Structural Properties of Index Coding Capacity Using Fractional Graph Theory

Fatemeh Arbabjolfaei and Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego
Email: {farbabjo, yhk}@ucsd.edu

Abstract—The capacity region of the index coding problem is characterized through the notion of confusion graph and its fractional chromatic number. Based on this multiletter characterization, several structural properties of the capacity region are established, some of which are already noted by Tahmasbi, Shahrasbi, and Gohari, but proved here with simple and more direct graph-theoretic arguments. In particular, the capacity region of a given index coding problem is shown to be simple functionals of the capacity regions of smaller subproblems when the interaction between the subproblems is none, one-way, or complete.

I. INTRODUCTION

Suppose that a sender wishes to communicate a tuple of n messages, $x^n = (x_1, \ldots, x_n)$, $x_j \in \{0,1\}^{t_j}$, to their corresponding receivers using a shared noiseless channel. Receiver $j \in [1:n] := \{1, 2, \ldots, n\}$ has prior knowledge of a subset $x(A_j) := \{x_i : i \in A_j\}$, $A_j \subseteq [1:n] \setminus \{j\}$, of the messages and wishes to recover x_j. It is assumed that the sender is aware of $A_1, \ldots, A_n$. The goal is to minimize the amount of information that should be broadcast from the sender to the receivers so that every receiver can recover its desired message.

Any instance of this problem, referred to collectively as the *index coding problem*, is fully specified by the side information sets $A_1, \ldots, A_n$. Equivalently, it can be specified by a side information graph G with n nodes, in which a directed edge $i \to j$ represents that receiver j has message i as side information, i.e., $i \in A_j$ (see Fig. 1(a)). Thus, we often identify an index coding problem with its side information graph and simply write "index coding problem G."

A $(t_1, \ldots, t_n, r)$ index code is defined by
- an encoder $\phi : \prod_{i=1}^{n} \{0,1\}^{t_i} \to \{0,1\}^r$ that maps n-tuple of messages x^n to an r-bit index and
- n decoders $\psi_j : \{0,1\}^r \times \prod_{k \in A_j} \{0,1\}^{t_k} \to \{0,1\}^{t_j}$ that maps the received index $\phi(x^n)$ and the side information $x(A_j)$ back to x_j for $j \in [1:n]$.

Thus, for every $x^n \in \prod_{i=1}^{n} \{0,1\}^{t_i}$,

$$\psi_j(\phi(x^n), x(A_j)) = x_j, \quad j \in [1:n].$$

A rate tuple $(R_1, \ldots, R_n)$ is said to be *achievable* for the index coding problem G if there exists a $(t_1, \ldots, t_n, r)$
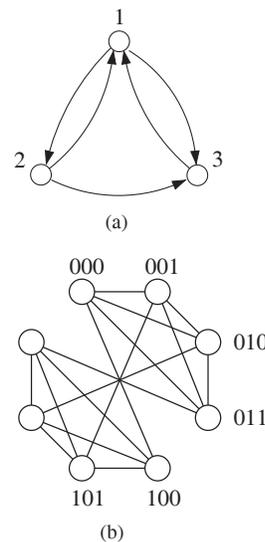

Fig. 1. (a) The graph representation for the index coding problem with $A_1 = \{2, 3\}$, $A_2 = \{1\}$, and $A_3 = \{1, 2\}$. (b) The confusion graph corresponding to the integer tuple $(t_1, t_2, t_3) = (1, 1, 1)$. Each node is labeled with the a message tuple.

index code such that

$$R_j \leq \frac{t_j}{r}, \quad j \in [1:n].$$

The *capacity region* $\mathscr{C}$ of the index coding problem is defined as the closure of the set of achievable rate tuples.

Since Birk and Kol [1] introduced the index coding problem in 1998, this simple yet fundamental problem attracted several research communities (see [2]–[5] for a subset of recent contributions). The capacity region has been established for all 9,608 index coding problems of $n = 5$ messages [6] (which includes all index coding problems *up to* five messages by taking projections). However, the coding schemes developed for small n prove to be suboptimal when n becomes large and there is no known computable characterization for the capacity region of a general index coding problem. On the theoretical side, there is no known algorithm even to approximate the capacity region within a factor of $O(n^{1-\epsilon})$ for $\epsilon \in (0, 1)$. On the computational side, the

number of index coding problems blows up quickly in n (for example, there are 1,540,944 distinct instances of index coding problems for $n = 6$) and it also becomes quite challenging to compare existing inner and outer bounds on the capacity region of each problem as n increases.

As an intermediate step towards characterizing the capacity region (analytically, approximately, or numerically), we study some structural properties of the capacity region. In particular, we show that the side information graph G can be partitioned into two vertex-induced subgraphs G_1 and G_2, then the capacity region $\mathscr{C}$ of the index coding problem G can be characterized as a simple functional of the capacity regions $\mathscr{C}_1$ and $\mathscr{C}_2$ of G_1 and G_2, respectively, provided that

1) there is no edge between G_1 and G_2, or
2) more generally, there is no edge from G_2 to G_1, or
3) every node in G_1 is connected to every node in G_2 and vice versa.

The immediate utility of these structural properties is that one can reduce the number of index coding problems that need to be studied. For example, we can check (details not shown) that 1,366,783 (89%) out of 1,540,944 index coding problems for $n = 6$ fall into one of the three aforementioned criteria or another simple case (Proposition 4 in Section III), significantly narrowing the set of problems that are worth further investigation.

We must note that the first two properties have been already established by Tahmasbi, Shahrasbi, and Gohari [7, Th. 2] using a somewhat convoluted argument based on joint typicality encoding and covering. In comparison, our approach is more direct and based on the definition of the capacity region itself. As discussed more precisely in Section III, our starting point is a graph-theoretic characterization of the index coding capacity region using the notion of *confusion graph*. This notion was introduced by Alon, Hassidim, Lubetzky, Stav, and Weinstein [8], who characterized the optimal broadcast rate (the reciprocal of the symmetric capacity) using the chromatic number of the confusion graph. We generalize and tighten their approach by connecting the capacity region with the *fractional* chromatic number of the confusion graph. This allows us to utilize well-known results from fractional graph theory [9] such as the identities on fractional chromatic numbers for graph products (see Section II) to establish several structural properties of the capacity region. Our approach based on confusion graph and fractional chromatic number seems to be broadly applicable beyond these structural results. Although it is not presented here, a similar method generalizes and tightens the recent result by Mazumdar [10] on the duality between index coding and distributed storage.

Throughout the paper, the base of logarithm is 2.

II. MATHEMATICAL PRELIMINARIES

A. Confusion Graphs

We generalize the notion of confusion graph, which was originally introduced in [8] for equal-length messages.

Given an index coding problem G, two tuples of n messages $x^n, z^n \in \prod_{i=1}^{n} \{0,1\}^{t_i}$ are said to be *confusable at receiver* $j \in [1 : n]$ if $x_j \neq z_j$ and $x_i = z_i$ for all $i \in A_j$. We simply say x^n and z^n are *confusable* if they are confusable at some receiver j. Given an index coding problem G and a tuple of message lengths $\mathbf{t} = (t_1, \ldots, t_n)$, the *confusion graph* $\Gamma_{\mathbf{t}}(G)$ is an undirected graph with $\prod_{i=1}^{n} 2^{t_i}$ vertices such that every vertex corresponds to a message tuple x^n and two vertices are connected iff (if and only if) the corresponding message tuples are confusable.

The confusion graph of the index coding problem with side information graph in Fig. 1(a) corresponding to $(t_1, t_2, t_3) = (1, 1, 1)$ is depicted in Fig. 1(b).

B. Graph Coloring

A (vertex) coloring of an undirected graph Γ is a mapping that assigns a color to each vertex such that no two adjacent vertices share the same color. The *chromatic number* $\chi(\Gamma)$ is the minimum number of colors such that a coloring of the graph exists.

More generally, a b-fold coloring assigns a set of b colors to each vertex such that no two adjacent vertices share the same color. The b-fold chromatic number $\chi^{(b)}(\Gamma)$ is the minimum number of colors such that a b-fold coloring exists. The *fractional chromatic number* of the graph is defined as

$$\chi_f(\Gamma) = \lim_{b \to \infty} \frac{\chi^{(b)}(\Gamma)}{b} = \inf_b \frac{\chi^{(b)}(\Gamma)}{b},$$

where the limit exists since $\chi^{(b)}(\Gamma)$ is subadditive. Consequently,

$$\chi_f(\Gamma) \leq \chi(\Gamma). \tag{1}$$

Let $\mathcal{I}$ be the collection of all independent sets in Γ (i.e., sets of vertices such that no two vertices are adjacent). The chromatic number and the fractional chromatic number are also characterized as the solution to the following optimization problem

$$\text{minimize} \quad \sum_{S \in \mathcal{I}} \rho_S$$
$$\text{subject to} \quad \sum_{S \in \mathcal{I} : j \in S} \rho_S \geq 1, \quad j \in [1 : n].$$

When the optimization variables ρ_S, $S \in \mathcal{I}$, take integer values $\{0, 1\}$, then the (integral) solution is the chromatic number. If this constraint is relaxed and $\rho_S \in [0, 1]$, then the (rational) solution is the fractional chromatic number [9].

C. Graph Products

Generally speaking, a graph product is a binary operation on two (undirected) graphs Γ_1 and Γ_2 that produces a graph Γ on the Cartesian product of the original vertex sets with the edge set constructed from the original edge sets according to certain rules. In this section, we review a few definitions of graph products and their (fractional) chromatic numbers. In the following, $v_1 \sim v_2$ denotes that there exists an edge between v_1 and v_2. The notation $V(\Gamma)$ means the vertex set of a graph Γ.

The *disjunctive product* $\Gamma = \Gamma_1 * \Gamma_2$ is defined as $V(\Gamma) = V(\Gamma_1) \times V(\Gamma_2)$ and $(u_1, u_2) \sim (v_1, v_2)$ iff

$$u_1 \sim v_1 \quad \text{or} \quad u_2 \sim v_2.$$

The fractional chromatic number of the disjunctive product is multiplicative.

Lemma 1 (Scheinerman and Ullman [9, Cor. 3.4.2]).

$$\chi_f(\Gamma_1 * \Gamma_2) = \chi_f(\Gamma_1)\chi_f(\Gamma_2).$$

Note that the chromatic number satisfies the following relationship [9, Prop. 3.4.4]:

$$\chi(\Gamma_1 * \Gamma_2) \leq \chi(\Gamma_1)\chi(\Gamma_2). \tag{2}$$

The chromatic and fractional chromatic numbers of the power of a graph scale in the same exponential rate.

Lemma 2 (Scheinerman and Ullman [9, Cor. 3.4.3]). *Let Γ^k be the k-th power of Γ in disjunctive product. Then*

$$\chi_f(\Gamma) = \lim_{k \to \infty} \sqrt[k]{\chi(\Gamma^k)} = \inf_k \sqrt[k]{\chi(\Gamma^k)}.$$

The *lexicographic product* $\Gamma = \Gamma_1 \bullet \Gamma_2$ is defined as $V(\Gamma) = V(\Gamma_1) \times V(\Gamma_2)$ and $(u_1, u_2) \sim (v_1, v_2)$ iff

$$u_1 \sim v_1 \quad \text{or} \quad (u_1 = v_1 \text{ and } u_2 \sim v_2).$$

Note that the lexicographic product of graphs is not commutative. Nonetheless, its fractional chromatic number is still multiplicative.

Lemma 3 (Scheinerman and Ullman [9, Cor. 3.4.5]).

$$\chi_f(\Gamma_1 \bullet \Gamma_2) = \chi_f(\Gamma_1)\chi_f(\Gamma_2).$$

Note that the chromatic number satisfies the following relationship [11, Th. 1]:

$$\chi(\Gamma_1 \bullet \Gamma_2) \leq \chi(\Gamma_1)\chi(\Gamma_2). \tag{3}$$

The *Cartesian product* $G = G_1 \wedge G_2$ is defined as $V(\Gamma) = V(\Gamma_1) \times V(\Gamma_2)$ and $(u_1, u_2) \sim (v_1, v_2)$ iff

$$(u_1 = v_1 \text{ and } u_2 \sim v_2) \quad \text{or} \quad (u_2 = v_2 \text{ and } u_1 \sim v_1).$$

This product does not increase the chromatic number.

Lemma 4 (Sabidussi [12, Lemma 2.6]).

$$\chi(\Gamma_1 \wedge \Gamma_2) = \max\{\chi(\Gamma_1), \chi(\Gamma_2)\}.$$

III. MAIN RESULTS

A. Capacity Region via the Confusion Graph

We first state a simple generalization of the result by Alon, et al. [8, Th. 1.1].

Proposition 1. *A rate tuple $(R_1, \ldots, R_n)$ is achievable for the index coding problem G iff there exists an integer tuple $\mathbf{t} = (t_1, \ldots, t_n)$ such that*

$$R_j \leq \frac{t_j}{\lceil \log(\chi(\Gamma_\mathbf{t}(G))) \rceil}, \quad j \in [1:n]. \tag{4}$$

Proof: Sufficiency (achievability). For a given tuple $\mathbf{t} = (t_1, \ldots, t_n)$, consider a coloring of the vertices of the confusion graph $\Gamma = \Gamma_\mathbf{t}(G)$ with $\chi(\Gamma)$ colors. This partitions the vertices of Γ into $\chi(\mathcal{G})$ independent sets. Now by the definition of the confusion graph, no two message tuples in each independent set are confusable and therefore assigning an index to each independent set yields a valid index code. The total number of codewords of this index code is $\chi(\Gamma)$, which requires $r = \lceil \log(\chi(\Gamma)) \rceil$ bits to be broadcast. This proves the existence of a $(t_1, \ldots, t_n, \lceil \log(\chi(\Gamma_\mathbf{t}(G))) \rceil)$ index code.

Necessity (converse). Consider any $(t_1, \ldots, t_n, r)$ index code, which assigns at most 2^r distinct indices to message tuples. By definition, all the message tuples mapped to an index form an independent set of the confusion graph $\Gamma = \Gamma_\mathbf{t}(G)$. Moreover, every message tuple is mapped to some index so that these independent sets partition $V(\Gamma)$. Thus, $\chi(\Gamma) \leq 2^r$, or equivalently, $r \geq \lceil \log(\chi(\Gamma)) \rceil$. Therefore, any achievable $(R_1, \ldots, R_n)$ must satisfy

$$R_j \leq \frac{t_j}{\lceil \log(\chi(\Gamma_\mathbf{t}(G))) \rceil}, \quad j \in [1:n],$$

for some $\mathbf{t} = (t_1, \ldots, t_n)$. ∎

The ceiling operation in (4), which results from the fact that the index is communicated in bits, is not essential. By using the code that maps $x^n \in \prod_{j=1}^{n}\{0,1\}^{t_j}$ to $[1 : \chi(\Gamma_\mathbf{t}(G))]$ repeatedly k times, one can easily construct a code that maps $\tilde{x}^n \in \prod_{j=1}^{n}\{0,1\}^{kt_j}$ to $[1 : \chi(\Gamma_\mathbf{t}(G))]^k$, thus achieving rates

$$R_j = \frac{kt_j}{k\log(\chi(\Gamma_\mathbf{t}(G))) + 1}$$
$$\leq \frac{kt_j}{\lceil k\log(\chi(\Gamma_\mathbf{t}(G))) \rceil}, \quad j \in [1:n]. \tag{5}$$

Letting $k \to \infty$ in (5) establishes the following.

Proposition 2. *The capacity region $\mathscr{C}$ of the index coding problem G is the closure of all rate tuples $(R_1, \ldots, R_n)$ such that*

$$R_j \leq \frac{t_j}{\log(\chi(\Gamma_\mathbf{t}(G)))}, \quad j \in [1:n], \tag{6}$$

for some $\mathbf{t} = (t_1, \ldots, t_n)$.

We now state a stronger result, in terms of the *fractional* chromatic number, which will prove to be useful in establishing structural properties of the capacity region.

Theorem 1. *The capacity region $\mathscr{C}$ of the index coding problem G is the closure of all rate tuples $(R_1, \ldots, R_n)$ such that*

$$R_j \leq \frac{t_j}{\log(\chi_f(\Gamma_{\mathbf{t}}(G)))}, \quad j \in [1:n], \qquad (7)$$

for some $\mathbf{t} = (t_1, \ldots, t_n)$.

Proof: The necessity follows by (1) and Proposition 1.

Let $\epsilon > 0$. For each $\mathbf{t} = (t_1, \ldots, t_n)$ and the corresponding confusion graph $\Gamma_{\mathbf{t}}(G)$, Lemma 2 implies that there exists an integer k such that

$$\sqrt[k]{\chi(\Gamma_{\mathbf{t}}^k(G))} \leq \chi_f(\Gamma_{\mathbf{t}}(G)) + \epsilon. \qquad (8)$$

It can be also checked that the set of edges of $\Gamma_{\mathbf{t}}^k(G)$ contains the set of edges of $\Gamma_{k\mathbf{t}}(G)$, which, when combined with (8), implies that $\sqrt[k]{\chi(\Gamma_{k\mathbf{t}}(G))} \leq \chi_f(\Gamma_{\mathbf{t}}(G)) + \epsilon$, or equivalently,

$$\frac{t_j}{\log(\chi_f(\Gamma_{\mathbf{t}}(G)) + \epsilon)} \leq \frac{k t_j}{\log(\chi(\Gamma_{k\mathbf{t}}(G)))}, \quad j \in [1:n].$$

Thus, by Proposition 2, if $(R_1, \ldots, R_n)$ satisfies

$$R_j \leq \frac{t_j}{\log(\chi_f(\Gamma_{\mathbf{t}}(G)) + \epsilon)}, \quad j \in [1:n],$$

then it must be in the capacity region. Since $\mathscr{C}$ is closed, taking $\epsilon \to 0$ completes the proof. ∎

B. Capacity Region via Confusion Graph Products

Throughout this subsection, we assume that G_1 and G_2 are two vertex-induced subgraphs of G such that $V(G_1) = [1:n_1]$ and $V(G_2) = [n_1 + 1 : n]$ partition $V(G) = [1:n]$. We denote the capacity regions of the index coding problems G, G_1 and G_2 by $\mathscr{C}$, $\mathscr{C}_1$ and $\mathscr{C}_2$, respectively.

Proposition 3. *If G has no edge between G_1 and G_2, then*

$$\mathscr{C} = \bigcup_{\alpha \in [0,1]} \{(\alpha \mathbf{R}_1, (1-\alpha)\mathbf{R}_2) \colon \mathbf{R}_1 \in \mathscr{C}_1, \mathbf{R}_2 \in \mathscr{C}_2\}.$$

In other words, the capacity region of G is achieved by time division between the optimal coding schemes for two disjoint subproblems G_1 and G_2.

Proof: It suffices to show that

$$\mathscr{C} \subseteq \bigcup_{\alpha \in [0,1]} \{(\alpha \mathbf{R}_1, (1-\alpha)\mathbf{R}_2) \colon \mathbf{R}_1 \in \mathscr{C}_1, \mathbf{R}_2 \in \mathscr{C}_2\}.$$

Let $x^n = (\mathbf{x}_1, \mathbf{x}_2)$ and $z^n = (\mathbf{z}_1, \mathbf{z}_2)$ be two message tuples, and $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2)$ be their common length tuple, where $\mathbf{x}_i$, $\mathbf{z}_i$, and $\mathbf{t}_i$ correspond to the subproblem

G_i, $i = 1, 2$. By the definition of confusability, x^n and z^n are confusable iff they are confusable at some receiver $j \in V(G_1)$ or confusable at some receiver $j \in V(G_2)$. Since there is no edge between G_1 and G_2, these local confusability conditions are equivalent to the confusability of $\mathbf{x}_1$ and $\mathbf{z}_1$ for the subproblem G_1 and the confusability of $\mathbf{x}_2$ and $\mathbf{z}_2$ for the subproblem G_2, respectively. In other words, x^n and z^n are confusable for G iff $\mathbf{x}_1$ and $\mathbf{z}_1$ are confusable for G_1 or $\mathbf{x}_2$ and $\mathbf{z}_2$ are confusable for G_2. Thus, $\Gamma_{\mathbf{t}}(G) = \Gamma_{\mathbf{t}_1}(G_1) * \Gamma_{\mathbf{t}_2}(G_2)$ and by Lemma 1 for disjunctive product,

$$\log(\chi_f(\Gamma_{\mathbf{t}}(G)))$$
$$= \log(\chi_f(\Gamma_{\mathbf{t}_1}(G_1))) + \log(\chi_f(\Gamma_{\mathbf{t}_2}(G_2))) =: l_1 + l_2.$$

We now let $\alpha = l_1/(l_1 + l_2)$ and apply Theorem 1. Before closure, any rate tuple in $\mathscr{C}$ should satisfy

$$R_j \leq \frac{t_j}{l_1 + l_2} = \begin{cases} \alpha \frac{t_j}{l_1}, & j \in V(G_1), \\ (1-\alpha)\frac{t_j}{l_2}, & j \in V(G_2). \end{cases}$$

But again by Theorem 1, $(t_j/l_1 : j \in V(G_1)) \in \mathscr{C}_1$ and $(t_j/l_2 : j \in V(G_2)) \in \mathscr{C}_2$, which completes the proof. ∎

We now state a stronger version of Proposition 3, originally established by Tahmasbi, Shahrasbi, and Gohari [7]; see also [5, Th. 8] for a related but much weaker statement.

Theorem 2 (Tahmasbi, Shahrasbi, and Gohari [7, Th. 2])**.** *If G has no edge from G_2 to G_1, then*

$$\mathscr{C} = \bigcup_{\alpha \in [0,1]} \{(\alpha \mathbf{R}_1, (1-\alpha)\mathbf{R}_2) \colon \mathbf{R}_1 \in \mathscr{C}_1, \mathbf{R}_2 \in \mathscr{C}_2\}.$$

Once again the capacity region is achieved by time division. Moreover, in light of Proposition 3 and the Farkas lemma [13, Th. 2.2] (that is, each edge in a directed graph either lies on a directed cycle or belongs to a directed cut but not both), Theorem 2 implies that removing edges of G that do not lie on a directed cycle does not change the capacity region.

Proof: Assume without loss of generality that there exists an edge from every node in G_1 to every node in G_2. Now, since every node in G_2 has every node (message) in G_1 as side information and no node in G_1 has any node in G_2 as side information, x^n and z^n are confusable for G iff $\mathbf{x}_1$ and $\mathbf{z}_1$ are confusable for G_1, or $\mathbf{x}_1 = \mathbf{z}_1$ and $\mathbf{x}_2$ and $\mathbf{z}_2$ are confusable for G_2 (recall the notation in the proof of Proposition 3). Thus, $\Gamma_{\mathbf{t}}(G) = \Gamma_{\mathbf{t}_1}(G_1) \bullet \Gamma_{\mathbf{t}_2}(G_2)$ and by Lemma 3 for lexicographic product,

$$\log(\chi_f(\Gamma_{\mathbf{t}}(G)))$$
$$= \log(\chi_f(\Gamma_{\mathbf{t}_1}(G_1))) + \log(\chi_f(\Gamma_{\mathbf{t}_2}(G_2))).$$

The rest of the proof follows the identical steps to that of Proposition 3. ∎

The only difference between Propposition 3 and Theorem 2 lies with which product of confusion subgraphs needs to be taken—the disjunctive product for two separate subproblems, while the lexicographic product for two subproblems dependent in only one direction. Note that the main tool we use from fractional graph theory (cf. Lemmas 1 and 3) is

$$\chi_f(\Gamma_{\mathbf{t}}(G)) \geq \chi_f(\Gamma_{\mathbf{t}_1}(G_1))\chi_f(\Gamma_{\mathbf{t}_2}(G_2)). \quad (9)$$

This implies that as long as an index coding problem can be partitioned into two subproblems and the corresponding (nonstandard) graph product of the confusion subgraphs satisfies (9), the capacity region has the same form as in Proposition 3 and Theorem 2. Also note that with the (integral) chromatic number, an inequality like (9) holds in the opposite direction (cf. (2) and (3)). This shows the major advantage of Theorem 1 over Proposition 2.

Next, we consider index coding problems with side information graphs that contain a complete bipartite graph as an edge-induced subgraph.

Theorem 3. *If there are edges from every node in G_1 to every node in G_2 and vice versa, then*

$$\mathscr{C} = \big\{(\mathbf{R}_1, \mathbf{R}_2) \colon \mathbf{R}_1 \in \mathscr{C}_1, \mathbf{R}_2 \in \mathscr{C}_2\big\}.$$

In other words, the capacity region of G is achieved by simultaneously using the optimal coding schemes for two disjoint subproblems G_1 and G_2.

Proof: Since every node in G_1 has every message in G_2 as side information and every node in G_2 has every message in G_1 as side information, x^n and z^n are confusable for G iff $\mathbf{x}_1 = \mathbf{z}_1$ and $\mathbf{x}_2$ and $\mathbf{z}_2$ are confusable for G_2, or $\mathbf{x}_2 = \mathbf{z}_2$ and $\mathbf{x}_1$ and $\mathbf{z}_1$ are confusable for G_1. Thus, $\Gamma_{\mathbf{t}}(G) = \Gamma_{\mathbf{t}_1}(G_1) \wedge \Gamma_{\mathbf{t}_2}(G_2)$ and by Lemma 4 for cartesian product,

$$\chi(\Gamma_{\mathbf{t}}(G)) = \max\{\chi(\Gamma_{\mathbf{t}_1}(G_1)), \chi(\Gamma_{\mathbf{t}_2}(G_2))\}. \quad (10)$$

By Proposition 2, before closure, any rate tuple in $\mathscr{C}$ should satisfy

$$R_j \leq \frac{t_j}{\log(\chi(\Gamma_{\mathbf{t}}(G)))}, \quad j \in [1:n],$$

for some $\mathbf{t} = (t_1, \ldots, t_n)$. Combining this with (10), we have for $i = 1, 2$

$$R_j \leq \frac{t_j}{\log(\chi(\Gamma_{\mathbf{t}_i}(G_i))} =: \frac{t_j}{l_i}, \quad j \in V(G_i).$$

By applying Proposition 2 once again, $(t_j/l_1 : j \in V(G_1)) \in \mathscr{C}_1$ and $(t_j/l_2 : j \in V(G_2)) \in \mathscr{C}_2$, which completes the proof. ∎

C. Capacity Region via Degraded Side Information Sets

Here we consider the index coding problem G with side information sets $A_1, \ldots, A_n$.

Proposition 4. *If $A_i \subseteq A_j$, then removing i from A_j does not decrease the capacity region.*

The proof is intuitively clear. Given any index code, receiver j can first recover x_i using A_i and then uses x_i along with $x(A_j \setminus \{i\})$ to recover x_j. Here is an alternative proof based on the notion of confusion graph.

Proof: Assume that there exist x^n and z^n confusable for the new problem G', but not for the original problem G. Then, they must be confusable at receiver j for G' (i.e., $x_j \neq z_j$ and $x(A_j \setminus \{i\}) = z(A_j \setminus \{i\})$). Now if $x_i = z_i$, then it contradicts the assumption that x^n and z^n are not confusable (at receiver j) for G. Alternatively, if $x_i \neq z_i$, then since $A_i \subseteq A_j$ and hence $x(A_i) = z(A_i)$, it again contradicts the assumption that x^n and z^n are not confusable (at receiver i) for G. Therefore, the confusion graphs must be the same and, by Theorem 1, so must be the capacity regions. ∎

REFERENCES

[1] Y. Birk and T. Kol, "Informed-source coding-on-demand (IS-COD) over broadcast channels," in *Proc. 17th Ann. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Francisco, CA, Mar. 1998, pp. 1257–1264.

[2] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5811–5823, Sep. 2013.

[3] K. Shanmugam, A. Dimakis, and M. Langberg, "Local graph coloring and index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1152–1156.

[4] H. Maleki, V. R. Cadambe, and S. A. Jafar, "Index coding: An interference alignment perspective," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, Jul. 2012, pp. 2236–2240.

[5] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1479–1494, Mar. 2011.

[6] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Sasoglu, and L. Wang, "On the capacity region for index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 962–966.

[7] M. Tahmasbi, A. Shahrasbi, and A. Gohari, "Critical graphs in index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, Jul. 2014, pp. 281–285.

[8] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, "Broadcasting with side information," in *49th Ann. IEEE Symp. Found. Comput. Sci.*, Philadelphia, PA, Oct. 2008, pp. 823–832.

[9] E. R. Scheinerman and D. H. Ullman, *Fractional Graph Theory, A Rational Approach to the Theory of Graphs*. New York: Dover Publications, 2011.

[10] A. Mazumdar, "On a duality between recoverable distributed storage and index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, Jul. 2014, pp. 1977–1981.

[11] N. Linial and V. Vazirani, "Graph products and chromatic numbers," in *30th Ann. IEEE Symp. Found. Comp. Sci*, Research Triangle Park, NC, Oct. 1989, pp. 124–128.

[12] G. Sabidussi, "Graphs with given group and given graph-theoretical properties," *Canad. J. Math.*, vol. 9, pp. 515–525, 1957.

[13] A. Bachem and W. Kern, *Linear Programming Duality, An Introduction to Oriented Matroids*. Berlin: Springer, 1992.