On the Relationship Between Edge Removal and Strong Converses

Oliver Kosut
School of Electrical, Computer and Energy Engineering
Arizona State University
Tempe, AZ 85287
Email: okosut@asu.edu

Jörg Kliewer
Department of Electrical and Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102
Email: jkliewer@njit.edu

***Abstract*—This paper explores the relationship between two ideas in network information theory: edge removal and strong converses. Edge removal properties state that if an edge of small capacity is removed from a network, the capacity region does not change too much. Strong converses state that, for rates outside the capacity region, the probability of error converges to 1. Various notions of edge removal and strong converse are defined, depending on how edge capacity and residual error probability scale with blocklength, and relations between them are proved. In particular, each class of strong converse implies a specific class of edge removal. The opposite direction is proved for deterministic networks, and some discussion is given for the noisy case.**

I. Introduction

Consider a general network communication scenario given an arbitrary collection of sources and sinks connected via arbitrary communication links. The sources are independent and each source is demanded by a subset of sinks, where this subset can be different for each sink. A general interest in network information theory is to determine the capacity of such networks, defined by the set of achievable rates for each source. As this problem is known to be challenging, we consider the simpler problem of how the capacity of these networks change if only a single edge is removed from the network. This problem has first been studied by [1], [2]. The authors have shown that for acyclic noiseless networks and a variety of demand types for which the cut-set bound is tight, removing an edge of capacity δ the capacity of every min-cut is reduced by at most δ in each dimension. Further, in [3] it has been shown for a noiseless multiple multicast demand that this edge removal property also holds for generalized network sharing [4] and linear programming [5] outer bounds. In addition, the existence of the edge removal property has for example been tied to the problem whether a network coding instance allows a reconstruction with ϵ or zero error [6], [7], respectively. Another example is the connection of edge removal to the equivalency between a network coding instance and a corresponding index coding problem [8]. Despite significant progress has been made to understand the scenarios in which the edge removal property holds, the solution to the general problem is open.

In this work, we address the connection of edge removal to the existence of strong converses, a connection that has been explored only briefly in [9, Chap. 3]. The strong converse theorem states that the error probability converges to one for large blocklengths n if the rate exceeds the capacity. This is in stark contrast to a weak converse which only indicates that the error probability is bounded away from zero if we operate at a rate beyond capacity. The benefit of strong converses are that they strengthen the interpretation of capacity as a sharp phase transition. They also allow the following interesting interpretation: if a strong converse exists for a given network instance, ϵ reliable codes (i.e., codes which allow reconstruction with ϵ error) must have rate tuples within the capacity region for $\epsilon \in [0, 1)$ and large n. For example, strong converses have been established for point-to-point settings, e.g., for discrete memoryless channels and quantum channels. Recently it has been shown that a strong converse holds for a subset of discrete memoryless networks with tight cut-set bounds [10], as deterministic relay networks with no interference or a certain class of wireless erasure networks.

In the following, we categorize the notions of edge removal and strong converses into different classes depending on how edge capacity and residual error probability, resp., scale with blocklength and demonstrate the relations between these instances. In particular, we show that each specific class of strong converses always implies a specific class of edge removal, independently whether the network consists of deterministic or probabilistic links. We also show that the opposite direction either holds only for deterministic networks or special cases of memoryless networks. See Fig. 1 for a summary of our results.

This material is based upon work supported by the National Science Foundation under grants CCF-1422358 and CNS-1526547.

II. Model and Definitions

Notation: For an integer k we define $[1 : k] = \{1, \dots, k\}$. All logarithms and exponentials have base 2. For sequences a_n, b_n, we write $a_n \doteq b_n$ if $\log(a_n)/n$ and $\log(b_n)/n$ have the same limit as $n \to \infty$.

A. Network Model

We begin with a very general network model. Many of our results apply only for discrete memoryless networks or deterministic networks, but some basic results apply in much more generality.

978-1-5090-1805-5/16/$31.00 ©2016 IEEE

Consider a network consisting of d nodes, where node $i \in [1:d]$ wishes to convey a message W_i at rate R_i to a set of destination nodes $\mathcal{D}_i \subseteq [1:d]$.[1] The channel model consists of:

- An input alphabet $\mathcal{X}_i$ for each $i \in [1:d]$,
- An output alphabet $\mathcal{Y}_i$ for each $i \in [1:d]$,
- For each time step t, a conditional probability measure given by

$$p(y_{1t},\dots,y_{dt}|y_1^{t-1},\dots,y_d^{t-1},x_1^t,\dots,x_d^t). \quad (1)$$

Definition 1: A channel model is *memoryless and stationary* if the probability measure in (1) can be written

$$p(y_{1t},\dots,y_{dt}|x_{1t},\dots,x_{dt})$$

and these distributions are the same for all t. A channel is *deterministic* if the probability measure in (1) is always deterministic.

For any $\mathbf{R}=(R_1,\dots,R_d)\in\mathbb{R}^d$, an $(\mathbf{R},n)$ code consists of

- for each node $i\in[1:d]$ and time $t\in[1:n]$, an encoder

$$\phi_{it}:[1:2^{nR_i}]\times\mathcal{Y}_i^{t-1}\to\mathcal{X}_i$$

- for each node $i\in[1:d]$, a decoder

$$\psi_i:[1:2^{nR_i}]\times\mathcal{Y}_i^n\to\prod_{j:i\in\mathcal{D}_j}[1:2^{nR_j}].$$

Assume messages W_i for $i=1,\dots,d$ are independent and each uniformly distributed over $[1:2^{nR_i}]$. The channel input from node i at time t is given by $X_{it}=\phi_{it}(W_i,Y_i^{t-1})$. For $i\in\mathcal{D}_j$, the estimate of W_j at node i is given by $\hat{W}_{ji}=\psi(W_i,Y_i^n)$. Given an $(\mathbf{R},n)$ code, the probability of error is

$$P_e^{(n)}=\mathbb{P}\{\hat{W}_{ji}\neq W_j \text{ for some } j\in[1:d], i\in\mathcal{D}_j\}.$$

For blocklength n and $\epsilon\in[0,1]$, let $\mathcal{R}^{(n)}(\mathcal{N},\epsilon)\subseteq\mathbb{R}^d$ be the set of rates $\mathbf{R}$ for which there exists an $(\mathbf{R},n)$ code with probability of error at most ϵ. For a sequence $\epsilon_n\in[0,1]$, let

$$\mathcal{R}(\mathcal{N},\epsilon_n)=\overline{\bigcup_{n_0\in\mathbb{N}}\bigcap_{n\geq n_0}\mathcal{R}^{(n)}(\mathcal{N},\epsilon_n)}$$

where $\overline{(\cdot)}$ indicates closure. We will assume throughout this paper that ϵ_n has the property that $\lim_{n\to\infty}-\frac{1}{n}\log(1-\epsilon_n)$ exists. In other words, there exists α where $1-\epsilon_n\doteq 2^{-n\alpha}$.

For fixed ϵ, $\mathcal{R}(\mathcal{N},\epsilon)$ denotes the capacity region with asymptotic error probability ϵ. With some abuse of notation, define the usual asymptotically-zero-error capacity region as

$$\mathcal{R}(\mathcal{N},0^+)=\bigcap_{\epsilon>0}\mathcal{R}(\mathcal{N},\epsilon).$$

B. Strong Converses

We define three different versions of the strong converse, beginning with the usual strong converse.

Definition 2: Network $\mathcal{N}$ satisfies the *strong converse* if $\mathcal{R}(\mathcal{N},\epsilon)=\mathcal{R}(\mathcal{N},0^+)$ for all $\epsilon\in(0,1)$.

The following definition originates from [9, Chap. 3].

[1]We assume for simplicity that at most one message originates at each node; all results can be easily generalized to the scenario in which multiple messages originate at each node.

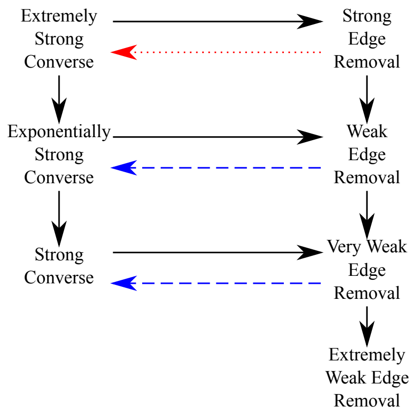

Fig. 1. Diagram showing the relationships between various strong converses and edge removal properties. Solid black lines represent implications that always hold. Dashed blue lines represent implications that we conjecture to hold for all memoryless networks, but have proven only for two special cases (namely, deterministic networks and those made up of independent point-to-ponit links). The dotted red line is an implication that holds for deterministic networks, but does not hold in general for noisy networks.

Definition 3: Network $\mathcal{N}$ satisfies the *exponentially strong converse* if, for any sequence $\epsilon_n\in(0,1)$ for which $-\log(1-\epsilon_n)\in o(n)$, $\mathcal{R}(\mathcal{N},\epsilon_n)\subseteq\mathcal{R}(\mathcal{N},0^+)$.

It is equivalent to say that the exponentially strong converse holds if, for any rate vector $\mathbf{R}\notin\mathcal{R}(\mathcal{N},0^+)$, the probability of error approaches 1 exponentially fast.

Definition 4: Network $\mathcal{N}$ satisfies the *extremely strong converse* if there exists a constant K depending only on the network $\mathcal{N}$ such that, for any ϵ_n satisfying $1-\epsilon_n\doteq 2^{-n\alpha}$ for $\alpha>0$, if $\mathbf{R}\in\mathcal{R}(\mathcal{N},\epsilon_n)$ then $\mathbf{R}-K\alpha\in\mathcal{R}(\mathcal{N},0^+)$.

In other words, the extremely strong converse holds if, for rates outside the capacity region, the probability of error approaches 1 at an exponential rate linear in the distance to the capacity region.

C. Edge Removal Properties

We consider a modified network $\mathcal{N}'$ defined as follows: Start with $\mathcal{N}$, and add two nodes denoted a and b.[2] For each node i in $\mathcal{N}$, add an infinite capacity link from i to a, and an infinite capacity link from b to i. Finally, add a link from a to b that can noiselessly transmit k_n bits total, where n is the blocklength of the code and k_n/n need not be constant.[3] Let $\mathcal{R}(\mathcal{N},\epsilon_n,k_n)\triangleq\mathcal{R}(\mathcal{N}',\epsilon_n)$ for this $\mathcal{N}'$. For any k_n, it is certainly true that $\mathcal{R}(\mathcal{N},\epsilon_n)\subseteq\mathcal{R}(\mathcal{N},\epsilon_n,k_n)$. Note that $\mathcal{R}(\mathcal{N},\epsilon_n,0)=\mathcal{R}(\mathcal{N},\epsilon_n)$,

Roughly, edge removal properties state that for small k_n, the capacity of network $\mathcal{N}'$ is not too different from that of $\mathcal{N}$. To be precise, we define four different versions of this property, beginning with the weakest.

Definition 5: Network $\mathcal{N}$ satisfies the *extremely weak edge removal property* if, for any constant integer k, $\mathcal{R}(\mathcal{N},0^+,k)=\mathcal{R}(\mathcal{N},0^+)$.

[2]These are special nodes in that messages do not originate at them. Thus the capacity region of $\mathcal{N}'$ has the same dimension as that of $\mathcal{N}$.

[3]We focus on the single link from a to b, rather than an arbitrary link in the network. Since node a has complete knowledge of every signal sent in the network, this link can be used to simulate any other small-capacity link.

Definition 6: Network $\mathcal{N}$ satisfies the *very weak edge removal property* if

$$\bigcap_{\epsilon>0} \bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N},\epsilon,k) = \mathcal{R}(\mathcal{N},0^+).$$

Definition 7: Network $\mathcal{N}$ satisfies the *weak edge removal property* if, for any $k_n \in o(n)$, $\mathcal{R}(\mathcal{N},0^+,k_n) = \mathcal{R}(\mathcal{N},0^+)$.

Definition 8: Network $\mathcal{N}$ satisfies the *strong edge removal property* if there exists a constant K depending only on the network $\mathcal{N}$ such that for all $\delta > 0$, if $\mathbf{R} \in \mathcal{R}(\mathcal{N},0^+,\delta n)$, then $\mathbf{R} - K\delta \in \mathcal{R}(\mathcal{N},0^+)$.

The following proposition justifies the names of the edge removal properties.

Proposition 1: The strong edge removal property implies the weak edge removal property, which implies the very weak edge removal property, which implies the extremely weak edge removal property.

Proof: We may write

$$\bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N},0^+,k) = \bigcup_{k\in\mathbb{N}} \bigcap_{\epsilon>0} \mathcal{R}(\mathcal{N},\epsilon,k) \tag{2}$$
$$\subseteq \bigcap_{\epsilon>0} \bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N},\epsilon,k) \tag{3}$$
$$\subseteq \bigcap_{\epsilon>0} \bigcap_{k_n:k_n\to\infty} \mathcal{R}(\mathcal{N},\epsilon,k_n) \tag{4}$$
$$= \bigcap_{k_n:k_n\to\infty} \mathcal{R}(\mathcal{N},0^+,k_n) \tag{5}$$
$$\subseteq \bigcup_{k_n\in o(n)} \mathcal{R}(\mathcal{N},0^+,k_n) \tag{6}$$
$$\subseteq \bigcap_{\delta>0} \mathcal{R}(\mathcal{N},0^+,\delta n) \tag{7}$$

where (2) holds by definition, (3) holds by reversing the order of the intersection and union, (4) holds since $k \le k_n$ for sufficiently large n for any $k \in \mathbb{N}$ and $k_n \to \infty$, (5) holds by definition, (6) holds because the intersection in (5) includes at least one sequence in $o(n)$, and (7) holds because $k_n \le \delta n$ for sufficiently large n for any $\delta > 0$ and $k_n \in o(n)$.

Note that if strong edge removal holds, for any $\mathbf{R}$ in the RHS of (7), $\mathbf{R} - K\delta \in \mathcal{R}(\mathcal{N},0^+)$ for all $\delta > 0$. Since $\mathcal{R}(\mathcal{N},0^+)$ is closed, this implies $\mathbf{R} \in \mathcal{R}(\mathcal{N},0^+)$. Therefore, since the RHS of (6) is contained in the RHS of (7), strong edge removal implies weak edge removal. Since the RHS of (3) is contained in the RHS of (6), weak edge removal implies very weak edge removal. Similarly, since the LHS of (2) is contained in the RHS of (3), very weak edge removal implies extremely weak edge removal. ■

III. Deriving Edge Removal Properties from Strong Converses

The following theorem states that each of the three strong converse properties implies one of the edge removal properties. This result holds for our most general network model.

Theorem 2: For any network:

1) The strong converse implies very weak edge removal.
2) The exponentially strong converse implies weak edge removal.
3) The extremely strong converse implies strong edge removal.

Statement (2) of this theorem was proved for noiseless networks in [9, Sec. 3.3]. Our proof is a generalized version of theirs. We begin with the following lemma.

Lemma 3: For any integers n and k and any $\epsilon \in [0,1]$,

$$\mathcal{R}^{(n)}(\mathcal{N},\epsilon,k) \subseteq \mathcal{R}^{(n)}(\mathcal{N},1-(1-\epsilon)2^{-k}). \tag{8}$$

Proof: Let $\mathbf{R} \in \mathcal{R}^{(n)}(\mathcal{N},\epsilon,k)$, so there is an n-length code with rate vector $\mathbf{R}$ and probability of error at most ϵ on network $\mathcal{N}'$ with bit-pipe between node a and node b carrying k bits. We convert this code to one on network $\mathcal{N}$ as follows. Under the code on $\mathcal{N}'$, let X_{ab} be the message sent on the link from node a to node b. Recall that $X_{ab} \in \{0,1\}^k$. Let $\mathcal{E}$ be the overall error event for network $\mathcal{N}'$. We have

$$1-\epsilon \le \mathbb{P}(\mathcal{E}^c) = \sum_{x_{ab}\in\{0,1\}^k} \mathbb{P}(X_{ab}=x_{ab})\mathbb{P}(\mathcal{E}^c|X_{ab}=x_{ab}).$$

There must be some $x^*_{ab} \in \{0,1\}^k$ for which

$$\mathbb{P}(X_{ab}=x^*_{ab})\mathbb{P}(\mathcal{E}^c|X_{ab}=x^*_{ab}) \ge (1-\epsilon)2^{-k}.$$

Construct a code for network $\mathcal{N}$ that behaves exactly like the original code on network $\mathcal{N}'$, except that all nodes assume that node b received the signal x^*_{ab}. Let P_e be the probability of error for this code. Note that with probability $\mathbb{P}(X_{ab}=x^*_{ab})$, the code's behavior will be just as if the code on $\mathcal{N}'$ were in effect. Thus

$$1-P_e \ge \mathbb{P}(X_{ab}=x^*_{ab})\mathbb{P}(\mathcal{E}^c|X_{ab}=x^*_{ab}) \ge (1-\epsilon)2^{-k}.$$

Therefore $\mathbf{R} \in \mathcal{R}^{(n)}(\mathcal{N},1-(1-\epsilon)2^{-k})$. ■

Proof of Theorem 2: We first show statement (1). Assume the strong converse holds. Thus

$$\bigcap_{\epsilon>0} \bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N},\epsilon,k) \subseteq \bigcap_{\epsilon>0} \bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N},1-(1-\epsilon)2^{-k}) \tag{9}$$
$$\subseteq \bigcup_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N},\tilde{\epsilon}) = \mathcal{R}(\mathcal{N},0^+) \tag{10}$$

where (9) follows from Lemma 3 and the equality in (10) follows from the strong converse. Therefore, very weak edge removal holds.

We now prove statement (2). Assume the exponentially strong converse holds. For any $k_n \in o(n)$, we have

$$\mathcal{R}(\mathcal{N},0^+,k_n) = \bigcap_{\epsilon>0} \mathcal{R}(\mathcal{N},\epsilon,k_n)$$
$$\subseteq \bigcap_{\epsilon>0} \mathcal{R}(\mathcal{N},1-(1-\epsilon)2^{-k_n}) \tag{11}$$
$$\subseteq \bigcup_{\epsilon_n:-\log(1-\epsilon_n)\in o(n)} \mathcal{R}(\mathcal{N},\epsilon_n) \tag{12}$$
$$\subseteq \mathcal{R}(\mathcal{N},0^+) \tag{13}$$

where (11) follows from Lemma 3, (12) from the fact that $k_n \in o(n)$, and (13) from the exponentially strong converse. Therefore weak edge removal holds.

We now prove statement (3). Assume the extremely strong converse holds. For any $\delta > 0$ we have

$$\mathcal{R}(\mathcal{N}, 0^+, \delta n) = \bigcap_{\epsilon>0} \mathcal{R}(\mathcal{N}, \epsilon, \delta n)$$
$$\subseteq \bigcap_{\epsilon>0} \mathcal{R}(\mathcal{N}, 1-(1-\epsilon)2^{-\delta n}) \quad (14)$$

where (14) follows from Lemma 3. Note that $(1-\epsilon)2^{-\delta n} \doteq 2^{-\delta n}$. Thus if $\mathbf{R} \in \mathcal{R}(\mathcal{N}, 0^+, \delta n)$, then, by the extremely strong converse, $\mathbf{R} - K\delta \in \mathcal{R}(\mathcal{N}, 0^+)$ for some constant K. Therefore strong edge removal holds. ■

IV. Deterministic Networks

For deterministic networks, each of the three implications in Theorem 2 is in fact an equivalence, as stated in the following.

Theorem 4: For any deterministic network:

1) The very weak edge removal property holds if and only if the strong converse holds.
2) The weak edge removal property holds if and only if the exponentially strong converse holds.
3) The strong edge removal property holds if and only if the extremely strong converse holds.

The proof of this theorem is along the lines of [11, Lemma 2]. We first prove the following lemma.

Lemma 5: Given a deterministic network $\mathcal{N}$, there exists a finite-valued function $\eta(\tilde{\epsilon})$ for all $\tilde{\epsilon} \in (0,1)$ and a constant C, such that, for any sequence $\epsilon_n \in [0,1]$ with $1-\epsilon_n \doteq 2^{-n\alpha}$, and any $\tilde{\epsilon} \in (0,1)$, for sufficiently large n

$$\mathcal{R}^{(n)}(\mathcal{N}, \epsilon_n) \subseteq \mathcal{R}^{(n)}(\mathcal{N}, \tilde{\epsilon}, \eta(\tilde{\epsilon}) - C\log(1-\epsilon_n)).$$

Proof: For brevity and clarity, we assume that $R_i > 2\alpha$ for all i. This assumption can be removed by slightly increasing the capacity of the extra edge from a to b.

Fix $\tilde{\epsilon} \in (0,1)$ and define with hindsight

$$k = d + \log\ln\frac{2d}{\tilde{\epsilon}} - \log(1-\epsilon_n).$$

We will show that for sufficiently large n, $\mathbf{R} \in \mathcal{R}^{(n)}(\mathcal{N}, \tilde{\epsilon}, 2dk)$. This will prove the lemma, because $2dk = \eta(\tilde{\epsilon}) - C\log(1-\epsilon_n)$ where $\eta(\tilde{\epsilon}) = 2d^2 + 2d\log\ln\frac{2d}{\tilde{\epsilon}}$ and $C = 2d$.

Since $\mathbf{R} \in \mathcal{R}^{(n)}(\mathcal{N}, \epsilon_n)$, there exists a code with rate $\mathbf{R}$, length n, and probability of error at most ϵ_n. For $i = 1, \ldots, d$, let $\mathcal{W}_i = [2^{nR_i}]$ be the message set for the ith message for this code, and let $\mathcal{W} = \prod_{i=1}^d \mathcal{W}_i$ be the set of complete message vectors $\mathbf{w} = (w_1, \ldots, w_d)$. Let $R = \sum_i R_i$, so $|\mathcal{W}| = 2^{nR}$. Since the network is deterministic, whether or not an error occurs depends entirely on the message vector $\mathbf{w} \in \mathcal{W}$. Let Γ be the subset of $\mathcal{W}$ of message vectors that do not lead to errors. Hence $|\Gamma| \geq 2^{nR}(1-\epsilon_n)$.

We employ a random binning argument in which, for each message i, we partition the 2^{nR_i} messages into partitions of size 2^k. In particular, randomly and uniformly choose a partition $P_i(1), \ldots, P_i(2^{nR_i-k})$, among all such partitions for which $|P_i(\tilde{w}_i)| = 2^k$ for all $\tilde{w}_i$. Defining $\tilde{\mathbf{R}} = \mathbf{R} - \frac{k}{n}$ ensures that the bin indices are elements of the set $\tilde{\mathcal{W}}_i = [1 : 2^{n\tilde{R}_i}]$. Note that since $R_i > 2\alpha$ and $k/n \to \alpha$ as $n \to \infty$, for sufficiently large n, $\tilde{R}_i \geq 0$ for all i. We will proceed to show that $\tilde{\mathbf{R}} \in \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, dk)$. To achieve the original rate $\mathbf{R}$, we can increase the capacity of the extra edge by dk bits and therefore transmit k additional bits for each message. This proves $\mathbf{R} \in \mathcal{R}^{(n)}(\mathcal{N}, \tilde{\epsilon}, 2dk)$.

The code on the extended network proceeds as follows. Each message $\tilde{W}_i$ for $i = 1, \ldots, d$ is transmitted to node a. Node a then looks for an element of

$$\Gamma \cap \prod_{i=1}^d P_i(\tilde{W}_i). \quad (15)$$

If there is no such element, declare an error. Otherwise, let $\mathbf{W} = (W_1, \ldots, W_d)$ be one such element. For each i, let $I_i \in \{1, \ldots, 2^k\}$ be the index of W_i in the set $P_i(\tilde{W}_i)$. Node a transmits $(I_1, \ldots, I_d)$ to node b, which requires dk bits.

At node i, W_i can be determined from $\tilde{W}_i$ and I_i. Subsequently, the code proceeds as if $\mathbf{W}$ were the true message vector in the original rate $\mathbf{R}$ code. When a destination node produces a message estimate $\hat{W}_i$, it determines $\widehat{\tilde{W}}_i \in \tilde{\mathcal{W}}_i$ such that $\hat{W}_i \in P_i\left(\widehat{\tilde{W}}_i\right)$. Since by assumption $\mathbf{W} \in \Gamma$, there is no error as long as (15) is not empty.

For $\tilde{\mathbf{w}} = (\tilde{w}_1, \ldots, \tilde{w}_d)$ let

$$q(\tilde{\mathbf{w}}) = \mathbb{P}\left(\Gamma \cap \prod_{i=1}^d P_i(\tilde{w}_i) = \emptyset\right)$$

where the probability is with respect to the random choice of partitions P_i. In Appendix A, we show that for all $\tilde{\mathbf{w}}$, $q(\tilde{\mathbf{w}}) \leq \tilde{\epsilon}$. This proves that there exists at least one code with probability of error $\tilde{\epsilon}$. ■

Proof of Theorem 4: Theorem 2 proves that each strong converse property implies the edge removal properties, so we only need to prove the opposite directions.

Suppose very weak edge removal holds. For any constant ϵ, applying Lemma 5 gives

$$\mathcal{R}(\mathcal{N}, \epsilon) \subseteq \bigcap_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, \eta(\tilde{\epsilon}) - C\log(1-\epsilon))$$
$$\subseteq \bigcap_{\tilde{\epsilon}>0} \bigcup_{k\in\mathbb{N}} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, k) = \mathcal{R}(\mathcal{N}, 0^+)$$

where the last equality holds by very weak edge removal. Therefore the strong converse holds.

Now suppose weak edge removal holds. For any ϵ_n with $-\log(1-\epsilon_n) \in o(n)$, applying Lemma 5 gives

$$\mathcal{R}(\mathcal{N}, \epsilon_n) \subseteq \bigcap_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, \eta(\tilde{\epsilon}) - C\log(1-\epsilon_n))$$
$$\subseteq \bigcap_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, \sqrt{n} - C\log(1-\epsilon_n))$$
$$= \mathcal{R}(\mathcal{N}, 0^+, \sqrt{n} - C\log(1-\epsilon_n)) = \mathcal{R}(\mathcal{N}, 0^+)$$

where the last equality follows from weak edge removal since $\sqrt{n} - C\log(1-\epsilon_n) \in o(n)$. Therefore the exponentially strong converse holds.

Finally, suppose strong edge removal holds. Take ϵ_n so that $1-\epsilon_n \doteq 2^{-n\alpha}$ for some $\alpha > 0$. Applying Lemma 5 gives

$$\mathcal{R}(\mathcal{N}, \epsilon_n) \subseteq \bigcap_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, \eta(\tilde{\epsilon}) + C\alpha n)$$

$$\subseteq \bigcap_{\tilde{\epsilon}>0} \mathcal{R}(\mathcal{N}, \tilde{\epsilon}, (C+1)\alpha n) = \mathcal{R}(\mathcal{N}, 0^+, (C+1)\alpha n).$$

Thus, if $\mathbf{R} \in \mathcal{R}(\mathcal{N}, \epsilon_n)$, by strong edge removal $\mathbf{R} - K(C+1)\alpha n \in \mathcal{R}(\mathcal{N}, 0^+)$. Therefore the extremely strong converse holds. ∎

V. Remarks on Noisy Networks

Extending our results for deterministic networks to noisy networks has proven difficult in general. We conjecture that two of the three equivalences hold for stationary memoryless networks, as stated below.

Conjecture 1: For stationary memoryless networks:

1) The very weak edge removal property holds if and only if the strong converse holds.
2) The weak edge removal property holds if and only if the exponentially strong converse holds.

We have proven this conjecture (to appear) for the special case of networks composed of independent point-to-point links; i.e. the setup of network equivalence [12].

Note that the above conjecture does not include equivalence between the extremely strong converse and strong edge removal. This is because this equivalence does *not* hold in general. Indeed, even for point-to-point noisy discrete memoryless channels, strong edge removal holds but the *extremely* strong converse does not. The latter fact is derived from [13], which characterizes a function $\alpha(R)$ where, if ϵ_n is the optimal probability of error for a code of rate R that is *greater* than the capacity C of the channel, $1 - \epsilon_n \doteq 2^{-\alpha(R)n}$. For noisy channels, the derivative $\alpha'(C) = 0$, whereas to satisfy the extremely strong converse we would need $\alpha'(C) > 0$.

References

[1] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies," in *Proc. Forty-Eighth Annual Allerton Conference*, Monticello, IL, Oct. 2010.

[2] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2011, pp. 1–5.

[3] E. J. Lee, M. Langberg, and M. Effros, "Outer bounds and a functional study of the edge removal problem," in *Proc. IEEE Information Theory Workshop*, Sevilla, Spain, Sep. 2013, pp. 1–5.

[4] S. U. Kamath, D. N. C. Tse, and V. Anantharam, "Generalized network sharing outer bound and the two-unicast problem," in *Proc. International Symposium on Network Coding (NetCod)*, Beijing, China, Jul. 2011.

[5] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1924–1934, Nov. 1997.

[6] M. Langberg and M. Effros, "Network coding: Is zero error always possible?" in *Proc. Forty-Nine Annual Allerton Conference*, Monticello, IL, Sep. 2011, pp. 1–8.

[7] T. H. Chan and A. Grant, "Network coding capacity regions via entropy functions," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5347–5374, Sep. 2014.

[8] M. F. Wong, M. Langberg, and M. Effros, "On a capacity equivalence between network and index coding and the edge removal problem," in *Proc. IEEE Int. Symposium on Inform. Theory*, Istanbul, Turkey, Jun. 2013, pp. 972–976.

[9] W. Gu, "On achievable rate regions for source coding over networks," Ph.D. dissertation, California Institute of Technology, 2009.

[10] S. L. Fong and V. Y. F. Tang, "Strong converse theorems for classes of multimessage multicast networks: A Rényi divergence approach," [Online] http://arxiv.org/abs/1407.2417, Jul. 2014.

[11] M. Langberg and M. Effros, "Source coding for dependent sources," in *Information Theory Workshop (ITW), 2012 IEEE*, Sept 2012, pp. 70–74.

[12] R. Koetter, M. Effros, and M. Medard, "A theory of network equivalence— part i: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, 2011.

[13] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, no. 1, pp. 82–85, Jan 1979.

Appendix A
Proof that $q(\mathbf{w}) \leq \tilde{\epsilon}$

Define

$$A_1 = \big\{w_1 \in \mathcal{W}_1 : |\{(w_2,..., w_d) : (w_1,..., w_d) \in \Gamma\}| \geq (1-\epsilon_n)2^{n(R_2+\cdots+R_d)-1}\big\}.$$

We have that

$$\begin{aligned}(1-\epsilon_n)2^{nR} &\leq |\Gamma| \\ &= \sum_{w_1 \in \mathcal{W}_1} |\{(w_2,\dots,w_d) : (w_1,\dots,w_d) \in \Gamma\}| \\ &\leq |A_1| 2^{n(R_2+\cdots+R_d)} + (2^{nR_1} - |A_1|)(1-\epsilon_n)2^{n(R_2+\cdots+R_d)-1}.\end{aligned}$$

Thus

$$|A_1| \geq \frac{(1-\epsilon_n)2^{nR} - (1-\epsilon_n)2^{nR-1}}{2^{n(R_2+\cdots+R_d)}(1-(1-\epsilon_n)2^{-1})} \geq (1-\epsilon_n)2^{nR_1-1}.$$

Now, for $i = 2, \dots, d$, define, for all $w_1, \dots, w_{i-1}$, the set

$$A_i(w_1,..., w_{i-1}) = \big\{w_i : |\{(w_{i+1},..., w_d) : (w_1,..., w_d) \in \Gamma\}| \geq (1-\epsilon_n)2^{n(R_{i+1}+\cdots+R_d)-i}\big\}.$$

Using induction and a similar argument as above, it follows that for all i and all $w_1 \in A_1$, $w_2 \in A_2(w_1)$, $\dots$, $w_{i-1} \in A_{i-1}(w_1, \dots, w_{i-2})$,

$$|A_i(w_1, \dots, w_{i-1})| \geq (1-\epsilon_n)2^{nR_i - i}.$$

Fix $\tilde{\mathbf{w}} = (\tilde{w}_1, \dots, \tilde{w}_d)$. For each $i = 1, \dots, d$, define

$$Q_i = \{(w_1, \dots, w_i) : w_j \in P_j(\tilde{w}_j) \cap A_j(w_1, \dots, w_{j-1}) \text{ for all } j \leq i\}.$$

Note that $Q_d \subset \Gamma \cap \prod_{i=1}^d P_i(\tilde{w}_i)$, so

$$\begin{aligned}q(\tilde{\mathbf{w}}) &\leq \mathbb{P}(Q_d = \emptyset) \leq \sum_{i=1}^d \mathbb{P}(Q_i = \emptyset | Q_{i-1} \neq \emptyset) \\ &= \sum_{i=1}^d \frac{\binom{2^{nR_i} - |A_i(w_1,\dots,w_{i-1})|}{2^k}}{\binom{2^{nR_i}}{2^k}} \\ &\leq \sum_{i=1}^d \frac{\binom{2^{R_i n} - (1-\epsilon_n)2^{R_i n - i}}{2^k}}{\binom{2^{R_i n}}{2^k}} \\ &\leq \sum_{i=1}^d \frac{(2^{R_i n} - (1-\epsilon_n)2^{R_i n - i})^{2^k}}{(2^{R_i n} - 2^k)^{2^k}} \\ &\leq \sum_{i=1}^d \exp\left\{-2^{k-d}(1-\epsilon_n)\log e - 2^k \log(1 - 2^{k - R_i n})\right\} \\ &= \sum_{i=1}^d \frac{\tilde{\epsilon}}{2d} \exp\left\{-2^k \log(1 - 2^{k-R_i n})\right\}.\end{aligned}$$

Note that $2^k \doteq 2^{\alpha n}$, so for all i, $2^{2k - R_i n} \doteq 2^{2\alpha n - R_i n} \to 0$ by the assumption that $R_i > 2\alpha$. Thus $2^k \log(1 - 2^{k - R_i n}) \to 0$. Therefore, for sufficiently large n, $q(\tilde{\mathbf{w}}) \leq \tilde{\epsilon}$.