Structural Properties of Index Coding Capacity

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

Abstract—The index coding capacity is investigated through its structural properties. First, the capacity is characterized in three new multiletter expressions involving the clique number, Shannon capacity, and Lovász theta function of the confusion graph, the latter notion introduced by Alon, Hassidim, Lubetzky, Stav, and Weinstein. The main idea is that every confusion graph can be decomposed into a small number of perfect graphs. The clique-number characterization is then utilized to show that the capacity is multiplicative under the lexicographic product of side information graphs, establishing the converse to an earlier result by Blasiak, Kleinberg, and Lubetzky. Second, sufficient and necessary conditions on the criticality of an index coding instance, namely, whether side information can be removed without reducing the capacity, are established based on the notion of unicycle, providing a partial answer to the question first raised by Tahmasbi, Shahrasbi, and Gohari. The necessary condition, along with other existing conditions, can be used to eliminate noncritical instances that do not need to be investigated. As an application of the established multiplicativity and criticality, only 10,634 (0.69%) out of 1,540,944 nonisomorphic six-message index coding instances are identified for further investigation, among which the capacity is still unknown for 119 instances.

I. INTRODUCTION

The index coding problem is a canonical problem in network information theory in which a server has a tuple of \( n \) messages \( x^A = (x_1, \ldots, x_n) \), \( x_j \in \{0,1\}^{t_j} \), and is connected to \( n \) receivers via a noiseless broadcast channel. Suppose that receiver \( j \in [n] := \{1,2,\ldots,n\} \) is interested in message \( x_j \) and has a set of other messages \( A_j := \{x_i, i \in A_j\}, A_j \subseteq [n] \setminus \{j\} \) as side information. Assuming that the server knows side information sets \( A_1, \ldots, A_n \), one wishes to characterize the minimum amount of information the server needs to broadcast and to find the optimal coding scheme that achieves this minimum.

More precisely, a \((t_1, \ldots, t_n, r)\) index code is defined by

1. an encoder \( \phi : \prod_{j=1}^{n} \{0,1\}^{t_j} \to \{0,1\}^{r} \) that maps \( n \)-tuple of messages \( x^n \) to an \( r \)-bit index and
2. \( n \) 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 \( j \in [n] \).

Thus, for every \( x^n \in \prod_{j=1}^{n} \{0,1\}^{t_j} \),

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

A \((t, \ldots, t, r)\) code is written as a \((t, r)\) code. A rate tuple \((R_1, \ldots, R_n)\) is said to be achievable for the index coding problem if there exists a \((t_1, \ldots, t_n, r)\) index code such that

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

The capacity region \( \mathcal{C} \) of the index coding problem is defined as the closure of the set of achievable rate tuples. The symmetric capacity (or the capacity in short) of the index coding problem is defined as

\[
C_{\text{sym}} = \max \{R : (R, \ldots, R) \in \mathcal{C}\},
\]

and its reciprocal \( \beta = 1/C_{\text{sym}} \) is referred to as the broadcast rate, which can be equivalently defined as

\[
\beta = \inf_{(t_1, r) \text{ codes}} \inf_{t \to \infty} \frac{r}{t} = \lim_{t \to \infty} \inf_{(t_1, r) \text{ codes}} \frac{r}{t},
\]

where the equality follows by Fekete’s lemma [1] and the subadditivity

\[
\inf_{(t_1+f_2, r) \text{ codes}} r \leq \inf_{(t_1, r_1) \text{ codes}} r_1 + \inf_{(t_2, r_2) \text{ codes}} r_2.
\]

The goal is to characterize the capacity region or the symmetric capacity for the general index coding problem and to determine the coding scheme that can achieve it.

Any instance of the index coding problem is fully determined by the side information sets \( A_1, \ldots, A_n \), and is represented compactly as \((j|A_j), j \in [n]\). For example, the 3-message index coding problem with \( A_1 = \{2,3\}, A_2 = \{1\} \), and \( A_3 = \{1,2\} \) is represented as

\((1|2,3), (2|1), (3|1,2)\).

The problem can be equivalently specified by a directed graph with \( n \) vertices, commonly referred to as the side information graph. Each vertex of the side information graph \( G = (V,E) \) corresponds to a receiver (and its associated message) and there is a directed edge \( i \to j \) if and only if (iff) receiver \( j \) knows message \( i \) as side information, i.e., \( i \in A_j \) (see Fig. 1).

Throughout the paper, we identify an instance of the index coding problem with its side information graph \( G \) and often write “index coding problem \( G \).” We also denote the broadcast rate and the capacity region of problem \( G \) with \( \beta(G) \) and \( \mathcal{C}(G) \) respectively.

The problem of broadcasting to multiple receivers with different side information traces back to the work by Celebi and Stette [2], Wyner, Wolf, and Willems [3], [4], Yeung [5], and Birk and Kol [6], [7]. The current problem formulation is due to the last. This problem has been shown to be closely related to many other important problems in network information theory such as network coding [8]–[10], locally
recoverable distributed storage [11]–[13], guessing games on directed graphs [8], [13], [14], and zero-error capacity of channels [15]. In addition, index coding has its own applications in diverse areas ranging from satellite communication [2]–[7] and multimedia distribution [16] to interference management [17] and coded caching [18], [19]. Due to this significance, the index coding problem has been broadly studied over the past two decades in several disciplines including graph theory, coding theory, and information theory, and various bounds have been established on the capacity region and the broadcast rate. Despite all these efforts, however, the problem is still open in general and the capacity in a computable (single-letter) expression is known only for a handful of special cases (see, for example, [14], [17], [20]–[31]).

Deviating from the common approach of finding the capacity by establishing tight upper and lower bounds, we take a more direct attack at the capacity itself by its structural properties using several graph-theoretic tools. The main contributions are summarized as follows:

- **A new multiletter characterization of the capacity (Theorem 2).** Paralleling the multiletter characterization of the capacity (broadcast rate) via the chromatic number of the confusion graph [32], we establish a multiletter characterization via the clique number of the confusion graph. As a corollary, we establish a nonasymptotic upper bound on the broadcast rate via the Lovász theta function of the confusion graph that can be computed more efficiently than the existing upper bound using the chromatic number.

- **Multiplicativity of the capacity under the lexicographic product (Theorem 5).** As another corollary of the aforementioned clique-number characterization, we show that if the side information graph is the lexicographic product of two graphs, the capacity is the product of the capacities of the two component graphs, completing an earlier result by Blasiak, Kleinberg, and Lubetzky [33]).

- **Conditions on the criticality of an index coding instance (Theorem 6 and Proposition 10).** Providing a partial answer to the question raised by Tahmasbi, Shahrasbi, and Gohari [34], we establish conditions under which the removal of an edge reduces the capacity. Both sufficient and necessary conditions are based on the notion of unicycle that is closely related to the maximum acyclic induced subgraph bound on the capacity.

The rest of the paper is organized as follows. Sections II and III review graph-theoretic preliminaries and some of the previously known bounds on the capacity, respectively. In Section IV, we introduce the notion of confusion graph associated with a given index coding problem and establish several properties including a tight bound on the chromatic number of a confusion graph in terms of its clique number. In Section V, we characterize the broadcast rate of a general index coding problem via asymptotic expressions involving the clique number, Shannon capacity, and Lovász theta function of the confusion graph. Nonasymptotic upper bounds on the broadcast rate are also established in terms of the Shannon capacity and Lovász theta function of the confusion graph. Based on the clique-number characterization, we prove in Section VI that the broadcast rate is multiplicative under the lexicographic product of side information graphs. In Section VII, we investigate the criticality problem and present sufficient and necessary conditions based on the notion of unicycle. Section VIII concludes the paper with an application of the established structural properties in computing the capacity for index coding problems with six messages.

**II. Mathematical Preliminaries**

Throughout the paper, a graph $G = (V, E)$ (without a qualifier) means a directed, finite, and simple graph, where $V = V(G)$ is the set of vertices (nodes) and $E = E(G) \subseteq V \times V$ is the set of directed edges. A graph $G = (V, E)$ is said to be unidirectional if $(i, j) \in E$ implies $(j, i) \notin E$. Similarly, $G$ is said to be bidirectional if $(i, j) \in E$ implies $(j, i) \in E$. Given $G$, its associated undirected graph $U = U(G)$ is defined by identifying $V(U) = V(G)$ and $E(U) = \{(i, j) : (i, j) \in E(G) \text{ or } (j, i) \in E(G)\}$. A bi-directional graph $G$ is sometimes identified with its undirected graph. The complement $\overline{G}$ of the graph $G$ is defined by $V(\overline{G}) = V(G)$ and $(i, j) \in E(\overline{G})$ iff $(i, j) \notin E(G)$. For any $S \subseteq V(G)$, $G[S]$ denotes the subgraph induced by $S$, i.e., $V(G[S]) = S$ and $E(G[S]) = \{(i, j) : i, j \in S\}$.

An independent set $I$ of a graph $G$ is a set of vertices with no edge among them. The independence number $\alpha(G)$ is the size of the largest independent set of the graph $G$. A clique $K$ of a graph $G$ is a set of vertices such that there is a (directed) edge from every vertex in $K$ to every other vertex in $K$. Thus, $K$ is a clique of $G$ iff it is an independent set of $\overline{G}$. The clique number $\omega(G)$ is the size of the largest clique of the graph $G$.

It is easy to see that

$$\omega(G) = \alpha(\overline{G})$$

for any directed or undirected graph $G$. A Hamiltonian cycle of a graph is a cycle that visits each vertex exactly once. A graph possessing a Hamiltonian cycle is said to be Hamiltonian.

**A. Chromatic Number**

A (vertex) coloring of an undirected (finite simple) graph $U$ is a mapping that assigns a color to each vertex such that no two adjacent vertices share the same color. The chromatic number $\chi(U)$ 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 \( \chi^{(b)}(U) \) is the minimum number of colors such that a \( b \)-fold coloring exists. The fractional chromatic number of the graph is defined as

\[
\chi_f(U) = \lim_{b \to \infty} \frac{\chi^{(b)}(U)}{b} = \inf_b \frac{\chi^{(b)}(U)}{b},
\]

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

\[
\chi_f(U) \leq \chi(U). \tag{3}
\]

Let \( I \) be the collection of all independent sets in \( U \). The chromatic number and the fractional chromatic number are also characterized via the following optimization problem

\[
\begin{align*}
\text{minimize} & \quad \sum_{S \in I} \rho_S \\
\text{subject to} & \quad \sum_{S \in I, j \in S} \rho_S \geq 1, \quad j \in V.
\end{align*}
\]

When the optimization variables \( \rho_S, S \in I \), take integer values in \( \{0, 1\} \), then the (integer) solution is the chromatic number, i.e., \( \chi(U) \). If this constraint is relaxed and \( \rho_S \in [0, 1] \), then the (rational) solution is the fractional chromatic number \([35]\). The (fractional) chromatic number can be related to the independence and clique numbers.

**Lemma 1** (Scheinerman and Ullman [35]). For any undirected graph \( U \) with \( n \) vertices,

\[
\frac{n}{\alpha(U)} \leq \chi_f(U).
\]

**Lemma 2.** For any graph \( U \) we have

\[
\omega(U) \leq \chi_f(U) \leq \chi(U).
\]

An undirected graph \( U = (V, E) \) is said to be perfect if for every induced subgraph \( U[S], S \subseteq V \), the clique number equals the chromatic number, i.e., \( \omega(U[S]) = \chi(U[S]) \). Perfect graphs can be characterized as follows.

**Proposition 1** (Chudnovsky, Robertson, Seymour, and Thomas [36]). A graph \( U \) is perfect iff no induced subgraph of \( U \) is an odd cycle of length at least five (odd hole) or the complement of one (odd antihole).

Let \( U = (V, E) \) be an undirected graph with \( V = [n] \). For each clique \( K \) of \( U \), the incidence vector is an \( n \)-dimensional vector whose \( j \)th component is equal to 1 if \( j \in K \) and 0 otherwise. The clique polytope of \( U \) is defined as

\[
P_K(U) = \{x \in \mathbb{R}^n_{\geq 0} : x \text{ is a convex combination of incidence vectors of cliques of } U\}. \tag{4}
\]

Another (convex) polytope associated with \( U \) is defined as

\[
P(U) = \{x \in \mathbb{R}^n_{\geq 0} : \sum_{i \in I} x_i \leq 1 \text{ for all independent sets } I\}. \tag{5}
\]

Since every incidence vector \( x \) of a clique satisfies \( \sum_{i \in I} x_i \leq 1 \) for an independent set \( I \), \( P_K(U) \subseteq P(U) \) for every \( U \).

Lovász’s perfect graph theorem states that equality holds iff \( U \) is perfect.

**Lemma 3** (Lovász [37]). For any graph \( U \) the following statements are equivalent:

- \( U \) is perfect.
- \( P_K(U) = P(U) \).
- \( \bar{U} \) is perfect.

We now state a result on chromatic numbers that will be useful later. The chromatic number of a graph can be upper bounded by decomposing it into smaller graphs. The following decomposition result will be proved in Appendix A.

**Lemma 4.** Let \( U_1 = (V, E_1) \) and \( U_2 = (V, E_2) \) be two undirected graphs on the set of vertices \( V \). Consider the graph \( U = (V, E_1 \cup E_2) \) defined on the same vertex set \( V \) in which each edge either belongs to \( E_1 \) or \( E_2 \). Then

\[
\chi(U) \leq \chi(U_1) + \chi(U_2).
\]

**B. Graph Products**

Generally speaking, a graph product is a binary operation on two graphs that produces a graph on the Cartesian product of the original vertex sets with the edge set constructed from the original edge sets according to certain rules. In the following, \( v_1 \sim v_2 \) denotes that there exists an edge between \( v_1 \) and \( v_2 \).

Given two undirected graphs \( U_1 \) and \( U_2 \), the disjunctive product \( U = U_1 \lor U_2 \) [35], [38] is defined as \( V(U) = V(U_1) \times V(U_2) \) and \( (u_1, u_2) \sim (v_1, v_2) \) iff

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

Throughout the paper, \( U_{\lor k} \) denotes the disjunctive product of \( k \) copies of \( U \).

Given two undirected graphs \( U_1 \) and \( U_2 \), the strong product \( U = U_1 \otimes U_2 \) [39] is defined as \( V(U) = V(U_1) \times V(U_2) \) and \( (u_1, u_2) \sim (v_1, v_2) \) iff

\[
(\begin{array}{c}
u_1 = v_1 \quad \text{and} \quad u_2 \sim v_2, \\
u_1 \sim v_1 \quad \text{and} \quad u_2 = v_2,
\end{array}) \quad \text{or} \quad (u_1 \sim v_1 \quad \text{and} \quad u_2 \sim v_2).
\]

Throughout the paper, \( U_{\otimes k} \) denotes the strong product of \( k \) copies of \( U \). The following lemma elucidates the relation between the disjunctive product and the strong product.

**Lemma 5. For any two undirected graphs \( U_1 \) and \( U_2 \)

\[
\overline{U_1 \lor U_2} = U_1 \otimes U_2.
\]

Given two graphs \( G_1 \) and \( G_2 \), the lexicographic product \( G = G_1 \circ G_2 \) [39] is defined as \( V(G) = V(G_1) \times V(G_2) \) and \( ((u_1, u_2), (v_1, v_2)) \in E(G) \) iff

\[
(u_1, v_1) \in E(G_1) \quad \text{or} \quad (u_1 = v_1 \quad \text{and} \quad (u_2, v_2) \in E(G_2)).
\]

The lexicographic product \( G_1 \circ G_2 \) can be thought of as replacing each vertex \( i \in V(G_1) \) with a copy of \( G_2 \). Therefore, the edges among the vertices of each copy of \( G_2 \) remain the same as in \( G_2 \) and there exists a directed edge from every vertex in copy \( i \) of \( G_2 \) to every vertex in copy \( j \) of \( G_2 \) iff \( (i, j) \in E(G_1) \) (see Fig. 2 for an example).
The lexicographic product can be generalized as follows. Let \( G_0 \) be a graph with \( m \) vertices and \( G_1, \ldots, G_m \) be \( m \) graphs with \( n_1, \ldots, n_m \) vertices respectively. The \textit{generalized lexicographic product} \( G_0 \circ (G_1, \ldots, G_m) \) is defined to be a graph on \( n_1 + \cdots + n_m \) vertices in which vertex \( i \in V(G_0) \) is replaced with \( G_i \), i.e., the edges among the vertices of \( G_i \) remain the same as before and there is a directed edge from every vertex of \( G_i \) to every vertex of \( G_j \) if \( (i, j) \in E(G_0) \).

\[ \text{Graphs } G_1 \text{ and } G_2 \text{ and their lexicographic product } G_1 \circ G_2. \]  

The bold arrows indicate that there is an edge from every vertex in the circle attached to the tail of the arrow to every vertex in the circle attached to the head of the arrow.

C. Shannon Capacity of a Graph and Lovász Function

Consider a graph \( U \) whose vertices represent input symbols of a noisy channel and two vertices are connected iff the corresponding channel inputs are confusable as they may result in the same channel output. The goal is to find the zero-error capacity of the channel represented by the graph \( U \). If we are limited to use the channel only once, then we can send up to \( \lceil \log(\alpha(U)) \rceil \) bits without an error. However, if we are allowed to use the channel \( t \) times, then we can construct the following graph to capture the confusabilities. Assign each \( t \)-tuple of the input symbols to a vertex and the vertices for two tuples \( x^t \) and \( z^t \) connect iff for every \( i, x_i = z_i \) or \( x_i \sim z_i \) in \( U \). We can easily check that the resulting graph is the strong product \( U^\otimes t \). Thus, by using the channel \( t \) times, we can send \( \lceil \log(\alpha(U^\otimes t)) \rceil \) bits without an error. Based on this observation [40], the Shannon capacity of a graph \( U \) is defined as

\[
\Theta(U) = \sup_t \sqrt[t]{\alpha(U^\otimes t)} = \lim_{t \to \infty} \sqrt[t]{\alpha(U^\otimes t)}. \tag{6}
\]

In other words, \( \log(\Theta) \) indicates the number of bits per input symbol that can be sent through the channel without error. By definition,

\[
\alpha(U) \leq \Theta(U). \tag{7}
\]

Shannon [40] showed that for perfect graphs \( \alpha(U) = \Theta(U) \).

The value of an orthonormal representation is defined as

\[
\min_{c:||c||=1} \max_{i \in [n]} \frac{1}{(c^Tv_i)^2}.
\]

The unit vector \( c \) attaining the minimum is referred to as the handle of the representation. The \textit{Lovász theta function} of \( U \), denoted as \( \vartheta(U) \), is defined to be the minimum value over all orthonormal representations of \( U \). A representation is said to be optimal if it attains this minimum.

Lemma 6 (Lovász [41]). For any undirected graph \( U \),

\[
\Theta(U) \leq \vartheta(U).
\]

By (2), (7), Lemma 6, and Theorem 10 in [41], the Lovász theta function is sandwiched by other graph-theoretic quantities that are NP-hard to compute.

Lemma 7. For any undirected graph \( U \),

\[
\omega(U) \leq \vartheta(U) \leq \chi(U).
\]

However, the Lovász theta function \( \vartheta(U) \) is polynomially computable in \( |V(U)| \) [42].

III. Bounds on the Capacity

The simplest approach to index coding is a coding scheme by Birk and Kol [6] that partitions the side information graph \( G \) by cliques and transmits the binary sums (parities) of all the messages in each clique.

Proposition 2 (Clique covering bound). Let \( \beta_{CC}(G) \) be the minimum number of cliques that partition \( G \), or equivalently, the chromatic number of \( U(G) \), which is the solution to the integer program

\[
\begin{align*}
\text{minimize} & \quad \sum_{S \in \mathcal{K}} \rho_S \\
\text{subject to} & \quad \sum_{S \in \mathcal{K} : j \in S} \rho_S \geq 1, \quad j \in V(G), \quad (8) \\
& \quad \rho_S \in \{0, 1\}, \quad S \in \mathcal{K},
\end{align*}
\]

where \( \mathcal{K} \) is the collection of all cliques in \( G \). Then for any index coding problem \( G \), \( \beta(G) \leq \beta_{CC}(G) \).

This bound, which is achieved by time division over a clique partition, has been extended in several directions. First, Birk and Kol [6] showed that one can use an MDS code over a finite field and perform time division over arbitrary subgraphs (partial cliques) instead of cliques. The number of parity symbols needed for a subgraph \( H \) is characterized by the difference \( \kappa(H) \) between the number of vertices in \( H \) and the minimum indegree within \( H \).

Proposition 3 (Partial clique covering bound). If \( G_1, \ldots, G_m \) partition \( G \), then the optimal broadcast rate is upper bounded by

\[
\beta_{PC}(G_1, \ldots, G_m) = \sum_{i=1}^{m} \kappa(G_i) \tag{9}
\]

and thus by

\[
\beta_{PC} = \min_{G_1, \ldots, G_m} \beta_{PC}(G_1, \ldots, G_m),
\]

where the minimum is over all partitions.
Lemma 3. If the graph $G$ with $n$ vertices is Hamiltonian, then the minimum indegree is at least one and thus $\beta_{PC} \leq n-1$, or equivalently, the symmetric rate $(\frac{1}{n-1}, \ldots, \frac{1}{n-1})$ is achievable for problem $G$.

By the standard time-sharing argument, Blasiak, Kleinberg, and Lubetzky [22] extended the clique covering bound to the fractional clique covering bound, which is equivalent to the fractional chromatic number of $U(G)$, namely, the solution to the linear program obtained by relaxing the integer constraint $\rho_S \in \{0, 1\}$ in (8) to $\rho_S \in [0, 1]$.

Remark 2. The integral, partial, and fractional clique covering bounds can be readily extended to the corresponding inner bounds on the capacity region. For example, by fractional clique covering, a rate tuple $(R_1, \ldots, R_n)$ is achievable for the index coding problem $(j | A_j)$, $j \in [n]$, if there exists $(\rho_S \in [0, 1], S \in K)$ such that

$$\sum_{S \in K} \rho_S \leq 1,$$
$$\sum_{S \in K : j \in S} \rho_S \geq R_j, \quad j \in V(G).$$  \hspace{1cm} (10)$$

Tighter bounds can be found in [25], [26], [28], [43]. In this paper, we only need the simpler integral, partial, and fractional clique covering bounds.

As for bounding the broadcast rate from below, Bar-Yossef, Birk, Jayram, and Kol [20] proposed the following.

Proposition 4 (Maximum acyclic induced subgraph (MAIS) bound). For any index coding problem $G$

$$\beta_{MAIS}(G) := \max_{S \subseteq V(G) : G[S] \text{ is acyclic}} |S| \leq \beta(G).$$

Remark 3. Since every independent set is acyclic, Proposition 4 implies that for any $G$, $\alpha(G) \leq \beta(G)$.

Remark 4. When $G$ is bidirectional (undirected) and perfect we have $\omega(G) = \alpha(G) = \chi(G)$. Hence, the upper bound of Proposition 2 matches the lower bound of Remark 3 and the broadcast rate is known [20].

Remark 5. The MAIS bound can be generalized to an outer bound $\beta_{MAIS}^*$ on the capacity region [25] as follows. If a rate tuple $(R_1, \ldots, R_n)$ is achievable for index coding problem $G$, then

$$\sum_{j \in S} R_j \leq 1$$  \hspace{1cm} (11)$$

for all $S$ such that $G[S]$ is acyclic. This bound is a special case of the polymatroidal outer bound [33], [44], [45].

Remark 6. When $G$ is bidirectional (undirected), the polytope associated with $G$ in (5) is equivalent to the MAIS outer bound in (11). It is also easy to see that the rate tuple given by each incidence vector of cliques in $G$ is achievable by clique covering and thus the polytope associated with $G$ in (4) is achievable by fractional clique covering. Therefore, by Lemma 3, if $G$ is bidirectional and perfect, then the capacity region is equal to the MAIS outer bound in (11), which is achieved by fractional clique covering [14].

IV. Confusion Graphs

The notion of confusion graph for the index coding problem was originally introduced by Alon, Hassidim, Lubetzky, Stav, and Weinstein [32]. In the context of guessing games, an equivalent notion was introduced independently by Gadow and Rii [46]. Consider a directed graph $G = (V, E)$ with $V = [n]$. Let $A_j = \{i \in V : (i, j) \in E\}$, $j \in [n]$, and let $t = (t_1, \ldots, t_n)$ be a length-$n$ integer tuple. Two $q$-ary $n$-tuples $x^n, z^n \in \prod_{i=1}^n \{0, \ldots, q-1\}^t$ are said to be confusable at position $l \in [t]$ of node $j \in [n]$ if $x_{jl} \neq z_{jl}$ and $x_i = z_i$ for all $i \in A_j$.

Given a directed graph $G$ and a length-$n$ integer tuple $t = (t_1, \ldots, t_n)$, the confusion graph $\Gamma_t(G)$ at position $l$ is an undirected graph with $\prod_{i=1}^n q^{t_i}$ vertices such that every vertex corresponds to a $q$-ary tuple $x^n$ and two vertices are connected iff the corresponding $q$-ary tuples are confusable at position $l$ of receiver $j$.

Aggregating over all positions, we say that $x^n, z^n \in \prod_{i=1}^n \{0, \ldots, q-1\}^t$ are confusable if they are confusable at some position $l$ of some node $j$. The confusion graph $\Gamma_t(G)$ is defined as before based on confusion between each pair of vertices, or equivalently,

$$E(\Gamma_t(G)) = \bigcup_{j=1}^n \bigcup_{l=1}^t E(\Gamma_t^{(jl)}(G)).$$  \hspace{1cm} (12)$$

If $t = (t, \ldots, t)$, then $\Gamma_t(G)$ is simply denoted by $\Gamma_t(G)$. Fig. 3 shows $\Gamma_t^{(11)}(G)$, $\Gamma_t^{(21)}(G)$, and $\Gamma_t^{(31)}(G)$ as well as $\Gamma_t(G)$ corresponding to $t = (1, 1, 1)$ for $G$ in Fig. 1.

By Lemma 4 and (12), the chromatic number of $\Gamma_t(G)$ can be upper bounded by those of its components.

Proposition 5. $\chi(\Gamma_t(G)) \leq \sum_{j=1}^n \sum_{l=1}^t \chi(\Gamma_t^{(jl)}(G)).$

Each component confusion graph $\Gamma_t^{(jl)}(G)$ has the following properties.

Lemma 8. $\Gamma_t^{(jl)}(G)$ does not have any chordless cycle of length greater than four.

Lemma 9. The complement of $\Gamma_t^{(jl)}(G)$ does not have any chordless cycle of length greater than four.

The proofs of the lemmas are given in Appendices B and C. By Proposition 1, Lemma 8, and Lemma 9, the following is immediate.

Proposition 6. $\Gamma_t^{(jl)}(G)$ is perfect.

As the main contribution of this section, we now establish an upper bound on the chromatic number of a confusion graph in terms of its clique number.

Theorem 1. Given a directed graph $G$, a length-$n$ integer tuple $t = (t_1, \ldots, t_n)$, and a positive integer $q$, the confusion graph $\Gamma_t(G)$ satisfies

$$\chi(\Gamma_t(G)) \leq \left( \sum_{j=1}^n t_j \right) \omega(\Gamma_t(G)).$$  \hspace{1cm} (13)$$
Proof: Consider

\[
\chi(\Gamma_t(G)) \leq \sum_{j=1}^{n} \sum_{i=1}^{t_j} \chi(\Gamma_t^{(j)}(G))
\]

(14)

\[
= \sum_{j=1}^{n} t_j \chi(\Gamma_t^{(j)}(G))
\]

\[
\leq \sum_{j=1}^{n} t_j \omega(\Gamma_t(G))
\]

(15)

\[
= \sum_{j=1}^{n} t_j \omega(\Gamma_t(G)),
\]

where (14) follows by Proposition 5, (15) follows by Proposition 6, and (16) follows by (12).

V. MULTILETTER CHARACTERIZATIONS OF THE CAPACITY

Consider an index coding problem \( G \). Using the notion of confusion graph introduced in Section IV, Alon, Hassidim, Lubetzky, Stav, and Weinstein [32] showed that

\[
\frac{\beta_t(G)}{t} := \inf_{(t,r) \text{ codes}} \frac{r}{t} = \frac{1}{t} \left[ \log(\chi(\Gamma_t(G))) \right].
\]

(17)

To prove this, consider a coloring of the vertices of the confusion graph \( \Gamma = \Gamma_t(G) \) with \( \chi(\Gamma) \) colors. This partitions the vertices of \( \Gamma \) into \( \chi(\Gamma) \) independent sets. By the definition of the confusion graph, no two message tuples in each independent set are confusable and therefore assigning a unique 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. Hence, \( \beta_t(G) \leq \lceil \log(\chi(\Gamma_t(G))) \rceil \). Conversely, consider any \((t,r)\) index code that 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 \( \Gamma = \Gamma_t(G) \). Moreover, every message tuple is mapped to some index so that these independent sets partition \( \Gamma \).

Thus, \( \chi(\Gamma) \leq 2^r \), or equivalently, \( r \geq \lceil \log(\chi(\Gamma)) \rceil \), and hence \( \beta_t(G) \geq \lceil \log(\chi(\Gamma_t(G))) \rceil \).

Based on (17), Alon, Hassidim, Lubetzky, Stav, and Weinstein [32] established the following upper bound on the broadcast rate

\[
\beta(G) \leq \frac{1}{t} \left[ \log(\chi(\Gamma_t(G))) \right],
\]

(18)

for every positive integer \( t \), and established a multiletter characterization of the broadcast rate as

\[
\beta(G) = \lim_{t \to \infty} \frac{1}{t} \log(\chi(\Gamma_t(G))).
\]

(19)

In our earlier work [47], this characterization was strengthened using the fractional chromatic number as

\[
\beta(G) = \lim_{t \to \infty} \frac{1}{t} \log(\chi_f(\Gamma_t(G))).
\]

(20)

We now further strengthen this result and characterize the broadcast rate in terms of the clique number of the confusion graph.

Theorem 2. For any side information graph \( G \),

\[
\beta(G) = \lim_{t \to \infty} \frac{1}{t} \log(\omega(\Gamma_t(G))).
\]

(21)

Proof: By setting \( t = (t_1, \ldots, t) \) in Theorem 1 and recalling Lemma 2, we have

\[
\omega(\Gamma_t(G)) \leq \chi(\Gamma_t(G)) \leq nt \cdot \omega(\Gamma_t(G)).
\]

(22)

Hence,

\[
\lim_{t \to \infty} \frac{1}{t} \log(\chi(\Gamma_t(G))) = \lim_{t \to \infty} \frac{1}{t} \log(\omega(\Gamma_t(G)));
\]

(23)

which, combined with (19), completes the proof.

Note that since \( \omega(\Gamma) \leq \chi_f(\Gamma) \leq \chi(\Gamma) \) for any graph \( \Gamma \), Equation (20) can be derived as a corollary of Theorem 2.
Combining (7), Lemma 6, and Lemma 7, we have for any positive integer \( t \)
\[
\omega(\Gamma_t) = \alpha(\overline{T}_t) \leq \Theta(\overline{T}_t) \leq \vartheta(\overline{T}_t) \leq \chi(\Gamma_t). \tag{24}
\]
Thus, we can characterize the broadcast rate in terms of the Shannon capacity and the Lovász theta function of the complement of the confusion graph.

**Corollary 1.**
\[
\beta(G) = \lim_{t \to \infty} \frac{1}{t} \log \left( \Theta\left( \overline{T}_t(G) \right) \right).
\]

In summary, the broadcast rate can be characterized as the first order in the exponent of six well-known graph theoretic quantities associated with \( \Gamma(G) \) and its complement, namely, \( \omega(\Gamma_t) \), \( \alpha(\overline{T}_t) \), \( \Theta(\overline{T}_t) \), \( \vartheta(\overline{T}_t) \), \( \chi(\Gamma_t) \), and \( \chi(\overline{T}_t) \).

In the following, we present nonasymptotic upper bounds on the broadcast rate \( \beta(G) \) in terms of the Shannon capacity and the Lovász theta function that hold for every positive integer \( t \) and, due to (24), are tighter than the upper bound in (18).

**Theorem 3.** For any side information graph \( G \) and any positive integer \( t \),
\[
\beta(G) \leq \frac{1}{t} \log \left( \Theta\left( \overline{T}_t(G) \right) \right). \tag{25}
\]

*Proof:* Consider
\[
\omega(\Gamma_{tk}) \leq \omega(\Gamma_{t^k}) = \alpha(\overline{T}_{t^k}) = \alpha(\overline{T}_t^k), \tag{26}
\]
where the inequality holds since the set of edges of \( \Gamma_{t^k} \) contains the set of edges of \( \Gamma_{tk} \), and the last equality follows by Lemma 5. Now for any \( t \),
\[
\beta(G) = \lim_{k \to \infty} \frac{1}{t^k} \log \left( \frac{\omega(\Gamma_k)}{\alpha(\overline{T}_{t^k})} \right), \tag{27}
\]
\[
= \lim_{k \to \infty} \frac{1}{t^k} \log \left( \frac{\alpha(\overline{T}_{t^k})}{\omega(\Gamma_k)} \right), \tag{28}
\]
\[
\leq \lim_{k \to \infty} \frac{1}{t^k} \log \left( \frac{\alpha(\overline{T}_{t^k})}{\alpha(\overline{T}_{t^k})} \right), \tag{29}
\]
\[
= \lim_{k \to \infty} \frac{1}{t^k} \log \left( \frac{\alpha(\overline{T}_{t^k})}{\alpha(\overline{T}_{t^k})} \right), \tag{30}
\]
where (27) follows by Theorem 2, (28) holds since the limit of a subsequence is equal to the limit of the sequence, (29) follows by (26), and (30) follows by the definition of the Shannon capacity in (6).

**Corollary 2.** For any side information graph \( G \) and any positive integer \( t \),
\[
\beta(G) \leq \frac{1}{t} \log \left( \Theta\left( \overline{T}_t(G) \right) \right). \tag{31}
\]

**Remark 7.** Unlike the upper bounds in (18) and (25) in terms of the chromatic number and the Shannon capacity, the upper bound in (31) can be computed in polynomial time in the number of vertices of the confusion graph (see [42]).

**Remark 8.** Equation (20) can be generalized to characterize the capacity region \( \mathcal{C} \) of the index coding problem \( G \) as the closure of all rate tuples \((R_1, \ldots, R_n)\) such that
\[
R_j \leq \frac{t_j}{\log(\chi_f(\Gamma_j(G)))}, \quad j \in [n], \tag{32}
\]
for some \( t = (t_1, \ldots, t_n) \) [47]. By a sandwich argument similar to (22), \( \mathcal{C} \) can also be characterized in terms of \( \omega(\Gamma_k(G)) \) asymptotically as \( t \to \infty \).

**Remark 9.** Similar to the index coding problem, the optimal rate region of the locally recoverable distributed storage problem with recovery graph \( G \) [11], [12] is characterized as the closure of all rate tuples \((R'_1, \ldots, R'_n)\) such that
\[
R'_j \geq \frac{t_j}{\log(\alpha(\Gamma_j(G)))}, \quad j \in [n], \tag{33}
\]
for some \( t = (t_1, \ldots, t_n) \) [11], [13]. Based on the vertex transitivity of \( \Gamma_k \) which, inter alia, implies that \( \log(\alpha(\Gamma_k)) = \sum_{i=1}^n t_i - \log(\chi_f(\Gamma_k)) \), the relationship between the index coding capacity region in (32) and the distributed storage optimal rate region in (33) can be made precise. See [13] for the details.

**VI. LEXICOGRAPHIC PRODUCT OF SIDE INFORMATION**

We first establish an upper bound on the broadcast rate of the index coding problem whose side information graph is a general lexicographic product (recall the definition in Section II).

**Theorem 4.** Let \( G = ([m], E) \) be a directed graph with \( m \) vertices and \( F_1, \ldots, F_m \) be \( m \) directed graphs with \( \beta(F_i) \leq \cdots \leq \beta(F_m) \). Then
\[
\beta(G \circ (F_1, \ldots, F_m)) \\
\leq \beta(F_1) \beta(G) + \sum_{i=1}^{m-1} (\beta(F_{i+1}) - \beta(F_i)) \beta(G|_{[m] \setminus \{i\}}) \tag{34}
\]

The proof of the theorem is given in Appendix D.

**Remark 10.** For the special case in which \( G \) has two vertices, the upper bound in Theorem 4 is tight [34], [46], [47]. In particular, if \( G \) has either no edges or one edge (see Fig. 4(a) and 4(b)), then \( \beta(G \circ (F_1, F_2)) = \beta(F_1) + \beta(F_2) \), and if \( G \) is a complete graph on two vertices (see Fig. 4(c)), then \( \beta(G \circ (F_1, F_2)) = \beta(F_2) = \max\{\beta(F_1), \beta(F_2)\} \).

The following states another special case for which the bound in Theorem 4 is tight.

**Theorem 5.** For any two directed graphs \( G \) and \( F \),
\[
\beta(G \circ F) = \beta(G) \beta(F) \tag{35}
\]

In words, the broadcast rate is multiplicative under the lexicographic product of index coding side information graphs. Achievability was shown by Blasiak, Kleinberg, and Lubetzky [33]. It also follows from Theorem 4 by setting \( F_1 = \cdots = F_m = F \).
Theorem 2. \[ (35) \text{ follows by Lemma 10, and } (36) \text{ follows by Lemma 10.} \]

Fig. 4. Graph examples with (a) no interaction, (b) one-way interaction, and (c) complete interaction among its two parts (+/− indicates there is a bidirectional edge between every vertex on the left and every vertex on the right).

\[ F_m = F. \] The proof of the converse is based on the clique-number characterization of the broadcast rate in Theorem 2 and the following.

Lemma 10. For any \((t, r)\) index code for problem \(G \circ F\),

\[ \log[(\omega(\Gamma_t(F)))] \leq r. \]

The proof of the lemma is relegated to Appendix E.

Proof of the converse for Theorem 5: Consider

\[
\beta(G \circ F) = \lim_{l \to \infty} \frac{1}{l} \inf_{G \circ F} r
\]

\[ \geq \lim_{l \to \infty} \frac{1}{l} \log(\omega(\Gamma_t(F))) \beta(G) \]

\[ = \lim_{l \to \infty} \frac{1}{l} \log(\omega(\Gamma_t(F))) \beta(G) \]

\[ = \beta(F) \beta(G), \]

where (35) follows by Lemma 10, and (36) follows by Theorem 2.

Example 1. The graph shown in Fig. 5(a) can be considered as the lexicographic product \(G \circ F\) of two smaller graphs \(G\) and \(F\) shown in Fig. 5(b) and 5(c) respectively with \(\beta(G) = 2\) and \(\beta(F) = 2\). By Theorem 5, instead of directly computing the broadcast rate for this six-message problem, we can use the known broadcast rates of smaller problems and get \(\beta(G \circ F) = 4\). Note that although this six-message problem has a certain symmetric structure, it does not fall into the class of cyclically symmetric index coding problems studied by Maleki, Cadambe, and Jafar [29].

The bound in Theorem 4 is not tight in general, as illustrated by the following.

Example 2. Consider the following 7-message index coding problem \(G\)

\[ (1|2, 3, 4, 6, 7), (2|1, 3, 6, 7), (3|1, 4, 5, 7), (4|1, 2, 5, 6), \]

\[ (5|3, 4, 6, 7), (6|2, 4, 5, 7), (7|2, 3, 5, 6), \]

for which \(\beta(G) = 2.5\) [17]. Let \(F_1, \ldots, F_6\) be 1-message problems and \(F_7\) be the 2-message problem

\[ (1|2), (2|\cdot). \]

Then Theorem 4 yields

\[ \beta(G \circ (F_1, \ldots, F_7)) \leq 1 \times 2.5 + (2 - 1) \times 1 = 3.5. \]

This bound is not tight since the composite coding scheme [25], [26] achieves the tighter upper bound of 10/3 on \(\beta(G \circ (F_1, \ldots, F_7))\).

Remark 11. The upper bound on the broadcast rate in Theorem 4 can be generalized to an inner bound on the capacity region as follows. Denoting the capacity regions of the index coding problems \(G \circ (F_1, \ldots, F_m)\), \(G\), and \(F_1, \ldots, F_m\) by \(C\), \(C_0\), and \(C_1, \ldots, C_m\) respectively, we have

\[ \bigcup_{\alpha \in \mathcal{C}_0} \{ (\alpha_1 R_1, \ldots, \alpha_m R_m) : R_i \in C_i, i \in [m] \} \subseteq \mathcal{C}. \]  

(37)

For the special case in which \(G\) has two vertices, the inner bound in (37) is tight [34], [47], generalizing the results in Remark 10. If \(G\) has either no edge or only one edge, then

\[ \mathcal{C} = \bigcup_{\alpha \in [0,1]} \{ (\alpha R_1, (1-\alpha) R_2) : R_1 \in C_1, R_2 \in C_2 \}. \]

In other words, in this case, the capacity region of \(G \circ (F_1, F_2)\) is achieved by time division between the optimal coding schemes for two subproblems \(F_1\) and \(F_2\). If \(G\) is a complete graph on two vertices, then

\[ \mathcal{C} = \{ (R_1, R_2) : R_1 \in C_1, R_2 \in C_2 \}. \]

In other words, the capacity region of \(G \circ (F_1, F_2)\) is achieved by simultaneously using the optimal coding schemes for \(F_1\) and \(F_2\).

VII. CRITICAL INDEX CODING INSTANCES

As Remark 11 suggests, if an edge \(e\) of the side information graph \(G\) belongs to a directed cut, removing \(e\) does not reduce the capacity region. The Farkas lemma [48, Th. 2.2] states that each edge in a directed graph either lies on a directed cycle or belongs to a directed cut but not both. Hence, if edge \(e\) does not lie on any directed cycle, it can be removed from \(G\).
without affecting the capacity region. This was first observed by Tahmasbi, Shahrasbi, and Gohari [34], who then asked for general conditions under which an edge of the side information graph can be removed without reducing the capacity region.

Let $e$ be an edge of side information graph $G = (V, E)$. We denote the graph resulting from removing $e$ from $G$ by $G_e$, i.e.,

$$V(G_e) = V(G) \text{ and } E(G_e) = E(G) \setminus \{e\}.$$  

Given the index coding problem $G$, the edge $e \in E$ is said to be critical if $\mathcal{C}(G_e) \neq \mathcal{C}(G)$, or in other words, if the removal of $e$ from $G$ strictly reduces the capacity region. The index coding problem $G$ itself is said to be critical if every $e \in E(G)$ is critical. Thus, each critical graph (= index coding problem) cannot be made “simpler” into another one of the same capacity region.

Remark 11 can be paraphrased into the following necessary condition for criticality.

**Proposition 7** (Union-of-cycles condition [34]). If $G$ is critical, then every edge belongs to a directed cycle.

This simple condition, however, is not sufficient. For the index coding problem shown in Fig. 1, although the edge $2 \to 3$ lies on a directed cycle, it can be shown that the capacity region is characterized by

$$R_1 \leq 1,$$

$$R_2 + R_3 \leq 1,$$

with or without this edge.

To observe another simple necessary condition for criticality, consider an index coding problem $G$ with side information sets $A_1, \ldots, A_n$. These sets are said to be degraded if there exist $i, j \in V(G)$ such that $i \in A_j$ and $A_i \subseteq A_j$. In this case, the edge $i \to j$ can be removed since $x_i$ can be recovered at node $j$. This observation leads to the following necessary condition.

**Proposition 8** (Nondegradedness condition). If $G$ is critical, then side information sets must be nondegraded.

Satisfying the above two necessary conditions at the same time is still not sufficient for criticality. As an example, it can be checked that the side information graph shown in Fig. 6 satisfies both union-of-cycles and nondegraded conditions. However, it is not a critical graph as the capacity region is characterized by

$$R_1 + R_3 \leq 1,$$

$$R_3 + R_4 \leq 1,$$

$$R_2 + R_4 \leq 1,$$

$$R_2 + R_5 \leq 1,$$

$$R_3 + R_5 \leq 1,$$

$$R_1 + R_2 + R_3 + R_4 + R_5 \leq 2,$$

with or without the edge $4 \to 1$.

In order to find a tighter necessary condition, we now focus on a sufficient condition. Given a graph $G = (V, E)$, the vertex induced subgraph $G[S]$ is referred to as a unicycle if its set of edges is a (chordless) Hamiltonian cycle over $S$. Note that if the subgraph $G[S]$ is a unicycle, then $G[S']$ cannot be a unicycle for any $S'$ that is a proper subset or superset of $S$. As an example, in Fig. 7(a), $G[\{1,2,3\}]$ is a unicycle, but $G$ itself is not a unicycle. As another example, for the graph in Fig. 7(b), $G[\{1,2,3\}]$ and $G[\{1,3,4\}]$ are both unicycles.

The following states a sufficient condition for the criticality of a problem.

**Theorem 6** (Union-of-unicycles condition). If every edge of $G$ belongs to a vertex induced subgraph that is a unicycle, then $G$ is critical.

**Proof:** It suffices to show that removing each edge of $G = (V, E)$ that belongs to a unicycle strictly reduces the capacity region. Let $e$ be an edge of $G[S]$, where $S \subseteq V$ and $G[S]$ is a unicycle. The rate tuple $(R_1, \ldots, R_n)$ such that

$$R_i = \begin{cases} 0, & i \notin S, \\ \frac{1}{|S|}, & i \in S, \end{cases} \quad (38)$$

is achievable for index coding problem $G$ by partial clique covering (see Proposition 3 and Remark 1). The vertex-induced subgraph $G_e[S]$, however, is acyclic (since the Hamiltonian cycle of $G[S]$ is broken and by definition there is no other cycle). Therefore, by the MAIS outer bound, any rate tuple $(R'_1, \ldots, R'_n) \in \mathcal{C}(G_e)$ must satisfy

$$\sum_{i \in S} R'_i \leq 1. \quad (39)$$

The rate tuple in (38), however, does not satisfy (39) and thus is not in $\mathcal{C}(G_e)$. This implies that removing edge $e$ from $G$ strictly reduces the capacity region ($\mathcal{C}(G_e) \neq \mathcal{C}(G)$) and hence $e$ is critical.

**Remark 12.** If a graph satisfies the union-of-unicycles condition, it trivially satisfies the union-of-cycles condition. We now argue that, as expected, satisfying the union-of-unicycles
condition also implies the nondegradedness condition. Assume that \( G \) has degraded side information sets. Then, there exists an edge \( i \to j \) such that \( A_i \subseteq A_j \). We show that this edge cannot belong to a unicycle. If the edge \( i \to j \) does not belong to any cycle, then trivially it does not belong to any unicycle. Otherwise, it suffices to show that none of the cycles that contain this edge is a unicycle. Assume that \( i \to j \) lies on a cycle \( C = (i,j,\ldots,v) \), which by degradedness must have at least three vertices. Then, by definition, \( v \in A_i \) and, by the assumption, \( v \in A_j \). Therefore, \((j,\ldots,v)\) is also a cycle and \( C \) is not a unicycle.

The converse to Theorem 6, however, does not hold in general.

**Example 3.** The capacity region of the index coding problem with side information graph shown in Fig. 8 is characterized by

\[
R_1 + R_2 \leq 1, \\
R_1 + R_3 \leq 1, \\
R_1 + R_4 \leq 1, \\
R_2 + R_4 \leq 1, \\
R_2 + R_5 \leq 1, \\
R_3 + R_5 \leq 1,
\]

which is achievable by the composite coding scheme [25]. Although the edge \( 2 \to 5 \) does not belong to any unicycle, removing it from the side information graph reduces the capacity region to

\[
R_1 + R_2 \leq 1, \\
R_1 + R_3 \leq 1, \\
R_1 + R_4 \leq 1, \\
R_2 + R_4 \leq 1, \\
R_2 + R_5 \leq 1, \\
R_3 + R_5 \leq 1,
\]

which is also achievable by the composite coding scheme [25].

![Fig. 8. A critical 5-message index coding problem. Although the edge 2 \( \to \) 5 does not belong to any unicycle, it is critical. The capacity region is achieved by composite coding [25] with or without the edge 2 \( \to \) 5.](image.png)

Proposition 9. Edge \( e \) belongs to a unicycle iff the MAIS bound \( \mathcal{R}_{MAIS}(G_e) \) on \( \mathcal{C}(G_e) \) is a proper subset of the MAIS bound \( \mathcal{R}_{MAIS}(G) \) on \( \mathcal{C}(G) \).

Proposition 9 implies the following partial converse to Theorem 6.

**Proposition 10.** If \( G = (V, E) \) is critical, then

1) every edge \( e \in E \) belongs to a unicycle, or
2) the MAIS bound is not tight for \( G_e \), i.e., \( \mathcal{R}_{MAIS}(G_e) \neq \mathcal{C}(G_e) \), for every \( e \in E \) that does not belong to any unicycle.

In other words, \( e \) is not critical if it does not belong to any unicycle and the MAIS bound is tight for \( G_e \).

**Proof:** It suffices to show that if \( G \) is critical and there exists an edge \( e \) that does not belong to any unicycle, then the MAIS bound is not tight for \( G_e \). Since \( G \) is critical, we have \( \mathcal{C}(G_e) \subseteq \mathcal{C}(G) \). Assume by contradiction that the MAIS bound is tight for \( G_e \). Then

\[
\mathcal{R}_{MAIS}(G_e) = \mathcal{C}(G_e) \subseteq \mathcal{C}(G) \subseteq \mathcal{R}_{MAIS}(G),
\]

which contradicts Proposition 9.

Recall that the edge \( 2 \to 5 \) in Fig. 8 is critical and does not belong to any unicycle. As is suggested by Proposition 10 and verified by (40), the MAIS bound is not tight for the side information graph resulting from removing this edge.

Remark 13. The three necessary conditions in Propositions 7, 8, and 10 can be rewritten as follows. For an index coding problem with the edge \( e \) from \( i \) to \( j \), if \( e \) does not belong to a directed cycle, or \( A_i \subseteq A_j \), or \( e \) does not belong to a unicycle and the MAIS bound is tight for \( G_e \), then \( e \) can be removed without reducing the capacity region.

The next three examples demonstrate that these necessary conditions are mutually independent.

**Example 4.** The six-message problem

\[
(1|5,6), (2|6), (3|6), (4|6), (5|1), (6|2,3,4,5)
\]
satisfies the union-of-cycles and nondegradedness conditions. However, it does not satisfy the necessary condition in Proposition 10, as the edge \( 5 \to 6 \) does not belong to any unicycle and the MAIS bound is tight (and is achieved by the composite coding scheme) after removing this edge.

**Example 5.** The six-message problem

\[
(1|4,5), (2|5,6), (3|5), (4|1,6), (5|1,2), (6|2,3,4,5)
\]
satisfies the union-of-cycles condition and the necessary condition in Proposition 10. However, \( A_3 \subseteq A_6 \) and thus it does not satisfy the nondegradedness condition.

**Example 6.** The six-message problem

\[
(1|4,6), (2|5,6), (3|5), (4|1,6), (5|1,2), (6|2,4,5)
\]
satisfies the nondegradedness condition and the necessary condition in Proposition 10. However, the edge \( 5 \to 3 \) does not belong to any cycle and thus the problem does not satisfy the union-of-cycles condition.
For the rest of this section, we present a few results that relate the capacity of index coding problem \( G \) and its MAIS bound to those of simpler problems. Consider the graph \( G = (V, E) \) and let \( G' \) be the graph resulting from removing all edges of \( G \) that do not belong to any unicycle, i.e.,

\[
V(G') = V(G),
\]
\[
E(G') = \{ e \in E(G) : e \text{ in a unicycle of } G \}. \tag{41}
\]

**Proposition 11.** \( \mathcal{R}_{\text{MAIS}}(G') = \mathcal{R}_{\text{MAIS}}(G) \).

In words, the set of edges of \( G \) that do not belong to any unicycle, is the (maximum) set of edges that can be removed from \( G \) without changing the MAIS bound. The proof of the proposition, which is implied by Proposition 9, is presented in Appendix G.

This observation leads to a condition under which the capacity of index coding problem \( G \) is equal to the capacity of the simpler problem \( G' \).

**Proposition 12.** If the MAIS bound is tight for \( G' \), then

\[
\mathcal{R}_{\text{MAIS}}(G') = \mathcal{C}(G') = \mathcal{C}(G) = \mathcal{R}_{\text{MAIS}}(G).
\]

Consequently, if the MAIS bound is tight for \( G' \), then \( G \) is not critical and all the edges that do not belong to any unicycle can be removed without reducing the capacity.

**Proof of Proposition 12:** Since

\[
\mathcal{R}_{\text{MAIS}}(G') = \mathcal{C}(G') \subseteq \mathcal{C}(G) \subseteq \mathcal{R}_{\text{MAIS}}(G),
\]

the proof follows by Proposition 11.

**Remark 14.** It can be similarly shown that the result of Proposition 12 also holds for the broadcast rate. If \( \beta_{\text{MAIS}}(G') = \beta(G') \), then \( \beta_{\text{MAIS}}(G) = \beta(G) = \beta(G') = \beta_{\text{MAIS}}(G') \).

**Example 7.** Consider the side information graph \( G \) shown in Fig. 9, where edges \( 5 \to 3, 3 \to 1, \) and \( 6 \to 5 \) do not belong to any unicycle. It can be shown that the capacity region for problem \( G' \) is achieved by composite coding [25] and is characterized by

\[
R_1 + R_3 + R_4 \leq 1, \\
R_1 + R_3 + R_5 \leq 1, \\
R_2 + R_3 + R_4 + R_6 \leq 1, \\
R_2 + R_3 + R_5 + R_6 \leq 1, \tag{42}
\]

which is equal to its MAIS bound. Thus, by Proposition 12, \( G \) is not critical and its capacity is also characterized by (42).

Proposition 12, together with Remark 6, implies the following.

**Proposition 13.** If \( G' \) is bidirectional and \( U(G') \) is perfect, then \( \mathcal{C}(G) = \mathcal{R}_{\text{MAIS}}(G) \) which is achieved by the fractional clique covering scheme.

This result can be recast to an earlier result by Yi, Sun, Jafar, and Gesbert [14], using the following two lemmas that are proved in Appendices H and I.

**Lemma 11.** Consider \( G = (V, E) \) and let \( G' \) be the graph as defined in (41). The following statements are equivalent.

(1) For each clique \( K \) in \( U(G') \), \( G|_K \) is acyclic.

(2) For each \( S \subseteq V(G) \), if \( G|_S \) contains a cycle, then there exists a bidirectional edge in \( G|_S \), i.e., \( \exists i, j \in S \) such that \( (i, j) \in E(G) \) and \( (j, i) \in E(G) \).

(3) No unidirectional edge of \( G \) belongs to a unicycle.

(4) \( G' \) is bidirectional.

**Lemma 12.** If \( G' \) is bidirectional, then \( \overline{U(G')} = U(G) \).

By Lemma 3 (\( U \) is perfect iff \( \overline{U} \) is perfect), Lemma 11, and Lemma 12, we can now restate Proposition 13 as follows.

**Proposition 14.** \( (Yi, Sun, Jafar, and Gesbert [14]) \). If \( U(G') \) is perfect and for each clique \( K \) in \( U(G) \), \( G|_K \) is acyclic, then \( \mathcal{C}(G) = \mathcal{R}_{\text{MAIS}}(G) \) which is achieved by the fractional clique covering scheme.

Note that this proposition includes Remark 6 as a special case.

As a concrete application of Proposition 12, consider a side information graph \( G \) satisfying

\[
A_j \subseteq \{ j - 1, j + 1 \}, \quad j \in [n].
\]

If \( A_j = \{ j - 1, j + 1 \} \) or \( A_j = \{ j - 1 \} \) for all \( j \in [n] \), then every edge belongs to a unicycle. (For these cases, the capacity is known [29] and is achieved by the fractional local clique covering scheme [23].) Otherwise, \( G' \) is a bidirectional (undirected) perfect graph (by Proposition 1). Therefore, by Remark 6, the MAIS bound is tight for index coding problem \( G' \) and Proposition 12 implies the following.

**Corollary 3.** For the class of index coding problems satisfying

\[
A_j \subseteq \{ j - 1, j + 1 \}, \quad j \in [n],
\]

any edge that does not belong to a unicycle can be removed without reducing the capacity region. Thus, for this class of index coding problems, the union-of-unicycles sufficient condition is also necessary for the problem to be critical.

**Example 8.** In the side information graph shown in Fig. 10 (a), edges \( 5 \to 4, 4 \to 3, \) and \( 2 \to 1 \) do not belong to any unicycle. Hence, the two side information graphs shown in Fig. 10 have the same capacity region.

**VIII. Application: Index Coding With Six Messages**

The number of instances of the index coding problem with \( n \) messages, which is equal to the number of nonisomorphic
directed graphs with \( n \) vertices [49, Seq. A000273], blows up quickly with \( n \). Even when \( n \) is as small as six, there are 1,540,944 nonisomorphic instances. In this section, we apply the structural properties discussed earlier to identify the 6-message index coding instances for which the capacity can be characterized via the capacities of “simpler” problems. By Theorem 5 and Remark 11, if \( G \) can be decomposed into smaller graphs, then the capacity of \( G \) can be expressed as a simple function of the capacities of smaller problems with smaller graphs, then the capacity of \( G \) can be decomposed into smaller problems, while the structural properties discussed earlier to identify the 6-message instances in Examples 4 to 6 have connected complement and thus do not satisfy property \( P_2 \). This proves that there are instances satisfying \((P_1 \land \neg P_2)\) or \((P_3 \land \neg P_1 \land \neg P_2 \land \neg P_4)\), or \((P_4 \land \neg P_1 \land \neg P_2 \land \neg P_3)\). Moreover, the six-message problem 
\[
(1\{6\}, 2\{6\}, 3\{6\}, 4\{6\}, 5\{6\}, 6\{1, 2, 3, 4, 5\}
\]
satisfies \( P_2 \) but not \( P_1, P_3, \) or \( P_4 \). Therefore, checking all of these four properties is useful in removing instances that do not need further investigation.

Among the remaining 10,634 instances that are not simplified, the polymatroidal bound [25] is achieved by a simplified form of composite coding [25], [26] for 10,515 instances. This leaves at most 119 instances that require further investigation via a tighter bound using non-Shannon inequalities [30] and more general coding schemes. In addition, there are at most 853 noncritical instances that simplify to one of the 119 critical instances of unknown capacity. In summary, the capacity is now fully characterized for at least 1,540,944–119–853 = 1,539,972 index coding problems (99.9%) with 6 messages.

**IX. Acknowledgments**

This work was supported by the National Science Foundation under Grant CCF-1320895 and the Korean Ministry of Science, ICT and Future Planning under the Institute for Information and Communications Technology Promotion Grant B0132-15-1005 (Development of Wired-Wireless Converged 5G Core Technologies).

**APPENDIX A**

**Proof of Lemma 4**

Let \( V' \) be the set of vertices incident to the edges in \( E_2 \setminus E_1 \) and let \( U' = (V', E_2 \setminus E_1) \). In order to color the vertices of \( U' \), we first color the vertices in \( V \setminus V' \) with \( \chi(U_1) \) colors using the optimal coloring for \( U_1 \). Next, we color \( U' \) with \( \chi(U_2) \) additional colors using the optimal coloring for \( U_2 \), which is valid since \( V' \subseteq V \) and \( E_2 \setminus E_1 \subseteq E_2 \). This guarantees that any pair of adjacent vertices are assigned different colors, whether both of them belong to \( V' \) or to \( V \setminus V' \) or one to each. Therefore, there exists a proper coloring of \( U \) with at most \( \chi(U_1) + \chi(U_2) \) colors and thus \( \chi(U) \leq \chi(U_1) + \chi(U_2) \).

**APPENDIX B**

**Proof of Lemma 8**

It suffices to show that every cycle of length greater than four has a chord. Let \( v_1^n, v_2^n, \ldots, v_k^n \) be the vertices (each associated with an \( n \)-message tuple) of a length-\( k \)-cycle of \( \Gamma_n^{(j)}(G) \) for \( k \geq 5 \). Then \( v_1^n \sim v_2^n, v_2^n \sim v_3^n, \ldots, v_{k-1}^n \sim v_k^n \). Therefore, \( v_{1j}(l) \neq v_{2j}(l), v_{2j}(l) \neq v_{3j}(l), \ldots, v_{(k-1)j}(l) \neq v_{kj}(l) \), and \( v_{1j} = v_{2j} = \cdots = v_{kj} \). If \( v_{1j}(l) \neq v_{3j}(l) \), then since \( v_{1j} = v_{3j} \), we have \( v_1^n \sim v_3^n \) and the length-\( k \) cycle has a chord. Otherwise, since \( v_{1j}(l) = v_{3j}(l) \neq v_{4j}(l) \) and \( v_{1j} = v_{4j} \), we have \( v_1^n \sim v_4^n \) and again the cycle has a chord.

**APPENDIX C**

**Proof of Lemma 9**

It suffices to show that every cycle of length greater than four has a chord. Let \( v_1^n, v_2^n, \ldots, v_k^n \) be the vertices of a length-\( k \) cycle of \( \Gamma = \Gamma_n^{(j)}(G) \) for \( k \geq 5 \). Then \( v_1^n \sim v_2^n, v_2^n \sim v_3^n, \ldots, v_{k-1}^n \sim v_k^n \) in \( \Gamma \). If \( v_{1j}(l) = \cdots = v_{kj}(l) \),

---

**Table I**

The number of 6-message index coding instances that satisfy properties \( P_1 - P_4 \).

| Structural Property | Number of six-message instances |
|---------------------|---------------------------------|
| \( P_1 \)           | 493,936                         |
| \( P_2 \)           | 10,101                          |
| \( P_3 \)           | \( \geq 1,513,890 \)             |
| \( P_4 \)           | 1,336,566                       |
| \( \neg(P_1 \lor P_2 \lor P_3 \lor P_4) \) | \( \leq 10,634 \) |
then $v_1^n, v_2^n, \ldots, v_k^n$ form a clique in $\Gamma$ and thus the cycle is not chordless. Hence, assume without loss of generality that $v_{ij}(l) \neq v_{2j}(l)$, which implies $v_{1A_j} \neq v_{2A_j}$. We now consider two cases.

Case 1 ($v_{2j}(l) = v_{3j}(l)$): In this case, if $v_{1A_j} \neq v_{3A_j}$, then $v_1^n \sim v_2^n$ in $\Gamma$ and the length-$k$ cycle has a chord. Suppose $v_{1A_j} = v_{3A_j}$ and consider $v_{4j}(l)$. If $v_{4j}(l) = v_{2j}(l)$, then $v_3^n \sim v_4^n$ in which is a chord for the length-$k$ cycle. Suppose $v_{4j}(l) \neq v_{2j}(l)$, then, since $v_3^n \sim v_4^n$ in $\Gamma$ we have $v_{3A_j} \neq v_{4A_j}$ and hence $v_{1A_j} \neq v_{4A_j}$. Therefore, $v_1^n \sim v_4^n$ in $\Gamma$ and the length-$k$ cycle has a chord.

Case 2 ($v_{2j}(l) \neq v_{3j}(l)$): In this case, if $v_{3j}(l) = v_{3j}(l)$, then $v_1^n \sim v_2^n$ in $\Gamma$ which is a chord. Suppose $v_{3j}(l) \neq v_{3j}(l)$. If $v_{1A_j} \neq v_{3A_j}$, then $v_1^n \sim v_3^n$ in $\Gamma$ which is a chord. Suppose $v_{1A_j} = v_{3A_j}$. If $v_{3j}(l) = v_{4j}(l)$, then the situation will be the same as case 1. Otherwise, we have that $v_{3A_j} \neq v_{4A_j}$ which implies $v_{1A_j} \neq v_{4A_j}$, and thus $v_1^n \sim v_4^n$ in $\Gamma$ which is a chord.

**APPENDIX D**

**PROOF OF THEOREM 4**

Fix $\epsilon > 0$. By the definition of the broadcast rate in (1) and (17), for sufficiently large $t$, there exists a $(t, r_i)$ index code for problem $F_i$ satisfying

$$\beta(F_i) \leq \frac{r_i}{t} \leq \beta(F_i) + \epsilon, \quad i \in [m].$$

(43)

Let $I := \{i \in [m - 1]: r_{i+1} - r_i > 0\}$, and $k$ be a sufficiently large integer such that there exist a $(kr_1, s_0)$ index code for problem $G$ satisfying

$$\frac{s_0}{kr_1} \leq \beta(G) + \epsilon,$$

(44)

and a $(k(r_{i+1} - r_i), s_i)$ index code for problem $G[m\setminus i]$ satisfying

$$\frac{s_i}{k(r_{i+1} - r_i)} \leq \beta(G[m\setminus i]) + \epsilon, \quad i \in I.$$

(45)

Consider the following coding scheme that consists of $m$ inner codes and at most $m$ outer codes. First, for each $i \in V(G)$, the $(t, r_i)$ index code for problem $F_i$ is applied to the messages indexed by $\{i\} \times V(F_i)$. This inner code is deployed $k$ times to generate $kr_i$ bits. As the second step, the outer codes are used to send these $k \sum_{i \in [m]} r_i$ bits. The $(kr_1, s_0)$ index code for problem $G$ is used to send the first $kr_1$ bits generated from copies of problems $F_1, \ldots, F_m$. (If $r_i < r_1$ for some $i$, zero-pad to get sufficient number of bits.) Next, for each $i \in I$, the $(k(r_{i+1} - r_i), s_i)$ index code for problem $G[m\setminus i]$ is used to send $k(r_{i+1} - r_i)$ bits generated from copies of $F_{i+1}, \ldots, F_m$ (if required, zero-pad to get sufficient number of bits). Using the above scheme, all the bits generated from messages indexed by $\{i\} \times V(F_i), i \in [m]$, are sent.

As for the decoding, first the decoders of the outer codes are utilized to recover the $kr_1$ bits corresponding to messages indexed by $\{i\} \times V(F_i)$. Next, each of the decoders of the inner codes is used $k$ times to recover the $kt$ bits of each message.

This coding scheme uses $s_0 + \sum_{i \in I} s_i$ bits to send $kt$ bits for each message of the problem $G \circ (F_1, \ldots, F_m)$. Now consider

$$s_0 + \sum_{i \in I} s_i$$

$$\leq kr_1(\beta(G) + \epsilon)$$

$$+ \sum_{i \in I} k(r_{i+1} - r_i)(\beta(G[m\setminus i]) + \epsilon)$$

(46)

$$\leq kt(\beta(F_1) + \epsilon)(\beta(G) + \epsilon)$$

$$+ \sum_{i \in [m-1]} k(\beta(F_{i+1}) - \beta(F_i) + \epsilon)(\beta(G[m\setminus i]) + \epsilon),$$

(47)

$$\leq kt(\beta(F_1) + \epsilon)(\beta(G) + \epsilon)$$

$$+ \sum_{i \in [m-1]} k(\beta(F_{i+1}) - \beta(F_i) + \epsilon)(\beta(G[m\setminus i]) + \epsilon),$$

(48)

where (46) follows by (44) and (45), and (47) follows by (43), and (48) follows by the assumption of the theorem. Letting $\epsilon \to 0$ completes the proof.

**APPENDIX E**

**PROOF OF Lemma 10**

Let $K = \{y_1, y_2, \ldots, y_{|K|}\}$ be a maximum clique in $\Gamma_1(F)$ and let $k = \lfloor \log(|K|) \rfloor = \lfloor \log(\omega(\Gamma_1(F))) \rfloor$. By the definition of the broadcast rate in (1), $\beta(G) \leq r_g/k$ for any $(k, r_g)$ index code for problem $G$. Hence, it suffices to show that given any $(t, r)$ index code for problem $G \circ F$, a $(k, r_g)$ index code for problem $G$ can be constructed such that $r_g \leq r$.

Let $m = |V(G)|$ and $n = |V(F)|$. We denote a tuple of $mn$ messages of problem $G \circ F$ by $x = (x_1, \ldots, x_m)$, where $x_i = (x_{i1}, \ldots, x_{im})$ and $x_{ij} \in \{0, 1\}^t$ for $i \in [m]$ and $j \in [n]$. Consider the one-to-one mapping

$$f : \{0, 1\}^k \to \{y_1, y_2, \ldots, y_{2^k}\}$$

that maps the $k$-bit binary representation of $i - 1$ to $y_i, i \in [2^k]$.

Let $\phi_{G \circ F}$ be the encoder associated with the $(t, r)$ index code for problem $G \circ F$. For any message tuple $v = (v_1, \ldots, v_m), v_i \in \{0, 1\}^{r_i}$, of problem $G$ define

$$\phi_G(v_1, \ldots, v_m) = \phi_{G \circ F}(f(v_1), \ldots, f(v_m)).$$

(49)

The function $\phi_G$ in (49) is the encoder of an index code for problem $G \circ F$ if any two message tuples to which the same codeword is assigned are nonconfusable. Hence, it suffices to show that if $\phi_G(v_1, \ldots, v_m) = \phi_G(v'_1, \ldots, v'_m)$, then $v_1, \ldots, v_m$ and $v'_1, \ldots, v'_m$ are nonconfusable for problem $G$.

Suppose $\phi_G(v_1, \ldots, v_m) = \phi_G(v'_1, \ldots, v'_m)$. Then $\phi_{G \circ F}(f(v_1), \ldots, f(v_m)) = \phi_{G \circ F}(f(v'_1), \ldots, f(v'_m))$. By the definition of the mapping $f$, for every $i \in [m]$, either $f(v_i) = f(v'_i)$ or $f(v_i) \sim f(v'_i)$ in $\Gamma_1(F)$. As $\phi_{G \circ F}$ is the encoder of an index code for problem $G \circ F$, $(f(v_1), \ldots, f(v_m))$ and $(f(v'_1), \ldots, f(v'_m))$ are nonconfusable for problem $G \circ F$ and thus, if $f(v_i) \sim f(v'_i)$ in $\Gamma_1(F)$, then $f(v_i) \neq f(v'_i)$ for some $j \in A_i(G)$. Hence, since $f$ is one-to-one, for every $i \in [m]$, either $v_i = v'_i$ or $v_i \neq v'_i$ for some $j \in A_i(G)$. Therefore, $(v_1, \ldots, v_m)$ and $(v'_1, \ldots, v'_m)$ are nonconfusable for problem $G$ and (49) defines the encoder of a $(k, r_g)$ index code for
problem $G$ such that the set of codewords is a subset of the set of codewords of the $(t, r)$ index code for problem $G \circ F$, which implies $r_G \leq r$.

**APPENDIX F**

**PROOF OF PROPOSITION 9**

**Sufficiency.** If the MAIS bound on $\mathcal{C}(G_e)$ is a proper subset of the MAIS bound on $\mathcal{C}(G)$, there exists a subset $S \subseteq V$ such that $G|_S$ contains a cycle and $G|_e$ is acyclic. Let $S_{\text{min}}$ be a minimal such subset. Then, $G|_{S_{\text{min}}}$ is a unicycle that contains $e$.

**Necessity.** Let $G|_S$, $S \subseteq V$, be a unicycle that contains $e$. By the definition of unicycle, $G_e|_S$ is acyclic. Therefore, by the MAIS outer bound, any rate tuple $(R_1, \ldots, R_n) \in \mathcal{C}(G_e)$ must satisfy

$$\sum_{j \in S} R_j \leq 1. \quad (50)$$

However, since $G|_S$ is not acyclic, (50) is not implied by the MAIS outer bound on $\mathcal{C}(G)$.

**APPENDIX G**

**PROOF OF PROPOSITION 11**

Proposition 9, together with the following, implies Proposition 11.

**Lemma 13.** If $e_1$ and $e_2$ do not belong to any unicycle of $G$, then $e_2$ does not belong to any unicycle of $G_{e_1}$.

**Proof:** If $e_2$ does not belong to any cycle of $G$, then it trivially does not belong to any unicycle of $G_{e_1}$. Suppose $e_2$ belongs to some cycle in $G$. It suffices to show that for every cycle $C$ of $G$ that contains $e_2$, $C \setminus e_1$ is not a unicycle of $G_{e_1}$. Let $e_1 = (u_1, u_2)$, $e_2 = (v_1, v_2)$, and $C = (v_1, \ldots, v_l)$ be a cycle of $G$ that contains $e_2$. By the assumption, $C$ is not a unicycle and thus $l \geq 3$. If $\{(u_1, u_2) \cap \{v_1, \ldots, v_l\}\} < 2$, then removing $e_1$ does not affect $C$ and hence $C \setminus e_1$ is not a unicycle of $G_{e_1}$. Suppose $\{(u_1, u_2) \cap \{v_1, \ldots, v_l\}\} = 2$ and consider three cases.

Case 1: $e_1 = (v_i, v_{i+1})$ for some $i \in [l-1]$. In this case, removing $e_1$ breaks the cycle $C$ and hence $C \setminus e_1$ is not a unicycle of $G_{e_1}$.

Case 2: $e_1 = (v_i, v_j)$ for some $1 \leq i < j \leq l$, $(i, j) \neq (1, l)$. In this case, $(v_1, \ldots, v_i, v_j, \ldots, v_l)$ is a cycle of $G$ that contains both $e_2$ and $e_1$ and thus, by the assumption, is not a unicycle and has a chord, which is also a chord of $C \setminus e_1$. Thus, $C \setminus e_1$ is not a unicycle of $G_{e_1}$.

Case 3: $e_1 = (v_i, v_j)$ for some $1 \leq i < j \leq l$, $(i, j) \neq (1, l)$. In this case, $(v_i, \ldots, v_j)$ is a cycle of $G$ that contains $e_1$ and thus, by the assumption, is not a unicycle and has a chord, which is also a chord of $C \setminus e_1$. Thus, $C \setminus e_1$ is not a unicycle of $G_{e_1}$.

**APPENDIX H**

**PROOF OF LEMA 11**

$(1) \Rightarrow (2):$ Assume that $(2)$ does not hold. Then there exists a subset $S$ such that $G|_S$ contains a cycle but does not have any bidirectional edge. By the definition of $U(G)$, $S$ is a clique of $U(G)$, which contradicts $(1)$.

$(2) \Rightarrow (1):$ Assume that $(1)$ does not hold. Then there exists a clique $K$ in $U(G)$ such that $G|_K$ has a cycle. By the definition of $U(G)$, $G|_K$ has no bidirectional edge, which contradicts $(2)$.

$(2) \Rightarrow (3):$ Assume that there exists a unidirectional edge $e$ and $S \subseteq V$, $|S| \geq 3$, such that $G|_S$ is a unicycle and $e \in E(G|_S)$. By the definition of unicycle, all of the edges of $G|_S$ are unidirectional, which contradicts $(2)$.

$(3) \Rightarrow (2):$ Assume that $(2)$ does not hold. Then there exists a subset $S$, $|S| \geq 3$ such that $G|_S$ has a cycle but does not have any bidirectional edge. A minimal such $S$ forms a unicycle and hence all of its unidirectional edges belong to a unicycle, which contradicts $(3)$.

$(3) \Rightarrow (4):$ To form $G'$, every edge of $G$ that do not belong to a unicycle is removed. Hence, if $(3)$ holds, then all unidirectional edges of $G$ are removed to form bidirectional $G'$.

$(4) \Rightarrow (3):$ $G'$ is formed by removing edges of $G$ that do not belong to any unicycle. Hence, $G'$ is bidirectional implies that no unidirectional edge of $G$ belongs to a unicycle.

**APPENDIX I**

**PROOF OF LEMA 12**

Since $G'$ is bidirectional and every bidirectional edge belongs to a unicycle, we have

$$\{i, j\} \in E(U(G')) \iff \{i, j\} \in E(G) \text{ and } (j, i) \in E(G).$$

By definition,

$$\{i, j\} \notin E(U(G)) \iff \{i, j\} \in E(G) \text{ and } (j, i) \in E(G).$$

Thus,

$$U(G') = U(G).$$

**REFERENCES**

[1] M. Fekete, “Über die verteilung der wurzeln bei gewissen algebraischen gleichungen mit ganzzahligen koeffizienten,” Mathematische Zeitschrift, vol. 17, pp. 228–249, 1923.

[2] M. Celebiler and G. Stette, “On increasing the down-link capacity of a regenerative satellite repeater in point-to-point communications,” Proc. IEEE, vol. 66, no. 1, pp. 98–100, Jan. 1978.

[3] F. M. J. Willems, J. K. Wolf, and A. D. Wyner, “Communicating via a processing broadcast satellite,” in IEEE/CAM Inf. Theory Workshop, Cornell, NY, 1989, pp. 3–1.

[4] A. D. Wyner, J. K. Wolf, and F. M. J. Willems, “Communicating via a processing broadcast satellite,” IEEE Trans. Inf. Theory, vol. 48, no. 6, pp. 1243–1249, 2002.

[5] R. W. Yeung, “Multilevel diversity coding with distortion,” IEEE Trans. Inf. Theory, vol. 41, no. 2, pp. 412–422, 1995.

[6] Y. Birk and T. Kol, “Infomation-source coding-on-demand (ISCOD) over broadcast channels,” in Proc. 17th Ann. IEEE Int. Conf. Comput. Commun. (INFOCOM), San Francisco, CA, Mar. 1998, pp. 1257–1264.

[7] ——, “Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients,” IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.

[8] S. Rüis, “Information flows, graphs and their guessing numbers,” Elec. J. Comb., vol. 14, no. R44, Jun. 2007.

[9] S. El Rouayheb, A. Sprintson, and C. Georghiades, “On the relation between the index coding and the network coding problems,” in Proc. IEEE Int. Symp. Inf. Theory, Toronto, ON, Jul. 2008, pp. 1823–1827.

[10] M. Effros, S. El Rouayheb, and M. Langberg, “An equivalence between network coding and index coding,” IEEE Trans. Inf. Theory, vol. 61, no. 5, pp. 2478–2487, May 2015.

[11] 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.
