Network Functional Compression

by

Soheil Feizi

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author ....
Department of Electrical Engineering and Computer Science
May 21, 2010

Certified by ....
Muriel Médard
Professor
Thesis Supervisor

Accepted by ....
Professor Terry P. Orlando
Chairman, Department Committee on Graduate Theses

Network Functional Compression

by

Soheil Feizi

Submitted to the Department of Electrical Engineering and Computer Science
on May 21, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

In this thesis, we consider different aspects of the functional compression problem. In functional compression, the computation of 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. In Chapter 2 of this Thesis, we consider this problem for an arbitrary tree network and asymptotically lossless computations. In particular, for one-stage tree networks, we compute a rate-region and for an arbitrary tree network, we derive a rate lower bound 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 Doshi *et al.*, this condition is necessary and sufficient for any achievable coding scheme based on colorings. We also show that, unlike entropy, 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 closely to the derived 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 random variables called chain rule proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily closely to the rate lower bound.

In Chapter 3 of this Thesis, we consider a multi-functional version of this problem with side information, where the receiver wants to compute several functions with different side information random variables and zero distortion. Our results are applicable to the case with several receivers computing different desired functions. We define a new concept named multi-functional graph entropy which is an extension of graph entropy defined by Körner. We show that the minimum achievable rate for this problem is equal to conditional multi-functional graph entropy of the source random variable given the side information. We also propose a coding scheme based on graph colorings to achieve this rate.

In these proposed coding schemes, one needs to compute the 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. However, in Chapter 4, 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, by having a non-zero joint probability condition on random variables' distributions, for any desired function, finding the minimum entropy coloring can be solved in polynomial time. In another case, we show that if the desired function is a quantization function, this problem is also tractable. We also consider this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we show that finding the minimum entropy coloring is heuristically equivalent to finding the maximum independent set of a graph. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to approximately solve the maximum independent set problem.

Next, in Chapter 5, we consider the effect of having feedback on the rate-region of the functional compression problem . If the function at the receiver is the identity function, this problem reduces to the Slepian-Wolf compression with feedback. For this case, having feedback does not make any benefits in terms of the rate. However, it is not the case when we have a general function at the receiver. By having feedback, one may outperform rate bounds of the case without feedback.

We finally consider the problem of distributed functional compression with distortion. The objective is to compress correlated discrete sources such that an arbitrary deterministic function of those sources can be computed up to a distortion level at the receiver. In this case, we compute a rate-distortion region and then, propose a simple coding scheme with a non-trivial performance guarantee.

Thesis Supervisor: Muriel Médard
Title: Professor

Acknowledgments

I want to thank my advisor Prof. Muriel Médard. I value our discussions on different research challenges as well as her insightful guidance. She has been a very supportive mentor, both professionally and personally. I also want to thank Prof. Michelle Effros and Prof. Devavrat Shah for their valuable comments on this Thesis. I am also grateful to Prof. Alan V. Oppenheim, my academic mentor, for his valuable advice during my master study.

I could not have come this far without the love of my family. I am always grateful to my parents and my brothers for their endless love and support.

Some parts of this Thesis have already been presented. Results discussed in Chapter 2 have been presented at 2009 Annual Allerton Conference on Communication, Control, and Computing [14]. Preliminary results explained in Chapter 3 have been presented at 2009 Globecom Conference on Communications [13]. Some parts of results discussed in Chapter 4 have been presented at 2010 IEEE International Symposium on Information Theory (ISIT, 2010) [15].

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, we consider different aspects of the functional compression problem over networks. 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). Traditional data compression schemes are special cases of functional compression, where their desired function is the identity function. In other words, in traditional data compression, sources are compressed so that the receiver can receive the whole sources' information. However, if the receiver is interested in computing a function (or some functions) of sources (and not the whole sources), they may be compressed further. In the rest of this chapter, first, we mention some major applications and motivations of this problem. Next, we review some prior relevant research and illustrate some research challenges of this problem through some motivating examples which will be discussed in the following chapters.

1.1 Applications and Motivations

In this section, we present two major applications for functional compression: data gathering and privacy issues. We explain these applications by two examples.

Example 1. *Consider Figure 1-1 as an example of a distributed sensor network. Some nodes have sensors to measure the temperature of their locations. These nodes are called source nodes. The network of this example has four sources, named* X_1, ...,

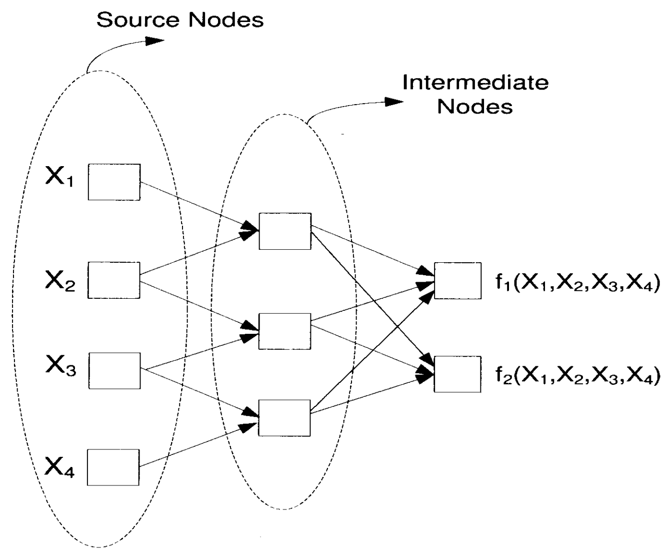


Figure 1-1: A example of a sensor network. Receivers want to compute different functions of source random variables. In example 1, $f_1(X_1, X_2, X_3, X_4) = \frac{X_1+X_2+X_3+X_4}{4}$ and $f_2(X_1, X_2, X_3, X_4) = \max(X_1, X_2, X_3, X_4)$.

X_4. *For this network, we have two receivers. One of the receivers desires to compute the average temperature of source locations, while the other one, the maximum temperature of sensors is desired. There are some nodes in this network which are neither a source, nor a receiver. We call them intermediate nodes. These intermediate nodes can perform some computations if required, but computations of functions at the receivers are only demands of this sensor network. Our objective here is to transmit as little data as possible so that receivers are able to compute their desired functions. In this thesis, we will investigate different aspects of this problem.*

Let us explain the second application, privacy issues, by another example.

Example 2. *Suppose a hospital has a database of medical records of patients such as their heights, weights, blood pressures, etc., which are private information. However, assume for a certain research program, some statistical information of these data (or, a certain function of these data) is required. To guarantee the privacy of patients' information, the database manager wants to release a certain amount of information which on one hand, researchers are able to compute their desired functions, and on*

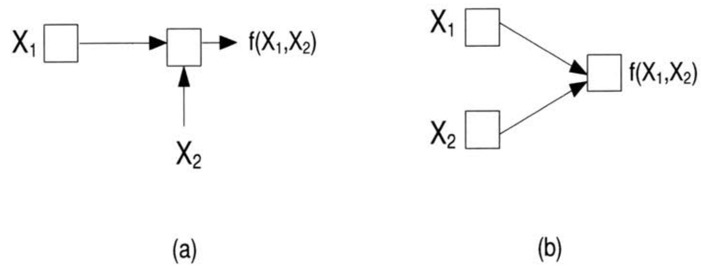


Figure 1-2: a) A network topology for functional compression with side information b) A network topology for distributed functional compression.

the other hand, the privacy of patients' medical records is being kept. In this thesis, we propose a framework to data-base managers to be able to deal with this problem.

Examples 1 and 2 demonstrate two major applications of the functional compression problem. There are also some other applications in the cloud computing, sensor networks, etc.

We proceed this chapter by investigating previous research progresses in this problem.

1.2 Prior Work in Functional Compression

We categorize prior work into the study in lossless functional compression and that in functional compression with distortion.

1.2.1 Lossless Functional Compression

By lossless computation, we mean asymptotically lossless computation of a function. In other words, we would like to have the error probability goes to zero as our block length goes to the infinity. We explain this concept later in more detail.

First, consider the network topology depicted in Figure 1-2-a which has two sources and a receiver. One of the sources is available at the receiver as a side information. Sometimes, this problem is referred to the side information problem.

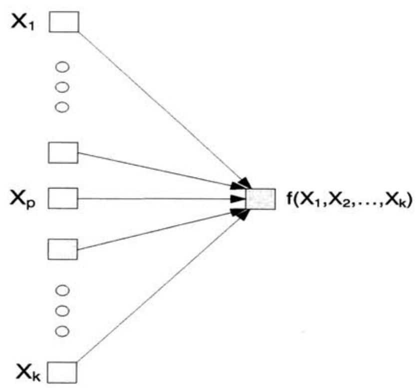


Figure 1-3: A general one-stage tree network with a desired function at the receiver.

Shannon was the first one who considered this problem in [23] for a special case when $f(X_1, X_2) = (X_1, X_2)$ (the identity function). For a general function, Orlitsky and Roche provided a single-letter characterization in [21]. In [11], Doshi et al. proposed an optimal coding scheme for this problem.

Now, consider the network topology depicted in Figure 1-2-b which has two sources and a receiver. This problem is a distributed compression problem. 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 [24]. Some other practical but suboptimal coding schemes have been proposed by Pradhan and Ramchandran in [22]. Also, a rate-splitting technique for this problem is developed by Coleman et al. in [7]. 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 [2], and Körner and Marton in [20], respectively. Under some special conditions on source distributions, Doshi *et al.* in [11] investigated this problem for a general function and proposed some achievable coding schemes.

Some parts of this Thesis consider different aspects of this problem (asymptotically lossless functional compression). In particular, we are going to answer to the following questions:

- For a one-stage tree network with one desired function at the receiver (as shown in Figure 1-3), what is a necessary and sufficient condition for any coding scheme

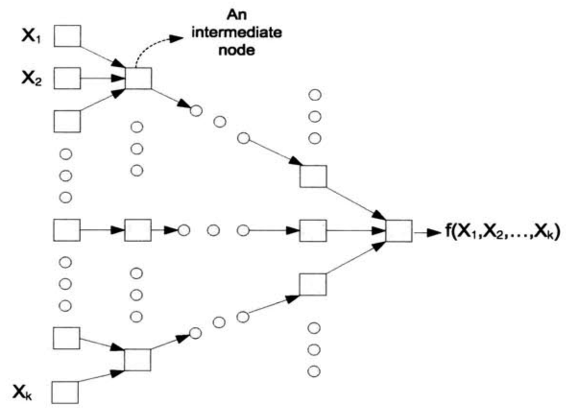


Figure 1-4: A general tree network setup with a desired function at the receiver.

to guarantee that the network is solvable (i.e., the receiver is able to compute its desired function)? What is a rate region for this network (a rate region is a set of rates for different links of the network under which the network is solvable)?

- For a general tree network with one desired function at the receiver (as shown in Figure 1-4), what is the optimal computation to be performed in an intermediate node? When do intermediate nodes need to perform some computations and what is a rate-region for this network?

- How do results extend to the case of having several desired functions at different receivers?

- Are there some special functions or some special source structures which lead to easy and practical coding schemes?

- What does happen if we have a feedback in our system?

1.2.2 Functional Compression with Distortion

In this section, we review prior results in functional compression for the case of being allowed to compute the desired function at the receiver within a distortion level.

First, consider the network topology depicted in Figure 1-2-a called the side information problem. Wyner and Ziv [27] considered this problem for computing the identity function at the receiver with distortion D. Yamamoto solved the side information problem for a general function $f(x, y)$ in [28]. Then, Doshi et al. gave a new characterization of the rate distortion function given by Yamamoto in [10].

For the network topology depicted in 1-2-b and for a general function, the rate-distortion region has been unknown, but some bounds have been given by Berger and Yeung [5], Barros and Servetto [4], and Wagner et al. [25], which considered a specific quadratic distortion function. Feng et al. [16] considered the side information problem for a general function at the receiver in the case the encoder and decoder have some noisy information.

In this Thesis, we characterize a rate-distortion function for a distributed network depicted in Figure 1-2-b. This proposed characterization is not a single letter characterization. However, we propose an achievable coding scheme for this problem.

In the rest of this chapter, we explain some research challenges of the functional compression problem by some examples. In the next chapters, we will explain these research challenges with more detail.

1.3 Research Challenges in Functional Compression

In this section, we address high level ideas of different research challenges in the functional compression problem. We use different simple examples to illustrate these ideas which will be carefully explained later in this Thesis.

Let us proceed by an example.

Example 3. *Consider the network shown in Figure 1-2-b which has two source nodes and a receiver. Suppose source nodes have two independent source random variables (RVs)* X_1 *and* X_2 *such that* X_1 *takes values from the set* $\mathcal{X}_1 = \{0, 1, 2, 3\}$*, and* X_2 *takes values from the set* $\mathcal{X}_2 = \{0, 1\}$*, both with equal probability. Suppose the receiver*

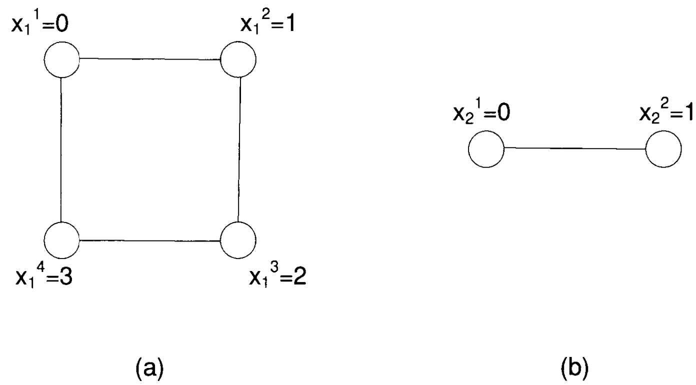


Figure 1-5: Characteristic graphs described in Example 3: a) G_{X_1}, b) G_{X_2}.

desires to compute a function $f(X_1, X_2) = (X_1 + X_2) \mod 2$.

If $X_1 = 0$ *or* $X_1 = 2$*, for all possible values of* X_2*, we have* $f(X_1, X_2) = X_2$*. Hence, we do not need to distinguish between* $X_1 = 0$ *and* $X_1 = 2$*. The same argument can be expressed for* $X_1 = 1$ *and* $X_1 = 3$*. However, cases* $X_1 = 0$ *and* $X_1 = 1$ *should be distinguished at a coding scheme, because for* $X_2 = 0$*, the function value is different when* $X_1 = 0$ *than the one when* $X_1 = 1$ *(i.e.,* $f(0,0) = 0 \neq f(1,0) = 1$*).*

In this, we notice that for each source random variable, depending on the function at the receiver and values of the other source random variable, we should *distinguish* some possible pair values. In other words, values of source random variables which *potentially* can cause a confusion at the receiver should be assigned to different *codes*. To determine which pair values of a random variable should be assigned to different codes, we make a graph for each RV, called the *characteristic graph* or the *confusion graph* of that RV. Vertices of this graph are different possible values of that RV. We connect two vertices if they should be distinguished. For the problem described in Example 3, the characteristic graph of X_1 (called G_{X_1}) is depicted in Figure 1-5-a. One may notice that we have not connected vertices which lead to the same function value for all values of X_2. The characteristic graph of X_2 (G_{X_2}) is shown in Figure 1-5-b.

Now, we seek to assign different codes to connected vertices. This code assignment is called a *graph coloring* where we assign different colors (codes) to connected vertices.

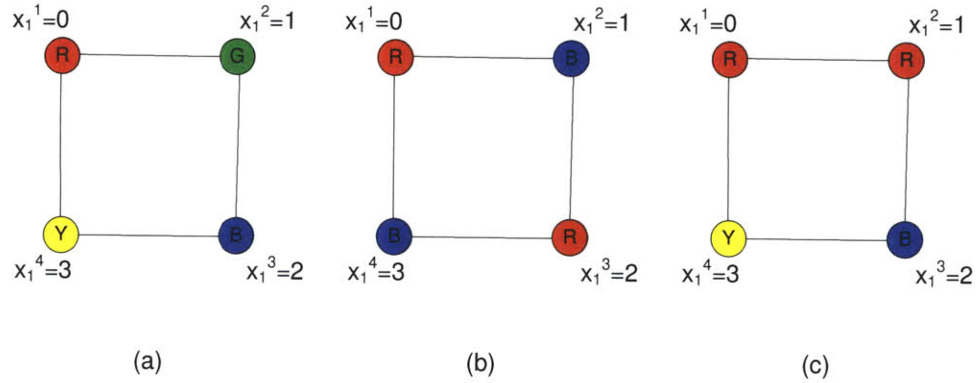


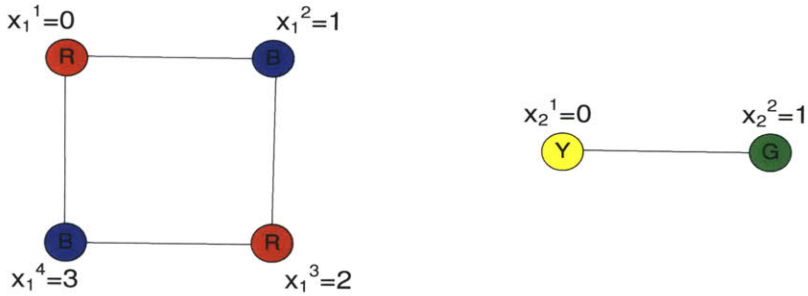
Figure 1-6: Graph coloring examples: (a) and (b) are valid colorings, while (c) is not a valid coloring. (For black and white prints, different letters written over graph vertices indicate different colors.)

Vertices that are not connected to each other can be assigned to the same or different colors (codes). Figure 1-6-(a,b) show two valid colorings for G_{X_1}, while Figure 1-6-c is not a valid coloring of G_{X_1}.

Now, we propose a possible coding scheme for this example. First, we choose valid colorings for G_{X_1} and G_{X_2}. Instead of sending source random variables, we send these coloring random variables. At the receiver side, we use a look-up table to compute the desired function value by using the received colorings. Figure 1-7 demonstrates this coding scheme.

However, this coloring-based coding scheme is not necessarily an achievable scheme. In other words, if we send coloring random variables instead of source random variables, the receiver may not be able to compute its desired function. Hence, we need some conditions to guarantee the achievability of coloring-based coding schemes. We explain this required condition by an example.

Example 4. *Consider the same network topology as explained in Example 3 shown in Figure 1-2-b. Suppose $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{0, 1\}$. The function values are depicted in Figure 1-8-a. In particular, $f(0, 0) = 0$ and $f(1, 1) = 1$. Dark squares in this figure represent points with zero probability. Figure 1-8-b demonstrates characteristic graphs of these source random variables. Each has two vertices, not connected to each*



(a) (b)

X_1 color	X_2 color	function value
r	y	0
r	g	1
b	y	1
b	g	0

(c)

Figure 1-7: a) G_{X_1} b) G_{X_2}, and c) a decoding look-up table for Example 3. (For black and white prints, different letters written over graph vertices indicate different colors.)

other. Hence, we can assign them to a same color. Figure 1-8-b shows these valid colorings for G_{X_1} and G_{X_2}. However, one may notice that if we send these coloring random variables instead of source random variables, the receiver would not be able to compute its desired function.

Example 4 demonstrates a case where a coloring-based coding scheme fails to be an achievable scheme. Thus, we need a condition to avoid these situations. We investigate this necessary and sufficient condition in Chapter 2. We call this condition the *coloring connectivity condition* or *C.C.C.* The situation of Example 4 happens

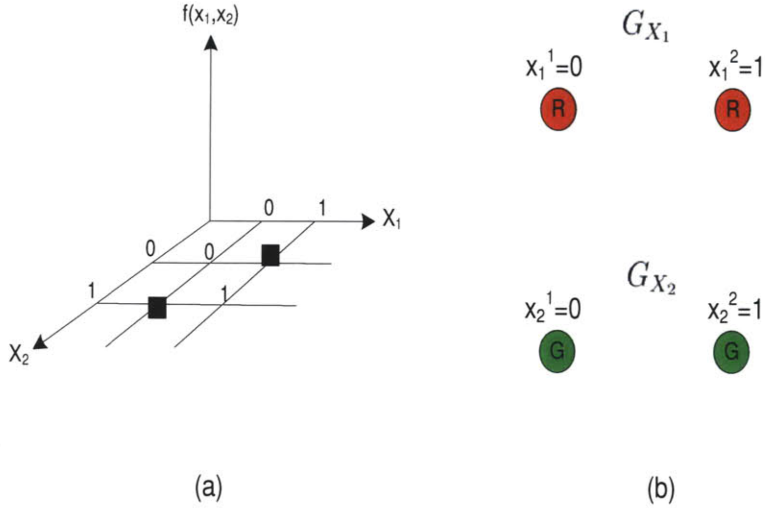


Figure 1-8: An example for colorings not satisfying C.C.C. (For black and white prints, different letters written over graph vertices indicate different colors.)

when we have a disconnected coloring class (i.e., a coloring class is a set of source pairs with the same color for each coordinates). C.C.C. is a condition to avoid this situation.

Hence, an achievable coding scheme can be expressed as follows. Sources instead of source random variables, send colorings of their random variables which satisfy C.C.C. Then, they perform a source coding on these coloring random variables. The receiver, by using these colors and a look-up table can compute its desired function.

However, there are some other coding schemes which are not included in this proposed scheme. In other words, we can compress source random variables more than the one of the proposed coloring-based coding scheme. We explain this fact by another example.

Example 5. *Consider the network shown in Figure 1-2-b. Suppose X_1 is uniformly distributed over $\mathcal{X}_1 = \{0, 1, 2, 3, 4\}$. Consider X_2 and $f(X_1, X_2)$ such that we have a graph depicted in Figure 1-9 for G_{X_1}. Figure 1-9 also demonstrates a valid coloring for this graph. Let us call this coloring random variable $c_{G_{X_1}}$. Hence, we have $H(c_{G_{X_1}}) \approx$ 1.52. Now, instead of X_1, suppose we encode $X_1 \times X_1$ (X_1^2), a random variable with*

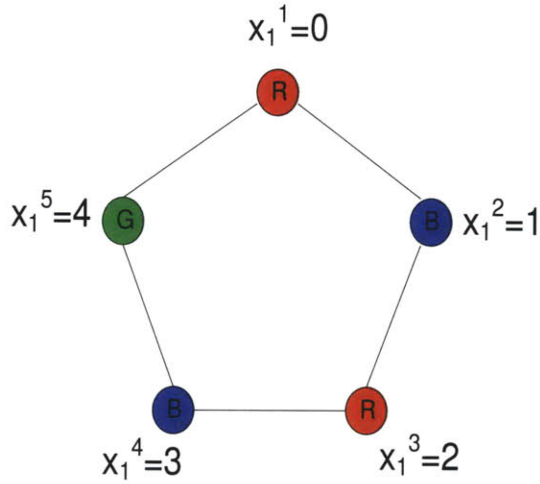


Figure 1-9: G_{X_1} described in Example 5. (For black and white prints, different letters written over graph vertices indicate different colors.)

25 possibilities ($\{00, 01, ..., 44\}$). To make its characteristic graph, we connect two vertices when at least one of their coordinates are connected in G_{X_1}. Figure 1-10 illustrates the characteristic graph of X_1^2 (referred by $G_{X_1}^2$ and called the second power of graph G_{X_1}). A valid coloring of this graph, called $c_{G_{X_1}^2}$ is shown in this figure. One may notice that we use eight colors to color this graph. We have,

$$\frac{1}{2}H(c_{G_{X_1}^2}) \approx 1.48 < H(c_{G_{X_1}}) \approx 1.52. \tag{1.1}$$

Example 5 demonstrates this fact that if we assign colors to a sufficiently large power graph of G_{X_1}, we can compress source random variables more. In Chapter 2, we show that sending colorings of sufficiently large power graphs of characteristic graphs which satisfy C.C.C. followed by a source coding (such as Slepian-Wolf compression) leads to an achievable coding scheme. On the other hand, any achievable coding scheme for this problem can be viewed as a coloring-based coding scheme satisfying C.C.C. In Chapter 2, we will explain these concepts with more detail.

Now, by another example, we explain some research challenges of this problem over tree networks.

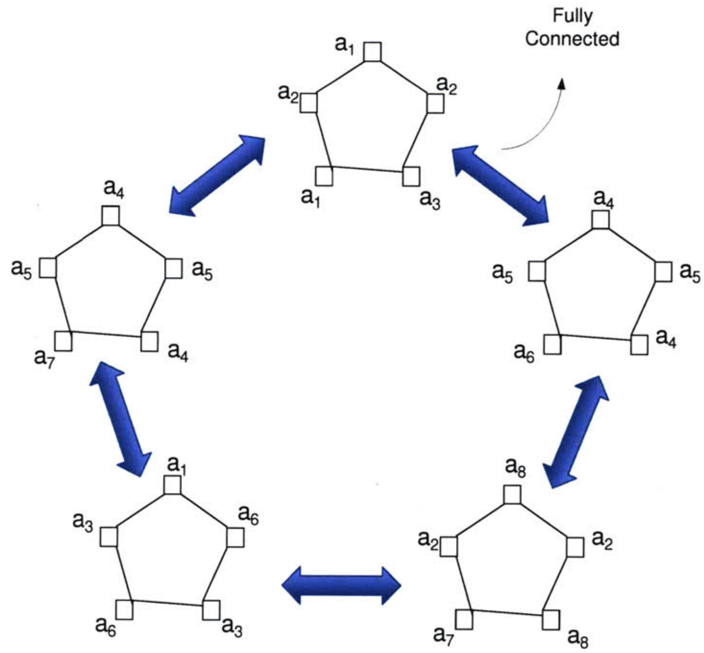


Figure 1-10: $G^2_{X_1}$, the second power graph of G_{X_1}, described in Example 5. Letters a_1,...,a_8 written over graph vertices indicate different colors. Two subsets of vertices are fully connected if each vertex of one set is connected to every vertex in the other set.

Example 6. *Consider the network topology depicted in Figure 1-11. This is a tree network with four sources, two intermediate nodes and a receiver. Suppose source random variables are independent, with equal probability to be zero or one. In other words, $\mathcal{X}_i = \{0,1\}$ for $i = 1,2,3,4$. Suppose the receiver wants to compute a parity check function $f(X_1, X_2, X_3, X_4) = (X_1 + X_2 + X_3 + X_4) \mod 2$. Also, intermediate nodes are allowed to perform some computations.*

In Example 6, first notice that characteristic graphs of source random variables are complete graphs. Hence, coloring random variables of sources are equal to source random variables. If intermediate nodes act like relays (i.e., no computations are performed at intermediate nodes), the following set of rates is an achievable scheme:

$$\begin{aligned} R_i &\geq 1 \text{ for } 1 \leq i \leq 4 \\ R_i &\geq 2 \text{ for } 5 \leq i \leq 6 \end{aligned} \tag{1.2}$$

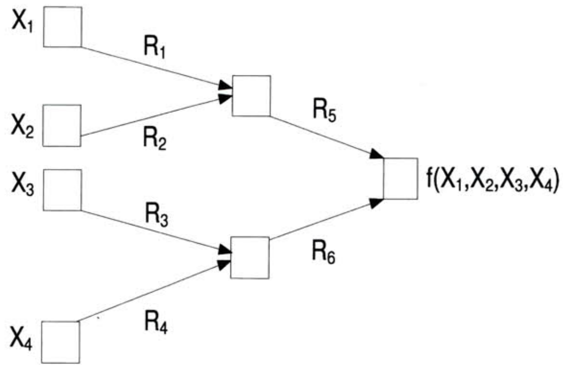


Figure 1-11: An example for a tree network with intermediate nodes and a desired function at the receiver.

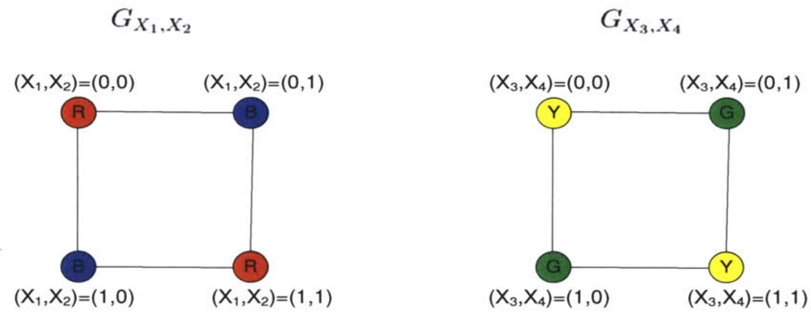
where R_i for $1 \leq i \leq 6$ are rates of different links depicted in Figure 1-11.

However, suppose intermediate nodes perform some computations. Assume source nodes send their coloring random variables satisfying C.C.C. (which in this case, they are equal to source random variables because characteristic graphs are complete). Then, each intermediate node makes its own characteristic graph and by using the received colors, it picks a corresponding color for its own characteristic graph and send that one. The receiver, by using the received colors of intermediate nodes' characteristic graphs and a look-up table, can compute its desired function. Figure 1-12 demonstrates this encoding/decoding scheme. Hence, for this example, intermediate nodes need to transmit one bit. Therefore, the following set of rates is achievable:

$$R_i \geq 1 \text{ for } 1 \leq i \leq 6. \tag{1.3}$$

It is worth to note that, in Example 6, by allowing intermediate nodes to compute, we can reduce transmission rates of some links. This problem is investigated in Chapter 2. In particular, we show what an optimal operation that an intermediate node is. Also, we show that for a family of functions and source random variables, intermediate nodes do not need to perform some computations and acting like relays is an optimal operation for them.

The problem of having more receivers than one with different desired functions



(a) (b)

(X_1,X_2) color	(X_3,X_4) color	function value
r	y	0
r	g	1
b	y	1
b	g	0

(c)

Figure 1-12: Characteristic graphs and a decoding look-up table for Example 6.

is considered in Chapter 3. For this problem, instead of a characteristic graph, we compute a new graph, called a *multi-functional characteristic graph*. This graph is basically an OR function of individual characteristic graphs with respect to different functions. In this chapter, we find a rate region and propose an achievable coding scheme for this problem.

In our proposed coding schemes for different functional compression problem, one needs to compute the 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 ([6]). However, in Chapter 4, 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. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to approximately solve the maximum independent set problem [8].

The effect of having a feedback on the rate-region of functional compression problem is investigated in Chapter 5. If the function at the receiver is the identity function, this problem is the Slepian-Wolf compression with feedback. For this case, having feedback does not give us any gain in terms of the rate. However, it is not the case when we have a general function f at the receiver. By having feedback, one may outperform rate bounds of the case without feedback.

Finally, we consider the problem of distributed functional compression with distortion in Chapter 6. The objective is to compress correlated discrete sources such that an arbitrary deterministic function of those sources can be computed at the receiver up to a distortion level. For this case, we compute a rate-distortion region and propose an achievable coding scheme. Conclusions and future work are expressed in Chapter 7.

Chapter 2

Functional Compression Over Tree Networks

In this chapter, 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 when sources are independent.

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, [12] considered a 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 [12], 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 random variables called chain-rule proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily close to the rate lower bound.

In this chapter, after giving the problem statement and reviewing previous results, we explain our main contributions in this problem.

2.1 Problem Setup

Consider k discrete memoryless random processes, $\{X_1^i\}_{i=1}^{\infty}$, ..., $\{X_k^i\}_{i=1}^{\infty}$, as source processes. Memorylessness is not necessary, and one can approximate a source by a memoryless one with an arbitrary precision [9]. Suppose these sources are drawn from finite sets $\mathcal{X}_1 = \{x_1^1, x_1^2, ..., x_1^{|\mathcal{X}_1|}\}$, ..., $\mathcal{X}_k = \{x_k^1, x_k^2, ..., x_k^{|\mathcal{X}_k|}\}$. These sources have a joint probability distribution $p(x_1, ..., x_k)$. We express n-sequences of these random variables as $\mathbf{X_1} = \{X_1^i\}_{i=l}^{i=l+n-1}$,..., $\mathbf{X_k} = \{X_k^i\}_{i=l}^{i=l+n-1}$ with the joint probability distribution $p(\mathbf{x}_1, ..., \mathbf{x}_k)$. 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 i^{th} element of $\mathbf{x}_j$ as x_{ji}. We use $\mathbf{x}_j^1$, $\mathbf{x}_j^2$,... as different n-sequences of $\mathbf{X}_j$. We shall drop the superscript when no confusion arises. Since the sequence $(\mathbf{x}_1, ..., \mathbf{x}_k)$ is drawn i.i.d. according to $p(x_1, ..., x_k)$, one can write $p(\mathbf{x}_1, ..., \mathbf{x}_k) = \prod_{i=1}^{n} p(x_{1i}, ..., x_{ki})$.

Consider an arbitrary tree network shown in Figure 2-1. 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 $\{X_j^i\}_{i=1}^{\infty}$. The receiver wishes to compute a deterministic function $f : \mathcal{X}_1 \times ... \times \mathcal{X}_k \to \mathcal{Z}$, or $f : \mathcal{X}_1^n \times ... \times \mathcal{X}_k^n \to \mathcal{Z}^n$, 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 located in leaves of the network. So, in the achieved

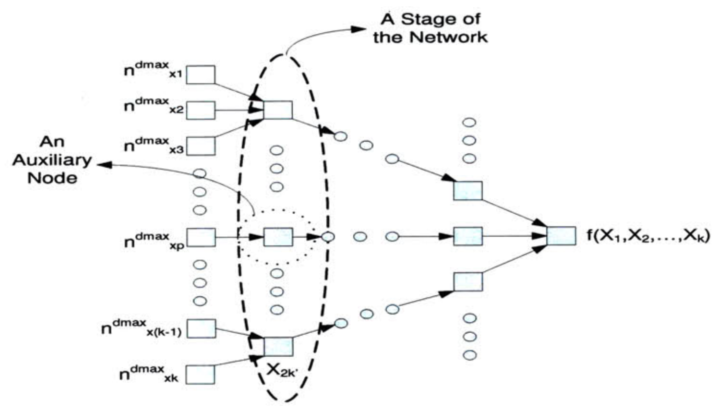


Figure 2-1: An arbitrary tree network topology.

network, sources are located in leaves (as an example, look at Figure 2-2).

Source node j encodes its message at a rate R_{X_j}. In other words, encoder en_{X_j} maps $en_{X_j} : \mathcal{X}_j^n \to \{1, ..., 2^{nR_{X_j}}\}$.

Suppose links connected to the receiver perform in rates $R^1_{\Xi_{1j}}$, $1 \le j \le w_1$, where w_1 is the number of links connected to the receiver (we explain these notations carefully in Section 2.4). The receiver has a decoder r which maps $r : \prod_j \{1, ..., 2^{nR^1_{\Xi_{1j}}}\} \to \mathcal{Z}^n$.

In other words, the receiver computes $r(\bigcup_{j=1}^{w_1} f^1_{\Xi_{1j}}) = r'(en_{X_1}(\mathbf{x}_1), ..., en_{X_k}(\mathbf{x}_k))$, where $\bigcup_{j=1}^{w_1} f^1_{\Xi_{1j}}$ is the information which the decoder gets at the receiver. 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 $P_e^n = Pr[(\mathbf{x}_1, ..., \mathbf{x}_k) : f(\mathbf{x}_1, ..., \mathbf{x}_k) \neq r'(en_{X_1}(\mathbf{x}_1), ..., en_{X_k}(\mathbf{x}_k))]$. A rate sequence is achievable iff 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.

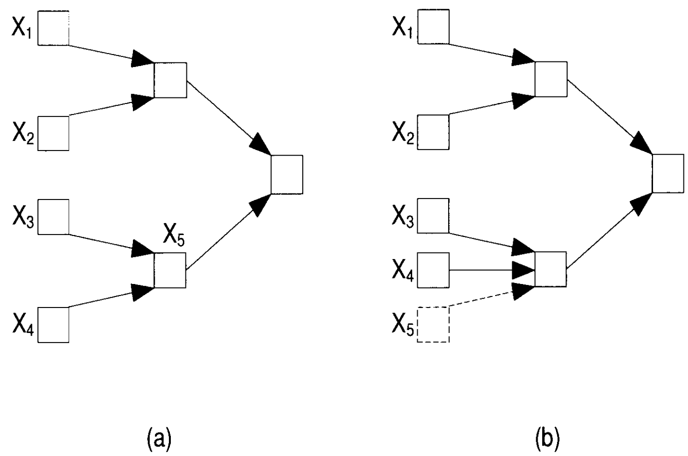


Figure 2-2: a) Sources are not necessarily located in leaves b) By adding some fake nodes, one can assume sources are in leaves of the modified tree.

2.2 Definitions and Prior Results

In this part, first we present some definitions used in formulating our results. We also review some prior results. Consider X_1 and X_2 as two random variables with the joint probability distribution $p(x_1, x_2)$. $f(X_1, X_2)$ is a deterministic function such that $f : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Z}$.

Definition 7. *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: $V_{X_1} = \mathcal{X}_1$ and an edge $(x_1^1, x_1^2) \in \mathcal{X}_1^2$ is in E_{X_1} iff there exists a $x_2^1 \in \mathcal{X}_2$ such that $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$ and $f(x_1^1, x_2^1) \neq f(x_1^2, x_2^1)$.*

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 [23]. Witsenhausen [26] used this concept to study a simplified version of our problem where one encodes X_1 to compute $f(X_1)$ with zero 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, ..., X_k are k random variables defined in Section 6.1.

Definition 8. *The characteristic graph $G_{X_1} = (V_{X_1}, E_{X_1})$ of X_1 with respect to random variables X_2,...,X_k, $p(x_1, ..., x_k)$, and $f(X_1, ..., X_k)$ is defined as follows: $V_{X_1} = \mathcal{X}_1$ and an edge $(x_1^1, x_1^2) \in \mathcal{X}_1^2$ is in E_{x_1} if there exist $x_j^1 \in \mathcal{X}_j$ for $2 \leq j \leq k$ such that $p(x_1^1, x_2^1, ..., x_k^1)p(x_1^2, x_2^1, ..., x_k^1) > 0$ and $f(x_1^1, x_2^1, ..., x_k^1) \neq f(x_1^2, x_2^1, ..., x_k^1)$.*

Example 9. *To illustrate the idea of confusability and the characteristic graph, consider two random variables X_1 and X_2 such that $\mathcal{X}_1 = \{0, 1, 2, 3\}$ and $\mathcal{X}_2 = \{0, 1\}$ where they are uniformly and independently distributed on their own supports. Suppose $f(X_1, X_2) = (X_1 + X_2) \bmod 2$ is to perfectly reconstructed at the receiver. Then, the characteristic graph of X_1 with respect to X_2, $p(x_1, x_2) = \frac{1}{8}$, and f is shown in Figure 1-5-a.*

The following definition can be found in [19].

Definition 10. *Given a graph $G_{X_1} = (V_{X_1}, E_{X_1})$ and a distribution on its vertices V_{X_1}, graph entropy is*

$$H_{G_{X_1}}(X_1) = \min_{X_1 \in W_1 \in \Gamma(G_{X_1})} I(X_1; W_1), \tag{2.1}$$

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

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 11. *Consider the scenario described in Example 9. For the characteristic graph of X_1 shown in Figure 1-5-a, the set of maximal independent sets is $W_1 = \{\{0, 2\}, \{1, 3\}\}$. To minimize $I(X_1; W_1) = H(X_1) - H(X_1|W_1) = \log(4) - H(X_1|W_1)$, 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, $H(X_1|w_1) = \log(2)$ bits, and since $p(w_1)$ is uniform, $H(X_1|W_1) = \log(2)$ bits. 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. In this example, functional compression saves us 1 bit in every 2 bits compared to the traditional data compression.

Witsenhausen [26] 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 [21] defined an extension of Körner's graph entropy, the *conditional graph entropy.*

Definition 12. *The conditional graph entropy is*

$$H_{G_{X_1}}(X_1|X_2) = \min_{\substack{X_1 \in W_1 \in \Gamma(G_{X_1}) \\ W_1 - X_1 - X_2}} I(W_1; X_1|X_2). \tag{2.2}$$

Notation $W_1 - X_1 - X_2$ indicates a Markov chain. If X_1 and X_2 are independent, $H_{G_{X_1}}(X_1|X_2) = H_{G_{X_1}}(X_1)$. To illustrate this concept, let us express an example borrowed from [21].

Example 13. *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 $x_2^1 \in \mathcal{X}_2$, the set $\{x_1^i : p(x_1^i, x_2^1) > 0\}$ of possible x_1^i 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 containing X_1 determine X_1. So,

$$\begin{aligned} H_{G_{x_1}}(X_1|X_2) &= \min_{\substack{X_1 \in W_1 \in \Gamma(G_{X_1}) \\ W_1 - X_1 - X_2}} I(W_1; X_1|X_2) \\ &= H(X_1|X_2) - \max_{X_1 \in W_1 \in \Gamma(G_{X_1})} H(X_1|W_1, X_2) \\ &= H(X_1|X_2). \end{aligned} \tag{2.3}$$

Definition 14. *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,*

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

Example 15. *Consider again the random variable* X_1 *described in Example 9, which its characteristic graph* G_{X_1} *is shown in Figure 1-5-a. Two valid colorings for* G_{X_1} *are shown in Figure* $1-6$*-(a,b). Figure* $1-6$*-c is not a valid coloring for this graph. Consider the coloring shown in Figure* $1-6$*-b. One can see that two connected vertices are assigned to different colors. Specifically,* $c_{G_{X_1}}(X_1) = \{r, b\}$ *standing for red and blue vertices (or equivalently, R, B letters in black and white print). So,* $p(c_{G_{X_1}}(x_1^i) = r) = p(x_1^i = 0) + p(x_1^i = 2)$, *and* $p(c_{G_{X_1}}(x_1^i) = b) = p(x_1^i = 1) + p(x_1^i = 3)$.

Definition 16. *The n-th power of a graph* G_{X_1} *is a graph* $G_{\mathbf{X}_1}^n = (V_{X_1}^n, E_{X_1}^n)$ *such that* $V_{X_1}^n = \mathcal{X}_1^n$ *and* $(\mathbf{x}_1^1, \mathbf{x}_1^2) \in E_{X_1}^n$ *when there exists at least one* i *such that* $(x_{1i}^1, x_{1i}^2) \in E_{X_1}$. *We denote a valid coloring of* $G_{\mathbf{X}_1}^n$ *by* $c_{G_{\mathbf{X}_1}^n}(\mathbf{X}_1)$.

Definition 17. *Given a non-empty set* $\mathcal{A} \subset \mathcal{X}_1 \times \mathcal{X}_2$, *define* $\hat{p}(x_1, x_2) = p(x_1, x_2)/p(\mathcal{A})$ *when* $(x_1, x_2) \in \mathcal{A}$, *and* $\hat{p}(x, y) = 0$ *otherwise.* $\hat{p}$ *is the distribution over* (x_1, x_2) *conditioned on* $(x_1, x_2) \in \mathcal{A}$. *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* $\hat{G}_{X_1} = (\hat{V}_{X_1}, \hat{E}_{X_1})$ *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* $\hat{G}_{X_2} = (\hat{V}_{X_2}, \hat{E}_{X_2})$. *Note that* $\hat{E}_{X_1} \subseteq E_{X_1}$ *and* $\hat{E}_{X_2} \subseteq E_{X_2}$. *Finally, we say that* $c_{G_{X_1}}(X_1)$ *and* $c_{G_{X_2}}(X_2)$ *are* ϵ*-colorings of* G_{X_1} *and* G_{x_2} *if they are valid colorings of* $\hat{G}_{X_1}$ *and* $\hat{G}_{X_2}$ *defined with respect to some set* $\mathcal{A}$ *for which* $p(\mathcal{A}) \geq 1 - \epsilon$.

In [3], the *Chromatic entropy* of a graph G_{X_1} is defined as,

Definition 18.

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

The chromatic entropy is a representation of the chromatic number of high probability subgraphs of the characteristic graph. In [12], the conditional chromatic entropy is defined as,

Definition 19.

$$H^{\chi}_{G_{X_1}}(X_1|X_2) = \min_{c_{G_{X_1}} \text{ is an } \epsilon\text{-coloring of } G_{X_1}} H(c_{G_{X_1}}(X_1)|X_2).$$

Regardless of ϵ, 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 ϵ-colorings, regardless of ϵ.

In general, these optimizations are NP-hard ([6]). But, depending on the desired function f, there are some cases that they are not NP-hard. For other cases, there exist some heuristic algorithms to approximate them. We will discuss these cases in Chapter 4.

Körner showed in [19] that, in the limit of large n, there is a relation between the chromatic entropy and the graph entropy.

Theorem 20.

$$\lim_{n\to\infty} \frac{1}{n} H^{\chi}_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1) = H_{G_{X_1}}(X_1). \tag{2.4}$$

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 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 $H_{G_{X_1}}(X_1)$, 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 $G^n_{\mathbf{X}_1}$, 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 [11].

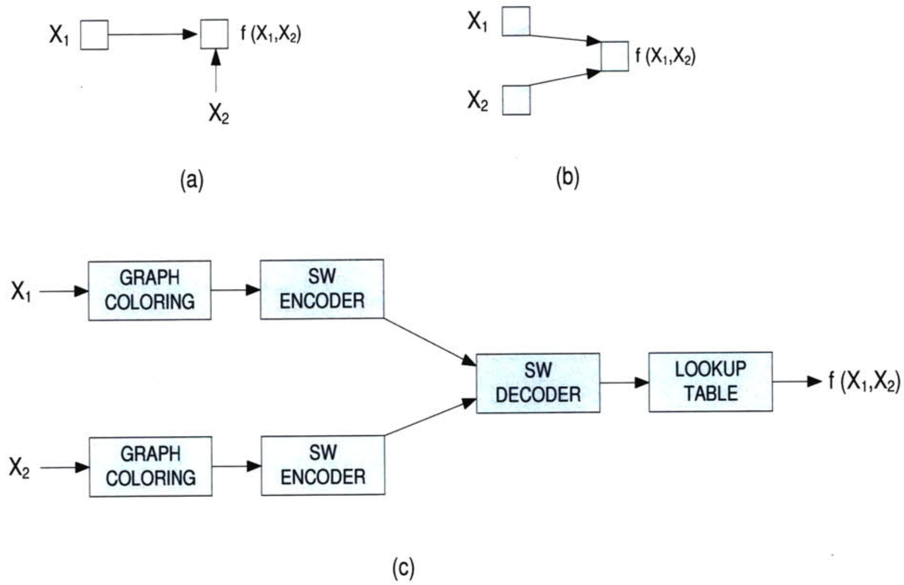


Figure 2-3: a) Functional compression with side information b) A distributed functional compression problem with two transmitters and a receiver c) An achievable encoding/decoding scheme for the functional compression.

Theorem 21.

$$\lim_{n\to\infty} \frac{1}{n} H^{\chi}_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1|\mathbf{X}_2) = H_{G_{X_1}}(X_1|X_2). \tag{2.5}$$

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 [21] 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 we first find the minimum entropy coloring of $G^n_{\mathbf{X}_1}$ for large enough n, and then use a compression scheme on the coloring random variable (such as Slepian-Wolf [24]) to achieve a rate arbitrarily close to $H(c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)|\mathbf{X}_2)$. This encoding scheme guarantees the computation of the function at the receiver with a vanishing probability of error.

All the mentioned results considered only functional compression with side information at the receiver (Figure 2-3-a). Consider the network shown in Figure 2-3-b. 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, [12] determined a rate-region of this network when source random variables satisfy a condition called the zigzag condition. In the following, we explain this condition.

We refer to the joint-typical set of sequences of random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ as T_ϵ^n. k is implied in this notation for simplicity. We explicitly mention k if some confusion arises. T_ϵ^n can be considered as a strong or weak typical set ([9]).

Definition 22. *A discrete memoryless source* $\{(X_1^i, X_2^i)\}_{i\in\mathbb{N}}$ *with a distribution* $p(x_1, x_2)$ *satisfies the zigzag condition if for any* ϵ *and some* n, $(\mathbf{x}_1^1, \mathbf{x}_2^1)$, $(\mathbf{x}_1^2, \mathbf{x}_2^2) \in T_\epsilon^n$, *there exists some* $(\mathbf{x}_1^3, \mathbf{x}_2^3) \in T_\epsilon^n$ *such that* $(\mathbf{x}_1^3, \mathbf{x}_2^i), (\mathbf{x}_1^i, \mathbf{x}_2^3) \in T_{\frac{\epsilon}{2}}^n$ *for each* $i \in \{1,2\}$, *and* $(x_{1j}^3, x_{2j}^3) = (x_{1j}^i, x_{2j}^{3-i})$ *for some* $i \in \{1,2\}$ *for each* j.

In fact, the zigzag condition forces many source sequences to be typical. We first explain the results of [12]. Then, in Section 2.3, we compute a rate-region of this network in a general case without having any restrictive conditions on source random variables (such as the zigzag condition). Then, We extend our results to the case of having k source nodes.

Reference [12] shows that, if the source random variables satisfy the zigzag condition, an 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. In [12], it is not claimed that this condition is necessary, but sufficient. In other words, [12] computed a rate-region only in the case that source random variables satisfy the zigzag condition. The zigzag condition is a restrictive condition which does not depend on the desired function at the receiver.

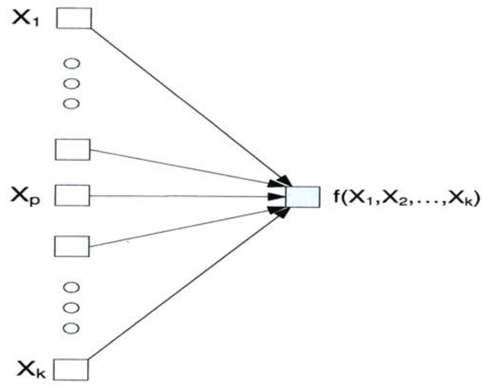


Figure 2-4: A general one-stage tree network topology.

2.3 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 2-4 with k sources.

Definition 23. *A path with length m between two points $Z_1 = (x_1^1, x_2^1, ..., x_k^1)$, and $Z_m = (x_1^2, x_2^2, ..., x_k^2)$ 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 23 can be expressed for two n-length vectors as follows.

Definition 24. *A path with length m between two points $\mathbf{Z}_1 = (\mathbf{x}_1^1, \mathbf{x}_2^1, ..., \mathbf{x}_k^1) \in T_\epsilon^n$, and $\mathbf{Z}_m = (\mathbf{x}_1^2, \mathbf{x}_2^2, ..., \mathbf{x}_k^2) \in T_\epsilon^n$ are determined by $m-1$ points $\mathbf{Z}_i$, $1 \leq i \leq m$ such that,*

i) $\mathbf{Z}_i \in T_\epsilon^n$, for all $1 \leq i \leq m$.

ii) $\mathbf{Z}_i$ and $\mathbf{Z}_{i+1}$ only differ in one of their coordinates.

Definition 25. *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 $J_C = \{j_c^1, ..., j_c^{n_{j_c}}\}$ where $j_c^i = \left\{(x_1^{i_1}, x_2^{i_2}, ..., x_k^{i_k}) : c_{G_{X_1}}(x_1^{i_1}) = c_{G_{X_2}}(x_2^{i_2}) = ... = c_{G_{X_k}}(x_k^{i_k})\right\}$ for any valid i_1,...i_k, and $n_{j_c} = |c_{G_{X_1}}| \times |c_{G_{X_2}}| \times ... \times |c_{G_{X_k}}|$. We call each j_c^i as a joint coloring class.*

Definition 25 can be expressed for random vectors $\mathbf{X}_1$,...,$\mathbf{X}_k$ with characteristic graphs $G^n_{\mathbf{X}_1}$,...,$G^n_{\mathbf{X}_k}$, and any valid ϵ-colorings $c_{G^n_{\mathbf{X}_1}}$,...,$c_{G^n_{\mathbf{X}_k}}$, respectively.

Definition 26. *Consider random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, and any valid colorings $c_{G_{X_1}}$, ..., $c_{G_{X_k}}$. We say these colorings satisfy the Coloring Connectivity Condition (C.C.C.) when, between any two points in $j^i_c \in J_C$, there exists a path that lies in j^i_c, or function f has the same value in disconnected parts of j^i_c.*

C.C.C. can be expressed for random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$, and any valid ϵ-colorings $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$, respectively.

Example 27. *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^1_c represents this joint coloring class. In other words, $j^1_c = \{(x_1^i, x_2^j)\}$, for all $1 \leq i, j \leq 2$ when $p(x_1^i, x_2^j) > 0$. Figure 2-5 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 2-5-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 2-5-b. In this case, we have $p(x_1^1, x_2^2) = 0$, $p(x_1^2, x_2^1) = 0$, 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^1_c. So, though these two points belong to a same joint coloring class, their corresponding function values can be different from each other. Thus, j^1_c does not satisfy C.C.C. for this example. Therefore, $c_{G_{X_1}}$ and $c_{G_{X_2}}$ do not satisfy C.C.C.*

Lemma 28. *Consider two random variables 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 $c_{G_{X_2}}(X_2)$ is a trivial coloring, assigning different colors to different vertices (to simplify the notation, we use $c_{G_{X_2}}(X_2) = X_2$ to refer to this coloring). These colorings satisfy C.C.C. Also, $c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)$ and $c_{G^n_{\mathbf{X}_2}}(\mathbf{x}_2) = \mathbf{X}_2$ satisfy C.C.C for any n.*

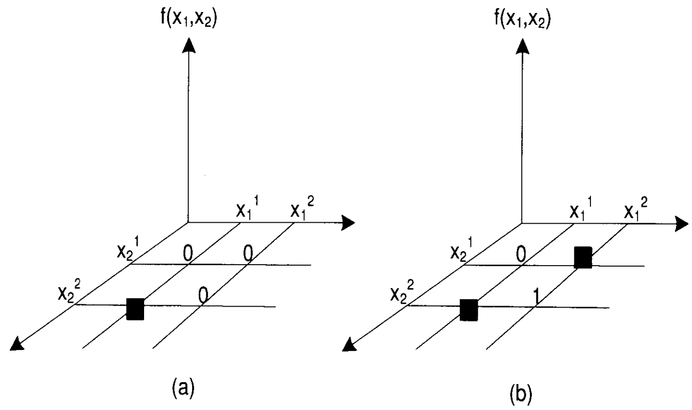


Figure 2-5: Two examples of a joint coloring class: a) satisfying C.C.C. b) not satisfying C.C.C. Dark squares indicate points with zero probability. Function values are depicted in the picture.

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 $J_C = \{j_c^1, ..., j_c^{n_{jc}}\}$ such that $j_c^1 = \{(x_1^i, x_2^1) : c_{G_{X_1}}(x_1^i) = \sigma_i\}$, where σ_i is a generic color. Any two points in j_c^1 are connected to each other with a path with length one. So, j_c^1 satisfies C.C.C. This arguments hold for any j_c^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 $c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)$ and $c_{G^n_{\mathbf{X}_2}}(\mathbf{X}_2) = \mathbf{X}_2$ is similar. □

Lemma 29. *Consider random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, and any valid colorings $c_{G_{X_1}}$, ..., $c_{G_{X_k}}$ with joint coloring class $J_C = \{j_c^i : i\}$. For any two points $(x_1^1, ..., x_k^1)$ and $(x_1^2, ..., x_k^2)$ in j_c^i, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$ if and only if j_c^i satisfies C.C.C.*

Proof. We first show that if j_c^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_c^i, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$. Since j_c^i satisfies C.C.C., either $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$, or there exists a path with length $m-1$ between these two points $Z_1 = (x_1^1, ..., x_k^1)$ and $Z_m = (x_1^2, ..., x_k^2)$, for some m. Two consecutive points Z_j and Z_{j+1} in this path, differ in just one of their coordinates. Without loss of generality, suppose they differ in their first coordinate. In other words, suppose $Z_j = (x_1^{j_1}, x_2^{j_2} ..., x_k^{j_k})$ and $Z_{j+1} = (x_1^{j_0}, x_2^{j_2} ..., x_k^{j_k})$. Since these two points belong to j_c^i,

$c_{G_{X_1}}(x_1^{j_1}) = c_{G_{X_1}}(x_1^{j_0})$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $x_1^{j_1}$ and $x_1^{j_0}$ 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 inductively 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_c^i does not satisfy C.C.C., it means that there exists at least two points Z_1 and Z_2 in j_c^i such that no path exists between them. So, the value of f can be different in these points. As an example, consider Figure 2-5-b. The value of the function can be different in two disconnected points in a same joint coloring class. $\square$

Lemma 30. *Consider random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$, and any valid ϵ-colorings $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$ with the joint coloring class $J_C = \{j_c^i : i\}$. For any two points $(\mathbf{x}_1^1, ..., \mathbf{x}_k^1)$ and $(\mathbf{x}_1^2, ..., \mathbf{x}_k^2)$ in j_c^i, $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$ if and only if j_c^i satisfies C.C.C.*

Proof. The proof is similar to Lemma 29. The only difference is to use the definition of C.C.C. for $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$. Since j_c^i satisfies C.C.C., either $f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2)$, or there exists a path with length $m-1$ between any two points $Z_1 = (\mathbf{x}_1^1, ..., \mathbf{x}_k^1) \in T_\epsilon^n$ and $Z_m = (\mathbf{x}_1^2, ..., \mathbf{x}_k^2) \in T_\epsilon^n$ in j_c^i, 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 $Z_j = (\mathbf{x}_1^{j_1}, \mathbf{x}_2^{j_2} ..., \mathbf{x}_k^{j_k}) \in T_\epsilon^n$ and $Z_{j+1} = (\mathbf{x}_1^{j_0}, \mathbf{x}_2^{j_2} ..., \mathbf{x}_k^{j_k}) \in T_\epsilon^n$. Since these two points belong to j_c^i, $c_{G_{X_1}}(\mathbf{x}_1^{j_1}) = c_{G_{X_1}}(\mathbf{x}_1^{j_0})$. If $f(Z_j) \neq f(Z_{j+1})$, there would exist an edge between $\mathbf{x}_1^{j_1}$ and $\mathbf{x}_1^{j_0}$ in $G^n_{\mathbf{X}_1}$ 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 $f(Z_1) = f(Z_2) = ... = f(Z_m)$. The converse part is similar to Lemma 29. $\square$

Next, we want to show that if X_1 and X_2 satisfy the zigzag condition mentioned in Definition 22, 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 [12] is not necessary, but sufficient.

Lemma 31. *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. 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 a joint coloring class j_c^i. 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. 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 28, C.C.C. always holds without having any condition. $\square$

Definition 32. *For random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, the joint graph entropy is defined as follows:*

$$H_{G_{X_1},...,G_{X_k}}(X_1, ..., X_k) \triangleq \lim_{n\to\infty} \min_{c_{G^n_{\mathbf{X}_1}},...,c_{G^n_{\mathbf{X}_k}}} \frac{1}{n} H(c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1), ..., c_{G^n_{\mathbf{X}_k}}(\mathbf{X}_k)) \tag{2.6}$$

in which $c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)$, ..., $c_{G^n_{\mathbf{X}_k}}(\mathbf{X}_k)$ are ϵ-colorings of $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$ satisfying C.C.C. We sometimes refer to the joint graph entropy by using $H_{\bigcup_{i=1}^k G_{X_i}}(X_1, ..., X_k)$. It is worth to note this limit exists because we have a monotonically decreasing sequence bounded below. Similarly, we can define the conditional graph entropy.

Definition 33. *For random variables X_1, ..., X_k with characteristic graphs G_{X_1}, ..., G_{X_k}, the conditional graph entropy can be defined as follows:*

$$\begin{aligned} &H_{G_{X_1},...,G_{X_i}}(X_1, ..., X_i | X_{i+1}, ..., X_k) \\ &\triangleq \lim_{n\to\infty} \min \frac{1}{n} H(c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1), ..., c_{G^n_{\mathbf{X}_i}}(\mathbf{X}_i) | c_{G^n_{\mathbf{X}_{i+1}}}(\mathbf{X}_{i+1}), ..., c_{G^n_{\mathbf{X}_k}}(\mathbf{X}_k)) \end{aligned} \tag{2.7}$$

in which minimization is over $c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)$, ..., $c_{G^n_{\mathbf{X}_k}}(\mathbf{X}_k)$, which are ϵ-colorings of $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$ satisfying C.C.C.

Lemma 34. *For $k = 2$, Definitions 12 and 33 are the same.*

Proof. By using the data processing inequality, we have

$$\begin{aligned}
H_{G_{X_1}}(X_1|X_2) &= \lim_{n\to\infty} \min_{c_{G^n_{\mathbf{X}_1}}, c_{G^n_{\mathbf{X}_2}}} \frac{1}{n} H(c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)|c_{G^n_{\mathbf{X}_2}}(\mathbf{X}_2)) \\
&= \lim_{n\to\infty} \min_{c_{G^n_{\mathbf{X}_1}}} \frac{1}{n} H(c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)|\mathbf{X}_2).
\end{aligned}$$

Then, Lemma 28 implies that $c_{G^n_{\mathbf{X}_1}}(\mathbf{X}_1)$ and $c_{G^n_{\mathbf{X}_2}}(\mathbf{x}_2) = \mathbf{X}_2$ satisfy C.C.C. A direct application of Theorem 21 completes the proof. □

By this definition, the graph entropy does not satisfy the chain rule.

Suppose $\mathcal{S}(k)$ denotes the power set of the set $\{1, 2, ..., k\}$ excluding the empty subset (this is the set of all subsets of $\{1, \ldots, k\}$ without the empty set). Then, for any $S \in \mathcal{S}(k)$,

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

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

Theorem 35. *A rate region of the network shown in Figure 2-4 is determined by these conditions:*

$$\forall S \in \mathcal{S}(k) \Longrightarrow \sum_{i\in S} R_{X_i} \geq H_{\bigcup_{i\in S} G_{X_i}}(X_S|X_{S^c}). \tag{2.8}$$

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 36. *Consider random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$, and any valid ϵ-colorings $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$ satisfying C.C.C., for sufficiently* *large n. There exists*

$$\hat{f} : c_{G^n_{\mathbf{X}_1}}(\mathcal{X}_1) \times ... \times c_{G^n_{\mathbf{X}_k}}(\mathcal{X}_k) \to \mathcal{Z}^n \tag{2.9}$$

such that $\hat{f}(c_{G^n_{\mathbf{X}_1}}(\mathbf{x}_1), ..., c_{G^n_{\mathbf{X}_k}}(\mathbf{x}_k)) = f(\mathbf{x}_1, ..., \mathbf{x}_k)$*, for all* $(\mathbf{x}_1, ..., \mathbf{x}_k) \in T^n_\epsilon$.

Proof. Suppose the joint coloring family for these colorings is $J_C = \{j^i_c : i\}$. We proceed by constructing $\hat{f}$. Assume $(\mathbf{x}^1_1, ..., \mathbf{x}^1_k) \in j^i_c$ and $c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^1_1) = \sigma_1$, ..., $c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^1_k) = \sigma_k$. Define $\hat{f}(\sigma_1, ...\sigma_k) = f(\mathbf{x}^1_1, ..., \mathbf{x}^1_k)$.

To show this function is well-defined on elements in its support, we should show that for any two points $(\mathbf{x}^1_1, ..., \mathbf{x}^1_k)$ and $(\mathbf{x}^2_1, ..., \mathbf{x}^2_k)$ in T^n_ϵ, if $c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^1_1) = c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^2_1)$, ..., $c_{G^n_{\mathbf{X}_k}}(\mathbf{x}^1_k) = c_{G^n_{\mathbf{X}_k}}(\mathbf{x}^2_k)$, then $f(\mathbf{x}^1_1, ..., \mathbf{x}^1_k) = f(\mathbf{x}^2_1, ..., \mathbf{x}^2_k)$.

Since $c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^1_1) = c_{G^n_{\mathbf{X}_1}}(\mathbf{x}^2_1)$, ..., $c_{G^n_{\mathbf{X}_k}}(\mathbf{x}^1_k) = c_{G^n_{\mathbf{X}_k}}(\mathbf{x}^2_k)$, these two points belong to a joint coloring class like j^i_c. Since $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$ satisfy C.C.C., by using Lemma 30, $f(\mathbf{x}^1_1, ..., \mathbf{x}^1_k) = f(\mathbf{x}^2_1, ..., \mathbf{x}^2_k)$. Therefore, our function $\hat{f}$ is well-defined and has the desired property. □

Lemma 36 implies that given ϵ-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 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 ϵ).

Lemma 37. *(Slepian-Wolf Theorem)*

A rate-region of the network shown in Figure 2-4 where $f(X_1, ..., X_k) = (X_1, ..., X_k)$ *can be determined by these conditions:*

$$\forall S \in \mathcal{S}(k) \Longrightarrow \sum_{i \in S} R_{X_i} \geq H(X_S | X_{S^c}). \tag{2.10}$$

Proof. See [24]. □

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 $\frac{\epsilon}{k}$. Then, the total error in the decoding of colorings at the receiver is less than ϵ. Therefore, the total error in the coding scheme of first coloring $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$, and then encoding those colors by using SW encoding/decoding scheme is upper by the sum of errors in each stage. By using Lemmas 36 and 37, the total error is less than ϵ, and goes to zero as n goes to infinity. By applying Lemma 37 on achieved coloring random variables, we have,

$$\forall S \in \mathcal{S}(k) \Longrightarrow \sum_{i \in S} R_{X_i} \geq \frac{1}{n} H(c_{G^n_{\mathbf{X}_S}} | c_{G^n_{\mathbf{X}_{S^c}}}), \tag{2.11}$$

where $c_{G^n_{\mathbf{X}_S}}$, and $c_{G^n_{\mathbf{X}_{S^c}}}$ are ϵ-colorings of characteristic graphs satisfying C.C.C. Thus, using Definition 33 completes the achievability part.

As an example, look at Figure 2-3-c. This network has two source nodes and a receiver. Source nodes compute ϵ-colorings of their characteristic graphs. These colorings should satisfy C.C.C. Then, an SW compression is performed on these colorings. The receiver, first, perform SW decoding to get the colors. Then, by using a look-up table, it can find the value of its desired function (As an example, look at Figure 1-7).

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

$$\mathcal{F}^n_\epsilon = \{\hat{f} : Pr[\hat{f}(\mathbf{X}_1, ..., \mathbf{X}_k) \neq f(\mathbf{X}_1, ..., \mathbf{X}_k)] < \epsilon\}. \tag{2.12}$$

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

Lemma 38. *Consider some function $f : \mathcal{X}_1 \times ... \times \mathcal{X}_k \to \mathcal{Z}$. 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,

$$f(x_1^1, ..., x_k^1) = f(x_1^2, ..., x_k^2) = r'(en_{X_1}(x_1^1), ..., en_{X_k}(x_k^1)). \tag{2.13}$$

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 random variables 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 $(x_2^1, ..., x_k^1)$ such that, $p(x_1^1, x_2^1, ..., x_k^1)p(x_1^2, x_2^1, ..., x_k^1) > 0$, $en_{X_1}(x_1^1) = en_{X_1}(x_1^2)$, and $f(x_1^1, x_2^1, ..., x_k^1) \neq f(x_1^2, x_2^1, ..., x_k^1)$. By taking $x_2^1 = x_2^2$, ..., $x_k^1 = x_k^2$ in (2.13), 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_c^i such that there is no path between them in j_c^i. So, Lemma 29 says that the function f can get different values in these two points. In other words, it is possible to have $f(x_1^1, ..., x_k^1) \neq f(x_1^2, ..., x_k^2)$, 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 (2.13). Thus, achieved colorings satisfy C.C.C. □

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

Lemma 39. *Consider random variables* $\mathbf{X}_1$, ..., $\mathbf{X}_k$. *All* ϵ*-achievable functional codes of these random variables induce* ϵ*-colorings on characteristic graphs satisfying*

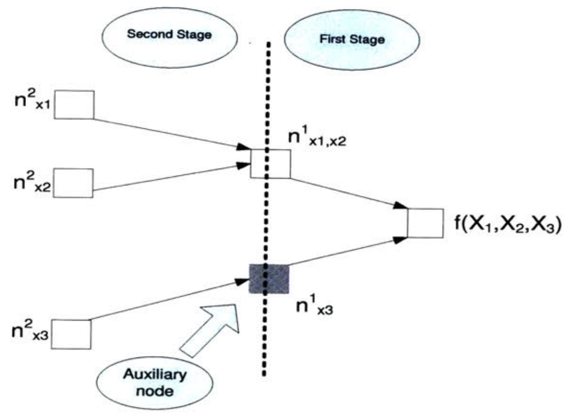


Figure 2-6: A simple tree network.

C.C.C.

Proof. Suppose $g(\mathbf{x}_1, ..., \mathbf{x}_k) = r'(en_{X_1}(\mathbf{x}_1), ..., en_{X_k}(\mathbf{x}_k)) \in \mathcal{F}_\epsilon^n$ is such a code. Lemma 38 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 $(\mathbf{x}_1, ..., \mathbf{x}_k)$ such that $g(\mathbf{x}_1, ..., \mathbf{x}_k) \neq f(\mathbf{x}_1, ..., \mathbf{x}_k)$ be denoted by $\mathcal{C}$. Since $g \in \mathcal{F}_\epsilon^n$, $Pr[\mathcal{C}] < \epsilon$. Therefore, functions en_{X_1}, ..., en_{X_k} restricted to $\mathcal{C}$ are ϵ-colorings of characteristic graphs satisfying C.C.C. (by definition). □

Lemmas 38 and 39 establish the converse part and complete the proof. □

If we have two transmitters ($k = 2$), Theorem 35 can be simplified as follows.

Corollary 40. *A rate region of the network shown in Figure 2-3-b is determined by these three conditions:*

$$\begin{aligned} R_{X_1} &\geq H_{G_{X_1}}(X_1|X_2) \\ R_{X_2} &\geq H_{G_{X_2}}(X_2|X_1) \\ R_{X_1} &+R_{X_2} \geq H_{G_{X_1},G_{X_2}}(X_1, X_2). \end{aligned} \tag{2.14}$$

2.4 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 2-1). We refer to other nodes of this tree as intermediate nodes. The receiver wishes to compute a deterministic function of source random variables. 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 41. *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 a tree standardization.*

These concepts are depicted in Figure 2-6. 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 42. *Any modified tree network with k source nodes with distance d_{max} from the receiver can be expressed by a connection set $S_T = \{s_t^i : 1 \leq i \leq d_{max}\}$ where $s_t^i = \{\Xi_{ij} : 1 \leq j \leq w_i\}$. 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 random variables is in Ξ_{ij} when paths of those source nodes have the last i common hops.*

For example, consider the network shown in Figure 2-6. Its connection set is $S_T = \{s_t^1, s_t^2\}$ such that $s_t^1 = \{(X_1, X_2), X_3\}$ and $s_t^2 = \{X_1, X_2, X_3\}$. 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 Ξ_{ij} as $n^i_{\Xi_{ij}}$. We call the outgoing link from this node $e^i_{\Xi_{ij}}$. The rate of this link is referred by $R^i_{\Xi_{ij}}$. For instance, nodes in the second stage of the network shown in Figure 2-6 are called $n^2_{X_1}$, $n^2_{X_2}$ and $n^2_{x_3}$ with outgoing links $e^2_{X_1}$, $e^2_{X_2}$ and $e^2_{X_3}$, respectively. Nodes in the first stage of this network are referred by $n^1_{X_1,X_2}$ and $n^1_{X_3}$ with outgoing links $e^1_{X_1,X_2}$ and $e^1_{X_3}$.

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 random variable which is transmitted in the link $e^i_{\Xi_{ij}}$ is called $f^i_{\Xi_{ij}}$. Also, we refer to the function computed in an intermediate node $n^i_{\Xi_{ij}}$ as $g^i_{\Xi_{ij}}$. For example, consider again the network shown in Figure 2-6. Random variables sent through links $e^2_{X_1}$, $e^2_{X_2}$, $e^2_{X_3}$, $e^1_{X_1,X_2}$ and $e^1_{X_3}$ are $f^2_{X_1}$, $f^2_{X_2}$, $f^2_{X_3}$, $f^1_{X_1,X_2}$ and $f^1_{X_3}$ such that $f^1_{X_1,X_2} = g^1_{X_1,X_2}(f^2_{X_1}, f^2_{X_2})$, and $f^1_{X_3} = g^1_{X_3}(f^2_{X_2})$.

2.4.1 A Rate Lower Bound

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

Theorem 43. *A rate lower bound of a tree network with the connection set* $S_T = \{s^i_t : i\}$ *can be determined by these conditions,*

$$\forall s_i \in \mathcal{S}(w_i) \Longrightarrow \sum_{j \in s_i} R^i_{\Xi_{ij}} \geq H_{\bigcup_{z \in s_i} G_{\Xi_{iz}}}(\Xi_{is_i} | \Xi_{is_i^c}) \tag{2.15}$$

for all $i = 1, ..., |S_T|$ *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, $n^i_{\Xi_{ij}}$ for $1 \leq j \leq w_i$. Suppose they are directly connected to the receiver. So, the information sent in links of this stage should be enough to compute the desired function. In the best case, suppose their parent nodes sent all their information without doing any compression. So, by direct application of Theorem 35, one can see that,

$$\forall s_i \in \mathcal{S}(w_i) \Longrightarrow \sum_{j \in s_i} R^i_{\Xi_{ij}} \geq H_{\bigcup_{z \in s_i} G_{\Xi_{iz}}}(\Xi_{is_i} | \Xi_{is_i^c}). \tag{2.16}$$

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

In the following, we express some cases under which we can achieve the derived rate lower bound of Theorem 43.

2.4.2 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 random variables $\mathbf{X}_1$, ..., $\mathbf{X}_k$ with characteristic graphs $G^n_{\mathbf{X}_1}$, ..., $G^n_{\mathbf{X}_k}$ are independent. Assume $c_{G^n_{\mathbf{X}_1}}$, ..., $c_{G^n_{\mathbf{X}_k}}$ are valid ϵ-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 random variables. Intermediate nodes first compute their parents' coloring random variables, and then, by using a look-up table, they find corresponding source values of their received colorings. Then, they compute ϵ-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 a source coding. The receiver first performs an entropy decoding 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.

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 $n^i_{\Xi_{ij}}$ in the i-th stage of the network. Since the corresponding source values Ξ_{ij} of its received colorings form an independent set on 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.

$$\forall s_i \in \mathcal{S}(w_i) \Longrightarrow \sum_{j \in s_i} R^i_{\Xi_{ij}} \geq H_{\bigcup_{z \in s_i} G_{\Xi_{iz}}}(\Xi_{is_i}). \tag{2.17}$$

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.

2.5 A Case When Intermediate Nodes Do not Need to Compute

Though the proposed coding scheme in Section 2.4.2 can perform arbitrarily close to the rate lower bound, it may require some computations at intermediate nodes.

Definition 44. *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 $s \in \mathcal{S}(k)$, $H_{\bigcup_{i \in s} G_{X_i}}(X_s) = H_{G_{X_s}}(X_s)$.*

Theorem 45. *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 to perform arbitrarily close to the rate lower bound mentioned in Theorem 43.*

Proof. Consider an intermediate node $n^i_{\Xi_{ij}}$ in the i-th stage of the network whose corresponding source random variables are X_s where $s \in \mathcal{S}(k)$ (In other words, $X_s = \Xi_{ij}$). Since random variables are independent, one can write up rate bounds of Theorem 43 as,

$$\forall s_i \in \mathcal{S}(w_i) \Longrightarrow \sum_{j \in s_i} R^i_{\Xi_{ij}} \geq H_{\bigcup_{z \in s_i} G_{\Xi_{iz}}}(\Xi_{is_i}). \tag{2.18}$$

Now, consider the outgoing link rate of the node $n^i_{\Xi_{ij}}$. If this intermediate node acts like a relay, we have $R^i_{\Xi_{ij}} = H_{\bigcup_{i \in s} G_{X_i}}(X_s)$ (since $X_s = \Xi_{ij}$). If $(f, X_1, ..., X_k)$ is a chain-rule proper set, we can write,

$$\begin{aligned} R^i_{\Xi_{ij}} &= H_{\bigcup_{i\in s} G_{X_i}}(X_s) \\ &= H_{G_{X_s}}(X_s) \\ &= H_{G_{\Xi_{ij}}}(\Xi_{ij}). \end{aligned} \tag{2.19}$$

For any intermediate node $n^i_{\Xi_{ij}}$ where $j \in s_i$ and $s_i \in \mathcal{S}(w_i)$, we can write a similar argument which lead to conditions (2.18). As mentioned in Theorem 43, to perform arbitrarily close to the rate lower bound, this node needs to compress its received information up to the rate $H_{G_{X_s}}(X_s)$. If this node acts as a relay and forward the received information from the previous stage, it would lead to an achievable rate $H_{\bigcup_{i\in s} G_{X_i}}(X_s)$ in the next stage which in general is not equal to $H_{G_{X_s}}(X_s)$. So, this scheme cannot achieve the rate lower bound. However, if for any $s \in \mathcal{S}(k)$, $H_{\bigcup_{i\in s} G_{X_i}}(X_s) = H_{G_{X_s}}(X_s)$, this scheme can perform arbitrarily close to the rate lower bound by having intermediate nodes as relays. □

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

Lemma 46. *Suppose X_1 and X_2 are independent and $f(X_1, X_2)$ is a deterministic function. If for any x_2^1 and x_2^2 in $\mathcal{X}_2$ we have $f(x_1^i, x_2^1) \neq f(x_1^j, x_2^2)$ for any possible i and j, then, (f, X_1, X_2) is a chain-rule proper set.*

Proof. We show that under this condition 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 2-7 which illustrates conditions of this lemma. Under these conditions, since all x_2 in $\mathcal{X}_2$ 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 $\mathcal{X}_2$. These subgraphs are fully connected to each other under conditions of Corollary 46. So, any coloring of this graph can be represented as two colorings

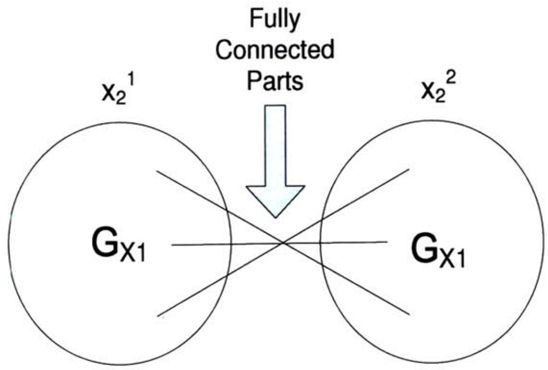


Figure 2-7: An example of G_{X_1,X_2} satisfying conditions of Lemma 46, when $\mathcal{X}_2$ has two members.

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, $H_{G_{X_1},G_{X_2}}(X_1, X_2) = H_{G_{X_1,X_2}}(X_1, X_2)$. □

2.6 Summary of Results

In this chapter, 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., [12]). 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 this 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 random variables called chain-rule proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily close to the rate lower bound.

Chapter 3

Multi-Functional Compression with Side Information

In this chapter, we consider the problem of multi-functional compression with side information. The problem is how we can compress a source X so that the receiver is able to compute some deterministic functions $f_1(X, Y_1)$, ..., $f_m(X, Y_m)$, where Y_i, $1 \leq i \leq m$, are available at the receiver as side information.

Results explained in Chapter 2 only consider the case where the receiver desires to compute one function (m=1). Here, we consider a case in which computations of several functions with different side information random variables and zero distortion are desired at the receiver. Our results do not depend on the fact that all desired functions are in one receiver and one can apply them to the case of having several receivers with different desired functions (i.e., functions are separable). We define a new concept named the *multi-functional graph entropy* which is an extension of the graph entropy defined by Körner in [19]. We show that the minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of the source random variable given side informations. We also propose a coding scheme based on graph colorings to achieve this rate.

In this chapter, after giving the problem statement, our main contributions are explained.

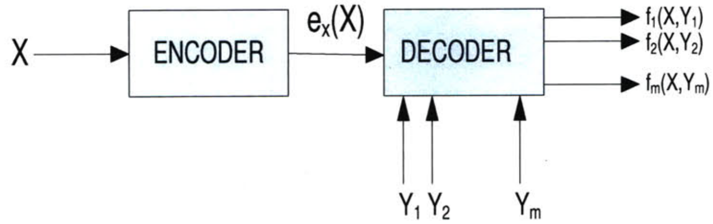


Figure 3-1: A multi-functional compression problem with side information

3.1 Problem Setup

Consider discrete memoryless sources (i.e., $\{X^i\}_{i=1}^{\infty}$ and $\{Y_k^i\}_{i=1}^{\infty}$) and assume that these sources are drawn from finite sets $\mathcal{X}$ and $\mathcal{Y}_k$ with a joint distribution $p_k(x, y_k)$. We express n-sequence of these random variables as $\mathbf{X} = \{X^i\}_{i=l}^{i=l+n-1}$ and $\mathbf{Y_k} = \{Y_k\}_{i=l}^{i=l+n-1}$ with joint probability distribution $p_k(\mathbf{x}, \mathbf{y}_k)$. Assume $k = 1, ..., m$ (we have m random variables as side information at the receiver).

The receiver wants to compute m deterministic functions $f_k : \mathcal{X} \times \mathcal{Y}_k \to \mathcal{Z}_k$ or $f_k : \mathcal{X}^n \times \mathcal{Y}_k^n \to \mathcal{Z}_k^n$, its vector extension. Without loss of generality, we assume $l = 1$ and to simplify the notations, n will be implied by the context. So, we have one encoder en_x and m decoders $r_1, ..., r_m$ (one for each function and its corresponding side information). Encoder en_x maps

$$en_x : \mathcal{X}^n \to \{1, ..., 2^{nR_x}\}, \tag{3.1}$$

and, each decoder r_k maps

$$r_k : \{1, ..., 2^{nR_x}\} \times \{1, ..., 2^n\} \to \mathcal{Z}_k^n, \tag{3.2}$$

The probability of error in decoding f_k is

$$P_{e_k}^n = Pr[(\mathbf{x}, \mathbf{y}_k) : f_k(\mathbf{x}, \mathbf{y}_k) \neq r_k(e_x(\mathbf{x}), \mathbf{y}_k)] \tag{3.3}$$

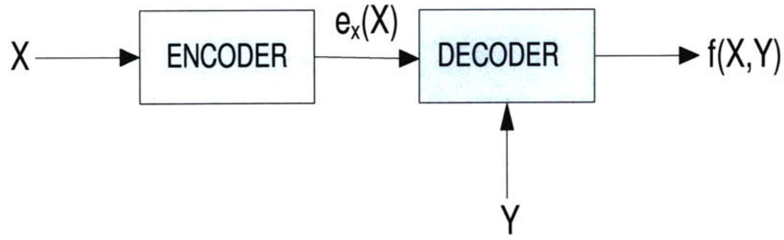


Figure 3-2: A functional compression problem with side information with one desired function at the receiver

and, total probability of error is

$$P_e^n = 1 - \prod_k (1 - P_{e_k}^n). \tag{3.4}$$

In other words, we declare an error when we have an error in the computation of at least one function. A rate R_X is achievable if $P_e^n \to 0$ when $n \to \infty$. Our aim here is to find the minimum achievable rate.

3.2 Main Results

Prior work in the functional compression problem consider a case when the computation of a function is desired at the receiver (m=1). In this chapter, we consider a case when computations of several functions are desired at the receiver. As an example, consider the network shown in Figure 3-1. The receiver wants to compute m functions with different side information random variables. We want to compute the minimum achievable rate for this case. Note that our results do not depend on the fact that all functions are desired in one receiver. In other words, one can apply them to the case of having several receivers with different desired functions (functions are separable).

First, let us consider the case $m = 2$ (i.e., the receiver has two different functions to compute). Then, we extend results to the case of arbitrary m. In this problem, the receiver wants to compute two deterministic functions $f_1(X, Y_1)$ and $f_2(X, Y_2)$; while, Y_1 and Y_2 are available at the receiver as side information. We want to find

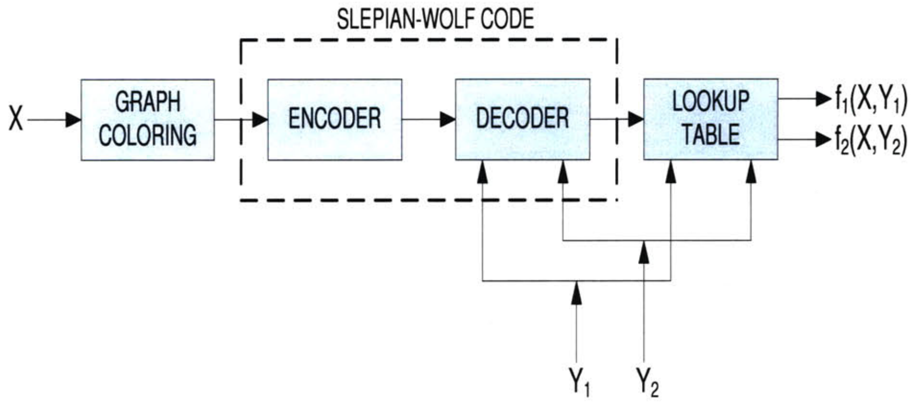


Figure 3-3: Source coding scheme for multi-functional compression problem with side information

the minimum achievable rate in which the source node X may encode its information so that the decoder is able to compute its desired functions.

Let us call $G_{X,f_1} = (V, E_{f_1})$ the characteristic graph of X with respect to Y_1, $p_1(x, y_1)$ and $f_1(X, Y_1)$, and $G_{X,f_2} = (V, E_{f_2})$ the characteristic graph of X with respect to Y_2, $p_2(x, y_2)$ and $f_2(X, Y_2)$. Now, define $G_{X,f_1,f_2} = (V, E_{f_1 f_2})$ such that $E_{f_1 f_2} = E_{f_1} \bigcup E_{f_2}$. In other words, G_{X,f_1,f_2} is the *or* function of G_{X,f_1} and G_{X,f_2}. We call G_{X,f_1,f_2} the *multi-functional characteristic graph* of X.

When we deal with one function, we drop f from the notations (as in Chapter 2).

Definition 47. *The multi-functional characteristic graph $G_{X,f_1,f_2} = (V, E_{f_1 f_2})$ of X with respect to Y_1, Y_2, $p_1(x, y_1)$, $p_2(x, y_2)$,and $f_1(x, y_1)$,$f_2(x, y_2)$ is defined as follows:*

$V = \mathcal{X}$ and an edge $(x_1, x_2) \in \mathcal{X}^2$ is in $E_{f_1 f_2}$ iff there exists a $y_1 \in \mathcal{Y}_1$ such that $p_1(x_1, y_1)p_1(x_2, y_1) > 0$ and $f_1(x_1, y_1) \neq f_1(x_2, y_1)$ or there exists a $y_2 \in \mathcal{Y}_2$ such that $p_2(x_1, y_2)p_2(x_2, y_2) > 0$ and $f_2(x_1, y_2) \neq f_2(x_2, y_2)$.

Similarly to Definition 20, we define the *multi-functional graph entropy* as follows.

Theorem 48.

$$H_{G_{X,f_1,f_2}}(X) = \lim_{n\to\infty} \frac{1}{n} H^{\mathcal{X}}_{G^n_{X,f_1,f_2}}(\mathbf{X}). \tag{3.5}$$

The *conditional multi-functional graph entropy* can be defined similarly to Definition 21 as follows,

Theorem 49.

$$H_{G_{X,f_1,f_2}}(X|Y) = \lim_{n\to\infty} \frac{1}{n} H^{\chi}_{G^n_{X,f_1,f_2}}(\mathbf{X}|\mathbf{Y}). \tag{3.6}$$

where G^n_{X,f_1,f_2} is the n-th power of G_{X,f_1,f_2}. Now, we can state the following theorem.

Theorem 50. *$H_{G_{X,f_1,f_2}}(X|Y_1,Y_2)$ is the minimum achievable rate for the network shown in Figure 3-1 when $m = 2$.*

Proof. To show this, we first show that $R_X \geq H_{G_{X,f_1,f_2}}(X|Y_1,Y_2)$ is an achievable rate (achievability), and no one can outperform this rate (converse). To do this, first, we show that any valid coloring of G^n_{X,f_1,f_2} for any n leads to an achievable encoding-decoding scheme for this problem (achievability). Then, we show that every achievable encoding-decoding scheme performing on blocks with length n, induces a valid coloring of G^n_{X,f_1,f_2} (converse).

Achievablity: According to [11], any valid coloring of G^n_{X,f_1} leads to successfully computing of $f_1(\mathbf{X},\mathbf{Y}_1)$ at the receiver. If $c_{G^n_{X,f_1}}$ is a valid coloring of G^n_{X,f_1}, there exists a function r_1 such that $r_1(c_{G^n_{X,f_1}}(\mathbf{X}),\mathbf{Y}_1) = f_1(\mathbf{X},\mathbf{Y}_1)$, with high probability. A similar argument holds for G_{X,f_2}. Now, assume that $c_{G^n_{X,f_1,f_2}}$ is a valid coloring of G^n_{X,f_1,f_2}. Since, $E^n_{f_1} \subseteq E^n_{f_1 f_2}$ and $E^n_{f_2} \subseteq E^n_{f_1 f_2}$, any valid coloring of G^n_{X,f_1,f_2} induces valid colorings for G^n_{X,f_1} and G^n_{X,f_2}. Thus, any valid coloring of G^n_{X,f_1,f_2} leads to successful computation of $f_1(\mathbf{X},\mathbf{Y}_1)$ and $f_2(\mathbf{X},\mathbf{Y}_2)$ at the receiver. So, $c_{G^n_{X,f_1,f_2}}$ leads to an achievable encoding scheme (i.e. there exist two functions r_1 and r_2 such that $r_1(c_{G^n_{X,f_1,f_2}}(\mathbf{X}),\mathbf{Y}_1) = f_1(\mathbf{X},\mathbf{Y}_1)$ and $r_2(c_{G^n_{X,f_1,f_2}}(\mathbf{X}),\mathbf{Y}_2) = f_2(\mathbf{X},\mathbf{Y}_2)$), with high probability.

For the case of having the identity function at the receiver (i.e., the receiver wants the whole information of the source node), Slepian and Wolf proposed a technique in [24] to compress source random variable X up to the rate $H(X|Y)$ when Y is available at the receiver. Here, one can perform Slepian-Wolf compression technique on the minimum entropy coloring of large enough power graph and get the given bound.

Converse: Now, we show that any achievable encoding-decoding scheme performing on blocks with length n, induces a valid coloring of G^n_{X,f_1,f_2}. In other words, we want to show that if there exist functions en, r_1 and r_2 such that $r_1(en(\mathbf{X}),\mathbf{Y}_1) = f_1(\mathbf{X},\mathbf{Y}_1)$ and $r_2(en(\mathbf{X}),\mathbf{Y}_2) = f_2(\mathbf{X},\mathbf{Y}_2)$, $en(\mathbf{X})$ is a valid coloring of G^n_{X,f_1,f_2}.

Let us proceed by contradiction. If $en(\mathbf{X})$ were not a valid coloring of G^n_{X,f_1,f_2}, there must be some edge in $E^n_{f_1 f_2}$ with both vertices with the same color. Let us call these two vertices $\mathbf{x}_1$ and $\mathbf{x}_2$ which take the same values (i.e., $en(\mathbf{x}_1) = en(\mathbf{x}_2)$), but also are connected. Since they are connected to each other, by definition of G^n_{X,f_1,f_2}, there exists a $\mathbf{y}_1 \in \mathcal{Y}_1$ such that $p_1(\mathbf{x}_1,\mathbf{y}_1)p_1(\mathbf{x}_2,\mathbf{y}_1) > 0$ and $f_1(\mathbf{x}_1,\mathbf{y}_1) \neq f_1(\mathbf{x}_2,\mathbf{y}_1)$ or there exists a $\mathbf{y}_2 \in \mathcal{Y}_2$ such that $p_2(\mathbf{x}_1,\mathbf{y}_2)p_2(\mathbf{x}_2,\mathbf{y}_2) > 0$ and $f_2(\mathbf{x}_1,\mathbf{y}_2) \neq f_1(\mathbf{x}_2,\mathbf{y}_2)$. Without loss of generality, assume that the first case occurs. Thus, we have a $\mathbf{y}_1 \in \mathcal{Y}_1$ such that $p_1(\mathbf{x}_1,\mathbf{y}_1)p_1(\mathbf{x}_2,\mathbf{y}_1) > 0$ and $f_1(\mathbf{x}_1,\mathbf{y}_1) \neq f_1(\mathbf{x}_2,\mathbf{y}_1)$. So, $r_1(en(\mathbf{x}_1),\mathbf{y}_1) \neq r_1(en(\mathbf{x}_2),\mathbf{y}_1)$. Since $en(\mathbf{x}_1) = en(\mathbf{x}_2)$, then, $r_1(en(\mathbf{x}_1),\mathbf{y}_1) \neq r_1(en(\mathbf{x}_1),\mathbf{y}_1)$. But, it is not possible. Thus, our contradiction assumption was not true. In other words, any achievable encoding-decoding scheme for this problem induces a valid coloring of G^n_{X,f_1,f_2} and it completes the proof. □

Now, let us consider the network shown in Figure 3-1 where the receiver wishes to compute m deterministic functions of source information having some side information.

Theorem 51. *$H_{G_{X,f_1,\ldots,f_m}}(X|Y_1,\ldots,Y_m)$ is the minimum achievable rate for the network shown in Figure 3-1.*

The argument here is similar to the case $m = 2$ mentioned in Theorem 50. So, we only sketch the proof. To show this, one may first show that any colorings of multi-functional characteristic graph of X with respect to desired functions ($G_{X,f_1,\ldots,f_m}$) leads to an achievable scheme. Then, showing that any achievable encoding-decoding scheme induces a coloring on $G_{X,f_1,\ldots,f_m}$ completes the proof.

3.3 Summary of Results

In this chapter, we considered the problem of multi-functional compression with side information. The problem is how we can compress a source X so that the receiver is able to compute some deterministic functions $f_1(X, Y_1)$, ..., $f_m(X, Y_m)$, where Y_i is available at the receiver as the side information.

In particular, we considered a case when the receiver wants to compute several deterministic functions with different side information random variables and zero distortion. Our results do not depend on the fact that all functions are desired in one receiver and one can apply them to the case of having several receivers with different desired functions (i.e., functions are separable). We defined a new concept named the *multi-functional graph entropy* which is an extension of the graph entropy defined by Körner in [19]. We showed that minimum achievable rate for this problem is equal to the conditional multi-functional graph entropy of the source random variable given side informations. We also proposed a coding scheme based on graph colorings to achieve this rate.

Chapter 4

Polynomial Time Cases for Finding the Minimum Entropy Coloring of a Characteristic Graph

In this chapter, we consider the problem of finding the minimum entropy coloring of a characteristic graph. This problem arises in the functional compression problem where the computation of a function of sources is desired at the receiver. We considered some aspects of this problem in Chapters 2 and 3 and proposed some coding schemes. In those proposed coding scheme, one needs to compute the 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. [6]) . However, in this chapter, 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. Therefore, the minimum entropy coloring can be solved in polynomial time. In another case, we show that if f is a quantization function, this problem is also tractable. We also consider this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we show that finding the minimum entropy coloring is heuristically equivalent to finding the maximum independent set of a graph. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to solve approximately the maximum independent set problem.

We proceed this chapter by stating the problem setup. Then, we explain our contributions to this problem.

4.1 Problem Setup

In some problems such as the functional compression problem, we need to find a coloring random variable of a characteristic graph which minimizes the entropy. The problem is how to compute such a coloring for a given characteristic graph. In other words, this problem can be expressed as follows. Given a characteristic graph G_{X_1} (or, its n-th power, $G^n_{\mathbf{X}_1}$), one can assign different colors to its vertices. Suppose $C_{G_{X_1}}$ is the collection of all valid colorings of this graph, G_{X_1}. Among these colorings, one which minimizes the entropy of the coloring random variable is called the minimum-entropy coloring, and we refer to it by $c^{min}_{G_{X_1}}$. In other words,

$$c^{min}_{G_{X_1}} = \arg\min_{c_{G_{X_1}} \in C_{G_{X_1}}} H(c_{G_{X_1}}). \tag{4.1}$$

The problem is how to compute $c^{min}_{G_{X_1}}$ given G_{X_1}.

4.2 Main Results

In this section, we consider the problem of finding the minimum entropy coloring of a characteristic graph. The problem is how to compute a coloring of a characteristic graph which minimizes the entropy. In general, finding $c^{min}_{G_{X_1}}$ is an NP-hard problem ([6]). 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,

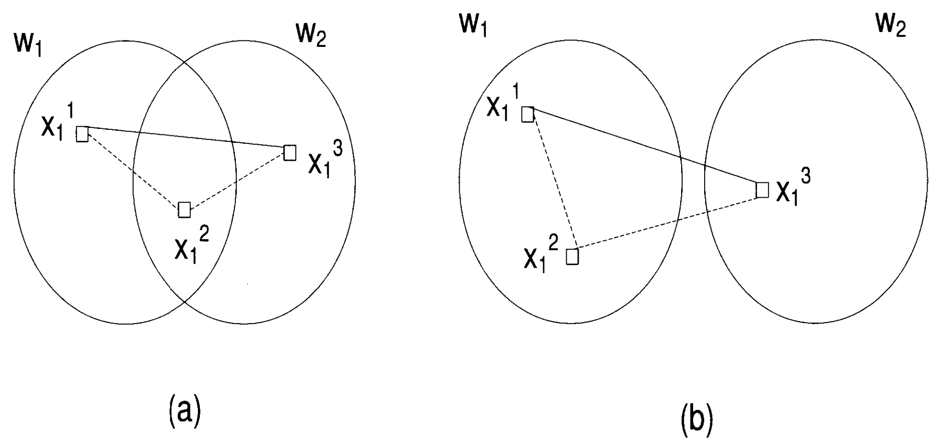


Figure 4-1: Having non-zero joint probability distribution, a) maximal independent sets cannot overlap with each other (this figure is to depict the contradiction) b) maximal independent sets should be fully connected to each other. In this figure, a solid line represents a connection, and a dashed line means no connection exists.

by having a non-zero joint probability condition on random variables' distributions, for any desired function f, finding $c^{min}_{G_{X_1}}$ can be solved in polynomial time. In another case, we show that, if f is a quantization function, this problem is also tractable. We also consider this problem in a general case. We show that by using Huffman or Lempel-Ziv coding notions, finding the minimum entropy coloring is heuristically equivalent to finding the maximum independent set of a graph.

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.

4.2.1 Non-Zero Joint Probability Distribution Condition

Consider the network shown in Figure 1-2-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 Chapter 2, we showed that in an achievable coding scheme, one needs to compute minimum entropy colorings of characteristic graphs. The question is how source nodes can compute minimum entropy colorings of their characteristic graphs G_{X_1} and G_{X_2} (or, similarly the minimum entropy colorings of $G^n_{\mathbf{X}_1}$ and $G^n_{\mathbf{X}_2}$, for some n). For an arbitrary graph, this problem is NP-hard ([6]). However, in certain cases, depending on the probability distribution or the desired function, the characteristic graph has some special structure which leads to a tractable scheme to find the minimum entropy coloring. In this section, we consider the effect of the probability distribution.

Theorem 52. *Suppose for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, $p(x_1, x_2) > 0$. Then, maximal independent sets of the characteristic graph G_{X_1} (and, its n-th power $G^n_{\mathbf{X}_1}$, 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.*

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 4-1-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. In other words, 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_1^2 and x_1^3. 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_1^1 and x_1^2, for any $x_2^1 \in \mathcal{X}_2$, $p(x_1^1, x_2^1)p(x_1^2, x_2^1) > 0$, and $f(x_1^1, x_2^1) = f(x_1^2, x_2^1)$. A similar argument can be expressed for x_1^2 and x_1^3. In other words, for any $x_2^1 \in \mathcal{X}_2$, $p(x_1^2, x_2^1)p(x_1^3, x_2^1) > 0$, and $f(x_1^2, x_2^1) = f(x_1^3, x_2^1)$. Thus, for all $x_2^1 \in \mathcal{X}_2$, $p(x_1^1, x_2^1)p(x_1^3, x_2^1) > 0$, and $f(x_1^1, x_2^1) = f(x_1^3, x_2^1)$. However, since x_1^1 and x_1^3 are connected to each other, there should exist a $x_2^1 \in \mathcal{X}_2$ such that $f(x_1^1, x_2^1) \neq f(x_1^3, x_2^1)$ 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 4-1-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_1^3) which is connected to one of elements in w_1 (call it x_1^1) and is not connected to another element of w_1 (call it x_1^2). By using a similar discussion to the one in the previous paragraph, we may show that it is not possible. Thus, x_1^3 should be connected to x_1^1. Therefore, if for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, $p(x_1, x_2) > 0$, 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 $\mathbf{x}_1^1$, $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$ are some typical sequences in $\mathcal{X}_1^n$. If $\mathbf{x}_1^1$ is not connected to $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$, it is not possible to have $\mathbf{x}_1^2$ and $\mathbf{x}_1^3$ connected. Therefore, one can apply a similar argument to prove the theorem for $G_{\mathbf{X}_1}^n$, for some n. This completes the proof. $\square$

One may note if the characteristic graph satisfying conditions of Theorem 52 is sparse, its power graph would also be sparse (a sparse graph with m vertices is a graph whose number of edges is much smaller than $\frac{m(m-1)}{2}$).

One should notice that the condition $p(x_1, x_2) > 0$, for all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$, is a necessary condition for Theorem 52. In order to illustrate this, consider Figure 4-2. In this example, x_1^1, x_1^2 and x_1^3 are in $\mathcal{X}_1$, and x_2^1, x_2^2 and x_2^3 are in $\mathcal{X}_2$. Suppose $p(x_1^2, x_2^2) = 0$. By considering the value of function f at these points depicted in the figure, one can see that, in G_{X_1}, x_1^2 is not connected to x_1^1 and x_1^3. However, x_1^1 and x_1^3 are connected to each other. Thus, Theorem 52 does not hold here.

It is also worthwhile to notice that the condition used in Theorem 52 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.

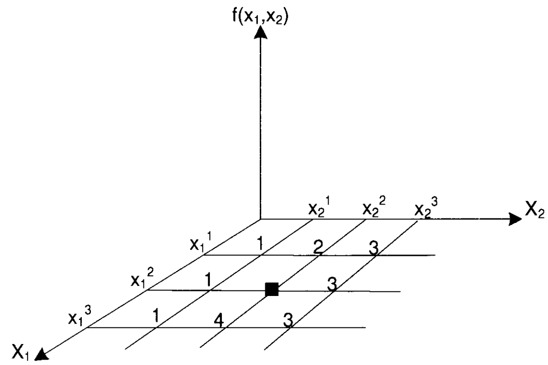


Figure 4-2: Having non-zero joint probability condition is necessary for Theorem 52. A dark square represents a zero probability point.

4.2.2 Quantization Functions

In Section 4.2.1, we introduced a condition on the joint probability distribution of random variables which leads to a specific structure of the characteristic graph such that finding the minimum entropy coloring is not NP-hard. In this section, we consider some special functions which lead to some graph structures so that one can easily find the minimum entropy coloring.

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 such that each rectangle corresponds to a different value of that function. Sides of these rectangles are parallel to the plane axes. Figure 4-3-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 $\mathcal{X}_1$ and $\mathcal{X}_2$. For example, in Figure 4-3-b, one of the function regions is distinguished by the shaded area.

Definition 53. *Consider two function regions $\mathcal{X}_1^1 \times \mathcal{X}_2^1$ and $\mathcal{X}_1^2 \times \mathcal{X}_2^2$. If for any $x_1^1 \in \mathcal{X}_1^1$ and $x_1^2 \in \mathcal{X}_1^2$, there exist x_2 such that $p(x_1^1, x_2)p(x_1^2, x_2) > 0$ and $f(x_1^1, x_2) \neq f(x_1^2, x_2)$, we say these two function regions are pairwise X_1-proper.*

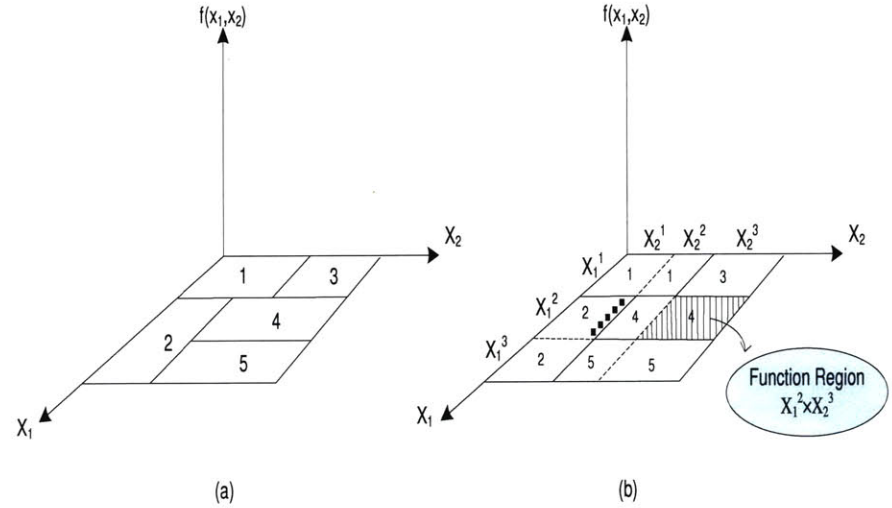


Figure 4-3: a) A quantization function. Function values are depicted in the figure on each rectangle. b) By extending sides of rectangles, the plane is covered by some function regions.

Theorem 54. *Consider a quantization function f such that its function regions are pairwise X_1-proper. Then, G_{X_1} (and $G^n_{\mathbf{X}_1}$, 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.*

Proof. We first prove it for G_{X_1}. Suppose $\mathcal{X}_1^1 \times \mathcal{X}_2^1$, and $\mathcal{X}_1^2 \times \mathcal{X}_2^2$ are two X_1-proper function regions of a quantization function f, where $\mathcal{X}_1^1 \neq \mathcal{X}_1^2$. We show that $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ are two non-overlapping fully-connected maximal independent sets. By definition, $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ are two non-equal partition sets of $\mathcal{X}_1$. 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 $\mathcal{X}_1^1$. If this partition set of $\mathcal{X}_1$ has only one element, this is a trivial case. So, suppose x_1^1 and x_1^2 are two elements in $\mathcal{X}_1^1$. By definition of function regions, one can see that, for any $x_2 \in \mathcal{X}_2$ such that $p(x_1^1, x_2)p(x_1^2, x_2) > 0$, then $f(x_1^1, x_2) = f(x_1^2, x_2)$. Thus, these two vertices are not connected to each other. Now, suppose x_1^3 is an element in $\mathcal{X}_1^2$. Since these function regions are X_1-proper, there should exist at least one $x_2 \in \mathcal{X}_2$, such that $p(x_1^1, x_2)p(x_1^3, x_2) > 0$, and $f(x_1^1, x_2) \neq f(x_1^3, x_2)$. Thus, x_1^1 and x_1^3 are connected to each other. Therefore, $\mathcal{X}_1^1$ and $\mathcal{X}_1^2$ 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 $G_{\mathbf{X}_1}^n$, for any n, is similar to the one mentioned in Theorem 52. This completes the proof. □

It is worthwhile to mention that without X_1-proper condition of Theorem 54, assigning different colors to different partitions still leads to an achievable coloring scheme. However, it is not necessarily the minimum entropy coloring. In other words, without this condition, maximal independent sets may overlap.

Corollary 55. *If a function f is strictly increasing (or, decreasing) with respect to X_1, and $p(x_1, x_2) \neq 0$, for all $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, then, G_{X_1} (and, $G_{\mathbf{X}_1}^n$ for any n) would be a complete graph.*

Under conditions of Corollary 55, functional compression does not give us any gain, because, in a complete graph, one should assign different colors to different vertices. Traditional compression in which f is the identity function is a special case of Corollary 55.

4.2.3 Minimum Entropy Coloring for an Arbitrary Graph

Finding the minimum entropy coloring of an arbitrary graph (called the chromatic entropy) is NP-hard ([6]). Specially, [6] showed that, even finding a coloring whose entropy is within $(\frac{1}{7} - \epsilon) \log m$ of its chromatic entropy is NP-hard, for any $\epsilon > 0$, where m is the number of vertices of the graph. That is a reason we have introduced some special structures on the characteristic graph in order to have some tractable and practical schemes to find the minimum entropy coloring. While cases investigated in Sections 4.2.1 and 4.2.2 cover certain practical cases, in this part, we want to consider this problem without assuming any special structure of the graph. In particular, we show that, by using a notion of an empirical Huffman coding scheme or a Lempel-Ziv coding scheme, one can heuristically have an equivalence between the minimum-entropy coloring problem and the maximum independent set problem. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to solve the maximum independent set problem [8].

Suppose G_{X_1} is the characteristic graph of X_1. Without loss of generality, in this section, we consider $n = 1$. All discussions can be extended to $G^n_{\mathbf{X}_1}$, for any n. Suppose $p(x_1)$ is the probability distribution of X_1. Let us define the adjacency matrix $A = [a_{ij}]$ for this graph as follows: $a_{ij} = 1$ when x_1^i and x_1^j are connected to each other in G_{X_1}, otherwise, $a_{ij} = 0$. One can see that the adjacency matrix is symmetric, with all zeros in its diagonal. A *one* in this matrix means that its corresponding vertices should be assigned to different colors.

Let us define a permutation matrix P with the same size of A. This matrix has only a *one* in each of its rows and columns. The matrix PAP^t would be a matrix such that rows and columns of A are reordered simultaneously, with respect to this permutation matrix P. For any valid coloring, there exists a permutation matrix P, such that PAP^t has zero square matrixes on its diagonal. This reordering is such that, vertices with the same color are adjacent to each other in PAP^t. Each of these zero square matrixes on the diagonal of PAP^t represents a maximal independent set, or equivalently a color class. One can see that there exists a bijective mapping between any valid coloring and any permutation matrix P which leads to have some zero square matrixes on the diagonal of PAP^t.

Example 56. *For an example, consider a coloring of a graph depicted in Figure 1-9. This coloring leads to the following PAP^t matrix.*

$$PAP^t = \begin{pmatrix} \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & D_1 & D_2 \\ D_1 & \begin{matrix} 0 & 0 \\ 0 & 0 \end{matrix} & D_3 \\ D_2 & D_3 & 0 \end{pmatrix} \tag{4.2}$$

where D_i, $i = 1, 2, 3$ are non-zero matrixes. Each of zero square matrixes on the diagonal represents a color class, or a maximal independent set of this graph. The permutation matrix P in this case is,

$$P = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.3}$$

Now, we want to take the probability distribution into account. To do this, we repeat each vertex x_1^i in the adjacency matrix, n_i times, such that $\frac{p(x_1^i)}{p(x_1^j)} = \frac{n_i}{n_j}$, for any valid i and j. We call the achieved matrix, the weighted adjacency matrix and show it by A_w. The above argument about the permutation remains the same. In other words, any valid coloring can be represented by a permutation matrix P such that PA_wP^t has some zero square matrixes on its diagonal. Since we represent the probability distribution of each vertex as its number of repetitions in A_w, the proportional sizes of zero square matrixes on the diagonal of PA_wP^t represent the corresponding probability of that color class. In other words, a color class of a larger zero square matrix has more probability than a color class with a smaller zero square matrix.

Now, one can heuristically use Huffman coding technique to find a coloring (or its corresponding permutation matrix) to minimize the entropy. To do this, we first find a permutation matrix which leads to the largest zero square matrix on the diagonal of PA_wP^t. Then, we assign a color to that class, and eliminate its corresponding rows and columns. We repeat this algorithm till all vertices are assigned to some colors. One can see that, finding the largest zero square matrix on the diagonal of PA_wP^t is equivalent to finding the maximum independent set of a graph. One should notice that, it is a heuristic algorithm, and does not necessarily reach to the minimum entropy coloring. The other point is that, here, we have assumed that the probability distribution of X_1 is known. If we do not know this probability distribution, one can use an empirical distribution, instead of the actual distribution. In that case, using a Lempel-Ziv coding notion instead of Huffman coding leads to a similar algorithm.

4.3 Summary of Results

In this chapter, we considered the problem of finding the minimum entropy coloring of a characteristic graph. This problem has been raised in the functional compression problem described in Chapters 2 and 3. In general, finding this coloring is an NP-hard problem ([6]).

Here, first we considered this problem under some conditions which make it to be solvable in polynomial time. We showed 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 showed that, by having a non-zero joint probability condition on random variables' distributions, for any desired function f, finding the minimum entropy coloring can be solved in polynomial time. In another case, we showed that, if f belongs to a natural and intuitive family of quantization functions, this problem is also tractable. Finally, we considered this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we showed that finding the minimum entropy coloring is heuristically equivalent to finding the maximum independent set of a graph.

Chapter 5

Feedback in Functional Compression

In this section, we investigate the effect of having feedback on the rate-region of the functional compression problem. If the function at the receiver is the identity function, this problem is the Slepian-Wolf compression with feedback. For this case, having feedback does not improve rate bounds. However, for a general desired function at the receiver, having feedback may improve rate bounds of the case without feedback.

5.1 Main Results

Consider a distributed functional compression problem with two sources and a receiver depicted in Figure 5-1-a. This network does not have feedback. In Chapter 2, we derived a rate-region for this network. In this section, we consider the effect of having feedback on the rate-region of the network. For simplicity, we consider a simple distributed network topology with two sources. However, one can extend all discussions to more complicated networks considered in Chapters 2 and 3.

Consider the network shown in Figure 5-1-b. In this network, sources can get some information from the receiver. If the desired function at the receiver is the identity function, this problem is the SW compression with feedback. For this case, having feedback does not change the rate region [24]. However, when we have a

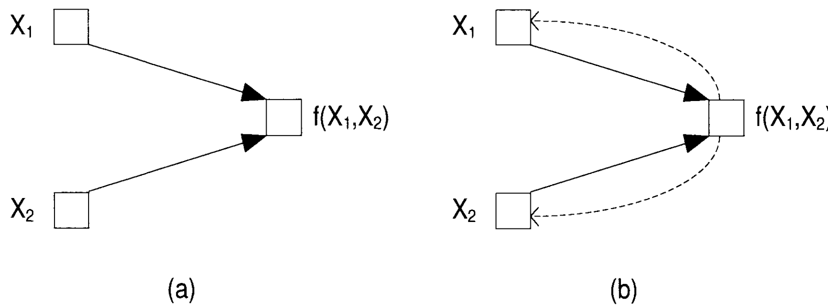


Figure 5-1: A distributed functional compression network a) without feedback b) with feedback.

general function at the receiver, by having feedback, one may improve rate bounds of Theorem 40.

Theorem 57. *Having feedback may improve rate bounds of Theorem 40.*

Proof. Consider a network without feedback depicted in Figure 5-1-a. In Chapter 2, we showed an achievable scheme where sources send their minimum entropy colorings of high probability subgraphs of their characteristic graphs satisfying C.C.C., followed by SW compression. This scheme performs arbitrarily closely to rate bounds derived in Theorem 40. Now, we seek to show that, in some cases, by having feedback, one can outperform these bounds. Consider source random variables X_1 and X_2 with characteristic graphs G_{X_1} and G_{X_2}, respectively. Suppose $S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$ and $S_{c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$ are two sets of joint colorings of source random variables defined as follows,

$$
\begin{aligned}
S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}} &= \arg\min_{(c_{G^n_{\mathbf{X}_1}},c_{G^n_{\mathbf{X}_2}})\in C_{G^n_{\mathbf{X}_1}}\times C_{G^n_{\mathbf{X}_2}}} \frac{1}{n}H(c_{G^n_{\mathbf{X}_1}},c_{G^n_{\mathbf{X}_2}}) \\
S_{c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}} &= \arg\min_{\substack{(c_{G^n_{\mathbf{X}_1}},c_{G^n_{\mathbf{X}_2}})\in C_{G^n_{\mathbf{X}_1}}\times C_{G^n_{\mathbf{X}_2}} \\ \text{satisfying C.C.C.}}} \frac{1}{n}H(c_{G^n_{\mathbf{X}_1}},c_{G^n_{\mathbf{X}_2}}).
\end{aligned}
\tag{5.1}
$$

Now, consider the case when $S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}} \cap S_{c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}} = \emptyset$, i.e., suppose any $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}} \in S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$ does not satisfy C.C.C. Thus, C.C.C. restricts the link sum rates of any

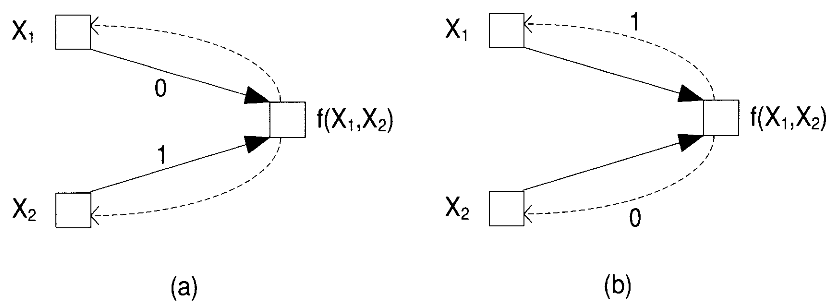


Figure 5-2: An example of the proposed feedback scheme. a) Since $\mathbf{x}_1 \notin A_{X_1}$, source X_1 sends 0. Since $\mathbf{x}_2 \in A_{X_2}$, source X_2 sends 1. b) The receiver forward signaling bits to the sources. Then, sources can use the coloring scheme $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$.

achievable scheme, because $H(c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}) < H(c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}})$ for any $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}} \in S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$ and $c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}} \in S_{c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$.

Choose any two joint colorings $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}} \in S_{c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$ and $c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}} \in S_{c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}}$. Suppose set A contains all points $(\mathbf{x}_1, \mathbf{x}_2)$ such that their corresponding colors in the joint-coloring class of $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$ do not satisfy C.C.C. Now, we propose a coding scheme with feedback which can outperform rate bounds of the case without having feedback. If sources know whether or not they have some sequences in A, they can switch between $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$ and $c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$ in their coding scheme with feedback. Since $H(c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}) < H(c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}})$, this approach outperforms the one without feedback in terms of rates. In the following, we present a possible feedback scheme.

Before sending each sequence, sources first check that if their sequences belong to A or not. To do this, say A_{X_1} is the set of all $\mathbf{x}_1$ where there exists a $\mathbf{x}_2$ such that $(\mathbf{x}_1, \mathbf{x}_2) \in A$. A_{X_2} is defined similarly. One can see that $A \subseteq A_{X_1} \times A_{X_2}$. So, instead of checking if a sequence is in A or not, sources check that if the sequence belong to $A_{X_1} \times A_{X_2}$ or not by exchanging some information. In order to do this, source X_1 sends a one to the receiver when $\mathbf{x}_1 \in A_{X_1}$. Otherwise, it sends a zero. Source X_2 uses a similar scheme. The receiver exchanges these bits using feedback channels. When a source sends a one, and receives a one from its feedback channel, it uses $c'_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$ as its joint coloring. Otherwise, it uses $c^{min}_{G^n_{\mathbf{X}_1},G^n_{\mathbf{X}_2}}$ in its coding scheme. Depending on which joint coloring scheme has been used by sources, the receiver uses a corresponding look-up table to compute the desired function. Hence, this scheme is achievable. An example of this scheme is depicted in Figure 5-2.

Since the length of sequences is arbitrarily large, one can ignore these four extra signaling bits in rate computations. If we did not have feedback, according to Theorem 40,

$$R_{X_1} + R_{X_2} \geq \frac{1}{n} H(c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}). \tag{5.2}$$

Say $P_a = Pr[(\mathbf{x}_1, \mathbf{x}_2) \in A_{X_1} \times A_{X_2}]$. Thus, for the proposed coding scheme with feedback, we have,

$$R^f_{X_1} + R^f_{X_2} \geq \frac{1}{n}[P_a H(c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}) + (1 - P_a) H(c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}})] \tag{5.3}$$

where $R^f_{X_i}$ is the rate of link X_i with feedback. Thus,

$$[R^f_{X_1} + R^f_{X_2}] - [R_{X_1} + R_{X_2}] \geq \frac{1}{n}(1 - P_a)[H(c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}) - H(c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}})]. \tag{5.4}$$

The right-hand side of (5.4) represents a gain in link sum rates one can achieve by having feedback. When, $P_a \neq 1$ and $c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}} \neq c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$, this is strictly positive which means the proposed coding scheme with feedback outperforms the one without having feedback in terms of rate bounds. For the identity function at the receiver, $c^{min}_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}} = c'_{G^n_{\mathbf{X}_1}, G^n_{\mathbf{X}_2}}$, and the proposed coding scheme with feedback does not improve rate bounds. Note that, for the identity function at the receiver, the SW compression can perform arbitrarily closely to min-cut max-flow bounds. □

5.2 Summary of Results

In this section, we investigated the effect of having feedback on the rate-region of a distributed functional compression problem. Particularly, we showed that for some functions in which the minimum entropy colorings of sources do not satisfy C.C.C., by having feedback, one may outperform rate bounds of the case without feedback. However, if the function at the receiver is the identity function, this problem is the

Slepian-Wolf compression with feedback for which having feedback does not improve rate bounds. In a general network, for cases where the minimum entropy colorings of sources satisfy C.C.C., it is not known whether or not feedback can improve rate bounds.

Chapter 6

A Rate-Distortion Region for Distributed Functional Compression

In this chapter, 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 those sources can be computed up to a distortion level at the receiver. In this chapter, we derive a rate-distortion region for a network with two transmitters and a receiver. All discussions can be extended to more general networks considered in Chapters 2 and 3.

A recent result is presented in [10] which computes a rate-distortion region for the side information problem (a case here one of sources is available at the receiver). The proposed result in [10] gives a characterization of Yamamoto's rate distortion function [28] in terms of a reconstruction function. Here, we extend it to the distributed functional compression problem. In this case, we compute a rate-distortion region and then, propose a practical coding scheme with a non-trivial performance guarantee.

In the rest of this chapter, first we express problem statement and previous results. Then, we explain our contributions in this problem.

6.1 Problem Setup

Consider two discrete memoryless sources (i.e.,$\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$) and assume these sources are drawn from finite sets X and Y with a joint distribution $p(x,y)$. We express a n-sequence of these random variables as $\mathbf{X} = \{X_i\}_{i=l}^{i=l+n-1}$ and $\mathbf{Y} = \{Y_i\}_{i=l}^{i=l+n-1}$ with joint probability distribution $p(\mathbf{x},\mathbf{y})$. To simplify the notations, n and l will be implied by the context. The receiver wants to compute a deterministic function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ or $f : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$, its vector extension with distortion D with respect to a given distortion function $d : \mathcal{Z} \times \mathcal{Z} \to [0,\infty)$. A vector extension of the distortion function is defined as follows:

$$d(\mathbf{z}_1,\mathbf{z}_2) = \frac{1}{n}\sum_{i=1}^{n} d(z_{1i}, z_{2i}) \tag{6.1}$$

in which $\mathbf{z}_1,\mathbf{z}_2 \in \mathcal{Z}^n$. As in [27], 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., $d(\mathbf{z}_1,\mathbf{z}_2) = 0$ if and only if $\mathbf{z}_1 = \mathbf{z}_2$).

Consider the network depicted in Figure 6-1-a. The sources encode their data at rates R_X and R_Y using encoders en_X and en_Y, respectively . The receiver decodes the received data using decoder r. Hence, we have:

$$en_X : \mathcal{X}^n \to \{1,...,2^{nR_X}\}$$
$$en_Y : \mathcal{Y}^n \to \{1,...,2^{nR_Y}\}$$

and a decoder map,

$$r : \{1,...,2^{nR_X}\} \times \{1,...,2^{nR_Y}\} \to \mathcal{Z}^n.$$

The probability of error is

$$P_e^n = \Pr[\{(\mathbf{x},\mathbf{y}) : d(f(\mathbf{x},\mathbf{y}), r(en_X(\mathbf{x}), en_Y(\mathbf{y}))) > D\}].$$

Table 6.1: Research progress on nonzero-distortion source coding problems

Problem types	$f(x, y) = (x, y)$	General $f(x, y)$
Side information	Wyner and Ziv [27]	Feng et al. [16] Yamamoto [28] Doshi et al. [10]
Distributed	Coleman et al. [7] Berger and Yeung [5] Barros and Servetto [4] Wagner et al. [25]	*

We say a rate pair (R_X, R_Y) is achievable up to a distortion D if there exist en_X, en_Y and r such that $P_e^n \to 0$ when $n \to \infty$.

Now, consider the network shown in Figure 6-1-a. Our aim is to find feasible rates for different links of this network when the receiver wants to compute $f(X, Y)$ up to a distortion D.

6.2 Prior Results

In this part, we overview prior relevant work. Consider the network shown in Figure 6-1-b. For this network, in [28], Yamamoto gives a characterization of a rate-distortion function for the side information functional compression problem (i.e., Y is available at the receiver). The rate-distortion function proposed in [28] is a generalization of the Wyner-Ziv side-information rate-distortion function [27]. Specifically, Yamamoto gives the rate distortion function as follows.

Theorem 58. *The rate distortion function for the functional compression problem with side information is*

$$R(D) = \min_{p \in \mathcal{P}(D)} I(W; X|Y)$$

where $\mathcal{P}(D)$ is the collection of all distributions on W given X such that there exists a $g : \mathcal{W} \times \mathcal{Y} \to \mathcal{Z}$ satisfying $E[d(f(\mathbf{X}, \mathbf{Y}), g(\mathbf{W}, \mathbf{Y}))] \leq D$.

This is an extension of the Wyner-Ziv rate-distortion result [27]. Further, the

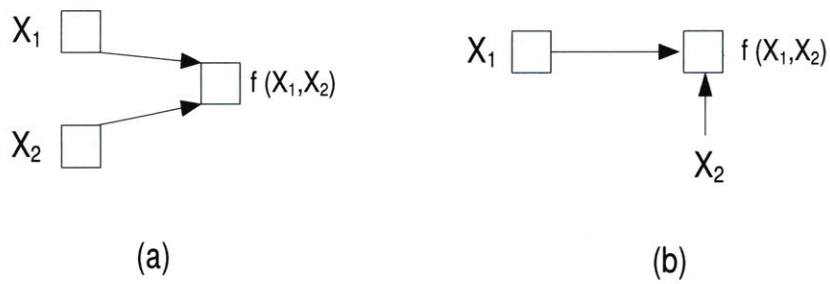


Figure 6-1: a)X and Y encode their data such that $\tilde{f}(X, Y)$, a representation of $f(X, Y)$ with distortion D, can be computed at the receiver. b) X encodes its data such that $\tilde{f}(X, Y)$, a representation of $f(X, Y)$ with distortion D, can be computed at the receiver.

variable $W \in \Gamma(G)$ in the definition of the Orlitsky-Roche rate, Definition 12, (a variable over the independent sets of G) can be seen as an interpretation of Yamamoto's auxiliary variable, W, for the zero-distortion case. In other words, when the distortion $D = 0$, the distributions on W given X for which there exists a reconstruction function g place nonzero probability only if w describes an independent set of G and $x \in w$.

A new characterization of the rate distortion function given by Yamamoto was discussed in [10]. It was shown in [10] that finding a suitable reconstruction function, $\hat{f}$, is equivalent to find g on $\mathcal{W} \times \mathcal{Y}$ from Theorem 58. Let $\mathcal{F}_m(D)$ denote the set of all functions $\hat{f}_m : \mathcal{X}^m \times \mathcal{Y}^m \to \mathcal{Z}^m$ such that

$$\lim_{n\to\infty} E[d(f(\mathbf{X}, \mathbf{Y}), \hat{f}_m(\mathbf{X}, \mathbf{Y}))] \leq D$$

and let $\mathcal{F}(D) = \bigcup_{m\in N} \mathcal{F}_m(D)$. Also, let $G_{X,\hat{f}}$ denote the characteristic graph of $\mathbf{X}$ with respect to $\mathbf{Y}$, $p(\mathbf{x}, \mathbf{y})$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$. For each m and all functions $\hat{f} \in \mathcal{F}(D)$, denote for brevity the normalized graph entropy $\frac{1}{m} H_{G_{X,\hat{f}}}(\mathbf{X}|\mathbf{Y})$ as $H_{G_{X,\hat{f}}}(X|Y)$. The following theorem was given in [10].

Theorem 59. *A rate distortion function for the network shown in Figure 6-1-b can* *be expressed as follows:*

$$R(D) = \inf_{\hat{f} \in \mathcal{F}(D)} H_{G_{X,\hat{f}}}(X|Y).$$

The problem of finding an appropriate function $\hat{f}$ is equivalent to finding a new graph whose edges are a subset of the edges of the characteristic graph. A graph parameterization by D was proposed in [10] to look at a subset of $\mathcal{F}(D)$. The resulting bound is not tight, but it provides a practical technique to tackle a very difficult problem.

Define the D-characteristic graph of X with respect to Y, $p(x, y)$, and $f(x, y)$, as having verticies $V = \mathcal{X}$ and the pair (x_1, x_2) is an edge if there exists some $y \in \mathcal{Y}$ such that $p(x_1, y)p(x_2, y) > 0$ and $d(f(x_1, y), f(x_2, y)) > D$. Denote this graph as $G_X(D)$. Because $d(z_1, z_2) = 0$ if and only if $z_1 = z_2$, the 0-characteristic graph is the characteristic graph (i.e., $G_X(0) = G_X$). So, the following corollary was given in [10].

Corollary 60. *The rate $H_{G_X(D)}(X|Y)$ is achievable.*

6.3 Main Results

This section contains our contributions in this problem. Our aim is to find a rate-distortion region for the network shown in Figure 6-1-a. Recall the Yamamoto rate distortion function (Theorem 58) and Theorem 59. These theorems explain a rate distortion function for the side information problem. Now, we are considering the case when we have distributed functional compression.

Again, for any m, let $\mathcal{F}_m(D)$ denote the set of all functions $\hat{f}_m : \mathcal{X}^m \times \mathcal{Y}^m \to \mathcal{Z}^m$ such that

$$\lim_{n\to\infty} E[d(f(\mathbf{X}, \mathbf{Y}), \hat{f}_m(\mathbf{X}, \mathbf{Y}))] \leq D.$$

In other words, we consider n blocks of m-vectors; thus, the functions in the expectation above will be on $\mathcal{X}^{mn} \times \mathcal{Y}^{mn}$. Let $\mathcal{F}(D) = \bigcup_{m\in N} \mathcal{F}_m(D)$. Let $G_{X,\hat{f}}$ denote the characteristic graph of $\mathbf{X}$ with respect to $\mathbf{Y}$, $p(\mathbf{x}, \mathbf{y})$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$ and $G_{Y,\hat{f}}$ denote the characteristic graph of $\mathbf{Y}$ with respect to $\mathbf{X}$, $p(\mathbf{x}, \mathbf{y})$, and $\hat{f}$ for any $\hat{f} \in \mathcal{F}(D)$. For each m and all functions $\hat{f} \in \mathcal{F}(D)$, denote for brevity the normalized graph entropy $\frac{1}{m}H_{G_{X,\hat{f}}}(\mathbf{X}|\mathbf{Y})$ as $H_{G_{X,\hat{f}}}(X|Y)$, $\frac{1}{m}H_{G_{Y,\hat{f}}}(\mathbf{Y}|\mathbf{X})$ as $H_{G_{Y,\hat{f}}}(Y|X)$ and $\frac{1}{m}H_{G_{X,\hat{f}},G_{Y,\hat{f}}}(\mathbf{X},\mathbf{Y})$ as $H_{G_{X,\hat{f}},G_{Y,\hat{f}}}(X,Y)$.

Now, for a specific function $\hat{f} \in \mathcal{F}(D)$, define $R_{\hat{f}}(D) = (R_X^{\hat{f}}(D), R_Y^{\hat{f}}(D))$ such that

$$\begin{aligned} R_X^{\hat{f}} &\geq H_{G_{X,\hat{f}}}(X|Y) \\ R_Y^{\hat{f}} &\geq H_{G_{Y,\hat{f}}}(Y|X) \\ R_X^{\hat{f}} + R_Y^{\hat{f}} &\geq H_{G_{X,\hat{f}},G_{Y,\hat{f}}}(X,Y). \end{aligned} \tag{6.2}$$

Theorem 61. *A rate-distortion region for the network shown in Figure 6-1-a is determined by* $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$.

Proof. We want to show that $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$ determines a rate-distortion region for the considered network. We first show this rate-distortion region is achievable for any $\hat{f} \in \mathcal{F}(D)$, and then we prove every achievable rate region is a subregion of it (converse).

According to Theorem 40, $R_{\hat{f}}(D)$ is sufficient to determine the function $\hat{f}(\mathbf{X},\mathbf{Y})$ at the receiver. Also, by definition,

$$\lim_{n\to\infty} E[d(f(\mathbf{X},\mathbf{Y}), \hat{f}(\mathbf{X},\mathbf{Y}))] \leq D.$$

Thus, for a specific $\hat{f} \in \mathcal{F}(D)$, $R_{\hat{f}}(D)$ is achievable. Therefore, the union of these achievable regions for different $\hat{f} \in \mathcal{F}(D)$ $\big($i.e., $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)\big)$ is also achievable.

Next, we show that any achievable rate region is a subregion of $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$. Assume that we have an achievable scheme in which X encodes its data to $en_X(\mathbf{X})$ and Y encodes its data to $en_Y(\mathbf{Y})$. At the receiver, we compute $r(en_X(\mathbf{X}), en_Y(\mathbf{Y}))$. Since it is an achievable scheme up to a distortion D, there exists $\hat{f} \in \mathcal{F}(D)$ such that $r(en_X(\mathbf{X}), en_Y(\mathbf{Y})) = \hat{f}(\mathbf{X},\mathbf{Y})$. Thus, considering Theorem 40, this achievable rate-distortion region is a subregion of $\bigcup_{\hat{f}\in\mathcal{F}(D)} R_{\hat{f}}(D)$. It completes the proof. □

Next, we present a simple scheme which satisfies Theorem 61. Again, the problem of finding an appropriate function $\hat{f}$ is equivalent to finding a new graph whose edges are a subset of the edges of the characteristic graph of random variables. This motivates Corollary 62 where we use a similar graph parameterization by D. Our scheme is as follows:

Define the D-characteristic graph of X with respect to Y, $p(x,y)$, and $f(x,y)$, as having verticies $V = \mathcal{X}$ and the pair (x_1, x_2) is an edge if there exists some $y \in \mathcal{Y}$ such that $p(x_1,y)p(x_2,y) > 0$ and $d(f(x_1,y), f(x_2,y)) > D$. Denote this graph as $G_X(D)$. Similarly, we define $G_Y(D)$. Following Corollary 60 and Theorem 61, we have the following Corollary.

Corollary 62. *If (R_X, R_Y) satisfies the following conditions, (R_X, R_Y) is achievable.*

$$\begin{aligned} R_X &\geq H_{G_X(D)}(X|Y) \\ R_Y &\geq H_{G_Y(D)}(Y|X) \\ R_X + R_Y &\geq H_{G_X(D),G_Y(D)}(X,Y). \end{aligned} \tag{6.3}$$

One may note when the number of vertices is small, constructing these graphs is not computationally difficult. Among all $\hat{f}$ which lead to the same D-characteristic graphs, one can choose a function which minimizes the distortion.

6.4 Summary of Results

In this chapter, we considered the problem of functional compression with distortion D. In this problem, a deterministic function of some correlated sources is desired at the receiver with distortion D. Some work ([10] and [28]) has addressed the side information version of this problem. Here, we considered distributed functional compression. We computed a feasible rate-distortion region and proposed an achievable scheme. For the case $D = 0$, our results are simplified to the lossless functional compression's results discussed in Chapter 2.

Chapter 7

Conclusions and Future Work

In this thesis, 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. In Chapter 2 of this thesis, we considered this problem for an arbitrary tree network and asymptotically lossless computations. For one-stage tree networks, we computed a rate-region and for an arbitrary tree network, we derived a rate lower bound based on graph entropy. We introduced a new condition on colorings of source random variables' characteristic graphs called the coloring connectivity condition (C.C.C.) and showed that, unlike the condition mentioned in Doshi *et al.*, this condition is necessary and sufficient for any achievable coding scheme based on colorings. We also showed that, unlike entropy, graph entropy does not satisfy the chain rule. We proposed a modularized coding scheme based on graph colorings which performs arbitrarily closely to the derived rate lower bounds for one stage trees, with correlated sources and general trees, with independent sources. We also 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 random variables called chain rule proper sets, it is sufficient to have intermediate nodes act like relays to perform arbitrarily closely to the rate lower bound.

The problem of having receivers with different desired functions was considered in Chapter 3. For this problem, we defined a new concept named multi-functional graph entropy which is an extension of graph entropy defined by Körner. We showed that the minimum achievable rate for this problem with side information is equal to conditional multi-functional graph entropy of the source random variable given the side information. We also proposed a coding scheme based on graph colorings to achieve this rate.

In these proposed coding schemes, one needs to compute the minimum entropy coloring of a characteristic graph. In general, finding this coloring is an NP-hard problem. However, in Chapter 4, we showed that depending on the characteristic graph's structure, there are certain cases where finding the minimum entropy coloring is not NP-hard, but tractable and practical. In one of these cases, we showed that, by having a non-zero joint probability condition on random variables' distributions, for any desired function, finding the minimum entropy coloring can be solved in polynomial time. In another case, we showed that if the desired function is a quantization function, this problem is also tractable. Then, we considered this problem in a general case. By using Huffman or Lempel-Ziv coding notions, we showed that finding the minimum entropy coloring is heuristically equivalent to finding the maximum independent set of a graph. While the minimum-entropy coloring problem is a recently studied problem, there are some heuristic algorithms to approximately solve the maximum independent set problem.

Next, in Chapter 5, we investigated the effect of having feedback on the rate-region of the functional compression problem. If the function at the receiver is the identity function, this problem reduces to the Slepian-Wolf compression with feedback, for which having feedback does not increase the rate. However, in general, feedback can improve rate bounds.

We finally considered the problem of distributed functional compression with distortion. The objective is to compress correlated discrete sources such that an arbitrary deterministic function of those sources can be computed up to a distortion level at the receiver. In this case, we computed a rate-distortion region and then, proposed a simple coding scheme for this problem.

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 [1], [18] and [17] under the name of network coding for multi-source multicast networks. Specially, [17] 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 [18]. This scheme is an achievable scheme; however it is not known whether it is optimal or not.

Throughout this thesis, we considered the 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. This problem is a communication complexity problem. One can use tools and schemes we have introduced in this thesis to attain some achievable schemes in the zero error computation case.

Bibliography

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. 46:1204–1216, 2000.

[2] Rudolf F. Ahlswede and János Körner. Source coding with side information and a converse for degraded broadcast channels. 21(6):629–637, November 1975.

[3] Noga Alon and Alon Orlitsky. Source coding and graph entropies. 42(5):1329–1339, September 1996.

[4] João Barros and Sergio Servetto. On the rate-distortion region for separate encoding of correlated sources. In *IEEE Symposium on Information Theory (ISIT)*, page 171, Yokohama, Japan, 2003.

[5] Toby Berger and Raymond W. Yeung. Multiterminal source encoding with one distortion criterion. 35(2):228–236, March 1989.

[6] Jean Cardinal, Samuel Fiorini, and Gwenal Joret. Tight results on minimum entropy set cover. 51(1):49–60, May 2008.

[7] Todd P. Coleman, Anna H. Lee, Muriel Médard, and Michelle Effros. Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression. 52(8):3546–3561, August 2006.

[8] Thomas Cormen. In *Introduction to algorithms*. MIT Press.

[9] I. Csiszar and J. Körner. In *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, New York, 1981.

[10] Vishal Doshi, Devavrat Shah, and Muriel Médard. Source coding with distortion through graph coloring. In *2007 International Symposium on Information Theory*, Nice, France, June 2007.

[11] Vishal Doshi, Devavrat Shah, Muriel Médard, and Sidharth Jaggi. Graph coloring and conditional graph entropy. In *2006 Asilomar Conference on Signals, Systems, and Computers*, pages 2137–2141, Asilomar, CA, October-November 2006.

[12] Vishal Doshi, Devavrat Shah, Muriel Médard, and Sidharth Jaggi. Distributed functional compression through graph coloring. In *2007 Data Compression Conference*, pages 93–102, Snowbird, UT, March 2007.

[13] Soheil Feizi and Muriel Médard. Multifunctional compression with side information. In *2009 Globecom Conference on Communications*, HI, US, November-December 2009.

[14] Soheil Feizi and Muriel Médard. When do only sources need to compute? on functional compression in tree networks. In *2009 Annual Allerton Conference on Communication, Control, and Computing*, September 2009.

[15] Soheil Feizi and Muriel Médard. Cases where finding the minimum entropy coloring of a characteristic graph is a polynomial time problem. In *2010 IEEE International Symposium on Information Theory, ISIT 2010*, TX, US, 2010.

[16] Hanying Feng, Michelle Effros, and Serap Savari. Functional source coding for networks with receiver side information. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, pages 1419–1427, September 2004.

[17] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. 52(10):4413–4430, October 2006.

[18] R. Koetter and M. Medard. An algebraic approach to network coding. *IEEE/ACM Trans. on networking*, 11(5):782–795, October 2003.

[19] János Körner. Coding of an information source having ambiguous alphabet and the entropy of graphs. In *6th Prague Conference on Information Theory*, pages 411–425, 1973.

[20] János Körner and Katalin Marton. How to encode the modulo-two sum of binary sources. 25(2):219–221, 1979.

[21] Alon Orlitsky and James R. Roche. Coding for computing. 47(3):903–917, March 2001.

[22] S. Sandeep Pradhan and Kannan Ramchandran. Distributed source coding using syndromes (DISCUS): design and construction. 49(3):626–643, March 2003.

[23] Claude E. Shannon. The zero error capacity of a noisy channel. 2(3):8–19, September 1956.

[24] David Slepian and Jack K. Wolf. Noiseless coding of correlated information sources. 19(4):471–480, July 1973.

[25] Aaron B. Wagner, Saurabha Tavildar, and Pramod Viswanath. Rate region of the quadratic gaussian two-terminal source-coding problem. Submitted February 2006.

[26] Hans S. Witsenhausen. The zero-error side information problem and chromatic numbers. 22(5):592–593, September 1976.

[27] Aaron Wyner and Jacob Ziv. The rate-distortion function for source coding with side information at the decoder. 22(1):1–10, January 1976.

[28] Hirosuke Yamamoto. Wyner-Ziv theory for a general function of the correlated sources. 28(5):803–807, September 1982.