When do only sources need to compute? On functional compression in tree networks

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ALLERTON.2009.5394799">http://dx.doi.org/10.1109/ALLERTON.2009.5394799</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Mar 16 18:47:01 EDT 2016</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/59303">http://hdl.handle.net/1721.1/59303</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td>The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.</td>
</tr>
</tbody>
</table>

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
When Do Only Sources Need to Compute? On Functional Compression in Tree Networks

Soheil Feizi, Muriel Médard
RLE at MIT
Emails: {sfeizi,medard}@mit.edu

Abstract—In this paper, we consider the problem of functional compression for an arbitrary tree network. Suppose we have \( k \) possibly correlated source processes in a tree network, and a receiver in its root wishes to compute a deterministic function of these processes. Other nodes of this tree (called intermediate nodes) are allowed to perform some computations to satisfy the node’s demand. Our objective is to find a lower bound on feasible rates for different links of this tree network (called a rate lower bound) and propose a coding scheme to achieve this rate lower bound in some cases.

The rate region of functional compression problem has been an open problem. However, it has been solved for some simple networks under some special conditions. For instance, [1] considered the rate region of a network with two transmitters and a receiver under a condition on source random variables. Here, we derive a rate lower bound for an arbitrary tree network based on the graph entropy. We introduce a new condition on colorings of source random variables’ characteristic graphs called the coloring connectivity condition (C.C.C.). We show that unlike the condition mentioned in [1], this condition is necessary and sufficient for any achievable coding scheme. We also show that unlike the entropy, the graph entropy does not satisfy the chain rule. For one stage trees with correlated sources, and general trees with independent sources, we propose a modularized coding scheme based on graph colorings to perform arbitrarily close to this rate lower bound. We show that in a general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform some computations. However, for a family of functions and RVs called coloring proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily close to the rate lower bound.

Index Terms—Functional compression, graph coloring, graph entropy.

I. INTRODUCTION

In this paper, we consider the problem of functional compression for an arbitrary tree network. While data compression considers the compression of sources at transmitters and their reconstruction at receivers, functional compression does not consider the recovery of whole sources, but the computation of a function of sources is desired at the receiver(s).

Assume we have a tree network with \( k \) possibly correlated source processes and a receiver in its root wishes to compute a deterministic function of these processes. Other nodes in this tree (called intermediate nodes) can compute some functions in demand. We want to find the set of feasible rates for different links of this tree network (called the rate region of this network) and propose a coding scheme to achieve these rates. We only consider the lossless computation of the function.

This problem has been an open problem in general. But, for some simple networks under some special conditions, it has been solved. For instance, [1] considered a rate region of a network with two transmitters and a receiver, under a condition on source random variables (RVs) called the zigzag condition. The zigzag condition forces source sequences to be mostly jointly typical. Our work extend their results in two senses: first, we expand the topology by considering arbitrary tree networks where intermediate nodes are allowed to perform computation, and second, we compute a rate lower bound without the need for the zigzag condition. We rely on a new definition of joint graph entropy of random variables. In some special cases, this general definition can be simplified to previous definitions proposed in [2], and [3]. We show that in general, the chain rule does not hold for the graph entropy. Our results show that for one stage trees with correlated sources, and general trees with independent sources, a modularized coding scheme based on graph colorings can perform arbitrarily close to this rate lower bound. We show that in the general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform some computations. However, for a family of functions and RVs called coloring proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily close to the rate lower bound.

II. FUNCTIONAL COMPRESSION BACKGROUND

In this section, after giving the problem statement, we explain our framework and previous results.

A. Problem Setup

Consider \( k \) discrete memoryless random processes, \( \{X_j\}_{j=1}^{\infty}, \{X_k\}_{k=1}^{\infty} \), as source processes. Memorylessness is not necessary, and one can approximate a source by a memoryless one with an arbitrarily precision [4]. Suppose these sources are drawn from finite sets \( X_1 = \{x_1^1, x_1^2, ..., x_1^{X_1}\} \), \( X_2 = \{x_2^1, x_2^2, ..., x_2^{X_2}\} \). These sources have a joint probability distribution \( p(x_1, ..., x_k) \). We express \( n \)-sequences of these RVs as \( X_j = \{X_j^{1}=i^{t+n-1}, ..., X_j^{k}=i^{t+n-1}\} \) with the joint probability distribution \( p(x_1, ..., x_k) \).

Without loss of generality, we assume \( n = 1 \), and to simplify notation, \( n \) will be implied by the context if no confusion arises. We refer to the \( t^{th} \) element of \( x_j \) as \( x_{jt} \). We use \( x_{t}^1, x_{t}^2, ..., x_{t}^{X_j} \) as different \( n \)-sequences of \( X_j \). We shall drop the superscript when no confusion arises. Since the sequence \( (x_1, ..., x_k) \) is drawn i.i.d. according to \( p(x_1, ..., x_k) \), one can write \( p(x_1, ..., x_k) = \prod_{i=1}^{k} p(x_{i1}, ..., x_{ik}) \).

Consider an arbitrary tree network shown in Figure 1. Suppose we have \( k \) source nodes in this network and a receiver in its root. We refer to other nodes of this tree as intermediate
nodes. Source node \( j \) has an input random process \( \{X_j^n\}_{n=1}^\infty \). The receiver wishes to compute a deterministic function \( f : X_1 \times \ldots \times X_k \to Z \), or \( f : X_1^n \times \ldots \times X_k^n \to Z^n \), its vector extension. It is worthwhile to notice that sources can be in 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 leaves of the network. So, in the achieved network, sources are located in leaves.

Source node \( j \) encodes its message at a rate \( R_{X_j} \). In other words, encoder \( e_{X_j} \), maps,

\[
en_{X_j} : X_j^n \to \{1, \ldots, 2^{nR_{X_j}}\}
\]

Suppose links connected to the receiver perform in rates \( R_{Z_{ij}} \), for \( 1 \leq j \leq w_1 \), where \( w_1 \) is the number of links connected to the receiver (we explain these notations more generally in Section IV). The receiver has a decoder \( r \) which maps,

\[
r : \prod_j \{1, \ldots, 2^{R_{Z_{ij}}}\} \to Z^n
\]

In other words, the receiver computes \( r(\bigcup_{j=1}^{w_1} f_{Z_{ij}}^j) = r'(e_{X_1}(x_1), \ldots, e_{X_k}(x_k)) \). We sometimes refer to this encoding/decoding scheme as an \( n \)-distributed functional code. Intermediate nodes are allowed to compute functions. However, they have no demand of their own. Computing the desired function \( f \) at the receiver is the only demand we permit in the network. For any encoding/decoding scheme, the probability of error is defined as \( P_{e}^n = Pr[(X_1, \ldots, X_k) : f(x_1, \ldots, x_k) \neq r'(e_{X_1}(x_1), \ldots, e_{X_k}(x_k))] \), where \( \bigcup_{j=1}^{w_1} f_{Z_{ij}}^j \) is the information which the decoder gets at the receiver. A rate sequence, \( R = (R_{X_1}, \ldots, R_{X_k}) \) is achievable if there exist \( k \) encoders in source nodes operating in these rates, and a decoder \( r \) at the receiver such 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, first we present some definitions prior to formulating our results. We also review some prior results. Consider \( X_1 \) and \( X_2 \) as two RVs with the joint probability distribution \( p(x_1, x_2) \). \( f(X_1, X_2) \) is a deterministic function such that \( f : X_1 \times X_2 \to Z \).

**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 follows: \( V_{X_1} = X_1 \) and an edge \((x_1^1, x_2^1) \in X_2^2 \) is in \( E_{X_1} \) iff there exists a \( x_2^1 \in X_2 \) such that \( p(x_1^1, x_2^1)p(x_2^1, x_2^2) > 0 \) and \( f(x_1^1, x_2^1) \neq f(x_1^2, x_2^2) \).

In other words, in order to avoid confusion about the function \( f(X_1, X_2) \) at the receiver, if \((x_1^1, x_2^1) \in E_{X_1} \), then descriptions of \( x_1^1 \) and \( x_2^1 \) must be different. Shannon first defined this when studying the zero error capacity of noisy channels [5]. Witsenhausen [6] used this concept to study a simplified version of our problem where one encodes \( X_1 \) to compute \( f(X_1) \) with 0 distortion. 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. 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 RVs \( X_2, \ldots, X_k \), \( p(x_1, \ldots, x_k) \), and \( f(X_1, \ldots, X_k) \) is defined as follows: \( v_{X_1} = X_1^\infty \) and an edge \((x_1^1, x_2^1) \in X_2^2 \) is in \( E_{X_1} \) if there exists \( x_2^1 \in X_2 \) for \( 2 \leq j \leq k \) such that \( p(x_1^1, x_2^1, \ldots, x_j^1)p(x_2^1, x_2^2, \ldots, x_k^1) > 0 \) and \( f(x_1^1, x_2^1, \ldots, x_k^1) \neq f(x_1^2, x_2^2, \ldots, x_k^2) \).

**Definition 3.** Given a graph \( G_{X_1} = (V_{X_1}, E_{X_1}) \) and a distribution on its vertices \( \Pi_{X_1} \), Körner [2] defines the graph entropy as follows:

\[
H_{G_{X_1}}(X_1) = \min_{X_1 \in \Pi_{W_1}} I(W_1; X_1),
\]

where \( \Gamma(G_{X_1}) \) is the set of all maximal independent sets of \( G_{X_1} \).

The notation \( X_1 \in \Pi_1 \in \Gamma(G_{X_1}) \) 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 \( G_{X_1} \).

Witsenhausen [6] showed that the graph entropy is the minimum rate at which a single source can be encoded such that a function of that source can be computed with zero distortion. Orlitsky and Roche [3] defined an extension of Körner’s graph entropy, the conditional graph entropy.

**Definition 4.** The conditional graph entropy is

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

Notation \( W_1 = X_1 - X_2 \) indicates a Markov chain. If \( X_1 \) \( X_2 \) are independent, \( H_{G_{X_1}}(X_1|X_2) = H_{G_{X_1}}(X_1) \). To illustrate this concept, let us express an example from [3].

**Definition 5.** 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, \( p(c_{G_{X_1}}(x_1^1)) = p(c_{G_{X_1}}(x_1^1|c_{G_{X_1}}(x_1^2))) \), where \( c_{G_{X_1}}(x_1^1) = \{x_1^j : c_{G_{X_1}}(x_1^j) = c_{G_{X_1}}(x_1^2)\} \) for all valid \( j \) is called a color class. We refer to the coloring which minimizes the entropy as the minimum entropy coloring. We also call the set of all valid colorings of a graph \( G_{X_1} \) as \( C_{G_{X_1}} \).

**Definition 6.** The \( n \)-th power of a graph \( G_{X_1} \) is a graph \( G_{X_1}^n = (V_{X_1}, E_{X_1}^n) \) 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 \((x_1^i, x_1^{i+1}) \in E_{X_1} \).
We denote a valid coloring of $G_{X_1}^n$, by $c_{G_{X_1}^n}(X_1)$.

**Definition 7.** Given a non-empty set $A \subset X_1 \times X_2$, 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 $f(X_1, X_2)$ as $\hat{G}_{z_1} = (\hat{V}_{z_1}, \hat{E}_{z_1})$ and the characteristic graph of $X_2$ with respect to $X_1$, $\hat{p}(x_1, x_2)$, and $f(X_1, X_2)$ as $\hat{G}_{z_2} = (\hat{V}_{z_2}, \hat{E}_{z_2})$. Note that $\hat{E}_{x_1} \subseteq E_{x_1}$ and $\hat{E}_{x_2} \subseteq E_{x_2}$. Finally, we say that $c_{G_{X_1}^n}(X_1)$ and $c_{G_{X_2}^n}(X_2)$ are $c$-colorings of $G_{X_1}$ and $G_{X_2}$ if they are valid colorings of $\hat{G}_{z_1}$ and $\hat{G}_{z_2}$ defined with respect to some set $A$ for which $p(A) \geq 1 - \epsilon$.

In [7], the Chromatic entropy of a graph $G_{X_1}$ is defined as,

**Definition 8.**

$$H_{G_{X_1}^n}(X_1) = \min_{c_{G_{X_1}^n}(X_1) \text{ is an } c\text{-coloring of } G_{X_1}} H(c_{G_{X_1}^n}(X_1)).$$

It means that the chromatic entropy is a representation of the chromatic number of high probability subgraphs of the characteristic graph. In [1], the conditional chromatic entropy is defined as

**Definition 9.**

$$H_{G_{X_1}^n}(X_1|X_2) = \min_{c_{G_{X_1}^n}(X_1) \text{ is an } c\text{-coloring of } G_{X_1}} H(c_{G_{X_1}^n}(X_1)|X_2).$$

Körner showed in [2] that, in the limit of large $n$, there is a relation between the chromatic entropy and the graph entropy.

**Theorem 10.**

$$\lim_{n \to \infty} \frac{1}{n} H_{G_{X_1}^n}(X_1) = H_{G_{X_1}^n}(X_1)$$

This theorem implies that the receiver can compute a deterministic function of a discrete memoryless source with a vanishing probability of error by first coloring a sufficiently large power of the characteristic graph of the source RV 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 $H_{G_{X_1}^n}(X_1)$, 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 $G_{X_1}$, 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 an entropy-rate compression module.

The conditional version of the above theorem is proved in [8].

**Theorem 11.**

$$\lim_{n \to \infty} \frac{1}{n} H_{G_{X_1}^n}(X_1|X_2) = H_{G_{X_1}^n}(X_1|X_2).$$

All the mentioned results considered only functional compression with side information at the receiver. Consider the network shown in Figure 1 when $d_{max} = 1$ and $k = 2$. It shows a network with two source nodes and a receiver which wishes to compute a function of the sources’ values. In general, the rate-region of this network has not been determined. However, in [1] determined the rate-region of this network when the source RVs satisfy a condition called the zigzag condition.

We first explain their results. Then, in Section III, we compute the rate-region of this network in a general case without having any restrictive conditions on the source RVs (such as the zigzag condition). Then, We extend our results to the case of having $k$ source nodes.

We refer to the joint-typical set of sequences of RVs $X_1, ..., X_n$ as $T^n_c$. $k$ is implied in this notation for simplicity. We explicitly mention $k$ if some confusion arises. $T^n_c$ can be considered as a strong or weak typical set ([4]).

**Definition 12.** A discrete memoryless source $\{(X_1, X_2)\}_{i \in \mathbb{N}}$ with a distribution $p(x_1, x_2)$ satisfies the Zigzag Condition if for any $c$ and some $n_i, (x_{1i}^c, x_{2i}^c)$, $(x_{1i}, x_{2i})$ is in $T^n_c$, there exist some $(x_{1i}^\ell, x_{2i}^\ell) \in T^n_\ell$ such that $(x_{1i}^\ell, x_{2i}^\ell, y_{1i}, y_{2i}) \in T^n_{\ell+c}$ for each $i \in \{1, 2\}$, and $(x_{1j}, x_{2j}) = (x_{1j}^2, x_{2j}^2)$ for some $i \in \{1, 2\}$ for each $j$. In fact, the zigzag condition forces many source sequences to be typical. If the source RVs satisfy this condition, the achievable rate region for this network is the set closure of the set of all rates that can be achieved through graph colorings. In other words, under the zigzag condition, any colorings of high probability subgraphs of sources’ characteristic graphs will allow the computation of the function with a vanishing probability of error. Among these colorings, some which allow us to reach the lower bound of the rate region are called minimum entropy colorings. In [1], it is not claimed that this condition is necessary, but sufficient. In other words, they computed the rate-region only in the case that source RVs satisfy the zigzag condition. The zigzag condition is a restrictive condition which does not depend on the desired function at the receiver.

III. A RATE REGION FOR ONE-STAGE TREE NETWORKS

In this section, we want to find a rate region for a general one-stage tree network without having any restrictive conditions such as the zigzag condition. Consider the network shown in Figure 1 when $d_{max} = 1$, with $k$ sources.

**Definition 13.** A path with length $m$ between two points $Z_i = (x_{1i}^m, x_{2i}^m, ..., x_{ki}^m)$, and $Z_m = (x_{1i}^m, x_{2i}^m, ..., x_{ki}^m)$ is determined by $m - 1$ points $Z_i$, $1 \leq i \leq m$ such that,

i) $P(Z_i) > 0$, for all $1 \leq i \leq m$.

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

Definition 13 can be expressed for two $n$-length vectors as follows.

**Definition 14.** A path with length $m$ between two points $Z_1 = (x_{11}^m, x_{21}^m, ..., x_{k1}^m) \in T^n_1$, and $Z_m = (x_{1m}^m, x_{2m}^m, ..., x_{km}^m) \in T^n_m$ are determined by $m - 1$ points $Z_i$, $1 \leq i \leq m$ such that,

i) $Z_i \in T^n_i$, for all $1 \leq i \leq m$.

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

**Definition 15.** A joint-coloring family $J_C$ for random variables $X_1, ..., X_k$ with characteristic graphs $G_{X_1}, ..., G_{X_k}$, and any valid colorings $c_{G_{X_1}}, ..., c_{G_{X_k}}$, respectively is defined as $J_C = \{j^i_{\ell} \colon (x_{1i}^\ell, x_{2i}^\ell, ..., x_{ki}^\ell) : c_{G_{X_1}}(x_{1i}^\ell) = c_{G_{X_2}}(x_{2i}^\ell) = ... = c_{G_{X_k}}(x_{ki}^\ell) \}$ for any valid $i_1, ..., i_k$, and $n_j = |c_{G_{X_1}}| \times |c_{G_{X_2}}| \times ... \times |c_{G_{X_k}}|$. We call each $j^i_{\ell}$ a joint coloring class.
Definition 15 can be expressed for RVs $X_1,...,X_k$ with characteristic graphs $G_{X_1}^n,...,G_{X_k}^n$, and any valid $e$-colorings $c_{G_{X_1}^n},...,c_{G_{X_k}^n}$, respectively.

Definition 16. Consider RVs $X_1,...,X_k$ with characteristic graphs $G_{X_1},...,G_{X_k}$, and any valid colorings $c_{G_{X_1}},...,c_{G_{X_k}}$. We say these colorings satisfy the Coloring Connectivity Condition (C.C.C.) when, between any two points in $j^e_i \in J_c$, there exists a path that lies in $j^e_i$, or function $f$ has the same value in disconnected parts of $j^e_i$.

C.C.C. can be expressed for RVs $X_1,...,X_k$ with characteristic graphs $G_{X_1}^n,...,G_{X_k}^n$, and any valid $e$-colorings $c_{G_{X_1}^n},...,c_{G_{X_k}^n}$, respectively.

Example 17. For example, suppose we have two random variables $X_1$ and $X_2$ with characteristic graphs $G_{X_1}$ and $G_{X_2}$. Let us assume $c_{G_{X_1}}$ and $c_{G_{X_2}}$ are two valid colorings of $G_{X_1}$ and $G_{X_2}$, respectively. Assume $c_{G_{X_1}}(x_1^1) = c_{G_{X_1}}(x_1^2)$ and $c_{G_{X_2}}(x_2^1) = c_{G_{X_2}}(x_2^2)$. Suppose $j^e_i$ represents this joint coloring class. In other words, $j^e_i = \{(x_1^i, x_2^j)\}$, for all $1 \leq i, j \leq 2$ when $p(x_1^i, x_2^j) > 0$. Figure 2 considers two different cases. The first case is when $p(x_1^1, x_2^2) = 0$, and other points have a non-zero probability. It is illustrated in Figure 2-a. One can see that there exists a path between any two points in this joint coloring class. So, this joint coloring class satisfies C.C.C. If other joint coloring classes of $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C., we say $c_{G_{X_1}}$ and $c_{G_{X_2}}$ satisfy C.C.C. Now, consider the second case depicted in Figure 2-b. In this case, we have $p(x_1^1, x_2^2) = 0$, $p(x_1^1, x_2^2) = 0$, and other points have a non-zero probability. One can see that there is no path between $(x_1^1, x_2^1)$ and $(x_1^2, x_2^2)$ in $j^e_i$. So, though these two points belong to a same joint coloring class, their corresponding function values can be different from each other. Thus, $j^e_i$ does not satisfy C.C.C. Therefore, $c_{G_{X_1}}$ and $c_{G_{X_2}}$ do not satisfy C.C.C.

Lemma 18. Consider two RVs $X_1$ and $X_2$ with characteristic graphs $G_{X_1}$ and $G_{X_2}$ and any valid colorings $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2)$ respectively, where $c_{G_{X_1}}(X_2)$ is a trivial coloring, assigning different colors to different vertices (to simplify the notation, we use $c_{G_{X_1}}(X_2) = X_2$ to refer to this coloring). These colorings satisfy C.C.C. Also, $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2) = X_2$ satisfy C.C.C for any $n$.

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 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 $J_{c} = \{j^e_i : x_{j^e_i} \in \sigma_i\}$ such that $j^e_i = \{(x_1^i, x_2^j) : c_{G_{X_1}}(x_1^i) = \sigma_i\}$, where $\sigma_i$ is a generic color. Any two points in $j^e_i$ are connected to each other with a path with length one. So, $j^e_i$ satisfies C.C.C. This argument holds for any $j^e_i$ for any valid $i$. Thus, $J_c$ 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.

Lemma 19. Consider RVs $X_1, ..., X_k$ with characteristic graphs $G_{X_1}, ..., G_{X_k}$, and any valid colorings $c_{G_{X_1}}, ..., c_{G_{X_k}}$ with joint coloring class $J_c = \{j^e_i : \}$. For any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in $j^e_i$, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$ if and only if $j^e_i$ satisfies C.C.C.

Proof: We first show that if $j^e_i$ satisfies C.C.C., then, for any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in $j^e_i$, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$. Since $j^e_i$ satisfies C.C.C., there exists a path with length $m-1$ between these two points $Z_1 = (x_1^i, x_k^i)$ and $Z_m = (x_1^1, x_k^2)$, for some $m$. Two consecutive points $Z_j$ and $Z_{j+1}$ in this path, just differ in one of their coordinates. Without loss of generality, suppose they differ in their first coordinate. In other words, $Z_j = (x_1^j, x_2^j, ..., x_k^j)$ and $Z_{j+1} = (x_1^{j+1}, x_2^j, ..., x_k^j)$. Since these two points belong to $j^e_i$, $c_{G_{X_1}}(x_1^j) = c_{G_{X_1}}(x_1^{j+1})$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $x_1^j$ and $x_1^{j+1}$ in $G_{X_1}$ and they could not have the same color. So, $f(Z_j) = f(Z_{j+1})$. By applying the same argument for all two consecutive points in the path between $Z_1$ and $Z_m$, one can get $f(Z_1) = f(Z_2) = ... = f(Z_m)$.

If $j^e_i$ does not satisfy C.C.C., it means that there exists at least two points $Z_1$ and $Z_2$ in $j^e_i$ such that no path exists between them. So, the value of $f$ can be different in these points. As an example, consider Figure 2-b. The value of the function can be different in two disconnected points in a same joint coloring class.

Lemma 20. Consider RVs $X_1, ..., X_k$ with characteristic graphs $G_{X_1}^n, ..., G_{X_k}^n$, and any valid $e$-colorings $c_{G_{X_1}^n}, ..., c_{G_{X_k}^n}$ with the joint coloring class $J_c = \{j^e_i : \}$. For any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in $j^e_i$, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$ if and only if $j^e_i$ satisfies C.C.C.

Proof: The proof is similar to Lemma 19. The only difference is to use the definition of C.C.C. for $c_{G_{X_1}^n}, ..., c_{G_{X_k}^n}$.

Next, we want to show that if $X_1$ and $X_2$ satisfy the zigzag condition mentioned in Definition 12, any valid colorings of their characteristic graphs satisfy C.C.C., but not vice versa. In other words, we want to show that the zigzag condition used in [1] is not necessary, but sufficient.

Lemma 21. If two RVs $X_1$ 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: 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 $(x_1^1, x_2^1)$ and $(x_1^2, x_2^2)$ in

---

**Fig. 2.** Two examples of a joint coloring class: a) satisfying C.C.C. b) not satisfying C.C.C. Dark squares indicate points with zero probability. Function values are depicted in the picture.
a joint coloring class $j_i$. The definition of the zigzag condition guarantees the existence of a path with length two between these two points. 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. In other words, the zigzag condition is not a necessary condition, but sufficient. To have an example, one can see that in a special case considered in Lemma 18, C.C.C. always holds without having any condition.

**Definition 22.** For RVs $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$, the joint graph entropy is defined as follows:

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

(7)

in which $c_{G_{X_1}}(X_1),\ldots,c_{G_{X_k}}(X_k)$ are $c$-colorings of $G_{X_1},\ldots,G_{X_k}$ satisfying C.C.C. We sometimes refer to the joint graph entropy by using $H_{\cup_{i=1}^k G_{X_i}}(X_1,\ldots,X_k)$.

Similarly, we can define the conditional graph entropy.

**Definition 23.** For RVs $X_1, \ldots, X_k$ with characteristic graphs $G_{X_1}, \ldots, G_{X_k}$, the conditional graph entropy can be defined as follows: $H_{G_{X_1},\ldots,G_{X_k}}(X_1,\ldots,X_k|X_1,\ldots,X_{k-1}) = \min_{c_{G_{X_1}},\ldots,c_{G_{X_k}}} \frac{1}{n} H(c_{G_{X_1}}(X_1),\ldots,c_{G_{X_k}}(X_k)|c_{G_{X_1}}(X_1),\ldots,c_{G_{X_{k-1}}}(X_{k-1}))$ in which minimization is over $c_{G_{X_1}}(X_1),\ldots,c_{G_{X_k}}(X_k)$ which are $c$-colorings of $G_{X_1},\ldots,G_{X_k}$ satisfying C.C.C.

**Lemma 24.** For $k=2$, definitions 4 and 23 are the same.

**Proof:** By using the data processing inequality, we have

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

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

Then, Lemma 18 implies that $c_{G_{X_1}}(X_1)$ and $c_{G_{X_2}}(X_2)$ satisfy C.C.C. So, a direct application of Theorem 11 completes the proof.

one can see 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 (this is the set of all subsets of $\{1,\ldots,k\}$ without the empty set). 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, $S(2) = \{\{\},\{1\},\{2\}\}$, and for $S = \{1,2\}$, we write $H_{\cup_{i \in S} G_{X_i}}(X_S)$ instead of $H_{G_{X_1},G_{X_2}}(X_1,X_2)$.

**Theorem 25.** A rate region of the network shown in Figure 1 when $d_{\max} = 1$ is determined by these conditions:

$$\forall S \in S(k) \implies \sum_{i \in S} R_{X_i} \geq H_{\cup_{i \in S} G_{X_i}}(X_S|X_{S^c})$$

(8)

**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 26.** Consider RVs $X_1, \ldots, X_k$ with characteristic graphs $G^a_{X_1}, \ldots, G^a_{X_k}$, and any valid $c$-colorings $c_{G_{X_1}}, \ldots, c_{G_{X_k}}$ satisfying C.C.C., for sufficiently large $n$. There exists

$$\hat{f}: c_{G_{X_1}}(X_1) \times \ldots \times c_{G_{X_k}}(X_k) \rightarrow Z^n$$

such that $\hat{f}(c_{G_{X_1}}(x_1),\ldots,c_{G_{X_k}}(x_k)) = f(x_1,\ldots,x_k)$, for all $(x_1,\ldots,x_k) \in T^n$.

**Proof:** Suppose the joint coloring family for these colorings is $J_c = \{j_i: i\}$. We proceed by constructing $f$. Assume $(x_1,\ldots,x_k) \in j_i^*$ and $c_{G_{X_i}}(x_i) = \sigma_i, \ldots, c_{G_{X_k}}(x_k) = \sigma_k$.

Define $f(\sigma_1,\ldots,\sigma_k) = f(x_1,\ldots,x_k)$.

To show this function is well-defined on elements in its support, we should show that for any two points $(x_1^1,\ldots,x_k^1)$ and $(x_1^2,\ldots,x_k^2)$ in $T^n$, if $c_{G_{X_i}}(x_i^1) = c_{G_{X_i}}(x_i^2)$, $\ldots$, $c_{G_{X_k}}(x_k^1) = c_{G_{X_k}}(x_k^2)$, then $f(x_1^1,\ldots,x_k^1) = f(x_1^2,\ldots,x_k^2)$. Therefore, our function $\hat{f}$ is well-defined and has the desired property.

**Lemma 26 implies that given $c$-colorings of characteristic graphs of RVs 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 the 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 is at what rates encoders can transmit these colors to the receiver faithfully (with a probability of error less than $\epsilon$).

**Lemma 27.** (Slepian-Wolf Theorem)

A rate region of the network shown in Figure 1 with $d_{\max} = 1$ where $f(X_1,\ldots,X_k) = (X_1,\ldots,X_k)$ can be determined by these conditions:

$$\forall S \in S(k) \implies \sum_{i \in S} R_{X_i} \geq H(X_S|X_{S^c})$$

(10)

**Proof:** See [9].

We now use the Slepian-Wolf (SW) encoding/decoding scheme on achieved coloring RVs. Suppose the probability of error in each decoder of SW is less than $\frac{\epsilon}{n}$. 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 $G_{X_1}, \ldots, G_{X_k}$, 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 26 and 27, it is less than $\epsilon$, and goes to zero as $n$ goes to infinity. Applying Lemma 27 on achieved coloring RVs, we have

$$\forall S \in S(k) \implies \sum_{i \in S} R_{X_i} \geq \frac{1}{n} H(c_{G_{X_S}}|c_{G_{X_{S^c}}})$$

(11)

where $c_{G_{X_S}}$ and $c_{G_{X_{S^c}}}$ are $c$-colorings of characteristic graphs satisfying C.C.C. Thus, using Definition 23 completes the achievability part.

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

$$\mathcal{F}_n^\epsilon = \{ f : \Pr[f(X_1, ..., X_k) \neq f(X_1, ..., X_k)] < \epsilon \}. \quad (12)$$

In other words, $\mathcal{F}_n^\epsilon$ is the set of all functions equal to $f$ with $\epsilon$ probability of error. For large enough $n$, all achievable functional source codes are in $\mathcal{F}_n^\epsilon$. We call these codes $\epsilon$-achievable functional codes.

**Lemma 28.** Consider some function $f : X_1 \times ... \times X_k \rightarrow \mathbb{Z}$. Any distributed functional code which reconstructs this function with zero error probability induces colorings on $G_{X_1}, ..., G_{X_k}$ with respect to this function, where these colorings satisfy C.C.C.

**Proof:** To show this lemma, let us assume we have a zero-error distributed functional code represented by encoders $en_{X_1}, ..., en_{X_k}$ and a decoder $r$. Since it is error free, for any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$, if $p(x_1^1, ..., x_k^1) > 0$, $p(x_1^2, ..., x_k^2) > 0$, $en_{X_1}(x_1^1) = en_{X_1}(x_1^2)$, ..., $en_{X_k}(x_k^1) = en_{X_k}(x_k^2)$, then,

$$f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2) = r'(en_{X_1}(x_1^1), ..., en_{X_k}(x_k^1)). \quad (13)$$

We want to show that $en_{X_1}, ..., en_{X_k}$ are some valid colorings of $G_{X_1}, ..., G_{X_k}$ satisfying C.C.C. We demonstrate this argument for $X_1$. The argument for other RVs is analogous. First, we show that $en_{X_1}$ 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 $en_{X_1}$ did not induce a coloring on $G_{X_1}$, there must be some edge in $G_{X_1}$ with both 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 path $(x_1^2, ..., x_1^1)$ such that, $p(x_1^2, ..., x_1^1)p(x_1^1, x_1^2, ..., x_1^1) > 0$, $en_{X_1}(x_1^1) = en_{X_1}(x_1^2)$, and $f(x_1^1, x_1^2, ..., x_1^1) \neq f(x_1^2, x_1^1, ..., x_1^1)$. By taking $x_2 = x_2^2$, ..., $x_k = x_k^2$ in (13), one can see that it is not possible. So, the contradiction assumption is wrong and $en_{X_1}$ induces a valid coloring on $G_{X_1}$.

Now, we should show that these induced colorings satisfy C.C.C. If it was not true, it means that there must exist two point $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in a joint coloring class $j^e_i$ such that there is no path between them in $j^e_i$. So, Lemma 19 says that the function $f$ can get different values in these two points. In other words, it is possible to have $f(x_1^1, x_1^2, ..., x_k^2) \neq f(x_1^2, x_1^1, ..., x_k^2)$, where $c_{G_{X_1}}(x_1^1) = c_{G_{X_1}}(x_1^2)$, ..., $c_{G_{X_k}}(x_k^1) = c_{G_{X_k}}(x_k^2)$, which is in contradiction with (13). Thus, achieved colorings satisfy C.C.C.

In the last step, we should show that any achievable functional code represented by $\mathcal{F}_n^\epsilon$ induces $e$-colorings on characteristic graphs satisfying C.C.C.

**Lemma 29.** Consider RVs $X_1, ..., X_k$. All $e$-achievable functional codes of these RVs induce $e$-colorings on characteristic graphs satisfying C.C.C.

**Proof:** Suppose $g(x_1, ..., x_k) = r'(en_{X_1}(x_1), ..., en_{X_k}(x_k)) \in \mathcal{F}_n^\epsilon$ be such a code. Lemma 28 says that a zero-error reconstruction of $g$ induces some colorings on characteristic graphs satisfying C.C.C., with respect to $g$. Suppose the set of all points $(x_1, ..., x_k)$ such that $g(x_1, ..., x_k) \neq f(x_1, ..., x_k)$ be denoted by $\mathcal{C}$. Since $g \in \mathcal{F}_n^\epsilon$, $\Pr[\mathcal{C}] < \epsilon$. Therefore, functions $en_{X_1}, ..., en_{X_k}$ restricted to $\mathcal{C}$ are $e$-colorings of characteristic graphs satisfying C.C.C. (by definition).

Lemmas 28 and 29 establish the converse part and complete the proof.

**Corollary 30.** A rate region of the network shown in Figure 1 when $d_{max} = 1$ and $k = 2$ is determined by these three conditions:

$$R_{x_1} \geq H_{G_{X_1}}(X_1|X_2)$$
$$R_{x_2} \geq H_{G_{X_2}}(X_2|X_1)$$
$$R_{x_1} + R_{x_2} \geq H_{G_{X_1}, G_{X_2}}(X_1, X_2) \quad (14)$$

**IV. A RATE LOWER BOUND FOR A GENERAL TREE NETWORK**

In this section, we seek to compute a rate lower bound of an arbitrary tree network with $k$ sources in its leaves and a receiver in its root (look at Figure 1). We refer to other nodes of this tree as intermediate nodes. The receiver wishes to compute a deterministic function of source RVs. Intermediate nodes have no demand of their own in terms of the functional compression, but they are allowed to perform some computations. Computing the desired function $f$ at the receiver is the only demand we permit in the network. Also, we show some cases in which we can achieve this lower bound.

First, we propose a framework to categorize any tree networks and their nodes.

**Definition 31.** For an arbitrary tree network,

- The distance of each node is the number of hops in the path between that node and the receiver.
- $d_{max}$ is the distance of the farthest node from the receiver.
- A standard tree is a tree such that all source nodes are in a distance $d_{max}$ from the receiver.
- An auxiliary node is a new node connected to a leaf of a tree and increases its distance by one. The added link is called an auxiliary link. The leaf in the original tree to which is added an auxiliary node is called the actual node corresponding to that auxiliary node. The link in the original tree connected to the actual node is called the actual link corresponding to that auxiliary link.
- For any given tree, one can make it to be a standard tree by adding some consecutive auxiliary nodes to its leaves with distance less than $d_{max}$. We call the achieved tree, the modified tree and refer to this process as the tree standardization.

These concepts are depicted in Figure 3. Auxiliary nodes in the modified tree network act like intermediate nodes. It
means one can imagine that they can compute some functions in demand. But, all functions computed in auxiliary nodes can be gathered in their corresponding actual node in the original tree. So, the rate of the actual link in the original tree network is the minimum of rates of corresponding auxiliary links in the modified network. Thus, if we compute the rate-region for the modiﬁed tree of any given arbitrary tree, we can compute the rate-region of the original tree. Therefore, in the rest of this section, we consider the rate-region of modiﬁed tree networks.

Definition 32. Any modiﬁed tree network with k source nodes with distance \(d_{\text{max}}\) from the receiver can be expressed by a connection set \(S_T = \{s_i^j : 1 \leq i \leq d_{\text{max}}\}\) where \(s_i^j = (\Xi_{ij} : 1 \leq j \leq w_i)\). \(w_i\) is the number of nodes with distance \(i\) from the receiver (called nodes in the \(i\)-th stage) and a subset of source RVs is in \(\Xi_{ij}\) when paths of those source nodes have the last \(i\) common hops.

For example, consider the network shown in Figure 3. Its connection set is \(S_T = \{s_1^1, s_2^1\}\) such that \(s_1^1 = (\{X_1, X_2\}, \{X_3\})\) and \(s_2^1 = (\{X_1, X_2, X_3\}\). In other words, \(\Xi_{11} = (X_1, X_2), \Xi_{12} = X_3, \Xi_{21} = X_1, \Xi_{22} = X_2\) and \(\Xi_{23} = X_3\). One can see that \(S_T\) completely describes the structure of the tree. In other words, there is a bijective map between any modiﬁed tree and its connection set \(S_T\). By using \(S_T\), we wish to assign some labels to nodes, links and their rates. We label each node with its corresponding \(\Xi_{ij}\) as \(n_{\Xi_{ij}}\).

We call the outgoing link from this node \(e_{ij}^s\). The rate of this link is referred by \(R_{\Xi_{ij}}^s\). For instance, nodes in the second stage of the network shown in Figure 3 are called \(n_{\Xi_{12}}^2, n_{\Xi_{22}}^2\) and \(n_{\Xi_{23}}^2\) with outgoing links \(e_{12}^2, e_{22}^2\) and \(e_{23}^2\), respectively. Nodes in the ﬁrst stage of this network are referred by \(n_{\Xi_{11}}^1, n_{\Xi_{21}}^1, n_{\Xi_{22}}^1, n_{\Xi_{23}}^1\) with outgoing links \(e_{11}^1, e_{12}^1, e_{13}^1, e_{21}^1, e_{22}^1, e_{23}^1\).

We have three types of nodes: source nodes, intermediate nodes and a receiver. Source nodes encode their messages by using some encoders and send encoded messages. Intermediate nodes can compute some functions of their received information. The receiver decodes the received information and wishes to be able to compute its desired function. The RV which is sent in the link \(e_{ij}^s\) is called \(R_{\Xi_{ij}}^s\). Also, we refer to the function computed in an intermediate node \(n_{\Xi_{ij}}^k\) as \(g_{\Xi_{ij}}^k\). For example, consider again the network shown in Figure 3. RVs sent through links \(e_{11}^2, e_{12}^2, e_{13}^2, e_{21}^2, e_{22}^2\) and \(e_{23}^2\) are \(f_{11}^2, f_{12}^2, f_{13}^2, f_{21}^2, f_{22}^2, f_{23}^2\) and \(f_{11}^2\), such that \(f_{11}^2, f_{12}^2, f_{13}^2, f_{21}^2, f_{22}^2, f_{23}^2 = g_{11}^k(f_{11}^2, f_{21}^2)\) and \(f_{11}^2 = g_{11}(f_{12}^2)\).

A. A Rate Lower Bound

Consider nodes in stage \(i\) of a tree network representing by \(\Xi_{ij}\) for \(j = \{1, 2, ..., w_i\}\) where \(w_i\) is the number of nodes in stage \(i\). \(S(w_i)\) is the power set of the set \(\{1, 2, ..., w_i\}\) and \(s_i \in S(w_i)\) is a non-empty subset of \(\{1, 2, ..., w_i\}\).

Theorem 33. A rate lower bound of a tree network with the connection set \(S_T = \{s_i^j : i\}\) can be determined by these conditions:

\[
\forall s_i \in S(w_i) \implies \sum_{j \in s_i} R_{\Xi_{ij}}^j \geq H_{\cup_{x \in \Xi_{ij}}^j} G_{\Xi_{ij}}(\Xi_{is_i}, \Xi_{is_i}) \tag{15}
\]

for all \(i = 1, ..., |ST|\) where \(\Xi_{is_i} = \bigcup_{j \in s_i} \Xi_{ij}\) and \(\Xi_{is_i}^c = \{X_1, ..., X_k\} - \Xi_{is_i}\).

Proof: In this part, we want to show that no coding scheme can outperform this rate region. Consider nodes in the \(i\)-th stage of this network, \(n_{\Xi_{ij}}^k\) for \(1 \leq j \leq w_i\). Suppose they are directly connected to the receiver. So, the information sent in links of these stages should be enough to compute the desired function. In the best case, suppose their parents sent all their information without doing any compression. So, by direct application of Theorem 25, one can see that,

\[
\forall s_i \in S(w_i) \implies \sum_{j \in s_i} R_{\Xi_{ij}}^j \geq H_{\cup_{x \in \Xi_{ij}}^j} G_{\Xi_{ij}}(\Xi_{is_i}, \Xi_{is_i}) \tag{16}
\]

This argument can be repeated for all stages. Thus, no coding scheme can outperform these bounds.

In the following, we express some cases under which we can achieve the derived rate lower bound of Theorem 33.

B. Tightness of the Rate Lower Bound for Independent Sources

In this part, we propose a functional coding scheme to achieve the rate lower bound. Suppose RVs \(X_1, ..., X_k\) with characteristic graphs \(G_{X_1}, ..., G_{X_k}\) are independent. Assume \(\epsilon G_{X_1}, ..., \epsilon G_{X_k}\) are valid \(\epsilon\)-colorings of these characteristic graphs satisfying C.C.C. The proposed coding scheme can be described as follows: 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 RVs. Intermediate nodes ﬁrst compute their parents’ coloring RVs, and then by using a look-up table, they 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 RV 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 a source coding. The receiver ﬁrst performs an entropy decoding on its received information and achieves coloring RVs. Then, it uses a look-up table to compute its desired function by using achieved colorings.

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. Consider the node \(n_{\Xi_{ij}}^k\) in the \(i\)-th stage of the network. Since the corresponding source values \(\Xi_{ij}\) of its received colorings form an independent set in its characteristic graph \(G_{\Xi_{ij}}\) and this node computes the minimum entropy of this graph, it is equivalent to the case that it would receive the exact source information because both of them lead to the same coloring RV. So, if all nodes of stage \(i\) were directly connected to the receiver, the receiver could compute its desired function and link rates would satisfy the following conditions:

\[
\forall s_i \in S(w_i) \implies \sum_{j \in s_i} R_{\Xi_{ij}}^j \geq H_{\cup_{x \in \Xi_{ij}}^j} G_{\Xi_{ij}}(\Xi_{is_i}, \Xi_{is_i}) \tag{17}
\]

Thus, by using a simple induction argument, one can see that the proposed scheme is achievable and it can perform arbitrarily close to the derived rate lower bound, while sources are independent.
Consider Figure 4 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 some subgraphs which have the same topology as \( G_{X_1} \), corresponding to each \( x_2 \) in \( X_2 \). These subgraphs are fully connected to each other under conditions of Corollary 36. So, any coloring of this graph can be represented as two colorings of \( G_{X_1} \) and \( G_{X_2} \) which is a complete graph. Thus, the minimum entropy coloring of \( G_{X_1,X_2} \) is equal to the minimum entropy coloring of \( (G_{X_1},G_{X_2}) \). Therefore,
\[
H_{G_{X_1},G_{X_2}}(X_1,X_2) = H_{G_{X_1},X_2}(X_1,X_2).
\]

\[\text{VI. CONCLUSION}\]

In this paper, we considered the problem of functional compression for an arbitrary tree network. In this problem, we have \( k \) possibly correlated source processes in a tree, and a receiver in its root wishes to compute a deterministic function of these processes. Intermediate nodes can perform some computations, but the computing of the desired function at the receiver is the only demand we permit in this tree network. The rate region of this problem has been an open problem in general. But, it has been solved for some simple networks under some special conditions (e.g., [1]). Here, we have computed a rate lower bound of an arbitrary tree network in an asymptotically lossless sense. We defined joint graph entropy of some random variables and showed that the chain rule does not hold for the graph entropy. For one stage trees with correlated sources, and general trees with independent sources, we proposed a modularized coding scheme based on graph colorings to perform arbitrarily close to the rate lower bound. We showed that in a general tree network case with independent sources, to achieve the rate lower bound, intermediate nodes should perform some computations. However, for a family of functions and RVs called coloring proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily close to the rate lower bound.

\[\text{REFERENCES}\]


