Conveying Discrete Memoryless Sources over Networks: When Are Zero-Rate Edges Dispensable?

Badri N. Vellambi

Institute for Telecommunications Research, University of South Australia
Email: badri.vellambi@unisa.edu.au

***Abstract*—This work investigates the problem of conveying discrete memoryless sources over general networks with specified lossless (vanishing error probability) and/or lossy (average per-symbol distortion) reconstruction demands. It presents two different sufficient conditions under which an edge carrying zero-rate messages is not crucial for the conveyance of the sources, i.e., the demands can be met even when the zero-rate edge is deleted and the rates on other edges are kept unchanged.**

I. Introduction

Single-letter characterizations of rate regions for several multi-terminal source coding problems are known [1]. Using these characterizations, several meaningful properties (e.g., continuity w.r.t source and demand parameters, convexity, slope) of the rate region, and optimal coding strategies can be identified. However, computable characterizations of rate regions for several networks such as the distributed lossy source coding problem are still unknown [2] [3, pp. 261-264].

Recently, Gu *et al.* have demonstrated the continuity of rate regions (w.r.t. both the source p.m.f. and the reconstruction demands) in many classes of networks [4], [5]. Their approach does not require explicit characterizations of the underlying rate regions. In the same vein, we have investigated *zero-rate removability* [6]. This property is similar to continuity; it determines if transmitting messages with asymptotically vanishing rate over an edge is crucial to conveying sources over networks. In other words, it ascertains the following: if a rate point $\mathbf{r} = \{r_e\}_{e\in E}$ with $r_{e^*} = 0$ is achievable for conveying a source over a network with a given set of demands, is $\mathbf{r}' = \{r_e\}_{e\in E\setminus\{e^*\}}$ achievable for conveying the same source over the network with the edge e^* deleted while meeting the same demands? This property was first formulated as the *vanishment conjecture* in [5]. In network coding, the edge-removal problem, which is closely related to zero-rate removability, has been explored in [7], [8]. However, general lossy source coding problems are not covered by these works.

In [6], we presented sufficient conditions for the removability of zero-rate links in the multi-terminal source coding (MTSC) problem. One condition for zero-rate removability was derived using a notion of code *stability* specifically for the MTSC problem. In [9], it was shown by the use of a wringing technique that zero-rate edges are always removable in the MTSC problem. In this work, we investigate zero-rate removability in general network problems involving the conveyance of discrete memoryless sources (DMSs) over a network with a prescribed set of reconstruction demands. We introduce a classification of network problems called *regularity*, which is based on both the network topology and the availability of sources at network nodes. Using this classification, two sufficient conditions for removability of zero-rate edges are derived. The first condition called weak concentration requires the existence of codes whose reconstructions are concentrated around the source realizations in a specific manner. The second condition is based on code stability introduced in [6]; stability requires the codes to meet the demands for any source whose p.m.f. is *close* to that of the given source. In this work, the notion of stability (Def. 1), the sufficient condition (Thm. 2), and its proof are extensions of those in [6, Def. 1, Thm. 2].

The rest of the paper is organized as follows. Section II details the notations used, Section III presents the problem setup and relevant terminologies. Finally, Section IV presents the new results and their proofs.

II. Notations

For $n \in \mathbb{N}$, $n_{\preccurlyeq} \triangleq \{1,\ldots,n\}$. Uppercase letters (e.g., X, Y) denote random variables (RVs), and the respective script versions (e.g., $\mathscr{X}$, $\mathscr{Y}$) denote their alphabets, which are assumed to be finite. Lowercase letters denote realizations of RVs (e.g., x, y). Subscripts denote components of vectors, e.g., $x_{n_{\preccurlyeq}}$ represents a vector of length n, $x_{n_{\preccurlyeq}} \triangleq (x_1,\ldots,x_n)$. For sources, superscripts within brackets refer to source indices, e.g., $X_{n_{\preccurlyeq}}^{(l_{\preccurlyeq})}$ represents l sources and at all times from 1 to n. Thus, $X_{n_{\preccurlyeq}}^{(l_{\preccurlyeq})} \triangleq \{(X_i^{(1)},\ldots,X_i^{(l)})\}_{i\in n_{\preccurlyeq}}$. Superscripts without brackets are used only with alphabets to denote Cartesian products, i.e., $\mathscr{X}^n = \mathscr{X}\times\cdots\times\mathscr{X}(n \text{ times})$. For RVs X,Y,Z, we denote $X - Y - Z$ iff $I(X;Z|Y) = 0$. We denote $\eth$ to be the normalized Hamming distance operator, i.e., for $\mathbf{x},\mathbf{y}\in\mathscr{X}^n$, $\eth(\mathbf{x},\mathbf{y}) = \frac{|\{i: x_i\neq y_i\}|}{n}$. As in [10], when $\eth$ operates over two p.m.f.'s p,q over the same set,

$$\eth(p,q) \triangleq \min_{(X,Y)\sim r_{XY}: r_X = p, r_Y = q} \mathbb{E}\,\eth(X,Y) \tag{1}$$

III. Problem Definition

A five-tuple $(G, p_{X^{(1)},\cdots,X^{(l)}}, \mathtt{S}, \mathtt{D}_0, \mathtt{D})$ defines a network problem. Here, $G = G(V,E)$ is the network modeled by a *directed acyclic graph* without parallel edges or self-loops. Edges E enable lossless and instantaneous communication between nodes V. A set of l correlated sources $X^{(1)},\ldots,X^{(l)}$ with p.m.f. $p_{X^{(1)},\cdots,X^{(l)}}$ define the DMS to be conveyed over

G. Source availability is given by $\mathtt{S}: V \to 2^{l_{\preccurlyeq}}$, i.e., $X^{(j)}$ is available at $w \in V$ iff $j \in \mathtt{S}(w)$. Reconstruction demands are given by two sets $\mathtt{D}_0$ and $\mathtt{D}$. A pair $(j, w) \in \mathtt{D}_0$ iff $X^{(j)}$ is required losslessly at $w \in V$. Similarly, $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$ iff $X^{(j)}$ is required at $w \in V$ with an average per-letter distortion not exceeding $\Delta_{j,w}$ under the distortion measure $d^{(j),w}$. For $w \in V$, $\mathtt{D}(w)$ denotes the sources required at w without referring to the details of reconstruction demands.

A. Class of Block Network Codes

Given a network problem $\mathtt{P} = (G, p_{X^{(1)} \cdots X^{(l)}}, \mathtt{S}, \mathtt{D}_0, \mathtt{D})$, we say that a rate point $\mathbf{r} = \{r_e\}_{e \in E} \in \mathbb{R}^{+|E|}$ is achievable for $\mathtt{P}$ if for any $\iota > 0$, we can find $\varepsilon, \delta \in (0, \iota)$ and an (ε, δ)-block code operating at $\mathbf{r}$. By an (ε, δ)-block code at $\mathbf{r}$, we mean a tuple $(n, T, T_h, T_t, \{f_i\}_{i=1}^{|E|}, \{g_{j,w}\}_{w \in V, j \in \mathtt{D}(w)})$ s.t.:

A1. $n \in \mathbb{N}$ refers to the blocklength of the code,

A2. $T_h, T_t : \{1, \ldots, |E|\} \to V$ are invertible functions and $T(i) \triangleq (T_h(i), T_t(i)) \in E$ for all $i \in \{1, \ldots, |E|\}$. For any instant i, $T_h(i)$, $T_t(i)$ denote the head and the tail of an edge over which a message will be sent, and $T(i)$ denotes this edge. These functions determine the order of generation and transmission of network messages.

A3. Sets $\mathscr{M}_t \subseteq \mathbb{N}$ for $t = 1, \ldots, |E|$ s.t. $\frac{\log_2 |\mathscr{M}_t|}{n} \leq r_{T(t)} + \varepsilon$,

A4. For each $i \in \{1, \ldots, |E|\}$, encoding functions f_i s.t.

$$\begin{aligned} f_i &: \Big(\bigtimes_{j \in \mathtt{S}(T_h(i))} (\mathscr{X}^{(j)})^n \Big) \to \mathscr{M}_i && \text{if } \mathcal{I}_i = \emptyset \\ f_i &: \Big(\bigtimes_{j \in \mathtt{S}(T_h(i))} (\mathscr{X}^{(j)})^n \Big) \times \Big(\bigtimes_{t \in \mathcal{I}_i} \mathscr{M}_t \Big) \to \mathscr{M}_i && \text{if } \mathcal{I}_i \neq \emptyset \end{aligned},$$

where $\mathcal{I}_i \triangleq T_t^{-1}(T_h(i)) \cap \{1, \ldots, i-1\}$; $\mathcal{I}_i$ denotes the set of edges ending in $T_h(i)$ whose messages have already been computed. Thus, at the i^{th} instant, the message M_i is computed using f_i from available source realizations, and incoming messages previously computed using the code. This message M_i is then sent on the edge $T(i)$.

A5. For $w \in V$ and source $j \in \mathtt{D}(w)$, reconstruction function $g_{j,w} : \Big(\bigtimes_{k \in \mathtt{S}(w)} (\mathscr{X}^{(k)})^n \Big) \times \Big(\bigtimes_{t \in T_t^{-1}(w)} \mathscr{M}_t \Big) \to \Big(\widehat{\mathscr{X}}^{(j),w} \Big)^n$ generates output $\widehat{X}_{n_{\preccurlyeq}}^{(j),w}$ satisfying

$$\Pr[\widehat{X}_{n_{\preccurlyeq}}^{(j),w} \neq X_{n_{\preccurlyeq}}^{(j)}] \leq \delta, \tag{2}$$

if $(j, w) \in \mathtt{D}_0$. Otherwise, $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$, and

$$\tfrac{1}{n} \sum_{i=1}^{n} \mathbb{E}\, d^{(j),w}(X_i^{(j)}, \widehat{X}_i^{(j),w}) \leq \Delta_{j,w} + \delta. \tag{3}$$

For a network problem, we define the rate region as the set of all rate points achievable in the above sense. Given a network problem, we say that $S \subseteq E$ is *zero-rate removable* (ZRR) at $\mathbf{r} = \{r_e\}_{e \in E}$, if (a) $\mathbf{r}$ is achievable; (b) $r_e = 0$ for each $e \in S$; and (c) $\mathbf{r}' \triangleq \{r_e\}_{e \in E \setminus \{S\}}$ is achievable for $G' = G(V, E \setminus S)$ and for the same $\mathtt{S}, \mathtt{D}_0, \mathtt{D}$. Before we present our results, the following terminologies will be necessary.

Given an edge $e = (u, v) \in E$ and a coding scheme $C = (n, T, T_h, T_t, \{f_i\}_{i=1}^{T}, \{g_{j,w}\}_{w \in V, j \in \mathtt{D}(w)})$, we say that the message $M_{T^{-1}(e')}$ generated by $f_{T(e')}$ and sent on the edge $e' = (u', v') \neq e$ is *altered* by e if for some $k \in \mathbb{N}$, there exists a path $(w_0 = v, w_1, \ldots, w_{k-1}, w_k = u', w_{k+1} = v')$ in G with $T^{-1}((w_s, w_{s+1})) < T^{-1}((w_t, w_{t+1}))$ for $0 \leq s < t \leq k$. Message $M_{T^{-1}(e)}$ is, by definition, unaltered by e. In other words, message M_i is altered by an edge e iff it is generated by a sequence of encoding functions, the arguments of (at least) one of which is the message $M_{T^{-1}(e)}$. This notion of alteration ignores the possibility that these functions might not actually use $M_{T^{-1}(e)}$ (i.e., $I(M_i; M_{T^{-1}(e)}) = 0$).

Given $e = (u, v) \in E$, let $\mathcal{P}_e \subseteq V$ denote the set of nodes parent to e, i.e., nodes that have a directed path to u in G. Similarly, define the set of children nodes $\mathcal{C}_e \subseteq V$ by $w \in \mathcal{C}_e \Leftrightarrow G$ contains a path from v to w. We call $\mathfrak{a}_e \triangleq \cup_{w \in \mathcal{P}_e \cup \{u\}} \mathtt{S}(w)$ as the sources *affected* by e. In simple terms, $\mathfrak{a}_e$ contains all the sources that collectively generate the message on the edge e. Similarly, for each $w \in V$, define $\mathcal{P}'_e(w)$ as the set of nodes having a directed path to w not involving e, and term $\mathfrak{u}_e(w) \triangleq \cup_{s \in \mathcal{P}'_e(w) \cup \{w\}} \mathtt{S}(s)$ as the set of sources *unaffected* at w by edge e. Note that a source can be both affected by e and unaffected by e (at some $w \in V$).

A network problem is termed *weakly regular* w.r.t a given $e \in E$ if for each $w \in \mathcal{C}_e$ with $\mathtt{D}(w) \neq \emptyset$, (a) the affected sources $\mathfrak{a}_e$ and the unaffected sources $\mathfrak{u}_e(w)$ are disjoint, i.e., $\mathfrak{a}_e \cap \mathfrak{u}_e(w) = \emptyset$; and (b) if $(j, w) \in \mathtt{D}_0$, then $j \in \mathfrak{u}_e(w)$, i.e, there exists a node with access to $X^{(j)}$ and has a path to w not involving e. Further, a network problem is termed *regular* w.r.t $e \in E$ if it is weakly regular w.r.t. $e \in E$, and for each $w \in \mathcal{C}_e$ with $\mathtt{D}(w) \neq \emptyset$, the conditional p.m.f. of $X^{(\mathfrak{u}_e(w))}$ given $X^{(\mathfrak{a}_e)}$ has full support. In network problems that are weakly regular or regular w.r.t. an edge, say $e \in E$, the set of affected sources $\mathfrak{a}_e$ are effectively *helper* sources, and the message sent on e *help* to meet the demands of nodes in $\mathcal{C}_e$.

We now present the sufficient conditions for removability of a single zero-rate edge. These results extend naturally to multiple zero-rate edges by induction.

IV. New Results

We begin with the following technical result. The result gives conditions to bound $D_{KL}(p_\Gamma p_\Lambda || p_{\Gamma\Lambda})$ for two random variables Γ and Λ satisfying $D_{KL}(p_{\Gamma\Lambda} || p_\Gamma p_\Lambda) \leq n\varepsilon$. This result will enable us to decorrelate messages with small rates from other messages under suitable conditions.

Lemma 1: Let $\{(A_i, B_i)\}_{i=1}^{n}$ be i.i.d. with p.m.f. p_{AB} s.t. $\mu = \min_{a, b: p_B(b) > 0} p_{A|B}(a|b) > 0$. Let $\Gamma = f(A_{n_{\preccurlyeq}})$, and $\Lambda = g(B_{n_{\preccurlyeq}})$, where $f : \mathscr{A}^n \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ for some $R > 0$, and $g : \mathscr{B}^n \to \{1, \ldots, \lfloor 2^{n\varepsilon} \rfloor\}$ for some $\varepsilon > 0$. Then,

$$\Delta_{\Gamma\Lambda} \triangleq D_{KL}(p_\Gamma p_\Lambda || p_{\Gamma\Lambda}) \leq 3n \sqrt[8]{\varepsilon} \log \tfrac{1}{\mu}. \tag{4}$$

Proof: Let $\mathcal{L} \triangleq \{\lambda : D_{KL}(p_{A_{n_{\preccurlyeq}} | \Lambda = \lambda} || p_{A_{n_{\preccurlyeq}}}) \leq n\sqrt{\varepsilon}\}$. Since $I(A_{n_{\preccurlyeq}}; \Lambda) = \sum_\lambda p_\Lambda(\lambda) D_{KL}(p_{A_{n_{\preccurlyeq}} | \Lambda = \lambda} || p_{A_{n_{\preccurlyeq}}}) \leq n\varepsilon$, we have by Markov's inequality [1],

$$\Pr[\Lambda \in \mathcal{L}^c] \leq \tfrac{1}{n\sqrt{\varepsilon}} I(A_{n_{\preccurlyeq}}; \Lambda) \leq \sqrt{\varepsilon}. \tag{5}$$

Fix $\lambda_0 \in \mathcal{L}$. Since $p_{A_{n_{\preccurlyeq}}}$ is a product distribution, [10, Prop. 1] ensures that there exist random variables U, V over $\mathscr{A}^n$ with a joint p.m.f. r_{UV} s.t. $U \sim p_{A_{n_{\preccurlyeq}} | \Lambda = \lambda_0}$, $V \sim p_{A_{n_{\preccurlyeq}}}$ and

$\mathbb{E}_{r_{UV}}\, \bar{d}(U,V) \leq \sqrt[4]{\varepsilon}$. Hence, $\Pr[\bar{d}(U,V) > \sqrt[8]{\varepsilon}] < \sqrt[8]{\varepsilon}$. For notational ease, let $\eta_{\lambda_0} \triangleq D_{KL}(p_{A_{n_\preccurlyeq}} || p_{B_{n_\preccurlyeq}|\Lambda=\lambda_0})$. Then,

$$\eta_{\lambda_0} = \sum_{\mathbf{u},\mathbf{v}\in\mathscr{A}^n} r_{UV}(\mathbf{u},\mathbf{v}) \log \frac{r_V(\mathbf{v})}{r_U(\mathbf{v})} \tag{6}$$

$$= -nH(A) + \sum_{\bar{d}(\mathbf{u},\mathbf{v})\leq \sqrt[8]{\varepsilon}} r_{UV}(\mathbf{u},\mathbf{v}) \log \frac{1}{r_U(\mathbf{v})} + \sum_{\bar{d}(\mathbf{u},\mathbf{v})> \sqrt[8]{\varepsilon}} r_{UV}(\mathbf{u},\mathbf{v}) \log \frac{1}{r_U(\mathbf{v})}. \tag{7}$$

Note that for any $\mathbf{u},\mathbf{v} \in \mathscr{X}^n$, $\lambda \in \{1,\ldots,\lfloor 2^{n\varepsilon}\rfloor\}$, we have

$$\frac{r_U(\mathbf{v})}{r_U(\mathbf{u})} = \frac{\sum\limits_{\mathbf{b}\in g^{-1}(\lambda)} \left(\prod\limits_{i=1}^{n} \frac{p_{AB}(v_i,b_i)p_{AB}(u_i,b_i)}{p_{AB}(u_i,b_i)}\right)}{\sum\limits_{\mathbf{b}\in g^{-1}(\lambda)} p_{A_{n_\preccurlyeq}B_{n_\preccurlyeq}}(\mathbf{u},\mathbf{b})} \geq \mu^{n\bar{d}(\mathbf{u},\mathbf{v})} \tag{8}$$

$$r_U(\mathbf{v}) = \sum_{\mathbf{b}\in g^{-1}(\lambda)} \frac{p_{A_{n_\preccurlyeq}B_{n_\preccurlyeq}}(\mathbf{v},\mathbf{b})}{p_\Lambda(\lambda)} \geq \sum_{\mathbf{b}\in g^{-1}(\lambda)} \frac{\mu^n p_{B_{n_\preccurlyeq}}(\mathbf{b})}{p_\Lambda(\lambda)} = \mu^n. \tag{9}$$

Using (8), (9) in the first and second sum of (7), we get

$$\eta_{\lambda_0} \leq -nH(A) + \left[\sum_{\mathbf{u}} r_U(\mathbf{u}) \log \frac{1}{r_U(\mathbf{u})} + n\sqrt[8]{\varepsilon}\log\frac{1}{\mu}\right] + \left[n\big(\Pr[\bar{d}(U,V) > \sqrt[8]{\varepsilon}]\big)\log\frac{1}{\mu}\right] \tag{10}$$

$$\leq H(A_{n_\preccurlyeq}|\Lambda=\lambda_0) - nH(A) + 2n\sqrt[8]{\varepsilon}\log\frac{1}{\mu}. \tag{11}$$

Now, an application of log-sum inequality [1] to $\Delta_{\Gamma\Lambda}$ yields

$$\Delta_{\Gamma\Lambda} \leq \textstyle\sum_\lambda p_\Lambda(\lambda) D_{KL}(p_{A_{n_\preccurlyeq}} || p_{A_{n_\preccurlyeq}|\Lambda=\lambda}), \tag{12}$$

Substituting (11) in (12), we get

$$\begin{aligned}\Delta_{\Gamma\Lambda} &= \sum_{\lambda\in\mathcal{L}} p_\Lambda(\lambda)\eta_\lambda + \sum_{\lambda\in\mathcal{L}^c}\sum_{\mathbf{a}} p_\Lambda(\lambda) p_{A_{n_\preccurlyeq}}(\mathbf{a}) \log \frac{p_{A_{n_\preccurlyeq}}(\mathbf{a})}{p_{A_{n_\preccurlyeq}|\Lambda}(\mathbf{a}|\lambda)} \\ &\overset{(iv)}{\leq} \sum_{\lambda\in\mathcal{L}} p_\Lambda(\lambda)\eta_\lambda + \sum_{\lambda\in\mathcal{L}^c}\sum_{\mathbf{a}} p_\Lambda(\lambda) p_{A_{n_\preccurlyeq}}(\mathbf{a}) \log \frac{p_{A_{n_\preccurlyeq}}(\mathbf{a})}{\mu^n} \\ &\overset{(v)}{\leq} H(A_{n_\preccurlyeq}|\Lambda) - nH(A) + 2n\sqrt[8]{\varepsilon}\log\frac{1}{\mu} + n\sqrt{\varepsilon}\log\frac{1}{\mu} \\ &\leq 3n\sqrt[8]{\varepsilon}\log\frac{1}{\mu},\end{aligned}$$

where (iv) follows from (9), and (v) from (5) and (11). ■

A. Weak Concentration of Source Reconstructions

In this section, we formalize a notion of concentration of source reconstructions observed in codes built using typicality arguments. Let $F, K : (0,1) \to (0,1)$ be continuous and $\lim\limits_{x\to 0} F(x) = \lim\limits_{x\to 0} K(x) = 0$. We say that (F,K)-weak concentration is attainable at rate point $\mathbf{r} = \{r_e\}_{e\in E}$ if for any $\iota > 0$, there exist $\varepsilon, \delta \in (0,\iota)$ and an (ε,δ)-code operating over n symbols at rate $\mathbf{r}$ satisfying A1-A5. Additionally, the reconstructions of the code are such that for $(j,w) \in \mathtt{D}_0$,

$$\Pr[\bar{d}(\widehat{X}^{(j),w}_{n_\preccurlyeq}, X^{(j)}_{n_\preccurlyeq}) > \delta + F(\varepsilon)] < e^{-nK(\varepsilon)}, \tag{13}$$

and for $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$,

$$\Pr\left[\sum_{i=1}^{n} \frac{d^{(j),w}(X^{(j)}_i, \widehat{X}^{(j),w}_i)}{n} > \Delta_{j,w} + \delta + F(\varepsilon)\right] < e^{-nK(\varepsilon)}. \tag{14}$$

Theorem 1: Let $F : (0,1) \to (0,1)$ with $\lim_{x\downarrow 0} F(x) = 0$, and $K(x) = x^\alpha$ for $\alpha \in (0, \frac{1}{8})$. Let network problem $(G, p_{X^{(1)}\ldots X^{(l)}}, \mathtt{S}, \mathtt{D}_0, \mathtt{D})$ be regular w.r.t. $e \in E$, and let (F,K)-weak concentration be attainable at an $\mathbf{r}$ with $r_e = 0$. Then, e is ZRR at $\mathbf{r}$.

Proof: Let $\iota > 0$. Let $\varepsilon, \delta \in (0,\iota)$ and code $C = (n, T, T_h, T_i\{f_i\}_{i=1}^{|E|}, \{g_{j,w}\}_{w\in V, j\in\mathtt{D}(w)})$ with (F,K)-concentration property be given. Let $i_e = T^{-1}(e)$. We may assume that $n\varepsilon > 1$, else C sends a constant message over e.

Let $C' = (n, T', T'_h, T'_t\{f'_i\}_{i=1}^{|E|}, \{g'_{j,w}\}_{w\in V, j\in\mathtt{D}(w)})$ be a randomized code for the network $G(V,E)$ with $T' = T$, $T'_h = T_h$, $T'_t = T_t$, $f'_i = f_i$ for $i \neq i_e$, and $g'_{j,w} = g_{j,w}$ for $w \in V$, $j \in \mathtt{D}(w)$. The only difference between C and C' is that $f'_{i_e} \neq f_{i_e}$. In C', the message M'_{i_e} sent over e is selected randomly from $\mathscr{M}_i$ using the p.m.f. $p_{M_{i_e}}$, and M'_{i_e} is independent of all sources (and messages M_t for $t < i_e$).

Now, consider $w \in V$ with $\mathtt{D}(w) \neq \emptyset$. If $w \in V \setminus \mathcal{C}_e$, then none of the messages received by w is altered by e. Then, for every realization of $X^{(l_\preccurlyeq)}_{n_\preccurlyeq}$, the reconstruction(s) generated at w using C' and C is/are identical. Hence, C' preserves the demand(s) at node w. Now assume that $w \in \mathcal{C}_e$ and that the code C is employed. Then one of the following cases hold:

Case 1: $(j,w) \in \mathtt{D}_0$. Then, due to the regularity of the problem, $j \notin \mathfrak{a}_e$ and $j \in \mathfrak{u}_e(w)$. In this case, the reconstruction $\widehat{X}^{(j),w}_{n_\preccurlyeq}$ generated by function $g_{j,w}$ can be viewed as a function of $(X^{(\mathfrak{u}_e(w))}_{n_\preccurlyeq}, M_{i_e})$. Let $\mathcal{G} \subseteq \bigtimes_{k\in\mathfrak{u}_e(w)} \big(\mathscr{X}^{(k)}\big)^n$ s.t. $\mathbf{x} \in \mathcal{G}$ iff

$$\Pr[\bar{d}(\widehat{X}^{(j),w}_{n_\preccurlyeq}, X^{(j)}_{n_\preccurlyeq}) > \delta + F(\varepsilon) | X^{(\mathfrak{u}_e(w))}_{n_\preccurlyeq} = \mathbf{x}] < e^{-n\frac{K(\varepsilon)}{2}}. \tag{15}$$

By an application of Markov's inequality on (13), we see that $\Pr[X^{(\mathfrak{u}_e(w))}_{n_\preccurlyeq} \notin \mathcal{G}] < e^{-n\frac{K(\varepsilon)}{2}}$. Now, define for $\mathbf{x} \in \mathcal{G}$,

$$\mathscr{N}_{\mathbf{x}} \triangleq \left\{m \in \mathscr{M}_{i_e} : \bar{d}(\mathbf{x}^{(j)}, \widehat{X}^{(j),w}_{n_\preccurlyeq}(\mathbf{x}, m)) \leq \delta + F(\varepsilon)\right\}. \tag{16}$$

Let $Y_i \triangleq X^{(\mathfrak{u}_e(w))}_i$ to simplify notation. From (15), we have

$$\mathbf{y} \in \mathcal{G} \Leftrightarrow \Pr[M_{i_e} \notin \mathscr{N}_{\mathbf{y}} | Y_{n_\preccurlyeq} = \mathbf{y}] < e^{-n\frac{K(\varepsilon)}{2}}. \tag{17}$$

Note that M_{i_e} is a function of $X^{(\mathfrak{a}_e)}_{n_\preccurlyeq}$. Due to regularity, the conditions of Lemma 1 are met for $A = Y = X^{(\mathfrak{u}_e(w))}$, $\Gamma = Y_{n_\preccurlyeq} = X^{(\mathfrak{u}_e(w))}_{n_\preccurlyeq}$, $B = X^{(\mathfrak{a}_e)}$ and $\Lambda = M_{i_e}$. Therefore, by Lemma 1, (18)-(20) hold. In these equations, σ is a positive constant depending only on the joint p.m.f. $p_{X^{(1)}\ldots X^{(l)}}$, (vi) follows from the log-sum inequality, and (vii) from (17).

In (20), the RHS contains a product of marginals instead of the joint probability $p_{Y_{n_\preccurlyeq}M_{i_e}}$. So, if $Y_{n_\preccurlyeq}$ and M_{i_e} were independent (i.e., M_{i_e} is selected randomly using $p_{M_{i_e}}$) as in C', $\frac{1}{n}\sum_{i=1}^n d^{(j),w}(X^{(j)}_i, \widehat{X}^{(j),w}_i)$ will exceed $\delta + F(\varepsilon)$ only if: (a) $Y_{n_\preccurlyeq} \notin \mathcal{G}$; or (b) $Y_{n_\preccurlyeq} \in \mathcal{G}$ and $M_{i_e} \notin \mathscr{N}_{Y_{n_\preccurlyeq}}$. Hence, $\Pr[\frac{1}{n}\sum_{i=1}^n d^{(j),w}(X^{(j)}_i, \widehat{X}^{(j),w}_i) > \delta + F(\varepsilon)]$, is no more than $\frac{3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma}+1}{0.5nK(\varepsilon)} + e^{-n\frac{K(\varepsilon)}{2}}$. Thus, for this demand, the code C' offers an average Hamming distortion of less than $\delta + F(\varepsilon) + \frac{3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma}+1}{0.5nK(\varepsilon)} + e^{-n\frac{K(\varepsilon)}{2}}$, which vanishes as $(\varepsilon,\delta) \to (0,0)$.

Case 2: $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$. In this case, it is possible that source $X^{(j)}$ is affected by e. If $j \in \mathfrak{a}_e$, Lemma 1 cannot be applied as is, since the conditional p.m.f. of $X^{(j)}_{n_\preccurlyeq}$ given M_{i_e} may not have full support. To bypass this, define a new

$$3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma} \geq \sum_{\mathbf{y},m} p_{Y_{n_\preccurlyeq}}(\mathbf{y}) p_{M_{i_e}}(m) \log \frac{p_{M_{i_e}}(m)}{p_{M_{i_e}|Y_{n_\preccurlyeq}}(m|\mathbf{y})} \tag{18}$$

$$\overset{(vi)}{\geq} \sum_{\mathbf{y}\in\mathcal{G}} p_{Y_{n_\preccurlyeq}}(\mathbf{y}) \Big[\Pr[M_{i_e} \in \mathscr{N}_{\mathbf{y}}] \log \frac{\Pr[M_{i_e}\in\mathscr{N}_{\mathbf{y}}]}{\Pr[M_{i_e}\in\mathscr{N}_{\mathbf{y}}|Y_{n_\preccurlyeq}=\mathbf{y}]} + \Pr[M_{i_e} \notin \mathscr{N}_{\mathbf{y}}] \log \frac{\Pr[M_{i_e}\notin\mathscr{N}_{\mathbf{y}}]}{\Pr[M_{i_e}\notin\mathscr{N}_{\mathbf{y}}|Y_{n_\preccurlyeq}=\mathbf{y}]} \Big]$$

$$\therefore \frac{3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma}+1}{0.5nK(\varepsilon)} \geq \sum_{\mathbf{y}\in\mathcal{G}} \frac{p_{Y_{n_\preccurlyeq}}(\mathbf{y})}{0.5nK(\varepsilon)} \Big[\Pr[M_{i_e} \in \mathscr{N}_{\mathbf{y}}] \log \frac{1}{\Pr[M_{i_e}\in\mathscr{N}_{\mathbf{y}}|Y_{n_\preccurlyeq}=\mathbf{y}]} + \Pr[M_{i_e} \notin \mathscr{N}_{\mathbf{y}}] \log \frac{1}{\Pr[M_{i_e}\notin\mathscr{N}_{\mathbf{y}}|Y_{n_\preccurlyeq}=\mathbf{y}]} \Big] \tag{19}$$

$$\geq \sum_{\mathbf{y}\in\mathcal{G}} \frac{p_{Y_{n_\preccurlyeq}}(\mathbf{y}) \Pr[M_{i_e}\notin\mathscr{N}_{\mathbf{y}}]}{0.5nK(\varepsilon)} \log \frac{1}{\Pr[M_{i_e}\notin\mathscr{N}_{\mathbf{y}}|Y_{n_\preccurlyeq}=\mathbf{y}]} \overset{(vii)}{\geq} \sum_{\mathbf{y}\in\mathcal{G}} p_{Y_{n_\preccurlyeq}}(\mathbf{y}) \Pr[M_{i_e} \notin \mathscr{N}_{\mathbf{y}}], \tag{20}$$

source $X^{(*)}$ by: (a) $\mathscr{X}^{(*)} = \mathscr{X}^{(j)}$; (b) $X^{(*)} - X^{(j)} - X^{(l_\preccurlyeq)}$; and

$$p_{X^{(*)}|X^{(j)}}(x'|x) = \begin{cases} 1-(|\mathscr{X}^{(j)}|-1)\phi(\varepsilon) & x'=x \\ \phi(\varepsilon) & x' \neq x \end{cases},$$

where $\phi(\varepsilon) = e^{-\frac{1}{\varepsilon^{\frac{1}{16}-\frac{\alpha}{2}}}}$. Let $M(\varepsilon) = \frac{\varepsilon^{\frac{\alpha}{2}}}{\phi(\varepsilon)}$. Then, by a simple application of Hœffding's inequality [3, p. 44], we have

$$\Pr\left[\frac{\mathscr{d}(X^{(j)}_{n_\preccurlyeq}, X^{(*)}_{n_\preccurlyeq})}{|\mathscr{X}^{(j)}|-1} > (M(\varepsilon)+1)\phi(\varepsilon)\right] \leq 2e^{-2n\varepsilon^{\alpha}}. \tag{21}$$

Combining (21) with (14), we obtain:

$$\Pr\Big[\sum_{i=1}^{n} d^{(j),w}(X_i^{(*)}, \widehat{X}_i^{(j),w}) \leq n\beta(\varepsilon)\Big] > 1 - 3e^{-n\varepsilon^{\alpha}}, \tag{22}$$

where $\beta(\varepsilon) \triangleq \Delta_{j,w} + \delta + F(\varepsilon) + |\mathscr{X}^{(j)}|(M(\varepsilon)+1)\phi(\varepsilon)$. Similar to Case 1, let $Y_i \triangleq X_i^{(\mathfrak{u}_e(w)\cup\{*\})}$. Then, $\widehat{X}^{(j),w}_{n_\preccurlyeq}$ generated by $g_{j,w}$ can be viewed as a function of $(Y_{n_\preccurlyeq}, M_{i_e})$. Define $\mathcal{G} \subseteq \bigtimes_{k\in\mathfrak{u}_e(w)} \left(\mathscr{X}^{(k)}\right)^n \times \left(\mathscr{X}^{(*)}\right)^n$ by $\mathbf{y}\in\mathcal{G}$ iff

$$\Pr\Big[\sum_{i=1}^{n} d^{(j),w}(X_i^{(*)}, \widehat{X}_i^{(j),w}) \geq n\beta(\varepsilon) \Big| Y_{n_\preccurlyeq} = \mathbf{y}\Big] \leq e^{-\frac{nK(\varepsilon)}{2}}.$$

Then, by Markov's inequality, $\Pr[Y_{n_\preccurlyeq} \notin \mathcal{G}] \leq 3e^{-\frac{nK(\varepsilon)}{2}}$. Define for each $\mathbf{y}\in\mathcal{G}$, $\mathscr{N}_{\mathbf{y}}$ by $m \in \mathscr{N}_{\mathbf{y}}$ iff

$$\sum_{i=1}^{n} d^{(j),w}(X_i^{(*)}, \widehat{X}_i^{(j),w}(\mathbf{y},m)) \leq n\beta(\varepsilon). \tag{23}$$

Applying Lemma 1 with $A = Y = X^{(\mathfrak{u}_e(w)\cup\{*\})}$, $\Gamma = Y_{n_\preccurlyeq} = X^{(\mathfrak{u}_e(w)\cup\{*\})}_{n_\preccurlyeq}$, $B = X^{(\mathfrak{a}_e)}$, and $\Lambda = M_{i_e}$, we see that (18)-(20) hold with σ denoting the smallest element in the conditional p.m.f. of $X^{(\mathfrak{u}_e(w)\cup\{*\})}$ given $X^{(\mathfrak{a}_e)}$. Hence, $\sigma = \phi(\varepsilon)\sigma_p$, where $\sigma_p > 0$ and depends only on $p_{X^{(1)}\cdots X^{(l)}}$. By the definition of $\phi(\varepsilon)$, we see that the LHS of (19) satisfies

$$\frac{3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma}+1}{0.5nK(\varepsilon)} = \frac{1-3n\sqrt[8]{\varepsilon}\log(\phi(\varepsilon)\sigma_p)}{0.5nK(\varepsilon)} = \Theta(\varepsilon^{\frac{1}{16}-\frac{\alpha}{2}}).$$

Hence, just as in Case 1, we note that the per-letter distortion between the reconstruction for $X^{(j)}$ generated by C' using $g'_{j,w}$, and $X^{(*)}_{n_\preccurlyeq}$, i.e., $\frac{1}{n}\sum_{i=1}^{n} d^{(j),w}(X_i^{(*)}, \widehat{X}_i^{(j),w})$, exceeds $\beta(\varepsilon)$ only if: (a) $Y_{n_\preccurlyeq} \notin \mathcal{G}$; or (b) $Y_{n_\preccurlyeq} \in \mathcal{G}$ and $M_{i_e} \notin \mathscr{N}_{Y_{n_\preccurlyeq}}$. Hence, $\Pr[\frac{1}{n}\sum_{i=1}^{n} d^{(j),w}(X_i^{(*)}, \widehat{X}_i^{(j),w}) > \beta(\varepsilon)]$ is bounded above by $3e^{-\frac{nK(\varepsilon)}{2}} + \frac{1-3n\sqrt[8]{\varepsilon}\log(\phi(\varepsilon)\sigma_p)}{0.5nK(\varepsilon)}$, which vanishes with ε. By triangle inequality, it follows that the average distortion between the reconstruction $\widehat{X}^{(j),w}_{n_\preccurlyeq}$ of C' and $X^{(j)}_{n_\preccurlyeq}$ is no more than $\beta(\varepsilon) + |\mathscr{X}^{(j)}|(M(\varepsilon)+1)\phi(\varepsilon)$ with high probability.

Note that the random code C' is a result of time-sharing $|\mathscr{M}_{i_e}|$ codes; each of these codes operates just as C' does, except that they transmit a constant message over e. Due to this time-sharing property and due to weak concentration offered by C', there must exist a code that transmits a constant message over e and offers asymptotically the same performance as C'. Let C'' now refer to the code that operates exactly as C does, except that some $m^* \in \mathscr{M}_{i_e}$ is transmitted constantly over e. This code C'' meets all lossy demands with a slightly larger asymptotically vanishing term than that offered by C. Note that the reconstructions for lossless demands at $w \in \mathcal{C}_e$ only meet the zero-distortion criterion under the respective Hamming measures. However, due to regularity of the network problem, for each $w \in \mathcal{C}_e$ with $(j,w) \in \mathbb{D}_0$, there exists a node $\nu(j,w) \in V$ s.t. $j \in \mathbb{S}(\nu(j,w))$, and $\nu(j,w)$ is connected to w by a path not involving e. By appending a Slepian-Wolf code between $\nu(j,w)$ and w for the source $p_{X^{(j)}_{n_\preccurlyeq}\widehat{X}^{(j),w}_{n_\preccurlyeq}}$, we can realize a lossless reconstruction of $X^{(j)}$ at w. This requires an additional rate of $\frac{1}{n}H(X^{(j)}_{n_\preccurlyeq}|\widehat{X}^{(j),w}_{n_\preccurlyeq})$ over the path between $\nu(j,w)$ and w. Since the average per-symbol Hamming distortion between $X^{(j)}_{n_\preccurlyeq}$ and $\widehat{X}^{(j),w}_{n_\preccurlyeq}$ is less than $\delta + F(\varepsilon) + \frac{3n\sqrt[8]{\varepsilon}\log\frac{1}{\sigma}+1}{0.5nK(\varepsilon)} + e^{-\frac{nK(\varepsilon)}{2}}$, the extra rate vanishes as $(\varepsilon,\delta) \to (0,0)$ (see [4, Lem. A.1, Cor. A.2]). Thus, we have a code that meets the demands and transmits a constant message on e. Lastly, since ι is arbitrary, e is ZRR. ∎

B. Stability

We begin by defining *stability*, a preliminary version of which was introduced in [6]. This notion of stability will be used to develop the next condition of sufficiency.

Definition 1: An achievable rate point $\mathbf{r}$ is (F,K)-*stable* for a given network problem if the following hold:

B1. $F, K : (0,1) \to (0,1)$ are continuous with $\lim_{x\to 0} F(x) = 0$ and $\lim_{x\to 0} K(x) = 0$.

B2. For any $\iota > 0$, there exist $n \in \mathbb{N}$, $\varepsilon, \delta \in (0,\iota)$ and an (ε,δ,F,K)-*stable* code C at $\mathbf{r}$ of length n, by which we mean:

(a) C is an (ε,δ)-block code at rate $\mathbf{r}$ satisfying A1-A5.

(b) If the joint source p is replaced by any p.m.f r over ${\mathscr{X}^{(1)}}^n \times \cdots \times {\mathscr{X}^{(l)}}^n$ with $\mathscr{d}(r, p_{X^{(1)}_{n_\preccurlyeq}\cdots X^{(l)}_{n_\preccurlyeq}}) \leq F(\varepsilon)$,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_r\big[d^{(j),w}(X_i^{(j)}, \widehat{X}_i^{(j),w})\big] \leq \Delta_{j,w} + \delta + K(\varepsilon),$$

for $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$, and if $(j, w) \in \mathtt{D}_0$

$$\mathbb{E}_r\left[\bar{d}(X^{(j)}_{n_\preccurlyeq}, \widehat{X}^{(j),w}_{n_\preccurlyeq})\right] \le \delta + K(\varepsilon).$$

In loose terms, Stable codes are *robust*, i.e., they meet the demands for any block source whose p.m.f. lies within a suitably small neighborhood of $p_{X^{(1)}_{n_\preccurlyeq}\cdots X^{(l)}_{n_\preccurlyeq}}$. This fact will be used to show zero-rate removability at stable rate points.

Theorem 2: Let $K : (0, 1) \to (0, 1)$ with $\lim_{x\downarrow 0} K(x) = 0$ be given. Let for some $\nu \in (0, \frac{1}{2})$, $F(x) \triangleq x^{\frac{1}{2}-\nu}$. Suppose that the network problem $\mathtt{P} = (G, p_{X^{(1)}\cdots X^{(l)}}, \mathtt{S}, \mathtt{D}_0, \mathtt{D})$ is weakly regular w.r.t. $e \in E$, and that a rate point $\mathbf{r}$ with $r_e = 0$ is (F, K)-stable. Then, the edge e is ZRR at $\mathbf{r}$.

Proof: This proof follows closely that of [6, Thm. 2] and only an outline is given here. Consider a weakened/modified network problem $\mathtt{P}' = (G, p_{X^{(1)}\cdots X^{(l)}}, \mathtt{S}, \mathtt{D}'_0, \mathtt{D}')$ with $\mathtt{D}'_0 = \emptyset$ and $\mathtt{D} \subseteq \mathtt{D}'$. In addition to $\mathtt{D}$, for each $(j, w) \in \mathtt{D}_0$, $\mathtt{D}'$ contains a zero Hamming distortion demand for the source $X^{(j)}$ at w. Clearly, $\mathbf{r}$ is also (F, K)-stable for $\mathtt{P}'$.

Let $\iota > 0$ and $\varepsilon, \delta \in (0, \iota)$. By concatenating C with itself sufficiently many times, we can construct for $\mathtt{P}'$ an $(\varepsilon, \delta, F_1, K_1)$-code C_1 of length n_1 with $n_1\varepsilon > \frac{1}{\varepsilon}$, $F_1(x) = 2\sqrt{x}$ and $K_1(x) = K(x) + 2x^\nu$ (see [6, Lemma 1]).

Let $\{f_i\}_{i=1}^{|E|}$ denote the encoding functions employed in C_1. Select $m^* \in \mathscr{M}_{i_e}$ s.t. $\Pr[M_{i_e} = m^*] > 2^{-\frac{2}{\log 2}n_1\varepsilon}$ and let $\mathcal{A}(m^*) \triangleq \{\mathbf{x} \in \times_{k\in\mathfrak{a}_e} (\mathscr{X}^{(k)})^{n_1} : M_{i_e}(\mathbf{x}) = m^*\}$. Here, we have used the fact that $X^{(\mathfrak{a}_e)}_{n_\preccurlyeq}$ uniquely determines M_{i_e}. Let

$$\mathcal{B}(m^*) \triangleq \bigcup_{\tilde{\mathbf{x}}\in\mathcal{A}(m^*)} \left\{\tilde{\mathbf{x}} \in \times_{k\in\mathfrak{a}_e} (\mathscr{X}^{(k)})^{n_1} : \bar{d}(\mathbf{x}, \tilde{\mathbf{x}}) < F_1(\varepsilon)\right\},$$

where $\bar{d}$-distance uses the Hamming distance measure over the set $\times_{k\in\mathfrak{a}_e} \mathscr{X}^{(k)}$. By [10, (1.2)], we have

$$\Pr[X^{(\mathfrak{a}_e)}_{n_{1\preccurlyeq}} \in \mathcal{B}(m^*)] \ge 1 - e^{-2n_1\varepsilon} > 1 - e^{-\frac{2}{\varepsilon}}. \quad (24)$$

Define functions $\xi : \times_{k\in\mathfrak{a}_e} (\mathscr{X}^{(k)})^{n_1} \to \times_{k\in\mathfrak{a}_e} (\mathscr{X}^{(k)})^{n_1}$, and $\Xi : \times_{k\in l_\preccurlyeq} (\mathscr{X}^{(k)})^{n_1} \to \times_{k\in\mathfrak{a}_e} (\mathscr{X}^{(k)})^{n_1}$ by

$$\begin{aligned} \xi(\mathbf{x}) &= \begin{cases} \mathbf{x} & \mathbf{x} \notin \mathcal{B}(m^*) \\ \arg\min_{\tilde{\mathbf{x}}\in\mathcal{A}(m^*)} \bar{d}(\mathbf{x}, \tilde{\mathbf{x}}) & \mathbf{x} \in \mathcal{B}(m^*) \end{cases} \\ \Xi(\mathbf{x}^{(\mathfrak{a}_e)}, \mathbf{y}^{(\mathfrak{a}_e^c)}) &= (\xi(\mathbf{x}^{(\mathfrak{a}_e)}), \mathbf{y}^{(\mathfrak{a}_e^c)}) \end{aligned}$$

Define a p.m.f. r on $\times_{k\in l_\preccurlyeq} (\mathscr{X}^{(k)})^{n_1}$ by $r(\mathbf{x}) = \sum_{\mathbf{y}:\Xi(\mathbf{y})=\mathbf{x}} p(\mathbf{y})$. Further, for any $\mathbf{x} \in \times_{k\in l_\preccurlyeq} (\mathscr{X}^{(k)})^{n_1}$, $\Xi(\mathbf{x})$ and $\mathbf{x}$ differ in less than $n_1 F_1(\varepsilon)$ positions. Hence, $\bar{d}(r, p_{X^{(1)}_{n_{1\preccurlyeq}},\cdots,X^{(l)}_{n_{1\preccurlyeq}}}) < F_1(\varepsilon)$.

Define a new code C' that: (a) uses the same encoding order as C_1; (b) uses the same encoding functions f_i as C_1 for $i \neq i_e$; (c) transmits constantly the message m^* over e; and (d) uses the same reconstruction functions as C_1.

Consider $(j, w, d^{(j),w}, \Delta_{j,w}) \in \mathtt{D}$. If $w \notin \mathcal{C}_e$, the messages received by w are not altered by e, and hence C' and C_1 generate the same reconstructions and hence offer the same performance. Now, let $w \in \mathcal{C}_e$. Let $\widehat{X}^{(j),w,C_1}_{n_{1\preccurlyeq}}(\mathbf{x})$, $\widehat{X}^{(j),w,C'}_{n_{1\preccurlyeq}}(\mathbf{x})$ denote the reconstructions of $X^{(j)}$ at w when $X^{(l_\preccurlyeq)}_{n_{1\preccurlyeq}} = \mathbf{x}$ and when C_1, C' are used, respectively. Then,

$$\widehat{X}^{(j),w,C'}_{n_{1\preccurlyeq}}(\mathbf{x}) = \widehat{X}^{(j),w,C_1}_{n_{1\preccurlyeq}}(\mathbf{x}), \qquad \mathbf{x}^{(\mathfrak{a}_e)} \in \mathcal{A}(m^*) \quad (25)$$

$$\widehat{X}^{(j),w,C'}_{n_{1\preccurlyeq}}(\mathbf{x}) = \widehat{X}^{(j),w,C'}_{n_{1\preccurlyeq}}(\Xi(\mathbf{x})),\ \mathbf{x} \in \times_{k\in l_\preccurlyeq} (\mathscr{X}^{(k)})^{n_1}. \quad (26)$$

Note that (26) follows since: (a) $X^{(j),w,C'}_{n_\preccurlyeq}$ is a function of $X^{(\mathfrak{u}_e(w))}_{n_{1\preccurlyeq}}$, (b) $\mathfrak{a}_e^c \supseteq \mathfrak{u}_e(w)$ due to weak regularity, and (c) Ξ does not alter the components corresponding to $\mathfrak{a}_e^c$. Now, let $\Delta_{C_1,j,w}(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n_1} d^{(j),w}(x_i, \widehat{X}^{(j),w,C_1}_i(\mathbf{x}))$ and let $\Delta_{C',j,w}(\mathbf{x})$ be defined similarly. Then, from (26), we have:

$$\Delta_{C',j,w}(\mathbf{x}) \le F_1(\varepsilon) + \Delta_{C',j,w}(\Xi(\mathbf{x})),\ \text{if } \mathbf{x}^{(\mathfrak{a}_e)} \in \mathcal{B}(m^*) \quad (27)$$

Now, consider $\lambda_{C'} = \sum_{\mathbf{x}} p(\mathbf{x})\Delta(\mathbf{x}, \widehat{X}^{(j),w}_{n_\preccurlyeq}(\mathbf{x}))$. Then,

$$\begin{aligned} \lambda_{C'} &\le \sum_{\mathbf{x}:\mathbf{x}^{(\mathfrak{a}_e)}\in\mathcal{B}(m^*)} p(\mathbf{x})\Delta_{C',j,w}(\mathbf{x}) + \Pr[X^{(\mathfrak{a}_e)}_{n_{1\preccurlyeq}} \notin \mathcal{B}(m^*)] \cdot 1 \\ &\overset{(viii)}{\le} F_1(\varepsilon) + \sum_{\mathbf{x}:\mathbf{x}^{(\mathfrak{a}_e)}\in\mathcal{B}(m^*)} p(\mathbf{x})\Delta_{C',j,w}(\Xi(\mathbf{x})) + e^{-\frac{2}{\varepsilon}} \quad (28) \\ &\overset{(ix)}{=} F_1(\varepsilon) + \sum_{\mathbf{x}:\mathbf{x}^{(\mathfrak{a}_e)}\in\mathcal{A}(m^*)} r(\mathbf{x})\Delta_{C',j,w}(\mathbf{x}) + e^{-\frac{2}{\varepsilon}} \quad (29) \\ &\overset{(x)}{=} F_1(\varepsilon) + \sum_{\mathbf{x}:\mathbf{x}^{(\mathfrak{a}_e)}\in\mathcal{A}(m^*)} r(\mathbf{x})\Delta_{C_1,j,w}(\mathbf{x}) + e^{-\frac{2}{\varepsilon}} \quad (30) \\ &\overset{(xi)}{\le} F_1(\varepsilon) + \Delta_{j,w} + \delta + K_1(\varepsilon) + e^{-\frac{2}{\varepsilon}}, \quad (31) \end{aligned}$$

where $(viii)$ follows from (24) and (27); (ix) from the definition of r; (x) from (25); and (xi) from the stability of C_1 at $\mathbf{r}$. Thus, C' asymptotically meets each demand in $\mathtt{D}'$. Since $\mathtt{P}$ is weakly regular, we can, as in Thm. 1, append to C', Slepian-Wolf codes between specific pairs of network nodes to additionally meet the lossless demands of $\mathtt{D}_0$. The resulting code then conveys a constant message on e and meets the demands of $\mathtt{P}$. Lastly, since ι is arbitrary, e is ZRR. ■

References

[1] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, no. 4-5, pp. 265–444, 2007.

[2] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge University Press, Jan. 2012.

[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Orlando, USA: Academic Press, 1982.

[4] W.-H. Gu, "On achievable rate regions for source coding over networks," Ph.D. dissertation, California Inst. of Tech., 2009.

[5] W.-H. Gu and M. Effros, "On the continuity of achievable rate regions for source coding over networks," *2007 IEEE Inform. Theory Workshop*, pp. 632–637, Sept. 2007.

[6] B. N. Vellambi, "The effect of zero-rate encoders in the multi-terminal source coding problem," *IEEE Intl. Symp. on Inform. Theory*, pp. 756–760, July 2012.

[7] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies," *Forty-Eighth Annual Allerton Conference on Commun., Control, and Computing, Monticello*, pp. 391–398, Sep 2010.

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

[9] B. N. Vellambi, "Multi-terminal source coding: Zero-rate encoders cannot enlarge the rate region," *2012 IEEE Inform. Theory Workshop*, pp. 317–321, Sept. 2012.

[10] K. Marton, "Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration," *The Annals of Probability*, vol. 24, no. 2, pp. 857–866, April 1996.