy coloring of this graph, it is equivalent to the case where it would receive the exact source information, because both of them lead to the same coloring random variable. Therefore, by having nodes 5 and 6 directly connected to the receiver, and a direct application of Theorem 29, the receiver is able to compute its desired function by using colorings of characteristic graphs of nodes 5 and 6.

I. Proof of Theorem 35

Proof: Here, we present the proof for the tree network structure depicted in Figure 4. However, all arguments can be extended to a general case. Suppose intermediate nodes 5 and 6 perform no computations and act as relays. Therefore, we have

$$R_5 = R_1 + R_2 = H_{G_{X_1}, G_{X_2}}(X_1, X_2 | X_3, X_4).$$

By using the chain-rule proper set condition, we can re-write it as

$$R_5 = H_{G_{X_1}, X_2}(X_1, X_2 | X_3, X_4),$$

which is the same condition as the one of Theorem 32. Repeating this argument for $R_6$ and $R_5 + R_6$ establishes the proof.

J. Proof of Lemma 36

Proof: To prove this lemma, it is sufficient to show that, under the conditions of this lemma, any colorings of the graph $G_{X_1, X_2}$ can be expressed as colorings of $G_{X_1}$ and $G_{X_2}$, and vice versa. The converse part is straightforward because any colorings of $G_{X_1}$ and $G_{X_2}$ can be viewed as a coloring of $G_{X_1, X_2}$.

Consider Figure 9 which illustrates conditions of this lemma. Under these conditions, since all $x_2$ in $X_2$ have different function values, graph $G_{X_1, X_2}$ can be decomposed to subgraphs which have the same topology as $G_{X_1}$ (i.e., isomorphism to $G_{X_1}$), corresponding to each $x_2$ in $X_2$. These subgraphs are fully connected to each other under conditions of this lemma. Thus, any coloring of this graph can be represented as two colorings of $G_{X_1}$ (within each subgraph) and $G_{X_2}$ (across subgraphs). Therefore, the minimum entropy coloring of $G_{X_1, X_2}$ is equal to the minimum entropy coloring of $(G_{X_1}, G_{X_2})$, I.e., $H_{G_{X_1}, G_{X_2}}(X_1, X_2) = H_{G_{X_1, X_2}}(X_1, X_2)$.

K. Proof of Theorem 37

Proof: Suppose $\Gamma(G_{X_1})$ is the set of all maximal independent sets of $G_{X_1}$. Let us proceed by contradiction. Consider Figure 8-a. Suppose $w_1$ and $w_2$ are two different non-empty maximal independent sets. Without loss of generality, assume $x_1^1$ and $x_1^2$ are in $w_1$, and $x_1^2$ and $x_1^3$ are in $w_2$. These sets have a common element $x_1^2$. Since $w_1$ and $w_2$ are two different maximal independent sets, $x_1^1 \notin w_2$ and $x_1^3 \notin w_1$. Since $x_1^1$ and $x_1^2$ are in $w_1$, there is no edge between them in $G_{X_1}$. The same argument holds for $x_2^1$ and $x_2^3$. But, we have an edge between $x_1^1$ and $x_1^3$, because $w_1$ and $w_2$ are two different
 maximal independent sets, and at least there should exist such an edge between them. Now, we want to show that it is not possible.

Since there is no edge between $x^1_i$ and $x^2_i$, for any $x^1_i \in X_2$, $p(x^1_i, x^1_j)p(x^2_i, x^2_j) > 0$, and $f(x^1_i, x^2_j) = f(x^2_i, x^1_j)$. A similar argument can be expressed for $x^1_i$ and $x^2_i$. In other words, for any $x^1_i \in X_2$, $p(x^1_i, x^1_j)p(x^2_i, x^2_j) > 0$, and $f(x^1_i, x^2_j) = f(x^2_i, x^1_j)$. Thus, for all $x^1_i \in X_2$, $p(x^1_i, x^1_j)p(x^2_i, x^2_j) > 0$, and $f(x^1_i, x^2_j) = f(x^2_i, x^1_j)$.

However, since $x^1_i$ and $x^2_i$ are connected to each other, there should exist a $x^2_i \in X_2$ such that $f(x^2_i, x^2_j) \neq f(x^2_i, x^2_j)$ which is not possible. So, the contradiction assumption is not correct and these two maximal independent sets do not overlap with each other.

We showed that maximal independent sets cannot have overlaps with each other. Now, we want to show that they are also fully connected to each other. Again, let us proceed by contradiction. Consider Figure 8-b. Suppose $w_1$ and $w_2$ are two different non-overlapping maximal independent sets. Suppose there exists an element in $w_2$ (call it $x^3_j$) which is connected to one of elements in $w_1$ (call it $x^1_i$) and is not connected to another element of $w_1$ (call it $x^1_k$).

By using a similar discussion to the one in the previous paragraph, we may show that it is not possible. Thus, $x^1_i$ should be connected to $x^1_k$. Therefore, if for all $(x^1_i, x^2_j) \in X_1 \times X_2$, $p(x^1_i, x^2_j) > 0$, then maximal independent sets of $G_{X_1}$ are some separate fully connected sets. In other words, the complement of $G_{X_1}$ is formed by some non-overlapping cliques. Finding the minimum entropy coloring of this graph is trivial and can be achieved by assigning different colors to these non-overlapping fully-connected maximal independent sets.

This argument also holds for any power of $G_{X_1}$. Suppose $x^1_{i_1}$, $x^2_{j_1}$ and $x^3_{k_1}$ are some typical sequences in $X_{1}^{n}$. If $x^1_{i_1}$ is not connected to $x^2_{j_1}$ and $x^3_{k_1}$, it is not possible to have $x^2_{j_1}$ and $x^3_{k_1}$ connected. Therefore, one can apply a similar argument to prove the theorem for $G_{X_1}$, for some $n$. This completes the proof.

L. Proof of Theorem 39

Proof: We first prove it for $G_{X_1}$. Suppose $X_{1}^1 \times X_{1}^1$, and $X_{1}^2 \times X_{1}^2$ are two $X_1$-proper function regions of a quantization function $f$, where $X_{1}^1 \neq X_{1}^2$. We show that $X_{1}^1$ and $X_{1}^2$ are two non-overlapping fully-connected maximal independent sets. By definition, $X_{1}^1$ and $X_{1}^2$ are two non-equal partition sets of $X_1$. Thus, they do not have any element in common.

Now, we want to show that vertices of each of these partition sets are not connected to each other. Without loss of generality, we show it for $X_{1}^1$. If this partition set of $X_1$ has only one element, this is a trivial case. So, suppose $x^1_{i_1}$ and $x^2_{j_1}$ are two elements in $X_{1}^1$. By definition of function regions, one can see that, for any $x^2_{j_1} \in X_2$ such that $p(x^1_{i_1}, x^2_{j_1})p(x^2_{i_1}, x^2_{j_1}) > 0$, then $f(x^1_{i_1}, x^1_{j_1}) = f(x^1_{i_1}, x^2_{j_1})$. Thus, these two vertices are not connected to each other. Now, suppose $x^3_{k_1}$ is an element in $X_{1}^2$. Since these function regions are $X_1$-proper, there should exist at least one $x^2_{j_1} \in X_2$, such that $p(x^1_{i_1}, x^2_{j_1})p(x^2_{i_1}, x^2_{j_1}) > 0$, and $f(x^1_{i_1}, x^2_{j_1}) \neq f(x^1_{i_1}, x^2_{j_1})$. Thus, $x^1_{i_1}$ and $x^3_{k_1}$ are connected to each other. Therefore, $X_{1}^1$ and $X_{1}^2$ are two non-overlapping fully-connected maximal independent sets. One can easily apply this argument to other partition sets. Thus, the minimum entropy coloring can be achieved by assigning different colors to different maximal independent sets (partition sets). The proof for $G_{X_1}^n$, for any $n$, is similar to the one mentioned in Theorem 37. This completes the proof.

M. Achievability Proof of Algorithm 41

Proof: Suppose $G_{X_1}$ has $P$ vertices labeled as $\{1, \ldots, P\}$ and sorted in a list. Say if a vertex $v$ has $d_v$ neighbors in $G_{X_1}$, the complexity of finding them is in the order of $O(d_v \log(P))$. Therefore, for a vertex $v$, the complexity of the first two steps of the algorithm is in the order of $O(d_v \log(P))$, where $O(\log(P))$ is the complexity of updating the list of un-colored vertices. Therefore, the overall worst case complexity of the algorithm is in the order of $O(P^2 \log(P))$. Since maximal independent sets of graph $G_{X_1}$ are non-overlapping and also fully connected, any valid coloring scheme should assign them to different colors. Therefore, the minimum number of required colors is equal to the number of non-overlapping maximal independent sets of $G_{X_1}$, which in fact is the number of colors used in Algorithm 41. This completes the proof.

N. Proof of Theorem 43

Proof: From Theorem 30, by sending colorings of high probability subgraphs of sources’ D/2-characteristic graphs satisfying C.C.C., one can achieve the rate region described in (14). For simplicity, we assume the power of the graphs is one. Extensions to an arbitrary power are analogous. Suppose the receiver gets two colors from sources (say $c_1$ from source 1, and $c_2$ from source 2). To show that the receiver is able to compute its desired function up to distortion level $D$, we need to show that for every $(x^1_{i_1}, x^2_{j_1})$ and $(x^2_{i_1}, x^2_{j_1})$ such that $C_{G_{X_1}(D/2)}(x^1_{i_1}, x^2_{j_1}) = C_{G_{X_1}(D/2)}(x^1_{i_1}, x^2_{j_1})$ and $C_{G_{X_1}(D/2)}(x^1_{i_1}, x^2_{j_1}) = C_{G_{X_1}(D/2)}(x^2_{i_1}, x^2_{j_1})$, we have $d(f(x^1_{i_1}, x^2_{j_1}), f(x^1_{i_1}, x^2_{j_1})) \leq D$. Since the distortion function $d$ is a metric, we have,

\[
d(f(x^1_{i_1}, x^2_{j_1}), f(x^1_{i_1}, x^2_{j_1})) \leq d(f(x^1_{i_1}, x^1_{i_1}), f(x^1_{i_1}, x^1_{i_1})) + d(f(x^1_{i_1}, x^1_{i_1}), f(x^1_{i_1}, x^1_{i_1})) + d(f(x^2_{i_1}, x^2_{i_1}), f(x^2_{i_1}, x^2_{i_1})) \leq D/2 + D/2 = D.
\]

This completes the proof.
