

The Capacity Region for Degree-2 K -pairs Three-layer Networks

Xijin Yan and Zhen Zhang
 Communication Sciences Institute
 Dept. of Electrical Engineering-Systems
 University of Southern California
 Los Angeles, CA 90089-2565, U.S.A.
 Email: xyan@usc.edu, zzhang@commsci1.usc.edu

Abstract—Ignited by the pioneering work of Ahlswede *et al* on the characterization of the capacity region for single-source multicast networks, a number of works have been devoted to determining the capacity regions for more general networks. However, except a few interesting outer bounds that have been derived so far, the exact characterization is still restricted to rather simple networks. In this paper, we determine the capacity region for a more general class of networks called degree-2 K -pairs three-layer networks. The result suggests a characterization technique for more general multi-source multi-sink networks.

I. INTRODUCTION

Consider a network in which several mutually independent information sources are generated at possibly different nodes, and the data of each information source is multicast to a specific set of sink nodes. We call this type of networks multi-source multi-sink networks. Specifically, if the source set contains only one source node whose data is multicast to the same set of sinks, then the network is called single-source multicast network; otherwise it is called a single-source non-multicast network. If the network contains more than one source and the data from all sources are sent to the same set of sinks, it is called a multi-source multi-sink multicast network and has been treated as a trivial single-source multicast problem [1]; otherwise it is a multi-source multi-sink network with different multicast requirement for each source, which we call the multi-source multi-sink non-multicast case.

In [1], Ahlswede, Cai, Li and Yeung showed that the capacity region of the single-source multicast network can be characterized by the max-flow min-cut bound. Motivated by this work, two natural questions that follow are “how far can we extend the max-flow min-cut bound to more general networks?” and “What is the capacity region for such general networks if the max-flow min-cut bound fails?”.

The earliest answer to the first question dated back to the work by Erez, Feder [5] and Ngai, Yeung [6], where they proved that the max-flow min-cut bound remains tight in single-source two-sink non-multicast networks. Surprisingly, it turns out that, although rather simple, these are by far the most complicated networks that the max-flow min-cut bound can be applied for an exact characterization. In fact, it is shown in [3] and [7], the max-flow min-cut bound is in general quite loose.

For this reason, the second question becomes a major focus of recent research. However, in a general multi-source multi-sink network, information from different sources is coded for subsequent transmissions, the underlying roles played by various types of *side information* [3] significantly complicate the problem. Although several improved bounds [2], [3] and [4] have been discovered recently, they still serve only as outer bounds in exceptional cases. Therefore, a complete answer to the second question is a very challenging task along this line of research.

In this paper, we propose an approach to the second question by a cut-based network decomposition which leads to a special type of networks called three-layer networks and a role-based information characterization which leads to the concept of the so-called side information. Based on this model, we determine the capacity region for a class of networks called degree-2 K -pairs three-layer networks. The obtained result suggests a proof technique that can be used to characterize the capacity region for more general multi-source multi-sink networks.

II. NETWORK MODEL

A. Some Notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an acyclic directed graph, where \mathcal{V} is the set of nodes and \mathcal{E} is the set of directed edges. A pair $(i, j) \in \mathcal{E}$ is an edge directed from node i to node j . Let

$$\begin{aligned} \mathcal{A}_j &= \{i \in \mathcal{V}, (i, j) \in \mathcal{E}\}, \\ \mathcal{B}_j &= \{i \in \mathcal{V}, (j, i) \in \mathcal{E}\}, \end{aligned} \quad (1)$$

and assume that $\mathcal{A}_{s_j} = \phi$ for any source node s_j and $\mathcal{B}_{t_j} = \phi$ for any sink node t_j . We define a cut \mathcal{U} in \mathcal{G} as a subset $\mathcal{U} \subseteq \mathcal{V}$ and the corresponding cut set $\mathcal{E}_{\mathcal{U}} = \{(i, j) \in \mathcal{E} : i \in \mathcal{U}, j \in \mathcal{V} \setminus \mathcal{U}\}$ as the set of edges across \mathcal{U} .

B. Network Model

In this paper, we analyze the type of networks called three-layer networks. A three-layer network is a special network of which all the nodes are lined up in a three-layer architecture. In a three-layer network, each directed edge is assumed to be error free and thus is called an error-free channel. All channels except the coding channels are considered straight

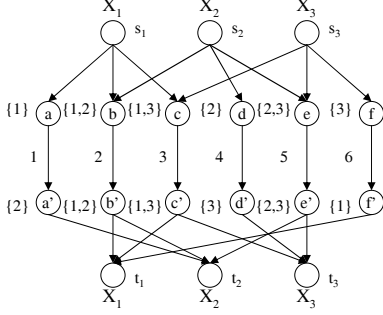


Fig. 1. An example of a degree-2 three-pairs three-layer network, the labeled sets next to nodes i and i' stand for \mathcal{A}_i and $\mathcal{B}_{i'}$ respectively.

connections, therefore there is no constraint on the capacities of these channels.

Definition 1: A three-layer network is specified by the following elements:

- 1) \mathcal{S} , the set of source nodes;
- 2) \mathcal{T} , the set of sink nodes;
- 3) \mathcal{I} , the set of coding channels;
- 4) $\mathcal{A} = \{\mathcal{A}_i : \mathcal{A}_i \subseteq \mathcal{S}, (i, j) \in \mathcal{I}\}$, the set of connections between the source nodes and the coding channels;
- 5) $\mathcal{B} = \{\mathcal{B}_j : \mathcal{B}_j \subseteq \mathcal{T}, (i, j) \in \mathcal{I}\}$, the set of connections between the coding channels and the sink nodes;
- 6) $\mathcal{C} = (C_{ij} : (i, j) \in \mathcal{I})$, the capacity constraints on the coding channels;
- 7) $\mathcal{D} = \{\mathcal{D}_i : t_i \in \mathcal{T}\}$, where $\mathcal{D}_i \in 2^{\mathcal{S}} \setminus \phi$, the reconstruction requirements of the sink nodes.

We assume the sources $\mathbf{X}_j = \{X_{jk}\}_{k=1}^{\infty}$ generated at source nodes $s_j, j = 1, \dots, |\mathcal{S}|$ are independent, where $X_{jk}, k = 1, 2, \dots$ are independently and identically distributed copies of a generic random variable X_j with alphabet $\mathcal{X}_j < \infty$.

Definition 2: A K -pairs three-layer network is a three-layer network with the additional constraints such that $|\mathcal{S}| = |\mathcal{T}| = K$ and $\mathcal{D}_i = \{s_i\}$.

Definition 3: A degree-2 K -pairs three-layer network is a K -pairs three-layer network with the additional constraints such that for all $(i, j) \in \mathcal{I}$,

- either $\mathcal{A}_i = \mathcal{B}_j$ and $|\mathcal{A}_i| \leq 2$,
- or $\mathcal{A}_i \cap \mathcal{B}_j = \phi$ and $|\mathcal{A}_i| = |\mathcal{B}_j| = 1$.

An example of a degree-2 three-pairs three-layer network is shown in Fig. 1.

Definition 4: For any $(i, j) \in \mathcal{I}$, if $\mathcal{A}_i \cap \mathcal{B}_j = \phi$, then the output of (i, j) is said to be *side information* for the sinks \mathcal{B}_j .

We emphasize the importance of the above model with the following observations:

- 1) A three-layer network serves essentially as a cut in a multi-source multi-sink network when cuts are taken to provide constraints on the total amount of information in transmission. Therefore, they provide a natural decomposition to an arbitrary acyclic multi-source multi-sink network.
- 2) A network with arbitrary sink requirements can be decomposed into a set of K -pairs subnetworks by a *sink-*

node decomposition technique in [3]. Although, whether this technique is tight is still an open question, it greatly simplifies the network topology.

- 3) The constraint on degree-2 cross coding gives the simplest case when coding among different sources is allowed. Therefore, it reduces the complexity of the problem and may provide insights into more general cases.

C. Capacity Region for Multi-source Multi-sink Networks

Consider the block code $(n, (\eta_{ij}, (i, j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$ defined in [8]. A rate vector $\mathbf{R} = (R_{ij}, (i, j) \in \mathcal{E})$ is admissible if for every $\epsilon > 0$, there exists for sufficiently large n an $(n, (\eta_{ij}, (i, j) \in \mathcal{E}), (\Delta_l, l \in \mathcal{T}))$ code such that

$$n^{-1} \log \eta_{ij} \leq R_{ij}, \forall (i, j) \in \mathcal{E} \text{ and } \Delta_l \leq \epsilon, \forall l \in \mathcal{T}.$$

Let

$$\mathcal{R} = \{\mathbf{R} : \mathbf{R} \text{ is admissible}\} \quad (2)$$

be the admissible rate region of the sources. If the capacity vector $\mathbf{C} \in \mathcal{R}$, then we say that the transmission problem of the sources over the network with the given transmission requirements at the sinks is resolvable. Let \mathcal{H} be the region consisting of all entropy vectors $\mathbf{H} = (H(X_k) : s_k \in \mathcal{S})$ for which the capacity vector is admissible, i.e.,

$$\mathcal{H} = \{\mathbf{H} : \mathbf{C} \text{ is admissible}\}. \quad (3)$$

We say that the entropy vectors in this region are achievable and call this region the capacity region.

III. MAIN RESULTS

Consider a degree-2 K -pairs three-layer network defined in the previous section, index the channels (i, j) by $(\mathcal{A}_i, \mathcal{B}_j)$. Let $\mathcal{J} = \{1, 2, \dots, K\}$ be the index set of both \mathcal{S} and \mathcal{T} . For convenience, we ignore the notation difference between \mathcal{S} , \mathcal{T} and \mathcal{J} . If two channels have the same indexing, then they are viewed as a single channel with capacity being the sum of the two capacities. Then the channel index is either of the form $(\{i, j\}, \{i, j\})$, (i, i) or of the form $(i, j), i \neq j$ by the assumptions for this particular model. Let the information rate on the first kind of channels be R_i, R_{ij} and the information rate on the second kind of channels be denoted by R_i^j , then we have the following theorem.

Theorem 1: For any degree-2 K -pairs three-layer network, the capacity region is given by

$$\mathcal{H} = \left\{ \mathbf{H} : \forall \gamma \subseteq \mathcal{J}, \sum_{i \in \gamma} H(X_i) \leq \sum_{i \in \gamma} R_i + \sum_{\substack{i \in \gamma, j \in \gamma, \\ i \prec_{\gamma} j}} (R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}) + \sum_{i \in \gamma, j \notin \gamma} R_{ij} \right\}, \quad (4)$$

where \prec_{γ} is an arbitrary linear order in γ .

IV. PROOF OF MAIN RESULT

A. Proof of Achievability:

Let \mathcal{H} be the capacity region in (4). This region is defined by $2^K - 1$ constraints, that is, for each non-empty subset γ of \mathcal{J} , there is a constraint which will be simply called constraint γ . An extremal point of the region can be specified by at least K active constraints (constraints with equality). We have the following result.

Lemma 1: For each extremal point of the region \mathcal{H} , there exists K distinct active constraints $\gamma_i: i = 1, \dots, K$ satisfying the condition that $\gamma_1 \subseteq \gamma_2 \subseteq \dots \subseteq \gamma_K$ where $|\gamma_i| = i$.

Proof: Let \mathbf{H} be an extremal point of $\mathcal{H} \subseteq R^K$.

Observation: Suppose at \mathbf{H} , there exist two sets $\alpha, \beta, \alpha \not\subseteq \beta, \beta \not\subseteq \alpha$, for which the corresponding constraints in (4) are active, i.e.,

$$\begin{aligned} \sum_{i \in \alpha} H(X_i) &= \Omega_\alpha + \Gamma_\alpha - \Pi_\alpha, \\ \sum_{i \in \beta} H(X_i) &= \Omega_\beta + \Gamma_\beta - \Pi_\beta, \end{aligned}$$

where

$$\begin{aligned} \Omega_\gamma &= \sum_{i \in \gamma} R_i, \\ \Gamma_\gamma &= \sum_{i \in \gamma} \sum_{j \in \mathcal{J}, j \neq i} R_{ij}, \\ \Pi_\gamma &= \sum_{\substack{i \in \gamma, j \in \gamma, \\ i \prec j}} (R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}). \end{aligned}$$

Then the two sets $\alpha \cup \beta$ and $\alpha \cap \beta$ are both active.

The observation is proved as follows. We have

$$\begin{aligned} & \sum_{i \in \alpha \cup \beta} H(X_i) + \sum_{i \in \alpha \cap \beta} H(X_i) \\ &= \sum_{i \in \alpha} H(X_i) + \sum_{i \in \beta} H(X_i) \\ &= \Omega_\alpha + \Omega_\beta + \Gamma_\alpha + \Gamma_\beta - \Pi_\alpha - \Pi_\beta \\ &\stackrel{(a)}{=} \Omega_{\alpha \cup \beta} + \Omega_{\alpha \cap \beta} + \Gamma_\alpha + \Gamma_\beta - \Pi_\alpha - \Pi_\beta \\ &\stackrel{(b)}{=} \Omega_{\alpha \cup \beta} + \Omega_{\alpha \cap \beta} + \Gamma_{\alpha \cup \beta} + \Gamma_{\alpha \cap \beta} - \Pi_\alpha - \Pi_\beta \\ &\stackrel{(c)}{\geq} \Omega_{\alpha \cup \beta} + \Omega_{\alpha \cap \beta} + \Gamma_{\alpha \cup \beta} + \Gamma_{\alpha \cap \beta} - \Pi_{\alpha \cup \beta} - \Pi_{\alpha \cap \beta}, \end{aligned}$$

which together with constraints

$$\begin{aligned} \sum_{i \in \alpha \cup \beta} H(X_i) &\leq \Omega_{\alpha \cup \beta} + \Gamma_{\alpha \cup \beta} - \Pi_{\alpha \cup \beta}, \\ \sum_{i \in \alpha \cap \beta} H(X_i) &\leq \Omega_{\alpha \cap \beta} + \Gamma_{\alpha \cap \beta} - \Pi_{\alpha \cap \beta}, \end{aligned}$$

implies the desired conclusion.

The noted steps are explained as follows:

(a) Define the indicator function of channel (i, i) as

$$1_\gamma = \begin{cases} 1, & \text{if } i \in \gamma, \\ 0, & \text{if } i \notin \gamma, \end{cases}$$

then $1_\alpha + 1_\beta = 1_{\alpha \cup \beta} + 1_{\alpha \cap \beta}$, i.e., the number of counts of (i, i) in α and β is the same as in $\alpha \cup \beta$ and $\alpha \cap \beta$, thus the sum of rates of such channels should be equal.

- (b) Let $Q_i = \sum_{j \in \mathcal{J}, j \neq i} R_{ij}$, thus $\Gamma_\gamma = \sum_{i \in \gamma} Q_i$. Using a similar indicator function for Q_i , we have $1_\alpha + 1_\beta = 1_{\alpha \cup \beta} + 1_{\alpha \cap \beta}$. Therefore, Q_i appears the same number of times in $\Gamma_\alpha + \Gamma_\beta$ as in $\Gamma_{\alpha \cup \beta} + \Gamma_{\alpha \cap \beta}$.
- (c) Define $\delta_{ij} = R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}$ as the excessive rate of R_{ij} over the side information rate R_i^j and R_j^i . To show the inequality holds, we list and compare the counts of δ_{ij} in $\Pi_\alpha + \Pi_\beta$ and $\Pi_{\alpha \cup \beta} + \Pi_{\alpha \cap \beta}$. To facilitate this procedure, we decompose the sets α, β and $\alpha \cup \beta, \alpha \cap \beta$ into common subregions, i.e., $I = \alpha \setminus \beta$, $II = \alpha \cap \beta$, $III = \beta \setminus \alpha$ and $IV = \alpha^c \cap \beta^c$, where γ^c stands for the complement of γ . Let A and B be two subregions, by $A \Leftrightarrow B$, we mean the number of counts of terms with indices $i \in A$ and $j \in B$. We proceed as in the following table.

$A \Leftrightarrow B$	$\Pi_\alpha + \Pi_\beta$	$\Pi_{\alpha \cup \beta} + \Pi_{\alpha \cap \beta}$
I \Leftrightarrow I	1	1
I \Leftrightarrow II	1	1
I \Leftrightarrow III	0	1
I \Leftrightarrow IV	0	0
II \Leftrightarrow II	2	2
II \Leftrightarrow III	1	1
II \Leftrightarrow IV	0	0
III \Leftrightarrow III	1	1
III \Leftrightarrow IV	0	0
IV \Leftrightarrow IV	0	0

Thus, in I \Leftrightarrow III, δ_{ij} is counted in $\Pi_{\alpha \cup \beta} + \Pi_{\alpha \cap \beta}$ but not in $\Pi_\alpha + \Pi_\beta$. This asserts that $\Pi_\alpha + \Pi_\beta \leq \Pi_{\alpha \cup \beta} + \Pi_{\alpha \cap \beta}$, where the equality holds only if $\delta_{ij} = 0$, for all $(\{i, j\}, \{i, j\})$ such that $i \in I, j \in III$.

The observation is proved. Using this observation and induction, it is a simple matter to prove the conclusion of the Lemma. That is, by replacing α, β with $\alpha \cap \beta$ and $\alpha \cup \beta$, we will eventually reach a set of active constraints satisfying the requirement of the lemma.

To prove the achievability of the region \mathcal{H} , we need only to prove the achievability of the extremal points. We simplify the exposition by assuming that sources are i.i.d. binary symmetric with rates per unit time being the entropies of the sources. This does not lose generality since i.i.d. sources with other alphabet size can be first compressed by binary source codes to "almost" i.i.d. binary data stream before transmission. By using block channel codes of block length n , without essential loss of generality, we assume that n times of the coding channel capacities are integer valued.

We distinguish different coding strategies:

- Butterfly Coding

Let $\hat{R}_{ij} = \min\{R_{ij}, R_i^j, R_j^i\}$. By assuming that $m = n\hat{R}_{ij}$ is an integer, we code two binary sequences $X_i^m = (X_i(1), \dots, X_i(m))$ of length m from source s_i and

$X_j^m = (X_j(1), \dots, X_j(m))$ of length m from source s_j by transmitting X_i^m and X_j^m over side information channels (i, j) and (j, i) respectively. Meanwhile, we transmit $X_i^m \oplus X_j^m$ over the main channel $(\{i, j\}, \{i, j\})$, where \oplus stands for the modulo 2 sum. At sink t_i , both $X_i^m \oplus X_j^m$ and X_j^m are received. Therefore, X_i^m can be decoded. For the same reason, X_j^m can be decoded at sink t_j .

- **Transmission without Coding**

If $\hat{R}_{ij} = R_{ij} - \tilde{R}_{ij} > 0$, by assuming that $l = n\hat{R}_{ij}$ is an integer, we can either transmit a binary sequence X_i^l of length l from source s_i over the channel $(\{i, j\}, \{i, j\})$ or a binary sequence X_j^l of length l from source s_j over the channel $(\{i, j\}, \{i, j\})$. In both cases, the transmitted data is received at sinks t_i and t_j . This coding method is also used for channel (i, i) where nR_i bits can be transmitted from source s_i to sink t_i .

Remarks:

- It is obvious that for channels $(\{i, j\}, \{i, j\})$, (i, j) and (j, i) , the two coding strategies can be performed simultaneously without violating the rate constraints for the channels.
- We always assume that the bits from the sources transmitted by distinct coding methods described above over channels are independent. Therefore, the total amount of data transmitted by several distinct methods is simply the summation of the amount of data transmitted by each of these methods.
- When use the transmission without coding over channel $(\{i, j\}, \{i, j\})$, we have two possible choices. That is, we can either transmit a binary sequence from source s_i or a binary sequence from source s_j . These two possible coding strategies are distinguished by saying that the rate \hat{R}_{ij} is assigned to source s_i when X_i^l is transmitted over the channel or assigned to source s_j otherwise.

The different ways of assigning rates \hat{R}_{ij} for all possible (i, j) to the sources give different information rate vectors for the sources. In the following, we describe a rate assignment strategy which proves the achievability of the extremal point with active constraints $\gamma_1 \subset \dots \subset \gamma_K$, where $|\gamma_k| = k$ for all k . There exists a linear order of the index set $\mathcal{J} : i_1 \prec_\gamma \dots \prec_\gamma i_K$ such that $\gamma_k = \{i_1, \dots, i_k\}$. We assign the rate \hat{R}_{ij} to s_i if and only if $i \prec_\gamma j$ and assign to s_j otherwise.

The total number of bits transmitted successfully from source s_i to sink t_i in n time slots, i.e. $nH(X_i)$, is calculated as follows:

$$\begin{aligned} nH(X_i) &= nR_i + n \sum_j \min\{R_{ij}, R_i^j, R_j^i\} \\ &\quad + n \sum_{j:i \prec_\gamma j} (R_{ij} - \min\{R_{ij}, R_i^j, R_j^i\}) \\ &= n[R_i + \sum_{j:i \prec_\gamma j} R_{ij} + \\ &\quad \sum_{j:j \prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\}]. \end{aligned}$$

Therefore,

$$H(X_i) = R_i + \sum_{j:i \prec_\gamma j} R_{ij} + \sum_{j:j \prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\}.$$

For any k ,

$$\begin{aligned} \sum_{i \in \gamma_k} H(X_i) &= \sum_{i \in \gamma_k} R_i + \sum_{i \in \gamma_k} \sum_{j:i \prec_\gamma j} R_{ij} + \\ &\quad + \sum_{i \in \gamma_k} \sum_{j:j \prec_\gamma i} \min\{R_{ij}, R_i^j, R_j^i\} \\ &= \sum_{i \in \gamma_k} R_i + \sum_{\substack{i, j \in \gamma_k, \\ i \prec_\gamma j}} (R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}) \\ &\quad + \sum_{i \in \gamma_k, j \notin \gamma_k} R_{ij}. \end{aligned}$$

This proves the achievability of the extremal point with active constraints $\gamma_1 \subset \gamma_2 \subset \dots \subset \gamma_K$ and therefore the achievability of the region \mathcal{H} by time sharing argument.

B. Proof of Converse Part:

Suppose that the rate vector $\mathbf{R} = (R_{ij} : (i, j) \in \mathcal{E})$ is admissible for sources with entropies $H(X_i) : i \in \mathcal{J}$. Then, for any $\epsilon > 0$ for sufficiently large n , there exists an $(n, (\eta_{ij}, (i, j) \in \mathcal{E}), (\Delta_l : l \in \mathcal{T}))$ code such that

$$n^{-1} \log \eta_{ij} \leq R_{ij}, \forall (i, j) \in \mathcal{E} \text{ and } \Delta_l \leq \epsilon, \forall l \in \mathcal{T}.$$

Let U_{ij} be the information flow on channel $(\{i, j\}, \{i, j\})$, U_i be the information flow on channel (i, i) and U_i^j be the information flow on channel $(i, j) : i \neq j$. We have the following facts:

- 1) $H(X_i^n | U_i, \{U_{ij}, \forall j \neq i\}, \{U_k^i, \forall k \neq i\}) = 0$, (5)

- 2) the coded information rate meets the constraint of the channel, i.e.,

$$H(U_{ij}) \leq nR_{ij}, H(U_i) \leq nR_i, H(U_i^j) \leq nR_i^j. \quad (6)$$

Thus, we have

$$\begin{aligned} &\sum_{i \in \gamma} nH(X_i) \\ &\stackrel{(a)}{=} \sum_{i \in \gamma} I(X_i^n; U_i, \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\}) \\ &\stackrel{(b)}{=} \sum_{i \in \gamma} I(X_i^n; U_i, \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\}) \\ &= \sum_{i \in \gamma} [I(X_i^n; \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\}) \\ &\quad + I(X_i^n; U_i | \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\})] \\ &\stackrel{(c)}{\leq} \sum_{i \in \gamma} [H(U_i) + I(X_i^n; \{U_{ij} : j \neq i\} | \{U_k^i : k \neq i\})] \end{aligned}$$

$$\begin{aligned}
&= \sum_{i \in \gamma} [H(U_i) + \\
&\quad \sum_{j \neq i} I(X_i^n; U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\})] \\
&= \sum_{i \in \gamma} [H(U_i) + \sum_{j \neq i} H(U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\}) \\
&\quad - \sum_{j \neq i} H(U_{ij} | X_i^n, \{U_{il} : l < j\}, \{U_k^i : k \neq i\})] \\
&\stackrel{(d)}{\leq} \sum_{i \in \gamma} [H(U_i) + \sum_{j \neq i} \{H(U_{ij} | U_j^i) - H(U_{ij} | X_i^n, U_j^i)\}] \\
&= \sum_{i \in \gamma} [H(U_i) + \sum_{j \neq i} I(U_{ij}; X_i^n | U_j^i)] \\
&= \sum_{i \in \gamma} H(U_i) + \sum_{i, j \in \gamma, i \prec j} [I(U_{ij}; X_i^n | U_j^i) \\
&\quad + I(U_{ij}; X_j^n | U_i^j)] + \sum_{i \in \gamma, j \notin \gamma} I(U_{ij}; X_i^n | U_j^i) \\
&\stackrel{(e)}{\leq} \sum_{i \in \gamma} nR_i + n \sum_{i, j \in \gamma, i \prec j} [R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}] \\
&\quad + n \sum_{i \in \gamma, j \notin \gamma} R_{ij}.
\end{aligned}$$

Dividing by n , Theorem 1 is proved.

In the proof, the marked steps are explained as follows:

- (a) follows from fact 1).
- (b) follows from the independence of X_i^n and $U_k^i : k \neq i$.
- (c) follows from the inequality

$$I(X_i^n; U_i | \{U_{ij} : j \neq i\}, \{U_k^i : k \neq i\}) \leq H(U_i).$$

- (d) follows from the following two facts:
 - $H(U_{ij} | \{U_{il} : l < j\}, \{U_k^i : k \neq i\}) \leq H(U_{ij} | U_j^i)$,
 - and

$$\begin{aligned}
&H(U_{ij} | X_i^n, \{U_{il} : l < j\}, \{U_k^i : k \neq i\}) \\
&= H(U_{ij} | X_i^n, U_j^i),
\end{aligned}$$

because of the conditional independence of U_{ij} and $\{U_{il} : l < j\}, \{U_k^i : k \neq i, k \neq j\}$ given X_i^n .

- (e) in this step, we use fact 2) and the following inequality,

$$\begin{aligned}
&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j) \\
&\leq n(R_{ij} + \min\{R_{ij}, R_i^j, R_j^i\}).
\end{aligned}$$

This inequality is proved below:

$$\begin{aligned}
&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j) \\
&= H(U_{ij} | U_j^i) + H(U_{ij} | U_i^j) \\
&\quad - H(U_{ij} | X_i^n, U_j^i) - H(U_{ij} | X_j^n, U_i^j) \\
&= H(U_{ij} | U_j^i) + I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j) \\
&\leq H(U_{ij}) + I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j).
\end{aligned}$$

Similarly, we can prove

$$\begin{aligned}
&I(U_{ij}; X_i^n | U_j^i) + I(U_{ij}; X_j^n | U_i^j) \\
&\leq H(U_{ij}) + I(U_{ij}; X_j^n, U_j^i; X_i^n | U_i^j).
\end{aligned}$$

We have,

- $I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j) \leq H(U_{ij}) \leq nR_{ij}$,
- $I(U_{ij}; X_j^n; X_i^n, U_j^i | U_i^j) \leq I(X_j^n; X_i^n, U_j^i | U_i^j) = I(X_j^n; U_j^i | X_i^n, U_i^j) = H(U_j^i) \leq nR_j^i$,
- similarly, we have

$$I(U_{ij}; X_j^n, U_j^i; X_i^n | U_i^j) \leq nR_{ij},$$

and

$$I(U_{ij}; X_j^n, U_j^i; X_i^n | U_j^i) \leq nR_j^i.$$

These inequalities imply the desired result. \square

V. CONCLUSION

In this paper, we proved the capacity region for a simple class of multi-source multi-sink networks called degree-2 K -pairs three-layer networks. we proposed a new approach to the multi-source multi-sink network capacity problem by introducing a cut-based network decomposition built on the three-layer networks and a role-based information characterization by analyzing the roles of side information.

For simplicity, we restricted to degree-2 cross coding. However, the extension to the case of degree-3 could be an important step toward a more general characterization.

Although the network model considered in this paper is very preliminary and the method used in the proof is not fully developed, we believe that it can be a useful tool to determine the capacity regions for more general multi-source multi-sink networks.

REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, July 2000.
- [2] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1111–1120, May 1999.
- [3] X. Yan, J. Yang, and Z. Zhang, "An outer bound for multisource multisink network coding with minimum cost consideration," *submitted to IEEE Trans. Inform. Theory*, 2005.
- [4] N. Harvey, R. Kleinberg, and A. Lehman, "On the capacity of information networks," *submitted to IEEE Trans. Inform. Theory*, 2005.
- [5] E. Erez and M. Feder, "Capacity region and network codes for two receivers multicast with private and common data," *Workshop on Coding, Cryptography and Combinatorics*, 2003.
- [6] C. K. Ngai and R. W. Yeung, "Multisource network coding with two sinks," in *International Conference on Communications, Circuits and Systems (ICCCAS)*, June 2004.
- [7] R. W. Yeung, *A First Course in Information Theory*. New York: Kluwer/Plenum, 2002.
- [8] X. Yan, J. Yang, and Z. Zhang, "An improved outer bound for multisource multisink network coding," in *First Workshop on Network Coding, Theory, and Applications (NETCOD)*, Italy, Apr. 2005.