Distributed Function Computation
Over a Rooted Directed Tree

Milad Sefidgaran and Aslan Tchamkerten
Department of Communications and Electronics
Telecom ParisTech
{sefidgaran, aslan.tchamkerten}@telecom-paristech.fr

Abstract

This paper establishes the rate region for a class of source coding function computation setups where sources of information are available at the nodes of a tree and where a function of these sources must be computed at the root. The rate region holds for any function as long as the sources’ joint distribution satisfies a certain Markov criterion. This criterion is met, in particular, when the sources are independent.

I. INTRODUCTION

Consider a directed tree network with \( k \geq 1 \) nodes where each edge points towards the root. An example of such a network is depicted in Fig. 1. Source \( X_u, u \in \{1, 2, \ldots, k\} \), is available at vertex \( u \) and a given function \( f(X_1, X_2, \ldots, X_k) \) must be computed at the root. Communication occurs in multiple hops, losslessly, from level one (composed of sources \( X_1, X_2, X_3, X_4 \) in the example) up to the root where the function is finally computed. Tree networks generalize some previously investigated settings including point-to-point [26], multiple access [17], [28], and cascade (relay-assisted) [8], [33], [31], [30], and can be used as backbones for computing functions over general networks [19].

Given a tree, a function \( f \), and a joint distribution over the sources \( (X_1, X_2, \ldots, X_k) \) we seek the corresponding rate region, \textit{i.e.}, we seek to characterize the least amounts of information that need to flow across each edge of the tree so that the function can be computed with arbitrarily high probability. In this paper, we first provide a cut-set outer bound to the rate region which generalizes the outer bounds established in [28, Corollary 2] for multiple access and in [33, Theorem 2] for cascade. Second, we establish an inner bound to the rate region which generalizes the inner bound for multiple access derived in [28, Proposition 1]. We then derive the main result which gives a sufficient condition on the sources’ joint distribution under which the inner and outer bounds are equal. This condition is satisfied, in particular, when the sources are independent.

This paper was presented in part at ITW 2012.

By multiple access we intend a noiseless multiple access channel where the receiver gets separate streams of data from each of the sources.
Related Works

Communication in distributed function computation has been investigated both under the zero error probability criterion and the asymptotic zero error probability criterion. We review related works separately as these criteria, although conceptually similar, can yield very different results and often involve different analysis—zero error problems are typically more combinatorial.

**Zero-error Probability:** Computational complexity has traditionally been measured in terms of the number of primitive operations required to compute a given function. When computation is carried out in a distributed fashion, Abelson [1] and Yao [35] proposed instead to measure complexity in terms of “data movement” between computing entities (processors) while ignoring local computations. In their model, one entity knows \( x_1 \) and another knows \( x_2 \), both \( x_1 \) and \( x_2 \) being length \( n \) (say, binary) vectors. The goal is for one of the entities to compute a given function \( f(x_1, x_2) \). Complexity is then defined as the minimum number of exchanged bits between the two entities. Communication in this setup involves no coding in the sense that protocols between entities allow to compute the function after each instance of the sources—\( x_1 \) and \( x_2 \) represent one instance of source 1 and one instance of source 2, respectively. This framework lead the foundations of communication complexity and has been widely studied ever since (see, e.g., [27], [7], [24], [25], [20]), though for “simple” networks involving no more than three nodes.

Coding for computing over multiple source instances was first considered by Ahlswede and Cai [2] for the Abelson-Yao’s setup. The non-interactive (one-way) version was subsequently considered by Alon and Orlitsky [4] and Koulgi et al. [18]. Recently, Shayevitz [31] investigated function computation over a cascade network where the transmitter can communicate to the receiver only via a relay.

Close to our setting is the one of Kowshik and Kumar [19] who investigated function computation over rooted directed trees and rooted directed acyclic graphs. Main results for rooted directed trees are necessary and sufficient
conditions on the nodes’ encoding procedures that allow function computation error free. When the sources’
distribution is positive, these conditions are independent and allow to compute the rate region. For more general
distributions these conditions appear hard to translate into bounds on the rate region.

Another closely related work is the one of Appuswamy et al. [5], [6] who derived bounds on the maximum
network computation rate for general directed acyclic graphs and independent sources. An extension to multiple
receivers was considered by Kannan and Viswanath [15].

ASYMPTOTIC ZERO-ERROR PROBABILITY: Slepian and Wolf [32] characterized the rate region for multiple
access networks and the identity function, i.e., when the receiver wants to recover the sources. For non-identity
functions the problem was considered by Körner and Marton [17] who investigated the rate region for the multiple
access network with binary sources and sum modulo two function. The rate region was established only for the
case of symmetric distributions and was obtained by means of Elias’s linear scheme [10]. Variations of this scheme
have later been used for computing linear functions over multiple access networks (see, e.g., [3], [13], [21], [14]).

An early and perhaps less known paper of Gel’fand and Pinsker [12] provides bounds for multiple access
networks and arbitrary functions. They showed that these bounds are tight in a general case which includes the case
of (conditionally) independent sources. As a byproduct they derived the optimal compression rate for the single
source and arbitrary function setting with side information at the receiver. For this latter setting, an equivalent
solution in terms of graph entropy was established by Orlitsky and Roche [26]. This graph entropy approach was
later used for multiple access networks in [28] and in [29] for the case of cooperative transmitters.²

In addition to multiple access networks, function computation over cascade networks have been investigated in
[8], [33], and [30] referenced here in increasing order of generality.

Beyond multiple access and cascade networks, collocated networks have been investigated by Ma, Ishwar and
Gupta [23] who established the rate region for independent sources.

Function computation over general networks remains challenging. As summarized in [19] such problems “combine
the complexity of source coding of correlated sources with rate distortion, together with the complications introduced
by the function structure.” Our results provide further insights by establishing the rate region for a general class of
networks with possibly dependent sources.

The paper is organized as follows. In Section II provides graph related preliminaries and Section III contains the
precise problem formulation. Results are presented in Section IV and their proofs are given in Section V.

II. PRELIMINARIES

We provide some graph theoretic background and introduce various notations which are summarized in Table I
to come.

We denote by $\mathcal{V}(G)$ and $\mathcal{E}(G)$ the vertex set and the edge set, respectively, of an undirected graph $G$. An
undirected edge between nodes $u$ and $v$ is denoted by $uv$ or $vu$. An independent set of a graph is a subset of its

²An early work on multiple access with cooperative transmitters is [11].
vertices no two of which are connected. A maximal independent set is an independent set that is not included in any other independent set. The set of independent sets of a graph $G$ and the set of maximal independent sets of $G$ are denoted by $\Gamma(G)$ and $\Gamma^*(G)$, respectively.

A path between two nodes $u$ and $v$ in a given graph $G$ is a sequence of nodes $u_1, \ldots, u_k$ where $u_1 = u, u_k = v,$ and $[u_iu_{i+1}] \in E(G)$ for $1 \leq i \leq k - 1$. A graph $G$ is connected if there exists a path between any two vertices $u, v \in V(G)$. A path $u_1, \ldots, u_k, k \geq 2$, with $u_1 = u_k$ is called a cycle. A graph is called acyclic if it contains no cycle. A tree is a connected acyclic graph.

A directed graph, denoted by $\overrightarrow{G}$, is a graph whose edges have a direction. We use $\overrightarrow{uv}$ to denote an edge from node $u$ to node $v$. A directed path from node $u$ to node $v$ is a sequence of nodes $u_1, \ldots, u_k$ where $u_1 = u, u_k = v$ and $[u_iu_{i+1}] \in E(\overrightarrow{G})$ for $1 \leq i \leq k - 1$.

The set of incoming neighbors of a node $u \in V(\overrightarrow{G})$, denoted by $\mathcal{N}_u$, is the set of nodes $v \in V(\overrightarrow{G})$ such that $\overrightarrow{vu} \in E(\overrightarrow{G})$. Their number, i.e., $|\mathcal{N}_u|$, is sometimes variously denoted by $n(u)$. For a vertex $u \in V(\overrightarrow{G})$, we denote by $Child(u)$ the set of all nodes $v$ such that there exists a directed path from $v$ to $u$, including $u$ itself, and by $Strangers(u)$ the set of vertices $v$ for which there is no directed path between $u$ and $v$, i.e.,

$$Strangers(u) \overset{\text{def}}{=} \{v : v \notin Child(u) \text{ and } u \notin Child(v)\}.$$ 

A rooted directed tree, denoted by $\overrightarrow{T}$, is a directed tree where all the edges point towards the root node $r$. The immediate (unique) vertex which $u$ is pointing to is denoted by $u_o$, whenever $u \neq r$.

For a rooted directed tree $\overrightarrow{T}$, an ordering

$$O_{\overrightarrow{T}} : V(\overrightarrow{T}) \to \{1, 2, \ldots, |V(\overrightarrow{T})|\}$$

is a one-to-one mapping from the set of vertices to the natural numbers $\{1, 2, \ldots, |V(\overrightarrow{T})|\}$ such that if for two vertices $u, v \in V(\overrightarrow{T})$

$$O_{\overrightarrow{T}}(u) > O_{\overrightarrow{T}}(v),$$

then the directed edge $\overrightarrow{uv}$ does not exist. The function

$$O^{-1}_{\overrightarrow{T}} : \{1, \ldots, |V(\overrightarrow{T})|\} \to V(\overrightarrow{T})$$

denotes the inverse of $O_{\overrightarrow{T}}$.

For any vertex $u$ and any ordering $O_{\overrightarrow{T}}$, $Sub_{O_{\overrightarrow{T}}}(u)$ and $Sup_{O_{\overrightarrow{T}}}(u)$ denote the set of vertices with lower and higher orderings than $u$, respectively:

$$Sub_{O_{\overrightarrow{T}}}(u) \overset{\text{def}}{=} \{v : v \in V(\overrightarrow{T}), O_{\overrightarrow{T}}(v) < O_{\overrightarrow{T}}(u)\}$$

$$Sup_{O_{\overrightarrow{T}}}(u) \overset{\text{def}}{=} \{v : v \in V(\overrightarrow{T}), O_{\overrightarrow{T}}(v) > O_{\overrightarrow{T}}(u)\}.$$
Sup\(\mathcal{T}(u)\) \(\overset{\text{def}}{=} \{ v : v \in \mathcal{V}(\mathcal{T}), O_{\mathcal{T}}(v) > O_{\mathcal{T}}(u) \}\).

In particular, we have
\[ \{u\} \cup \text{Sub}_{\mathcal{T}}(u) \cup \text{Sup}_{\mathcal{T}}(u) = \mathcal{V}(\mathcal{T}) \]
for any \(u \in \mathcal{V}(\mathcal{T})\).

Finally, for any vertex \(u\) and any ordering \(O_{\mathcal{T}}\) define

\[ \text{Roots}_{\mathcal{T}}(u) \overset{\text{def}}{=} \{ v : v \in \text{Sub}_{\mathcal{T}}(u), v \notin \text{Sub}_{\mathcal{T}}(u) \cup \{u\} \} \]
i.e., \(\text{Roots}_{\mathcal{T}}(u)\) represents the set of nodes \(v\) whose order is lower than \(u\) but for which there exists no directed path from \(v\) to \(\text{Sub}_{\mathcal{T}}(u) \cup \{u\}\).

The definition of \(\text{Roots}_{\mathcal{T}}(u)\) can be interpreted as follows. Consider the restriction of \(\mathcal{T}\) to the set of vertices \(\text{Sub}_{\mathcal{T}}(u) \cup \{u\}\). This subgraph is composed of some disconnected rooted directed trees\(^5\) whose roots are \(\text{Roots}_{\mathcal{T}}(u) \cup \{u\}\).

**Example 1.** Consider the rooted directed tree \(\mathcal{T}\) depicted in Fig. 1 with root \(r = 10\) being the root. For vertex \(2\), the unique outgoing neighbor is \(5\) and the set of incoming neighbors is \(\mathcal{N}_8 = \{5, 6\}\).

Also we have \(\text{Child}(8) = \{1, 2, 5, 6\}\) and \(\text{Strangers}(8) = \{3, 4, 7, 9\}\).

A possible ordering is the ordering given by the labels of the nodes (which already satisfies the ordering definition):

\[ O_{\mathcal{T}}(i) = i \quad 1 \leq i \leq 10. \]

For this ordering we have

\[ \text{Sub}_{\mathcal{T}}(7) = \{1, 2, 3, 4, 5, 6\} \]
\[ \text{Sup}_{\mathcal{T}}(7) = \{8, 9, 10\} \]
\[ \text{Roots}_{\mathcal{T}}(7) = \{5, 6\}. \]

Conditional characteristic graph plays a key role in coding for computing. We give here a general definition:

**Definition 1** (Conditional Characteristic Graph). Let \((L, K, S) \sim p(l, s, k)\) be a triple of random variables taking on values over some finite alphabet \(L \times K \times S\). Let \(f : S \rightarrow \mathbb{R}\) be a function such that \(H(f(S)|L, K) = 0\). The conditional characteristic graph \(G_{L|K}(f)\) of \(L\) given \(K\) with respect to \(f(s)\) is the graph whose vertex set is \(L\) and such that \(l_1 \in L\) and \(l_2 \in L\) are connected if for some \(s_1, s_2 \in S\), and \(k \in K\)

i. \(p(l_1, k, s_1) \cdot p(l_2, k, s_2) > 0\),

ii. \(f(s_1) \neq f(s_2)\).

When \(f(s)\) is known by the context, the above conditional characteristic graph is simply denoted by \(G_{L|K}\).

\(^5\)A graph composed of a single node is considered a (degenerate) tree.
Remark 1. When $L = S = X$ and $K = \emptyset$, Definition 1 reduces to the definition of the characteristic graph introduced by Körner in [16] and when $S = (X, Y)$, $L = X$, and $K = Y$ Definition 1 reduces to the definition of conditional characteristic graph introduced by Witsenhausen in [34].

Definition 1 can be interpreted as follows. Suppose a transmitter has access to random variable $L$ and a receiver has access to random variable $K$ and wants to compute function $f(S)$. The condition $H(f(S)|L, K) = 0$ guarantees that knowing $L$ and $K$ the receiver can compute $f(S)$. Moreover, in the characteristic graph $G_{L|K}$, given $K = k$, the knowledge of an independent set of $G_{L|K}$ that includes the realization $L = l$ suffices for the receiver to compute $f(S)$ since no two vertices in an independent set can produce different function outputs. Hence, for computing $f(S)$ the receiver needs only to know an independent set that includes $L$.

Example 2. Let $X$ and $Y$ be random variables defined over the alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, with

$$\mathcal{X} = \mathcal{Y} = \{1, 2, 3, 4\}.$$ 

Further, suppose that $P(X = Y) = 0$ and that $(X, Y)$ takes on values uniformly over the pairs $(i, j) \in \mathcal{X} \times \mathcal{Y}$ with $i \neq j$. Let $f(x, y)$ be defined as

$$f(x, y) = \begin{cases} 0 & \text{if } x < y, \\ 1 & \text{if } x > y. \end{cases}$$

In Definition 1, let $S = (X, Y)$, $L = X$, and $K = Y$. Fig. 2 depicts $G_{X|Y}$ and we have

$$\Gamma(G_{X|Y}) = \{\{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{2, 3\}, \{3, 4\}\}$$

and

$$\Gamma^*(G_{X|Y}) = \{\{1, 2\}, \{2, 3\}, \{3, 4\}\}.$$ 

In this example, the maximal independent sets overlap with each other and do not partition the vertices of the graph. The following lemma, whose proof is deferred to Appendix A, provides a sufficient condition under which the set of maximal independent sets forms a partition of the vertices of $G_{L|K}$.

Lemma 1. Let

$$(L, K, S_1, S_2) \sim p(l, k, s_1, s_2) = p(l, s_1) \cdot p(k, s_2)$$
and $f : S_1 \times S_2 \to \mathbb{R}$ be a function such that $H(f(S_1, S_2) | L, K) = 0$. Then $\Gamma^*(G_{L|K})$ is a partition of the set $L$. In other words, each $l \in L$ is included in exactly one maximal independent set.

A multiset of a set $S$ is a collection of elements from $S$ possibly with repetitions, e.g., if $S = \{0, 1\}$, then $\{0, 1, 1\}$ is a multiset. We use $M(S)$ to denote the collection of all multisets of $S$.

**Definition 2 (Conditional Graph Entropy [26]).** Given $(L, K, S) \sim p(l, k, s)$ and $f : S \to \mathbb{R}$ such that $H(f(S) | L, K) = 0$, the conditional graph entropy $H(G_{L|K}(f))$ is defined as

$$H(G_{L|K}(f)) \overset{\text{def}}{=} \min_{L \in V \in \mathcal{M}(\Gamma(G_{L|K}(f)))} I(V; L|K) = \min_{L \in V \in \mathcal{M}(\Gamma(G_{L|K}(f)))} I(V; L|K).$$

When the function $f(s)$ is known by the context, the above conditional graph entropy is simply denoted by $H(G_{L|K})$. Note that we always have $H(G_{L|K}) \leq H(L|K)$.

**Example 3.** Consider Example 2. According to Definition 2, for computing $H(G_X|Y)$ we can restrict the minimization of $I(V; X|Y)$ to be over all $V$ that take values over maximal independent sets, i.e.,

$$V = \{v_1 = \{1, 2\}, v_2 = \{2, 3\}, v_3 = \{3, 4\}\}.$$

Moreover, from the condition $X \in V$ and the symmetries of the pair $(1, 4)$ and the pair $(2, 3)$ it can be deduced that the $p(v|x)$ that minimizes the mutual information $I(V; X|Y)$ is given by

$$p(v_1|2) = p(v_3|4) = \delta,$$

$$p(v_2|2) = p(v_2|3) = 1 - \delta,$$

$$p(v_1|1) = p(v_4|3) = 1,$$

$$p(v_2|1) = p(v_3|1) = p(v_1|4) = p(v_2|4) = 0$$

for some $\delta \in [0, 1]$. This gives

$$I(V; X|Y) = \frac{1}{2} \cdot (H(\delta/3, (1 + \delta)/3, (2 - 2\delta)/3) + H(1/3, (1 - \delta)/3, (1 + \delta)/3) - h_b(\delta))$$

where $h_b(\delta)$ denotes the binary entropy $-\delta \log \delta - (1 - \delta) \log (1 - \delta)$. It can be checked that $I(V; X|Y)$ is minimized for $\delta = 1$. Hence, $H(G_X|Y) \approx 0.92 < 1.58 \approx H(X|Y)$ and the alphabet of the optimal $V$ is $V = \{v_1 = \{1, 2\}, v_3 = \{3, 4\}\}$ since $p(v_2|2) = p(v_2|3) = 1 - \delta = 0$.

The following table summarizes the main notations used throughout the paper.
TABLE I: Notation list

| Notation          | Definition                                                                 |
|-------------------|---------------------------------------------------------------------------|
| $G$               | Graph                                                                      |
| $\mathcal{V}(G)$  | Set of vertices of $G$                                                   |
| $E(G)$            | Set of edges of $G$                                                       |
| $\Gamma(G)$       | Set of independent sets of $G$                                           |
| $\Gamma^*(G)$     | Set of maximal independent sets of $G$                                   |
| $\mathcal{M}(\Gamma(G))$ | Multiset of independent sets of $G$                                      |
| $\overrightarrow{T}$ | A rooted directed tree                                                  |
| $r$               | Root of a rooted directed tree                                           |
| $N_u$             | Set of vertices whose outgoing edges directly point to vertex $u$         |
| $u_o$             | The outgoing neighbor of vertex $u$                                      |
| Child($u$)        | Set of vertices with directed path to $u$, including $u$ itself          |
| Strangers($u$)    | Set of vertices $v$ with no directed path between $u$ and $v$            |
| $O_{\overrightarrow{T}}$ | An ordering                                                              |
| Sub$_{O_{\overrightarrow{T}}}$($u$) | Set of vertices with lower ordering than $u$                          |
| Sup$_{O_{\overrightarrow{T}}}$($u$) | Set of vertices with higher ordering than $u$                           |
| $\text{Roots}_{O_{\overrightarrow{T}}}(u)$ | Roots (except from $u$) of the restriction of $\overrightarrow{T}$ to the set of vertices $\text{Sub}_{O_{\overrightarrow{T}}}(u) \cup \{u\}$ |
| $G_{L|K}(f)$      | Conditional characteristic graph of $L$ given $K$ with respect to function $f$ |
| $H(G_{L|K}(f))$   | Conditional graph entropy                                                |

III. Setup

Consider a rooted directed tree $\overrightarrow{T}$ with root $r$. Let $X_{\mathcal{V}(\overrightarrow{T})} \overset{\text{def}}{=} \{X_u,u \in \mathcal{V}(\overrightarrow{T})\}$ and $f : X_{\mathcal{V}(\overrightarrow{T})} \rightarrow \mathcal{F}$ where $\mathcal{F}$ is a finite set. Node $u \in \mathcal{V}(\overrightarrow{T})$ has access to random variable $X_u \in X_u$. Let $\{(x_{\mathcal{V}(\overrightarrow{T})})_i\}_{i \geq 1}$, be independent instances of random variables $X_{\mathcal{V}(\overrightarrow{T})}$ taking values over $X_{\mathcal{V}(\overrightarrow{T})}$ and distributed according to $p(x_{\mathcal{V}(\overrightarrow{T})})$.

To simplify notation, in the following we shall often avoid any explicit reference to the underlying tree $\overrightarrow{T}$ and will write, for instance, simply $O$ and $\mathcal{V}$ instead of $O_{\overrightarrow{T}}$ and $\mathcal{V}(\overrightarrow{T})$, respectively.

**Definition 3** (Code). A $((2^{nR_u})_{u \in \mathcal{V}\setminus\{r\}},n)$ code consists of encoding functions

$$
\varphi_u : X_u^n \times \{1, \ldots, 2^{nR_u}\} \times \cdots \times \{1, \ldots, 2^{nR_{u_2}}\} \rightarrow \{1, \ldots, 2^{nR_u}\}
$$

at nodes $u \in \mathcal{V}\setminus\{r\}$, where $\{u_1, \ldots, u_{n(u)}\} \overset{\text{def}}{=} N_u$, and a decoding function

$$
\psi : X_r^n \times \{1, \ldots, 2^{nR_r}\} \times \cdots \times \{1, \ldots, 2^{nR_{r_2}}\} \rightarrow \mathcal{F}^n
$$

at the root $r$, where $\{r_1, \ldots, r_{n(r)}\} = N_r$. 

December 13, 2013 DRAFT
Recall that by definition of $\text{Child}(u)$ we have

$$
\text{Child}(u) = \{u\} \cup \text{Child}(u_1) \cup \cdots \cup \text{Child}(u_n(u)).
$$

This allows to recursively define

$$
\varphi_u(X_{\text{Child}(u)}) \overset{\text{def}}{=} \varphi_u(X_u, \varphi_{u_1}(X_{\text{Child}(u_1)}), \cdots, \varphi_{u_n(u)}(X_{\text{Child}(u_n(u)))}).
$$

Throughout the paper we use bold fonts to denote length $n$ vectors. In the above expression, for instance, $X_{\text{Child}(u)}$ denotes a block of $n$ independent realizations of $X_{\text{Child}(u)}$.

The (block) error probability of a code (averaged over the sources’ outcomes) is defined as

$$
P(\psi(X_r, \varphi_{r_1}(X_{\text{Child}(r_1)}), \cdots, \varphi_{r_n(r)}(X_{\text{Child}(r_n(r))})) \neq f(X_V))
$$

where, with a slight abuse of notation, we wrote $f(X_V)$ to denote $n$ (independent) realizations of $f(X_V)$.

**Definition 4** (Rate Region). A rate tuple $(R_u)_{u \in V \setminus \{r\}}$ is achievable if, for any $\varepsilon > 0$ and all $n$ large enough, there exists a $((2^nR_u)_{u \in V \setminus \{r\}}, n)$ code whose error probability is no larger than $\varepsilon$. The rate region is the closure of the set of achievable rate tuples $(R_u)_{u \in V \setminus \{r\}}$.

In this paper we seek to characterize the rate region for given $\overrightarrow{T}$, $f$, and $p(x_V)$.

**IV. RESULTS**

The following cut-set outer bound is an immediate extension of the single source result [26, Theorem 1]:

**Theorem 1** (Outer Bound). For any vertex $u \in V$ and nonempty subset $S \subseteq N_u$ we have

$$
\sum_{v \in S} R_v \geq H(G_{X_{\text{Child}(S)}|X_{\text{Child}(S)}^c})
$$

where $\text{Child}(S) \overset{\text{def}}{=} \bigcup_{v \in S} \text{Child}(v)$.

It can be easily checked that the above outer bound implies [28, Corollary 2] when $\overrightarrow{T}$ is a multiple access network and [33, Theorem 2] when $\overrightarrow{T}$ is a cascade network.

Theorem 2 to come provides an inner bound to the rate region. For a given ordering, the scheme used for establishing this inner bound applies the scheme proposed in [28, Proof of Proposition 1] for the multiple access configuration in an iterative fashion. To describe the main idea, consider the network depicted in Fig.1 where $f$ is a function of $X_{10}$ and consider the natural ordering given by the labels of the nodes.

**Step 1:** Vertex 1 chooses a message $W_1 \in \mathcal{W}_1^n$ such that each realization of $f(X_{10})$ is computable from the corresponding values in $W_1$ and $X_{30}$. Vertex 2 chooses a message $W_2 \in \mathcal{W}_2^n$ such that each realization of $f(X_{10})$ is computable from the corresponding values of $W_1$, $W_2$, and $X_{30}$.

**Step 2:** Both vertices 1 and 2 transmit their messages to vertex 5 through a Slepian-Wolf coding such that vertex 5 can decode $W_1$ and $W_2$ having access to side information $X_5$. 

December 13, 2013 DRAFT
Step 3: Remove vertices 1 and 2 and all edges connected to them, and replace $X_5$ by $(X_5, W_1, W_2)$. The resulting tree is depicted in Fig. 3.

Step 4: Repeat Steps 1, 2, and 3 until the root receives the messages $W_8$ and $W_9$ from which it can compute the function reliably.

**Theorem 2** (Inner Bound). An inner bound to the rate region is

$$\sum_{u \in S} R_u \geq I(X_S, W_{N_S}; W_S|X_u, W_{S^c})$$

for any $u \in V, \emptyset \neq S \subseteq N_u, S^c = N_u \setminus S$, where random variables $(W_u)_{u \in V \setminus \{r\}}$ satisfy the Markov chain conditions

$$W_u = (X_u, W_{N_u^c}) - (X_{Child(u)}^c, W_{Sub_O(u)} \setminus Child(u))$$

as well as the condition

$$(X_u, W_{N_u^c}) \in W_u \in M(\Gamma(G_{X_u^c, W_{N_u^c}}|X_{Sup_O(u)}, W_{Root_O(u)})),$$

for an ordering $O$. Moreover, the inner bound is the same regardless of the ordering $O$.

Note that in the above iterative strategy transmissions at any given node depend on the ordering. For instance, another possible ordering is the one obtained by swapping nodes 1 and 2 in Fig.1, i.e., $O(1) = 2, O(2) = 1,$ and $O(i) = i, i \in \{3,4,\ldots,10\}$. For this ordering, Vertex 2 chooses a message $W_2$ such that each realization of $f(X_{10}^1)$ is computable from the corresponding values of $W_2$ and $(X_1, X_{10}^3)$. As a consequence, it may seem that the rate region achieved by the strategy depends on the ordering we impose on transmissions. As shown in Theorem 2, the rate region is the same regardless of the ordering. Indeed, later we shall see that if a set of auxiliary random variables satisfies (2) and (3) for a specific ordering, then it also satisfies these equations for any other
ordering. Since (1) is independent of the ordering, this means that any two orderings give the same achievable rate tuples.

Let us explain the terms (1), (2), and (3). Random variable $W_u$ is interpreted as the message sent by vertex $u$ and the Markov condition (2) reflects the fact that this message can depend only on the available side information $X_u$ and the set of incoming messages $W_{N_u}$. Once vertex $u$ has transmitted its data, the aggregate information in the resulting tree is $(W_u, X_{\sup_{\mathcal{O}}(u)}, W_{\text{Roots}_{\mathcal{O}}(u)})$. Choosing the alphabet of the message $W_u$ as in (3) guarantees that the knowledge of $W_u$ and $(X_{\sup_{\mathcal{O}}(u)}, W_{\text{Roots}_{\mathcal{O}}(u)})$ suffices for computing $f$ error free. Finally, the rate condition (1) guarantees that $W_u$ can be reliably decoded at the outgoing neighbor of vertex $u$.

The main result, stated in Theorem 3 to come, characterizes the rate region when the sources satisfy the following Markov property:

**Definition 5** (Markov Property). Consider a vertex $u$ in a rooted directed tree with sources $X_V$ available at its nodes. Remove vertex $u$ from the tree together with its incoming and outgoing edges. The resulting graph is locally Markovian if the remaining sets of connected sources are independent given the value of $X_u$, i.e., if

$$(X_{\text{Child}(u_1)}, \cdots, X_{\text{Child}(u_n(u))}, X_{\text{Child}(u)^c})$$

are independent given $X_u$, where $\{u_1, \cdots, u_n(u)\} = N_u$.

A directed tree satisfies the Markov Property if it is locally Markovian for every $u \in \mathcal{V}$.

**Remark 2.** It can be verified that a rooted directed tree satisfies the Markov property if and only if the joint probability distribution of $X_V$ is of the form

$$p(x_V) = p(x_r) \cdot \prod_{u \neq r} p(x_u | x_{u_r}).$$

(4)

The Markov Property thus holds, in particular, when all the sources $X_V$ are independent.

**Theorem 3.** For a rooted directed tree $\overrightarrow{T}$ that satisfies the Markov property, the inner and outer bounds given by Theorem 1 and Theorem 2 are tight and the rate region is given by

$$R_u \geq H(G_{X_{\text{Child}(u)}} | X_{\text{Child}(u)^c}) \quad u \in \mathcal{V} \setminus \{r\}.$$  

(5)

The following corollary follows from Lemma 1 and Theorem 3.

**Corollary 1.** For a rooted directed tree $\overrightarrow{T}$ with independent sources $X_V$, the rate region is given by

$$R_u \geq H(W_u^*) \quad u \in \mathcal{V} \setminus \{r\},$$

where $\mathcal{W}_u^* = \Gamma^*(G_{X_{\text{Child}(u)}} | X_{\text{Child}(u)^c})$ and for any $w_u^* \in \mathcal{W}_u^*$

$$p(W_u^* = w_u^*) = \sum_{x_{\text{Child}(u)} \in w_u^*} p(x_{\text{Child}(u)}).$$
V. PROOFS

For notational simplicity we leave out any explicit reference to the ordering and write, for instance, $Sub(u)$ instead of $Sub_O(u)$.

Proof of Theorem 1: Reveal $X_{Child(S)}$ to all the vertices in $S$ and reveal $X_{Child(S)^c}$ to vertex $u$. Since all the vertices in $S$ have access to the same information, the sum rate constraint is the same as for the case where only one of them sends information to $u$ which has access to information $X_{Child(S)^c}$. Using the single source result [26, Theorem 1] completes the proof.

Proof of Theorem 2: Suppose random variables $W_{V \setminus \{r\}}$ satisfy (2) and (3). These random variables together with $X_V$ are distributed according to some $p(x_V, w_{V \setminus \{r\}})$. For each $u \in V \setminus \{r\}$, independently generate $2^{n I(X_u:W_{N_u};W_u)}$ sequences $w_u(i) = (w_u^{(i)}_1, w_u^{(i)}_2, \ldots, w_u^{(i)}_{n}) \quad i \in \{1, 2, \ldots, 2^{n I(X_u:W_{N_u};W_u)}\}$, in an i.i.d. manner according to the marginal distribution $p(w_u)$; randomly and uniformly bin these sequences into $2^{nR_u}$ bins; and reveal the bin assignments $\phi_u$ to vertices $u$ and $u_o$.

Encoding/decoding at intermediate nodes and leaves: Given an ordering $O$, the encoding is done sequentially at vertices $O^{-1}(1), O^{-1}(2), \ldots, O^{-1}(|V| - 1)$.

Let $0 < \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_{|V| - 1} < \varepsilon_{|V|} = \varepsilon \leq 1$.

We distinguish leaves from intermediate nodes.\(^6\)

If $u$ is a leaf, i.e., $N_u = \emptyset$, the corresponding encoder finds a sequence $w_u$ such that $(x_u, w_u) \in A_{\varepsilon_i}^{(n)}(X_u, W_u)$ where $i = O(u)$ and sends the index of the bin that contains it, i.e., $\phi_u(w_u)$, to vertex $u_o$.

If $u$ is not a leaf, then the corresponding decoder first decodes the set of $n(u)$ incoming messages as follows. Given $x_u$ and the incoming messages' indices $(\phi_{u_1}(w_{u_1}), \ldots, \phi_{u_{n(u)}}(w_{u_{n(u)}}))$, vertex $u$ declares $(\tilde{w}_{N_u})$ if it is the unique $(\tilde{w}_{N_u})$ such that $(x_u, \tilde{w}_{N_u}) \in A_{\varepsilon_i}^{(n)}(X_u, W_{N_u})$ where $i = O(u)$.

\(^6\)By intermediate node we intend any node that is not root or a leaf.
and such that
\[(\phi_{u_1}(\hat{w}_{u_1}), \cdots, \phi_{u_{n(u)}}(\hat{w}_{u_{n(u)}})) = (\phi_{u_1}(w_{u_1}), \cdots, \phi_{u_{n(u)}}(w_{u_{n(u)}})).\]

Having decoded \(\hat{w}_{\mathcal{N}_r}\), vertex \(u\) finds a sequence \(w_u\) such that
\[(x_u, \hat{w}_{\mathcal{N}_u}, w_u) \in A_{\xi}^{(n)}(X_u, W_{\mathcal{N}_u}, W_u)\]
and sends the index of the bin that contains it, i.e., \(\phi_u(w_u)\), to vertex \(u_o\).

**Decoding at the root:** Given \(x_r\) and the indices \((\phi_{r_1}(w_{r_1}), \cdots, \phi_{r_{n(r)}}(w_{r_{n(r)}}))\), where
\[\{r_1, r_2, \cdots, r_{n(r)}\} = \mathcal{N}_r,\]
the root first declares
\[(\hat{w}_{\mathcal{N}_r})\]
if it is the unique \(\hat{w}_{\mathcal{N}_r}\) such that
\[(x_r, \hat{w}_{\mathcal{N}_r}) \in A_{\xi}^{(n)}(X_r, W_{\mathcal{N}_r})\]
and such that
\[(\phi_{r_1}(\hat{w}_{r_1}), \cdots, \phi_{r_{n(r)}}(\hat{w}_{r_{n(r)}})) = (\phi_{r_1}(w_{r_1}), \cdots, \phi_{r_{n(r)}}(w_{r_{n(r)}})).\]

**Probability of error:** Before computing the error probability let us observe that if for all \(u \in \mathcal{V} \setminus \{r\}\) message \(w_u\) is correctly encoded at vertex \(u\) and correctly decoded at vertex \(u_o\), the function \(f(X_V)\) can be computed with no error. To see this note first that at each step message \(w_u\) is chosen such that \((x_u, \hat{w}_{\mathcal{N}_u}, w_u)\) is jointly typical. Due to Claim c.ii. of Lemma 2, this implies that \(p(x_u, i, w_{\mathcal{N}_u, i}, w_{u, i}) > 0\) for any \(1 \leq i \leq n\). This together with (3) implies that \((x_u, i, w_{\mathcal{N}_u, i}) \in w_{u, i}\), i.e., each component \(w_{u, i}\) is an independent set in the graph
\[G_{X_u, W_{\mathcal{N}_u}, X_{\text{Sup}(u)}, W_{\text{Roots}(u)}}\]
that includes \((x_u, j, w_{\mathcal{N}_u, i})\). Moreover, due to the definition of conditional characteristic graph (Definition 1), by choosing the random variables \(W_u\) recursively as in (3), at each step
\[(W_u, W_{\text{Roots}(u)}, X_{\text{Sup}(u)})\]
is sufficient for computing the function \(f(X_V)\). Taking \(u = O^{-1}(|\mathcal{V}| - 1)\) implies that the root can compute the function by knowing \((W_{\mathcal{N}_r}, X_r)\).

We now show that for any node \(u \neq r\) the probability that message \(w_u\) is incorrectly encoded at vertex \(u\) or incorrectly decoded at vertex \(u_o\) can be made arbitrarily small by taking \(n\) large enough. A union bound over the nodes then implies that the root can compute the function with arbitrarily high probability.

Equivalently, we show that the following two events happen with arbitrarily low probability. The first event happens when some of the (incoming) messages in \(w_{\mathcal{N}_u}\) are incorrectly decoded assuming that they all have been
correctly encoded at nodes \( \mathcal{N}_u \). The second event happens when message \( w_u \) is incorrectly encoded, \textit{i.e.}, when no \( w_u \) is jointly typical with \((x_u, \hat{w}_{\mathcal{N}_u})\).\(^7\)

The probability of the second event is negligible for \( n \) large enough due to the covering lemma (Lemma 6).

We now bound the probability of the first event assuming that the incoming neighbors correctly encoded their messages. By symmetry of the encoding and decoding procedures, the probability of this event, averaged over sources outcomes, over \( w_v \)'s, and over the binning assignments, is the same as the average probability conditioned on vertex \( v \) correctly selecting \( W_{v}^{(1)} \), \( v \in \mathcal{N}_u \). Hence, we compute the probability of the event

\[
\{ \hat{W}_{\mathcal{N}_u} \neq W_{\mathcal{N}_u}^{(1)} \} \tag{6}
\]

assuming that each vertex \( v \in \mathcal{N}_u \) has previously selected \( W_{v}^{(1)} \) such that

\[
(W_{\mathcal{N}_u}, X_v, W_{v}^{(1)}) \in \mathcal{A}^{(n)}_{\mathcal{N}_u} (W_{\mathcal{N}_u}, X_v, W_v). \tag{7}
\]

Denote the elements of a set \( S \subseteq \mathcal{N}_u \) by \( u_1, u_2, \ldots, u_{n(u)} \) and let \( j_l \) be a natural number such that

\[ 1 \leq j_l \leq 2^{nI(X_{u_l}; W_{\mathcal{N}_{u_l}}; W_{u_l})} \quad 1 \leq l \leq n(u) \]

where \( \{u_1, u_2, \ldots, u_{n(u)}\} = \mathcal{N}_u \).

Define event \( \mathcal{E}(j^{n(u)}) \) as

\[
\mathcal{E}(j^{n(u)}):= \{ (W_{\mathcal{N}_u}^{(j_{n(u)}(u))}, X_u) \in \mathcal{A}^{(n)}_{\mathcal{N}_u} (W_{\mathcal{N}_u}, X_u), \phi_{u_1}(W_{u_1}^{(j_1)}(u_1)) = \phi_{u_1}(W_{u_1}^{(1)}), \phi_{u_2}(W_{u_2}^{(j_2)}(u_2)) = \phi_{u_2}(W_{u_2}^{(1)}), \ldots, \phi_{u_{n(u)}}(W_{u_{n(u)}}^{(j_{n(u)}(u))}(u_{n(u)})) = \phi_{u_{n(u)}}(W_{u_{n(u)}}^{(1)}) \}
\]

where

\[
W_{\mathcal{N}_u}^{(j^{n(u)})} \triangleq (W_{u_1}^{(j_1)}, W_{u_2}^{(j_2)}, \ldots, W_{u_{n(u)}}^{(j_{n(u)})}).
\]

\(^7\) For leaves there is only the second event.
The probability of the event (6) is upper bounded by

$$P(\mathbf{W}_{N_u} \neq \mathbf{W}_{N_u}^{(1)}) = P\left(\mathcal{E}^c((1,1,\cdots,1)) \cup \bigcup_{j^{(n(u))} \neq (1,1,\cdots,1)} \mathcal{E}(j^{(n(u))})\right)$$

$$= P\left(\mathcal{E}^c((1,1,\cdots,1)) \cup \bigcup_{\emptyset \neq S \subseteq N_u} \bigcup_{j^{(n(u))} \neq (1,1,\cdots,1), j_{S_1} \neq 1, j_{S_2} \neq 1, \cdots, j_{S_{|S|}} \neq 1} \mathcal{E}(j^{(n(u))})\right)$$

$$\leq P(\mathcal{E}^c((1,1,\cdots,1))) + \sum_{\emptyset \neq S \subseteq N_u} \sum_{j^{(n(u))} \neq 1, j_{S_1} \neq 1, j_{S_2} \neq 1, \cdots, j_{S_{|S|}} \neq 1} P(\mathcal{E}(j^{(n(u))})). \quad (8)$$

We bound each of the two terms on the right-hand side of (8). For the first term, according to (7) and the properties of jointly typical sequences (Lemmas 2, 3, 4, and 5), we have

$$P(\mathcal{E}^c(1,1,\cdots,1)) \leq \delta(\varepsilon_i).$$

where $\delta(\varepsilon_i) \xrightarrow{\varepsilon_i \to 0} 0$.

Now for the second term. For any $S$ such that $\emptyset \neq S \subseteq N_u$ and any $j^{(n(u))}$ such that

$$j_{S_1} \neq 1, j_{S_2} \neq 1, \cdots, j_{S_{|S|}} \neq 1$$

and

$$j_{S^c} = (1,1,\cdots,1)$$

we have

$$P(\mathcal{E}(j^{(n(u))})) \leq 2^{-n} \sum_{v \in \mathbb{Z}} R_v \cdot 2^{-n(I(W_{us_1};W_{us_2}) - \delta_i(\varepsilon_i))} \cdot 2^{-n(I(W_{us_1};W_{us_3};W_{us_3}) - \delta_i(\varepsilon_i))} \cdots$$

$$\cdots 2^{-n(I(W_{us_1};W_{us_2};\cdots;W_{us_{|S|}};W_{us_{|S|}}) - \delta_i(\varepsilon_i))} \cdot 2^{-n(I(W_{us_1};W_{us_2};\cdots;W_{us_{|S|+1}};W_{us_{|S|+1}}) - \delta_i(\varepsilon_i))}.$$

Since

$$\sum_{j^{(n(u))}} 1 \leq 2^{nI(X_{us_1};X_{us_2};W_{us_2})} \cdots 2^{nI(X_{us_{|S|};W_{us_{|S|}}})} \cdot 2^{nI(X_{us_{|S|+1};W_{us_{|S|+1}}})}.$$
the second term on the right-hand side of (8) is negligible for \( n \) large enough provided that

\[
\sum_{v \in S} R_v \geq I(X_{u_1}, W_{N_{u_1}}; W_{u_1}) + \cdots + I(X_{u_{|S|}}, W_{N_{u_{|S|}}}; W_{u_{|S|}})
\]

\[
- I(W_{u_1}; W_{u_2}) - \cdots - I(W_{u_1}, W_{u_2}, \ldots, W_{u_{|S|}-1}; W_{u_{|S|}})
\]

\[
- I(W_S; X_u, W_{S^c})
\]

\[
= I(X_{u_1}, W_{N_{u_1}}; W_{u_1}) + \cdots + I(X_{u_{|S|}}, W_{N_{u_{|S|}}}; W_{u_{|S|}})
\]

\[
- I(W_{u_1}; W_{u_2}, \ldots, W_{u_{|S|}-1}; X_u, W_{S^c})
\]

\[
- I(W_{u_2}; W_{u_3}, \ldots, W_{u_{|S|}-1}; X_u, W_{S^c})
\]

\[\vdots\]

\[
- I(W_{u_{|S|-1}}; X_u, W_{S^c})
\]

\[
= H(W_{u_1} | W_{u_2}, \ldots, W_{u_{|S|}}, X_u, W_{S^c})
\]

\[
- H(W_{u_1} | W_{N_{u_1}}, X_u, W_{S^c})
\]

\[
+ H(W_{u_2} | W_{u_3}, \ldots, W_{u_{|S|}-1}, X_u, W_{S^c})
\]

\[
- H(W_{u_2} | W_{u_2}, W_{S^c})
\]

\[\vdots\]

\[
+ H(W_{u_{|S|-1}} | X_u, W_{S^c})
\]

\[
- H(W_{u_{|S|-1}} | X_{u_{|S|-1}}, W_{N_{u_{|S|-1}}})
\]

\[
\overset{(a)}{=} I(X_S, W_{N_S}; W_{u_1} | W_{u_2}, \ldots, W_{u_{|S|-1}}, X_u, W_{S^c})
\]

\[
+ I(X_S, W_{N_S}; W_{u_2} | W_{u_3}, \ldots, W_{u_{|S|-1}}, X_u, W_{S^c})
\]

\[\vdots\]

\[
+ I(X_S, W_{N_S}; W_{u_{|S|-1}} | X_u, W_{S^c})
\]

\[
= I(X_S, W_{N_S}; W_{u_{|S|-1}} | X_u, W_{S^c})
\]

where \((a)\) holds due to the following claim whose proof is deferred to Appendix B:

Claim 1. The Markov chains (2) are equivalent to the following Markov chains

\[
W_u - (X_u, W_{N_u}) - (X_{\text{Child}(u)}, W_{\text{Strangers}(u)}) \quad u \in V \setminus \{r\}, \tag{9}
\]

which do not depend on the ordering.

\^Note that the summation over the sets \( S \) in the second term on the right-hand side of (8) involves a constant number of elements that does not depend on \( n \).
This completes the achievability part of the Theorem.

It remains to show that for different orderings the corresponding achievable regions are the same. For this it is sufficient to show that any \( W_{\mathcal{V}\backslash\{r\}} \) that satisfy conditions (2) and (3) for an ordering \( O \) also satisfy these conditions for any other ordering \( O' \). Claim 1 says that \( W_{\mathcal{V}\backslash\{r\}} \) satisfy (2) for any ordering \( O' \). The following claim, whose proof is deferred to Appendix C, completes the proof:

Claim 2. \( W_{\mathcal{V}\backslash\{r\}} \) satisfies (3) for any other ordering \( O' \). □

\textbf{Proof of Theorem 3:} Suppose that random variables \( X_{\mathcal{V}} \) satisfy the Markov property (Definition 5). We show that the inner bound in Theorem 2 is tight with an outer bound obtained from the outer bound of Theorem 1. Without loss of generality, we suppose that the set of vertices are \( \{1, 2, \ldots, m + 1\} \) and that the ordering is the natural ordering given by \( O(u) = u \), for \( 1 \leq u \leq m + 1 \), with \( r = m \).

\textbf{Outer bound}

Consider the following constraints in Theorem 1

\[ R_u \geq H(G_{X_{\text{Child}(u)}|X_{\text{Child}(u)^c}}) \quad u \in \mathcal{V} \setminus \{r\} \]  

which are derived by letting \( S \) be a single vertex. Considering only these constraints gives a weaker outer bound than the one of Theorem 1.

\textbf{Inner bound}

We show that (10) is achievable using Theorem 2, thereby completing the proof of the theorem. This is done in a number of steps.

- \textbf{Simplifying the rate constraints:} We first simplify the rate constraints (1) in Theorem 2 using the following claim whose proof is deferred to Appendix D:

Claim 3. Suppose that the random variables \( X_{\mathcal{V}} \) satisfy the Markov property and that the random variables \( W_{\mathcal{V}\backslash\{r\}} \) satisfy the Markov chain conditions (2). Then, the set of pairs of random variables

\[ ((X_{\text{Child}(u_1)}, W_{\text{Child}(u_1)}), \ldots, (X_{\text{Child}(u_{n(u)})}, W_{\text{Child}(u_{n(u)})}), (X_{\text{Child}(u)^c}, W_{\text{Sub}(u)^c}) \backslash \text{Child}(u)) \]

are jointly independent given \( X_u \) for \( u \in \mathcal{V} \), where \( \{u_1, \ldots, u_{n(u)}\} = \mathcal{N}_u \). In particular, this implies that the pair

\[ (X_{\text{Child}(u)} \backslash \{u\}, W_{\text{Child}(u)} \backslash \{u\}) \]

is independent of the pair

\[ (X_{\text{Child}(u)^c}, W_{\text{Sub}(u)^c}) \backslash \text{Child}(u)) \]

given the value of \( X_u \), for any \( u \in \mathcal{V} \setminus \{r\} \). □
Consider the rate constraints (1) in Theorem 2. Claim 3 implies that for the terms on the right-hand side of (1) we have

\[ I(X_S;W_{N_S};W_S|X_u,W_{S^e}) = \sum_{u \in S} I(X_u;W_{N_u};W_u|X_u). \]  (11)

Hence, the rate constraints (1) reduce to the following constraints:

\[ R_u \geq I(X_u;W_{N_u};W_u|X_u) \quad u \in V \setminus \{r\}. \]  (12)

Therefore, we may consider the constraints (12) instead of (1).

*To prove:* To show that the inner bound matches the outer bound it is sufficient to show that

\[ I(X_u;W_{N_u};W_u|X_u) \leq H(G_{X_{\text{Child}(u)}|X_{\text{Child}(u)^c}}) \quad u \in V \setminus \{r\} \]  (13)

for a specific choice of \( W_{V\setminus \{r\}} \) that satisfy the constraints (2) and (3) in Theorem 2. Rewrite inequalities (13) as

\[
I(X_1;W_{N_1};W_1|X_{1_u}) \leq H(G_{X_{\text{Child}(1)}|X_{\text{Child}(1)^c}})
\]

\[
I(X_2;W_{N_2};W_2|X_{2_u}) \leq H(G_{X_{\text{Child}(2)}|X_{\text{Child}(2)^c}})
\]

\[
\cdots
\]

\[
I(X_m;W_{N_m};W_m|X_{m_u}) \leq H(G_{X_{\text{Child}(m)}|X_{\text{Child}(m)^c}}). \]  (14)

Note that the first \( u - 1 \) inequalities do not depend on \( W_u \), for \( 1 \leq u \leq m \). Using induction, we show that for any \( 1 \leq k \leq m \), the first \( k \) inequalities hold for some \( W_1^*, \cdots , W_k^* \) that satisfy conditions (2) and (3), and such that \( W_u^*, 1 \leq u \leq k \), takes values only over maximal independent sets of

\[ G_{X_u,W_{N_u}|X_{\text{Sup}(u)};W_{\text{Root}(u)}}. \]  (15)

*Induction base:* For \( k = 1 \), we have \( N_1 = \emptyset \) and \( \text{Child}(1) = \{1\} \). Moreover, we have \( I(X_1;W_1|X_{1_u}) = I(X_1;W_1|X_{2_u}^m) \) due to Claim 3. Hence, to show that the first inequality in (14) holds it suffices to show that there exists \( W_1 \) such that

\[ I(X_1;W_1|X_{2_u}^m) \leq H(G_{X_1|X_{2_u}^m}). \]

A natural choice is to pick \( W_1 = W_1^* \) as the random variable that achieves \( H(G_{X_1|X_{2_u}^m}) \), i.e., the one that minimizes

\[ I(X_1;W_1|X_{2_u}^m), \]

among all \( W_1 \)'s such that

\[ X_1 \in W_1 \in \Gamma^*(G_{X_1|X_{2_u}^m}) \]

\[ W_1 - X_1 - X_{2_u}^m. \]

Trivially conditions (2) and (3) are satisfied by \( W_1^* \). Since \( \Gamma^*(G_{X_1|X_{2_u}^m}) \) corresponds to the maximal independent sets of the conditional characteristic graph (15), the case \( k = 1 \) is proved.
• **Induction step:** Suppose that the first $k-1$ inequalities in (14) hold for some $W_1^*, \ldots, W_{k-1}^*$ that satisfy conditions (2) and (3), and such that $W_u^*$, $1 \leq u \leq k-1$, take values over the maximal independent sets of

$$G_{X_u,W_u^*|X_{Sup}(u),W_{Roots}(u)}.$$  

We now show how to choose a proper $W_k^*$ such that the $k$-th inequality holds. Note that random variable $W_k$ does not appear in the first $k-1$ inequalities (however, some of the $W_i$, $i < k$, appear in the $k$th inequality).

The following claim, whose proof is deferred to Appendix E, says that the graph entropy term on the right-hand side of the $k$-th inequality in (14) is equal to another graph entropy that we shall analyze here below:

**Claim 4.** Suppose that the random variables $X_V$ satisfy the Markov property and that the random variables $W_{V\setminus \{r\}}$ satisfy the conditions (2) and (3). Then,

$$H(G_{X_{Child(k)}|X_{Child(k)'}}) = H(G_{X_{Child(k)}|X_{Sup(k)},W_{Roots(k)}}).$$  

We show that this inequality holds for a proper choice of $W_k^*$ which completes the proof of the induction step, and hence the proof of the tightness of the inner and the outer bounds under the Markov property.

We first introduce a random variable $W_k'$ which satisfies the $k$-th inequality and condition (2). By a change of alphabet we then define $W_k^*$ which, in addition, takes values over the maximal independent sets of $G_{X_k,W_k^*|X_{Sup}(k),W_{Roots(k)}}$ and satisfies condition (3).

- **Defining $W_k'$:** Let $W_k'$ be the random variable that achieves $H(G_{X_{Child(k)}|X_{Sup(k)},W_{Roots(k)}})$, *i.e.*, the one that minimizes

$$I(X_{Child(k)}; W|X_{Sup(k)},W_{Roots(k)})$$

among all $W$'s such that

$$X_{Child(k)} \in W \in \Gamma^*(G_{X_{Child(k)}|X_{Sup(k)},W_{Roots(k)}})$$

$$W - X_{Child(k)} = (X_{Sup(k)}, W_{Roots(k)}).$$  

Suppose that $(W_k', X_V, W_{Roots(k)})$ and $(X_V, W_{Sub(k)})$ are distributed according to some joint distribution

$$p(W_k', X_V, W_{Roots(k)})$$

and

$$p(X_V, W_{Sub(k)})$$

Dec 13, 2013 DRAFT
respectively. Note that, by definition, \( \text{Roots}(k) \subseteq \text{Sub}(k) \), hence \( W^*_{\text{Roots}(k)} \) involves a subset of the random variables \( W^*_\text{Sub}(k) \).

Now define the joint distribution of \((W'_k, X_V, W^*_\text{Sub}(k))\) as
\[
p(W'_k, X_V, W^*_\text{Sub}(k)) \stackrel{\text{def}}{=} p(X_V, W^*_\text{Sub}(k)) \cdot p(W'_k | X_{\text{Child}(k)}) \cdot p(w'_k | X_{\text{Child}(k)}).
\] (21)

Note that this distribution keeps the marginals (19) and (20). Moreover, Definition (21) yields the following Markov chains

\[
W'_k - X_{\text{Child}(k)} - (X_{\text{Child}(k)^{\neq}}, W^*_\text{Sub}(k))
\]
\[
W'_k - (X_k, W^*_N_k) - (X_{\text{Child}(k)^{\neq}}, W^*_\text{Sub}(k) \setminus \text{Child}(k)),
\]

where the second Markov chain holds because of Claim 5 whose proof is deferred to Appendix F. These Markov chains imply that inequality (17) holds, i.e.
\[
I(X_k, W^*_N_k, W'_k | X_{\text{Sup}(k)}; W^*_{\text{Roots}(k)}) \leq I(X_{\text{Child}(k)}; W'_k | X_{\text{Sup}(k)}, W^*_{\text{Roots}(k)}).
\] (22)

Claim 5. Condition (2) holds for \( W_k = W'_k \).

- Defining \( W^*_k \) from \( W'_k \): For \( w'_k \in W'_k \) define
\[
\mathcal{B}_{w'_k} \stackrel{\text{def}}{=} \{(w_{k_1}, \ldots, w_{k_n(k)}, x_k) | (w_{k_1}, \ldots, w_{k_n(k)}, x_k) \in (W^*_k, \ldots, W^*_n | X_{\text{Sub}(k)}) \},
\]
\[
\exists x_{\text{Child}(k)} \in w'_k : p(x_{\text{Child}(k)}, w_{k_i}) > 0, \forall i \in \{1, 2, \ldots, n(k)\}
\]

where \( \{k_1, \ldots, k_n(k)\} = N_k \) and \( W^*_N_k \subseteq \Gamma^*(G_{X_k, W^*_N_k}, X_{\text{Sub}(k)}, W^*_{\text{Roots}(k)}) \).

First we show that \( w'_k \) and \( \mathcal{B}_{w'_k} \) are in one-to-one correspondence, i.e., there is no \( w_1, w_2 \in W'_k \) with \( w_1 \neq w_2 \) such that \( \mathcal{B}_{w_1} = \mathcal{B}_{w_2} \). This can be deduced from the following claim whose proof is deferred to Appendix G:

Claim 6. If \((w_{k_1}, \ldots, w_{k_n(k)}, x_k) \in \mathcal{B}_w, w \in W'_k, \) and \( p(x_{\text{Child}(k)}, w_{k_i}) > 0, \) then
\[
x_{\text{Child}(k)} = (x_{\text{Child}(k_1)}, \ldots, x_{\text{Child}(k_n(k))}, x_k) \in w.
\]

Let random variable \( W^*_k \) take values over the set
\[
\mathcal{W}^*_k = \{ \mathcal{B}_{w'_k} : w'_k \in W'_k \}
\]

with conditional distribution
\[
p(w^*_k | x_V, w^*_\text{Sub}(k)) = p(w^*_k | x_{\text{Child}(k)}) = p(w'_k | x_{\text{Child}(k)}),
\]

where \( w'_k \in \mathcal{W}^*_k \) is the unique value such that \( \mathcal{B}_{w'_k} = w^*_k \).

- Showing that the \( k \)-th inequality holds: We first show that \( W^*_k \) satisfies conditions (2) and (3) and that \( W^*_k \) takes values only over maximal independent sets. This can be deduced from parts a. and b. of the following claim whose proof is deferred to Appendix H.

Claim 7. We have the following relations for \( W^*_k \):
a. The Markov chain

\[ W_k^* - (X_k, W_{N_k}^*) - (X_{\text{Child}(k)\setminus \text{Child}(k)}, W_{\text{Roots}(k)}) \]

holds (equivalently, condition (2) holds);

b. \( (X_k, W_{N_k}^*) \in W_k^* \in \Gamma^*(G_{X_k, W_{N_k}^*}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}) \);

c. \( I(X_k, W_{N_k}^*; W_k^*|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}) = I(X_k, W_{N_k}^*; W_k^'|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}). \)

\( \square \)

Claim 7.c, together with (22), and the fact that

\[ I(X_{\text{Child}(k)}; W_k^*|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}) = H(G_{X_{\text{Child}(k)}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}) \]

implies that the \( k \)-th inequality holds. This completes the proof.

\[ \square \]

REFERENCES

[1] H. Abelson. Lower bounds on information transfer in distributed computations. In 19th Annual Symposium on Foundations of Computer Science, pages 151–158, October 1978.

[2] R. Ahlswede and N. Cai. On communication complexity of vector-valued functions. Information Theory, IEEE Transactions on, 40(6):2062–2067, November 1994.

[3] R. Ahlswede and T. Han. On source coding with side information via a multiple-access channel and related problems in multi-user information theory. Information Theory, IEEE Transactions on, 29(3):396–412, May 1983.

[4] N. Alon and A. Orlitsky. Source coding and graph entropies. Information Theory, IEEE Transactions on, 42(5):1329–1339, September 1996.

[5] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger. Network coding for computing: Cut-set bounds. Information Theory, IEEE Transactions on, 57(2):1015–1030, February 2011.

[6] R. Appuswamy, M. Franceschetti, N. Karamchandani, and K. Zeger. Linear codes, target function classes, and network computing capacity. Information Theory, IEEE Transactions on, 59(9):5741–5753, 2013.

[7] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In Foundations of Computer Science, 1986., 27th Annual Symposium on, pages 337–347. IEEE, 1986.

[8] P. Cuff, H. I. Su, and A. El Gamal. Cascade multiterminal source coding. In Information Theory Proceedings (ISIT), 2009 IEEE International Symposium on, pages 1199–1203, June 2009.

[9] A. El Gamal and Y. H. Kim. Network Information Theory. Cambridge University Press, 2012.

[10] P. Elias. Coding for noisy channels. IRE Conv. Rec, 3(pt 4):37–46, 1955.

[11] T. Ericson and J. Körner. Successive encoding of correlated sources. Information Theory, IEEE Transactions on, 29(3):390–395, May 1983.

[12] S. Gel’fand and M. Pinsker. Coding of sources on the basis of observations with incomplete information. Problemy Peredachi Informatsii, 15(2):115–125, 1979.

[13] T. S. Han and K. Kobayashi. A dichotomy of functions f(x,y) of correlated sources (x,y) from the viewpoint of the achievable rate region. Information Theory, IEEE Transactions on, 33:69–76, January 1987.

[14] S. Huang and M. Skoglund. Polynomials and computing functions of correlated sources. In Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, pages 771–775, July 2012.
A. Proof of Lemma 1

Each vertex is contained in at least one maximal independent set. Suppose, by way of contradiction, that there exists a vertex \( l \in L \) that belongs to two maximal independent sets \( w_1, w_2 \in \Gamma^*(G_{L|K}) \). This means that there
exist some $l_1 \in w_1$ and $l_2 \in w_2$ such that $l_1$ and $l_2$ are connected in $G_{L|K}$, i.e., there exist some $s_1, s'_1 \in S_1$, $s_2, s'_2 \in S_2$ and $k \in K$ such that

$$p(s_1, s_2, l_1, k) \cdot p(s'_1, s'_2, l_2, k) = p(s_1, l_1) \cdot p(s'_1, l_2) \cdot p(k, s_2) \cdot p(k, s'_2) > 0$$  \hspace{1cm} (23)

$$f(s_1, s_2) \neq f(s'_1, s'_2).$$  \hspace{1cm} (24)

Now take any $s''_1 \in S_1$ such that $p(s''_1, l) > 0$. This, together with (23), and the fact that both vertex pairs $(l_1, l)$ and $(l_2, l)$ are disconnected in $G_{L|K}$ implies that

$$f(s_1, s_2) = f(s''_1, s_2) = f(s'_1, s'_2),$$

which contradicts (24).

### B. Proof of Claim 1

Note that trivially the Markov chains (9) imply the Markov chains (2) since the set of random variables $W_{\text{Sub}(u) \setminus \text{Child}(u)}$ is contained in the set of random variables $W_{\text{Strangers}(u)}$.

We first show the reverse implication through the example in Fig. 1 with the natural ordering given by the labels of the nodes. We show the implication for vertex 5. For this vertex the Markov chain (2) becomes

$$W_5 - (X_5, W^2_1) - (X^4_3, X^{10}_6, W^4_3).$$  \hspace{1cm} (25)

Also, for vertex 6 the Markov chain (2) corresponds to

$$W_6 - X_6 - (X^5_1, X^{10}_7, W^5_1).$$  \hspace{1cm} (26)

Combining (25) and (26) yields the Markov chain

$$W_5 - (X_5, W^2_1) - (X^4_3, X^{10}_6, W^4_3, W_6).$$

Similarly, from this Markov chain and the corresponding Markov chains for vertices 7 and 9 in (2) we get

$$W_5 - (X_5, W^2_1) - (X^4_3, X^{10}_6, W^4_3, W_6, W_7, W_9)$$  \hspace{1cm} (27)

which corresponds to (9) with $u = 5$.

In general, to show that the Markov chains (9) hold, we observe that (2) and (9) have the generic forms

$$A - B - C$$  \hspace{1cm} (28)

and

$$A - B - (C, W_{d_1}, W_{d_2}, \ldots, W_{d_q})$$  \hspace{1cm} (29)

respectively, where

$$\{d_1, d_2, \ldots, d_q\} = \text{Strangers}(u) \setminus \text{Sub}(u)$$

Notice that random variables $A, B, C, D$ satisfy $A - B - (C, D)$ if and only if $A - B - C$ and $A - (B, C) - D$ hold.
and where, without loss of generality, the ordering is such that

\[ O(d_1) < O(d_2) < \ldots < O(d_q). \]

To show that (2) implies (9) one first shows that

\[ A - B - (C, W_{d_1}) \] (30)

holds by using (28) and (2) for the vertex \( d_1 \)—in the example above \( d_1 = 6 \). Then one shows that

\[ A - B - (C, W_{d_1}, W_{d_2}) \] (31)

holds using (30) and (2) for the vertex \( d_2 \)—in the example above \( d_2 = 7 \). The argument is iterated for \( d_3, \ldots, d_q \) completing the proof.

C. Proof of Claim 2

As for the previous claim, consider first the particular network depicted in Fig.1 and let \( O \) be the natural ordering given by the labels of the nodes and let \( O' \) be obtained from \( O \) by swapping the orders of the vertices 1 and 2, i.e.

\[ O'(1) = O(2) = 2, O'(2) = O(1) = 1, O'(i) = O(i), i \in \{3, 4, \ldots, 10\}. \]

We need to show that

\[ X_1 \in W_1 \in M(\Gamma(G_{X_1|W_2,X_{10}^u})) \]
\[ X_2 \in W_2 \in M(\Gamma(G_{X_2|X_1,X_{10}^u})) \]

holds assuming that \( W_{10}^9 \) satisfy (2) and (3) for ordering \( O \).

Since \( W_{10}^9 \) satisfy (2), from Claim 1 we have

\[ W_1 - X_1 - (X_{10}^2, W_4, W_6^7, W_9) \] (32)
\[ W_2 - X_2 - (X_1, X_{10}^1, W_1, W_3^4, W_6^7, W_9) \] (33)

and since \( W_{10}^9 \) satisfy (3) we have

\[ X_1 \in W_1 \in M(\Gamma(G_{X_1|X_2,X_{10}^u})) \]
\[ X_2 \in W_2 \in M(\Gamma(G_{X_2|X_1,X_{10}^u})). \]

To prove that \( W_1 \in M(\Gamma(G_{X_1|W_2,X_{10}^u})) \), we need to show that for any \( w_1 \in W_1, x_1, x'_1 \in w_1, x_2, x'_2 \in X_2, x_{10}^1 \in X_{10}^1, \) and \( w_2 \in W_2 \) such that

\[ p(x_1, x_2, X_{10}^1, w_2) \cdot p(x'_1, x'_2, X_{10}^1, w_2) > 0 \] (34)

we have

\[ f(x_1, x_2, x_{10}^1) = f(x'_1, x'_2, x_{10}^1). \] (35)
Note that (34), the fact that \(x_1, x'_1 \in w_1\), and the Markov chain (32) imply that
\[
p(x_1, x_2, x_3^{10}, w_1) \cdot p(x'_1, x'_2, x_3^{10}, w_1) > 0.
\]
This together with the facts that \(x_2, x'_2 \in w_2\) (which can be deduced from \(X_2 \in W_2\) and (34)) and \(W_2 \in M(G_{X_2|W_1,X_3^{10}})\) implies (35).

- To prove that \(W_2 \in M(G_{X_2|X_1,X_3^{10}})\), we need to show that for any \(w_2 \in W_2, x_2, x'_2 \in w_2, x_1 \in X_1, \) and \(x_3^{10} \in X_3^{10}\) such that
\[
p(x_1, x_2, x_3^{10}) \cdot p(x_1, x'_2, x_3^{10}) > 0,
\]
we have
\[
f(x_1, x_2, x_3^{10}) = f(x_1, x'_2, x_3^{10}).
\]
Since \(P(X_1 \in W_1) = 1\), there exists \(w_1 \in W_1\) such that \(p(w_1|x_1) > 0\). Then, using (36) and Markov chain (32) yields
\[
p(x_1, x_2, x_3^{10}, w_1) \cdot p(x_1, x'_2, x_3^{10}, w_1) > 0.
\]

From the definition of \(G_{X_2|W_1,X_3^{10}}\) we then deduce that equality (37) holds.

In general, to show that \(W_{V \setminus \{r\}}\) satisfies condition (3) for any ordering \(O'\) it suffices to use the same arguments as above repeatedly. In more details, suppose \(W_{V \setminus \{r\}}\) satisfy (3) for an ordering \(O\) (over some given tree). Observe that any \(O'\) can be obtained from \(O\) by a sequence of neighbors’ swaps (transpositions)—in the above example \(O'\) is obtained from \(O\) with one swap. To show that (3) also holds for an ordering \(O'\) one repeats the same arguments as above over the sequence of neighbor swaps that brings \(O\) to \(O'\). This completes the proof.

D. Proof of Claim 3

For notational simplicity, for a set \(S\), define \(XW_S \overset{\text{def}}{=} (X_S, W_S)\) and \(xWS \overset{\text{def}}{=} (x_S, W_S)\). To prove the claim, we show that the Markov chain
\[
W_{\text{Child}(u)} - X_{\text{Child}(u)} - (X_{\text{Child}(u)}\cdot W_{\text{Sub}(u)\setminus \text{Child}(u)})
\]
holds for any vertex \(u\). Having shown this, we get
\[
p(xW_{\text{Child}(u)}, \ldots, xW_{\text{Child}(u^{n(u)})}, x_{\text{Child}(u)}, \cdot W_{\text{Sub}(u)\setminus \text{Child}(u)}|x_u)
\]
\[
= p(x_{\text{Child}(u)}, \ldots, x_{\text{Child}(u^{n(u)})}, x_{\text{Child}(u)}|x_u) \cdot \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u) \cdot \prod_{i=1}^{n(u)} \frac{p(x_{\text{Child}(u^{i})}|x_u, x_{\text{Child}(u^{i})})}{p(x_{\text{Child}(u^{i})}|x_u)}
\]
\[
\overset{(a)}{=} (\prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u)) \cdot \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u) \cdot \prod_{i=1}^{n(u)} \frac{p(x_{\text{Child}(u^{i})}|x_u, x_{\text{Child}(u^{i})})}{p(x_{\text{Child}(u^{i})}|x_u)}
\]
\[
\overset{(b)}{=} \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u) \cdot \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u) \cdot \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u, x_{\text{Child}(u^{i})})
\]
\[
= \prod_{i=1}^{n(u)} p(x_{\text{Child}(u^{i})}|x_u) \cdot p(x_{\text{Child}(u)}, \cdot W_{\text{Sub}(u)\setminus \text{Child}(u)}|x_u)
\]
where (a) follows from the Markov property (Definition 5) and where (b) follows from a repeated use of (38) for the vertices in
\[ N_u \cup \{ \text{Sub}(u) \setminus \text{Child}(u) \} \]
with respect to their ordering values. This completes the proof.

We now establish that (38) holds for any \( u \) by induction. For \( u = 1 \), the Markov chain (38) reduces to
\[ W_u - X_u - X_{\text{Child}(u)} = \]
which is the same as the Markov chain (2) for \( u = 1 \).

Assuming (38) holds for \( u = i, 1 \leq i \leq k - 1 \), we show that the Markov chain (38) holds for \( u = k \).

Write \( \text{Sub}(u) \setminus \text{Child}(u) \) as
\[ \text{Sub}(u) \setminus \text{Child}(u) = \text{Child}(v_1) \cup \text{Child}(v_2) \cup \cdots \cup \text{Child}(v_l) \]
with
\[ \text{Child}(v_i) \cap \text{Child}(v_j) = \emptyset \quad 1 \leq i < j \leq l \]
where \( l \) depends on \( u \) and the ordering. We then have
\[
p(w_{\text{Child}(u)}, w_{\text{Sub}(u) \setminus \text{Child}(u)} | x, y) \\
= p(w_{\text{Child}(u)} \setminus \{u\}, w_{\text{Sub}(u) \setminus \text{Child}(u)} | x, y) \cdot p(w_u | w_{\text{Child}(u)} \setminus \{u\}, w_{\text{Sub}(u) \setminus \text{Child}(u)}, x, y) \\
\stackrel{(a)}{=} p(w_{\text{Child}(u)} \setminus \{u\}, w_{\text{Sub}(u) \setminus \text{Child}(u)} | x, y) \cdot p(w_u | w_{\text{Child}(u)} \setminus \{u\}, x_{\text{Child}(u)}) \\
= p(w_{\text{Child}(u) \setminus \{u\}}, \cdots, w_{\text{Child}(u) \setminus \{u\}}, w_{\text{Child}(v_1)}, \cdots, w_{\text{Child}(v_l)} | x, y) \cdot p(w_u | w_{\text{Child}(u)} \setminus \{u\}, x_{\text{Child}(u)}) \\
\stackrel{(b)}{=} p(w_{\text{Child}(u) \setminus \{u\}}, \cdots, w_{\text{Child}(u) \setminus \{u\}} | x_{\text{Child}(u)}) \cdot p(w_{\text{Child}(v_1)}, \cdots, w_{\text{Child}(v_l)} | x, y) \cdot p(w_u | w_{\text{Child}(u)} \setminus \{u\}, x_{\text{Child}(u)}) \\
= p(w_{\text{Child}(u)} | x_{\text{Child}(u)}) \cdot p(w_{\text{Sub}(u) \setminus \text{Child}(u)} | x, y)
\]
which implies that
\[
p(w_{\text{Child}(u)} | x, w_{\text{Sub}(u) \setminus \text{Child}(u)}) = p(w_{\text{Child}(u)} | x_{\text{Child}(u)})
\]
which shows the validity of the Markov chain (38) for \( u = k \).

Equality (a) holds because of (2) and equality (b) follows from a repeated use of (38) for vertices \( u_1, \cdots, u_{n(u)}, v_1, \cdots, v_l \) with respect to their ordering values—these Markov chains hold by the induction assumption. This completes the proof of Claim 3.

E. Proof of Claim 4

Suppose that
\[ G_{X_{\text{Child}(k)} | X_{\text{Sub}(k)}, W_{\text{Root}(k)}} = G_{X_{\text{Child}(k)} | X_{\text{Child}(k)}^c} \]

\[ (40) \]
holds. Then, we get

\[
H(G_{\text{Child}(k)}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)})
\]

\[
= \min_{V - X_{\text{Child}(k)} - (X_{\text{Sup}(k)}, W_{\text{Roots}(k)})} I(X_{\text{Child}(k)}; V|X_{\text{Sup}(k)}, W_{\text{Roots}(k)})
\]

\[
= \min_{V - X_{\text{Child}(k)} - (X_{\text{Sup}(k)}, W_{\text{Roots}(k)})} H(V|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}) - H(V|X_{\text{Child}(k)})
\]

\[
= \min_{V - X_{\text{Child}(k)} - (X_{\text{Sup}(k)}, W_{\text{Roots}(k)})} H(V|X_{\text{Sup}(k)}) - H(V|X_{\text{Child}(k)})
\]

Equality \((a)\) holds since the Markov chains

\[
V - X_{\text{Child}(k)} - (X_{\text{Sup}(k)}, W_{\text{Roots}(k)})
\]

and

\[
X_{\text{Child}(k)} - X_{\text{Sup}(k)} - W_{\text{Roots}(k)},
\]

which are due to Claim 3, imply the Markov chain \(V - X_{\text{Sup}(k)} - W_{\text{Roots}(k)}\). Equality \((b)\) holds by the Markov property (Definition 5). Finally \((c)\) holds due to \((40)\).

We now show the graph equality \((40)\). First observe that the vertex sets in these two graphs are the same and equal to \(\mathcal{X}_{\text{Child}(k)}\). It remains to show that any two vertices \(x_{\text{Child}(k)}\) and \(x'_{\text{Child}(k)}\) in \(G_{X_{\text{Child}(k)}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}}\) are connected if and only if they are connected in \(G_{X_{\text{Child}(k)}|X_{\text{Child}(k)}^c}\).

- Suppose that \(x_{\text{Child}(k)}\) and \(x'_{\text{Child}(k)}\) are connected in \(G_{X_{\text{Child}(k)}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}}\). Due to Definition 1 this means that there exist \(x_{\text{Child}(k)}^c\) and \(w_{\text{Roots}(k)}\) such that

\[
p(x_{\text{Child}(k)}, x_{\text{Child}(k)}^c, w_{\text{Roots}(k)}) \cdot p(x'_{\text{Child}(k)}, x_{\text{Child}(k)}^c, w_{\text{Roots}(k)}) > 0 \tag{41}
\]

and

\[
f(x_{\text{Child}(k)}, x_{\text{Child}(k)}^c) \neq f(x'_{\text{Child}(k)}, x_{\text{Child}(k)}^c). \tag{42}
\]

Inequality \((41)\) yields

\[
p(x_{\text{Child}(k)}, x_{\text{Child}(k)}^c) \cdot p(x'_{\text{Child}(k)}, x_{\text{Child}(k)}^c) > 0
\]

which, together with \((42)\), implies that \(x_{\text{Child}(k)}\) and \(x'_{\text{Child}(k)}\) are connected in \(G_{X_{\text{Child}(k)}|X_{\text{Child}(k)}^c}\).
• Suppose that \(x_{\text{Child}(k)}\) and \(x'_{\text{Child}(k)}\) are connected in \(G_{X_{\text{Child}(k)}|X_{\text{Child}(k)}'}\). Due to Definition 1, this means that there exists \(x_{\text{Child}(k)}'\) such that

\[
p(x_{\text{Child}(k)}, x_{\text{Child}(k)}') \cdot p(x'_{\text{Child}(k)}, x_{\text{Child}(k)}') > 0 \quad (43)
\]

and

\[
f(x_{\text{Child}(k)}, x_{\text{Child}(k)}') \neq f(x'_{\text{Child}(k)}, x_{\text{Child}(k)}'). \quad (44)
\]

Inequality (43) and the Markov chain

\[X_{\text{Child}(k)} - X_{\text{Child}(k)}' - W_{\text{Roots}(k)}\]

obtained from Claim 3 imply that there exists \(w_{\text{Roots}(k)}\) such that

\[p(x_{\text{Child}(k)}, x_{\text{Child}(k)}', w_{\text{Roots}(k)}) \cdot p(x'_{\text{Child}(k)}, x_{\text{Child}(k)}', w_{\text{Roots}(k)}) > 0.\]

This together with (44) implies that \(x_{\text{Child}(k)}\) and \(x'_{\text{Child}(k)}\) are connected in

\[G_{X_{\text{Child}(k)}|X_{\text{Sup}(k)}, W_{\text{Roots}(k)}'}\].

F. Proof of Claim 5

The Markov chain follows from

\[
p(w_k' | w_{N_k}^*, x_k, x_{\text{Child}(k)}', w_{\text{Sub}(k)} \setminus \text{Child}(k))
\]

\[= \sum_{x_{\text{Child}(k)} \setminus \{k\}} p(w_k' | w_{N_k}^*, x_{\text{Child}(k)}, x_{\text{Child}(k)}', w_{\text{Sub}(k)} \setminus \text{Child}(k)) \cdot p(x_{\text{Child}(k)} \setminus \{k\} | w_{N_k}^*, x_k, x_{\text{Child}(k)}', w_{\text{Sub}(k)} \setminus \text{Child}(k))
\]

\[(a) = \sum_{x_{\text{Child}(k)} \setminus \{k\}} p(w_k' | w_{N_k}^*, x_{\text{Child}(k)}) \cdot p(x_{\text{Child}(k)} \setminus \{k\} | w_{N_k}^*, x_k, x_{\text{Child}(k)}', w_{\text{Sub}(k)} \setminus \text{Child}(k))
\]

\[(b) = \sum_{x_{\text{Child}(k)} \setminus \{k\}} p(w_k' | w_{N_k}^*, x_{\text{Child}(k)}) \cdot p(x_{\text{Child}(k)} \setminus \{k\} | w_{N_k}^*, x_k)
\]

\[= p(w_k' | w_{N_k}^*, x_k),
\]

where (a) is true due to the Markov chain

\[W_k' = X_{\text{Child}(k)} - (X_{\text{Child}(k)}', W_{\text{Sub}(k)}^*)\]

which can be deduced from Definition (21); and where (b) follows from the Markov chains (2) and Claim 3 applied to the vertices \(\text{Sub}(k)\).

G. Proof of Claim 6

Suppose \((w_{k_1}, \ldots, w_{k_{n(k)}}, x_k) \in B_w\) and \(p(x_{\text{Child}(k_i)}, w_{k_i}) > 0, 1 \leq i \leq n(k)\). The first term together with the definition of \(B_w\) implies that there exists \(x'_{\text{Child}(k_i)} \setminus \{k_i\} \in X_{\text{Child}(k_i) \setminus \{k_i\}}\) such that \(p(x'_{\text{Child}(k_i)}, w_{k_i}) > 0, 1 \leq i \leq n(k)\), and \((x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(n(k))}, x_k) \in w'\).
For any \( x_{\text{Child}(k)^c} \in X_{\text{Child}(k)^c} \) and \( w_{\text{Roots}(k)} \in W_{\text{Roots}(k)}^n \) such that

\[
p(x_{\text{Child}(k_1)}, \ldots, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}, w_{\text{Roots}(k)}) \cdot p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}, w_{\text{Roots}(k)}) > 0,
\]

we have

\[
f(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}) \overset{(a)}{=} f(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c})
\]

\[
= f(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-2})}, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c})
\]

\[
= f(x_{\text{Child}(k_1)}, \ldots, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}).
\]

We justify equality \((a)\) since the other can be deduced similarly. Inequality (45) yields

\[
p(x_{\text{Child}(k_1)}, \ldots, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}) \cdot p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}) > 0.
\]

Due to the Markov property (Definition 5), the above inequality can be re-written as

\[
p(x_k) \prod_i p(x_{\text{Child}(k_i)} | x_k) \cdot \prod_i p(x'_{\text{Child}(k_i)} | x_k) \cdot p(x_{\text{Child}(k)^c} | x_k) > 0
\]

which implies

\[
p(x_k) \cdot p(x_{\text{Child}(k_{n-1})} | x_k) \cdot \prod_i p(x'_{\text{Child}(k_i)} | x_k) \cdot p(x_{\text{Child}(k)^c} | x_k) > 0.
\]

(47)

Using the Markov property, (47) can be re-written as

\[
p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}) \cdot p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}) > 0.
\]

(48)

By combining (48), the fact that

\[
p(x_{\text{Child}(k_{n-1})}, w_{k_{n-1}}) \cdot p(x'_{\text{Child}(k_{n-1})}, w_{k_{n-1}}) > 0
\]

and the Markov chain

\[
W_{k_{n-1}} - X_{\text{Child}(k_{n-1})} - X_{\text{Child}(k_{n-1})^c}
\]

deduced from Claim 3, imply

\[
p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}, w_{k_{n-1}}) \cdot p(x'_{\text{Child}(k_1)}, \ldots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)^c}, w_{k_{n-1}}) > 0.
\]

(49)

Inequality (49) and the Markov chain

\[
W_{\text{Roots}(k_{n-1})} - X_{\text{Child}(k_{n-1})^c} - (W_{k_{n-1}}, X_{\text{Child}(k_{n-1})})
\]
deduced from Claim 3, imply that there exists \( w_{\text{Roots}(k_n(k))} \in W^*_n \) such that
\[
p(x'_{\text{Child}(k_1)}, \cdots, x'_{\text{Child}(k_{n-1})}, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)}), w_{k_n(k)}, w_{\text{Roots}(k_n(k))}) > 0.
\]

From this inequality and
\[
w_{k_n(k)} \in \Gamma(G_{X_{k_n(k)}, W^*_n|_{\text{Roots}(k_n(k))}}, w_{\text{Roots}(k_n(k))})
\]
we get
\[
f(x'_{\text{Child}(k_1)}, \cdots, x'_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)}) = f(x'_{\text{Child}(k_1)}, \cdots, x'_{\text{Child}(k_{n-1})}, x_{\text{Child}(k_{n-1})}, x_k, x_{\text{Child}(k)})
\]
This justifies equality (a) in (46).

From (45) and (46) the vertices
\[
(x_{\text{Child}(k_1)}, \cdots, x_{\text{Child}(k)})
\]
and
\[
(x'_{\text{Child}(k_1)}, \cdots, x'_{\text{Child}(k)})
\]
are not connected in \( G_{X_{\text{Child}(k)}|_{\text{Sup}(k)}}, W^*_n \). From Claim 8 stated thereafter (and proved in Appendix I) we deduce that any maximal independent set in \( G_{X_{\text{Child}(k)}|_{\text{Sup}(k)}}, W^*_n \) that includes \((x'_{\text{Child}(k_1)}, \cdots, x'_{\text{Child}(k)}, x_k)\) should also include \((x_{\text{Child}(k_1)}, \cdots, x_{\text{Child}(k)}, x_k)\). Hence we have \((x_{\text{Child}(k_1)}, \cdots, x_{\text{Child}(k)}, x_k) \in w\).

**Claim 8.** Suppose that
\[
(x_{\text{Child}(k)}, x_k), (x'_{\text{Child}(k)}), (x''_{\text{Child}(k)}) \in X_{\text{Child}(k)},
\]
that
\[
(x_{\text{Child}(k)}, x_k) \quad \text{and} \quad (x'_{\text{Child}(k)}, x_k)
\]
are not connected in \( G_{X_{\text{Child}(k)}|_{\text{Sup}(k)}}, W^*_n \), and that
\[
p(x_{\text{Child}(k)}, x_k) \cdot p(x'_{\text{Child}(k)}, x_k) \cdot p(x''_{\text{Child}(k)}) \geq 0.
\]
Then, \((x_{\text{Child}(k)}, x_k)\) and \((x''_{\text{Child}(k)})\) are connected in the graph \( G_{X_{\text{Child}(k)}|_{\text{Sup}(k)}}, W^*_n \) if and only if

**H. Proof of Claim 7**

The distribution of \( W^*_n \) and the fact that \( w^*_n \) and \( B_{w^*_n} \) are in one-to-one correspondence guarantee that \( W^*_n \) satisfies a. and c. We now show \( W^*_n \) also satisfies b.

- \((W^*_n, X_k) \in W^*_n\): We show that \((w^*_n, x_k) \in B_{w^*_n} \) assuming that
\[
p(W^*_n = B_{w^*_n}|W^*_n = w^*_n, X_k = x_k) > 0.
\]
By the definitions of $W^*_k$ and $W'_k$ we have

$$
p(W_k^* = B_{w_k'} | W^*_N = w_N^*, X_k = x_k) = p(W_k' = w_k' | W^*_N = w_N^*, X_k = x_k) > 0.
$$

Claim 5 says that

$$W_k' = (X_k, W^*_N) - (X_{Child(k)} \subseteq W^*_N \setminus Child(k),$$

and Claim 3 then implies that there exists $x_{Child(k)} \subseteq k$ such that $x_{Child(k)} \subseteq w_k'$ and

$$p(x_{Child(k)}, w_k') > 0, i \in \{1, 2, \ldots, n(k)\}.$$

Hence, $(w_N^*, x_k) \in B_{w_k'}$. 

- $W^*_k \in \Gamma^*(G_{W^*_N \cdot X_k | X_{Sup(k)}}, W^*_N)$: Consider $w_k^* = B_{w_k'} \in \mathcal{W}^*$ and

$$p(w_N^*(1), x_k, w_k^*) \cdot p(w_N^*(2), x_k', w_k^*) > 0. \quad (50)$$

We now show that for any $x_{Sub(k)}, x'_{Sub(k)} \in \mathcal{X}_{Sub(k)}, X_{Sup(k)} \subseteq \mathcal{X}_{Sup(k)}$, and $w_{Roots(k)} \subseteq \mathcal{W}_{Roots(k)}$ such that

$$p(w_N^*(1), x_{Sub(k)}, x_k, x_{Sup(k)}, w_{Roots(k)}) \cdot p(w_N^*(2), x'_{Sub(k)}, x_k', x_{Sup(k)}, w_{Roots(k)}) > 0, \quad (51)$$

we have

$$f(x_{Sub(k)}, x_k, x_{Sup(k)}) = f(x'_{Sub(k)}, x_k', x_{Sup(k)}). \quad (52)$$

Note that (50) and the distribution of $W^*_k$ imply that

$$p(w_N^*(1), x_k, w_k') \cdot p(w_N^*(2), x_k', w_k') > 0.$$

This, (51), and the Markov chain

$$W_k' = (X_k, W^*_N) - (X_{Child(k)} \subseteq W^*_N \setminus Child(k))$$

obtained from Claim 5, imply that

$$p(w_k', x_{Sub(k)}, x_k, x_{Sup(k)}, w_{Roots(k)}) \cdot p(w_k', x_k', x_{Sub(k)}, x_k', x_{Sup(k)}, w_{Roots(k)}) > 0.$$

From this inequality and the fact that $w_k' \in \Gamma^*(G_{X_{Child(k)} | X_{Sup(k)}}, W^*_N)$ we deduce (52). We just showed that $w_k^*$ is an independent set. We now show that it is maximal by way of contradiction.

Let $w'$ be a maximal independent set in

$$G_1 \overset{\text{def}}{=} G_{X_{Child(k)} | X_{Sup(k)}, W^*_N}$$

such that

$$w_k^* = B_{w_k'}.$$

Suppose that $w \overset{\text{def}}{=} w_k^*$ is a subset of vertices that is not maximal in the graph

$$G_2 \overset{\text{def}}{=} G_{W^*_N \cdot X_k | X_{Sup(k)}, W^*_N}.$$. 
This means that $G_2$ contains a vertex $v \notin w$ that is not connected to any of the vertices in $w$. The fact that $v \notin w$ together with the definition of $B_{w'}$ implies that there exists a vertex $q$ in $G_1$ such that $q \notin w'$ and $p(q,v) > 0$. Because of the latter and since $v$ is not connected to any of the vertices in $w$ we deduce that $q$ is not connected to any vertex in $w'$ from the definition of $G_2$ and Claim 3. Finally, since $q \notin w'$ and since $q$ is not connected to any vertex in $w'$, we deduce that the set of vertices

$$w' \cup \{q\}$$

is an independent set, a contradiction since $w'$ was supposed to be a maximal independent set.

I. Proof of Claim 8

Suppose that $(x_{Child(k)} \backslash \{k\}, x_k)$ and $(x''_{Child(k)}) \in \mathcal{X}_{Child(k)}$ are connected in $G_{X_{Child(k)}|X_{\text{top}(k)},W_{\text{Roots}(k)}}^\tau$. This means that for some $x_{Child(k)^c} \in \mathcal{X}_{Child(k)^c}$ and $w_{\text{Roots}(k)} \in W_{\text{Roots}(k)}^\tau$ such that

$$p(x_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}, w_{\text{Roots}(k)}) \cdot p(x''_{Child(k)}, x_{Child(k)^c}, w_{\text{Roots}(k)}) > 0, \quad (53)$$

we have

$$f(x_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}) \neq f(x''_{Child(k)}, x_{Child(k)^c}). \quad (54)$$

Note that (53) implies

$$p(x_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}, w_{\text{Roots}(k)}) > 0$$

which, using Claim 3, can be re-written as

$$p(x_{Child(k)} \backslash \{k\}, x_k) \cdot p(x_{Child(k)^c}, w_{\text{Roots}(k)} | x_k) > 0.$$  \(\text{This inequality and the claim’s assumption that } p(x'_{Child(k)} \backslash \{k\}, x_k) > 0 \text{ imply}

$$p(x'_{Child(k)} \backslash \{k\}, x_k) \cdot p(x_{Child(k)} \backslash \{k\}, x_k) \cdot p(x_{Child(k)^c}, w_{\text{Roots}(k)} | x_k) > 0,$$

which, using Claim 3, can be re-written as

$$p(x_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}, w_{\text{Roots}(k)}) \cdot p(x'_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}, w_{\text{Roots}(k)}) > 0. \quad (55)$$

Using the claim’s assumption that $(x_{Child(k)} \backslash \{k\}, x_k)$ and $(x'_{Child(k)} \backslash \{k\}, x_k)$ are not connected in

$$G_{X_{Child(k)}|X_{\text{top}(k)},W_{\text{Roots}(k)}^\tau}$$

we get

$$f(x_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}) = f(x'_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}). \quad (56)$$

Now, (54) and (56) yield

$$f(x'_{Child(k)} \backslash \{k\}, x_k, x_{Child(k)^c}) \neq f(x''_{Child(k)}, x_{Child(k)^c}), \quad (57)$$
and (53) and (55) yield
\[ p(x'_{\text{Child}(k) \setminus \{k\}}, x_k, x_{\text{Child}(k)^c}, w_{\text{Roots}(k)}) \cdot p(x''_{\text{Child}(k)}, x_{\text{Child}(k)^c}, w_{\text{Roots}(k)}) > 0. \] (58)

From (57) and (58) we conclude that \((x'_{\text{Child}(k) \setminus \{k\}}, x_k)\) and \((x''_{\text{Child}(k)}) \in \mathcal{X}_{\text{Child}(k)}\) are also connected.

**J. Jointly Typical Sequences**

Let \((x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n\). Define the empirical probability mass function of \((x^n, y^n)\) (or its type) as
\[ \pi_{x^n, y^n}(x, y) \overset{\text{def}}{=} \left\{ \frac{1}{n} : (x_i, y_i) = (x, y) \right\} \] \[ (x, y) \in (\mathcal{X}, \mathcal{Y}). \]

Let \((X, Y) \sim p(x, y)\). The set of jointly \(\varepsilon\)-typical \(n\)-sequences is defined as
\[ \mathcal{A}_{\varepsilon}^{(n)}(X, Y) \overset{\text{def}}{=} \{(x^n, y^n) : |\pi_{x^n, y^n}(x, y) - p(x, y)| \leq \varepsilon \cdot p(x, y) \text{ for all } (x, y) \in (\mathcal{X}, \mathcal{Y})\}. \]

Also define the set of conditionally \(\varepsilon\)-typical \(n\)-sequences as
\[ \mathcal{A}_{\varepsilon}^{(n)}(Y|x^n) \overset{\text{def}}{=} \{y^n : (x^n, y^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y)\}. \]

Jointly typical sequences satisfy the following properties:

**Lemma 2** ([26, Corollary 2], [9, Page 27]). For any \(\varepsilon > 0\) the following claims hold:

a. Let \((X^n, Y^n) \sim \prod_{i=1}^n p_{X,Y}(x_i, y_i)\). Then, for \(n\) large enough we have
\[ P((X^n, Y^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y)) \geq 1 - \delta(\varepsilon) \]
where \(\delta(\varepsilon) \to 0\) as \(\varepsilon \to 0\).

b. For \(n\) large enough we have \((1 - \delta(\varepsilon))2^n H(X, Y)(1 - \varepsilon) \leq |A_{\varepsilon}^{(n)}(X, Y)| \leq 2^n H(X, Y)(1 + \varepsilon)\).

c. Let \(p(x^n, y^n) = \prod_{i=1}^n p_{X,Y}(x_i, y_i)\). Then, for each \((x^n, y^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y)\)
   i. \(x^n \in \mathcal{A}_{\varepsilon}^{(n)}(X)\) and \(y^n \in \mathcal{A}_{\varepsilon}^{(n)}(Y)\);
   ii. \(p_{X,Y}(x_i, y_i) > 0\) for all \(1 \leq i \leq n\);
   iii. \(2^{-n H(X,Y)(1+\varepsilon)} \leq p(x^n, y^n) \leq 2^{-n H(X,Y)(1-\varepsilon)}\);
   iv. \(2^{-n H(X|Y)(1+\varepsilon)} \leq p(x^n|y^n) \leq 2^{-n H(X|Y)(1-\varepsilon)}\).

**Lemma 3** (Conditional Typicality Lemma, [26, Lemma 22], [9, Page 27]). Fix \(0 < \varepsilon' < \varepsilon\), let \((X,Y) \sim p(x,y)\) and suppose that \(x^n \in \mathcal{A}_{\varepsilon}^{(n)}(X)\) and \(Y^n \sim p(y^n|x^n) = \prod_{i=1}^n p_{Y|X}(y_i|x_i)\). Then, for \(n\) large enough
\[ P((x^n, Y^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y)) \geq 1 - \delta(\varepsilon, \varepsilon') \]
where \(\lim_{\varepsilon \to 0} \lim_{\varepsilon' \to 0} \delta(\varepsilon, \varepsilon') = 0\).

**Lemma 4** (Markov Lemma, [26, Lemma 23]). Let \(X - Y - Z\) form a Markov chain. Suppose that \((x^n, y^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y)\) and \(Z^n \sim p(z^n|y^n) = \prod_{i=1}^n p_{Z|Y}(z_i|y_i)\). Then, for \(\varepsilon > \varepsilon'\) and \(n\) large enough
\[ P((x^n, y^n, Z^n) \in \mathcal{A}_{\varepsilon}^{(n)}(X, Y, Z)) \geq 1 - \delta(\varepsilon, \varepsilon'). \]
**Lemma 5** ([26, Corollary 4]). Let \( p_{X,Y}(x, y) \) have marginal distributions \( p_X(x) \) and \( p_Y(y) \) and let \( (X, Y) \sim p_{X,Y}(x, y) \). Let \( (X', Y') \sim \prod_{i=1}^{n} p_X(x'_i) \cdot p_Y(y'_i) \). Then, for \( n \) large enough

\[
(1 - \delta(\varepsilon)) \cdot 2^{-n(I(X;Y) + 2\varepsilon H(Y))} \leq P((X', Y') \in \mathcal{A}_n^{(n)}(X, Y)) \leq 2^{-n(I(X;Y) - 2\varepsilon H(Y))}.
\]

**Lemma 6** (Covering Lemma, [9, Lemma 3.3]). Let \( (X, \hat{X}) \sim p_{X,\hat{X}}(x, \hat{x}) \). Let \( X^n \sim \prod_{i=1}^{n} p_X(x_i) \) and

\[
\{\hat{X}^n(m), m \in \mathcal{B}\} \quad \text{with} \quad |\mathcal{B}| \geq 2^{nR}
\]

be a set of random sequences independent of each other and of \( X^n \), each distributed according to \( \prod_{i=1}^{n} p_{\hat{X}}(\hat{x}(m)) \). Then,

\[
\lim_{n \to \infty} P((X^n, \hat{X}^n(m)) \notin \mathcal{A}_n^{(n)}(X, \hat{X}^n) \text{ for all } m \in \mathcal{B}) = 0,
\]

if

\[
R > I(X; \hat{X}) + \delta(\varepsilon).
\]