Index coding and local partial clique covers

Abhishek Agarwal and Arya Mazumdar

(Invited Paper)

***Abstract*—Index coding, or broadcasting with side information, is a network coding problem of most fundamental importance. In this problem, given a directed graph, each vertex represents a user with a need of information, and the neighborhood of each vertex represents the side information availability to that user. The aim is to find an encoding to minimum number of bits (optimal rate) that, when broadcasted, will be sufficient to the need of every user. Not only the optimal rate is intractable, but it is also very hard to characterize with some other well-studied graph parameter or with a simpler formulation, such as a linear program. Recently there have been a series of works that address this question and provide explicit schemes for index coding as the optimal value of a linear program with rate given by well-studied properties such as local chromatic number or partial clique-covering number. There has been a recent attempt to combine these existing notions of local chromatic number and partial clique covering into a unified notion denoted as the *local partial clique cover* (Arbabjolfaei and Kim, 2014).**

We present a generalized novel upper-bound (encoding scheme) - in the form of the minimum value of a linear program - for optimal index coding. Our bound also combines the notions of local chromatic number and partial clique covering into a new definition of the local partial clique cover, which outperforms both the previous bounds, as well as beats the previous attempt to combination.

I. Introduction

Index coding is a multiuser communication problem in which the broadcaster reduces the total number of transmissions by taking advantage of the fact that a user may have side information about other user's data. We consider the index coding problem formulation of [4]. Suppose, a set of users are assigned bijectively to a set of vectors that they want to know. However, instead of knowing the assigned vector, each knows a subset of other vectors. This scenario can be depicted by a so-called side-information graph where each vertex represents a user and there is an edge from user A to user B, if A knows the vector assigned to B. Given this graph, how much information should a broadcaster transmit, such that each vertex can deduce its assigned vector?

Consider a group of n users with the side information represented by the directed graph $\mathcal{G}$, such that the out-neighbors of vertex $v_i \in V(\mathcal{G}) = \{v_1, v_2, \dots, v_n\}$ denote the nodes whose data is available to v_i. The data requested by node v_i is represented by a vector $\mathbf{x}_i \in \mathbb{F}_q^\ell$. Let $N(v, \mathcal{G}) \subseteq V(\mathcal{G}) \setminus \{v\}$ denote the out-neighbors of vertex v ($v \in \overline{N(v, \mathcal{G})}$). For example, an index coding scenario has been depicted in fig. 1.

Abhishek Agarwal is with the Department of ECE, University of Minnesota, Minneapolis, MN 55455. Arya Mazumdar is with the Computer Science Department of University of Massachusetts, Amherst, MA 01003. email: abhiag@umn.edu, arya@cs.umass.edu. Research supported by NSF grants CCF 1318093 and CAREER CCF 1453121.

978-1-5090-1746-1/16/$31.00 © 2016 IEEE

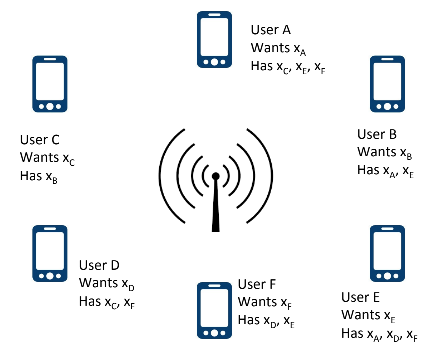

Fig. 1: Example of an Index Coding problem. The corresponding side-information graph is depicted in fig. 2

The side information graph for this scenario is given in fig. 2. In the graph $\mathcal{G}$ of fig. 2, $N(A, \mathcal{G}) = \{C, E, F\}, N(B, \mathcal{G}) = \{A, E\}, N(C, \mathcal{G}) = \{B\}, N(D, \mathcal{G}) = \{C, F\}, N(E, \mathcal{G}) = \{A, D, F\}, N(F, \mathcal{G}) = \{D, E\}$. The aim is to broadcast the minimum amount of data (in-terms of $\mathbb{F}_q^\ell$-symbols) such that vertex v_i is able to reconstruct $\mathbf{x}_i$ from $\mathbf{x}_{N(v_i, \mathcal{G})}$ (vectors assigned to the neighbors) and the broadcast for all $i \in \{1, \dots, n\}$. The amount of broadcasted data is referred to as the *broadcast rate* of the index coding problem.

This simple formulation attracted substantial attention recently [2], [6], [13]. In particular, it has been shown that any network coding problem can be cast as an index coding problem [7], [8]. Furthermore, index coding has been identified as the hardest instance of all of network coding [12], as well as related to locally recoverable codes [14]. For up to $n = 5$, the optimal broadcast rates of all index coding problems have been tabulated in [11].

If the recovery functions for the nodes, as well as the encoding function at the broadcaster are all linear over $\mathbb{F}_q$, then the optimal (minimum) broadcasting rate is given by a quantity called the *minrank* of the graph [4], [10]. In general, minrank is hard to compute for arbitrary graphs, and also does not give an achievable scheme in terms of simpler well known graph parameters.

There have been a series of works that aim to characterize the index coding broadcast rate in terms of other possibly intractable graph parameters or cast the index coding rate as the optimal solution to a linear program. To start with, an immediate achievable scheme in terms of the *clique cover* number the graph or the *chromatic number* of the complementary graph was provided in [5]. Their simple idea is to partition the graph into minimum number of vertex-disjoint cliques. And then for each clique, the broadcaster transmits only once, that is the sum (as $\mathbb{F}_q$-vectors) of the data in all the vertices of the clique. This provides an index coding solution equal to the chromatic number of the complementary graph.

The above approach then was extended in [15]. In particular, in [15] an interference alignment perspective of the linear index coding problem was given. It was shown that, it is possible to achieve an index coding solution with number of transmissions equal to the *local chromatic number* [9] of the graph. There also exists an orthogonal approach to generalize the clique covering bound (or the chromatic number bound). In this approach, appearing as early as in [5], one finds a *partial clique* cover (covering with subgraphs instead of cliques) and uses maximum-distance separable (MDS) codes for each of the partial cliques. A directed graph $\mathcal{G}$ on n vertices is a k-partial clique if every vertex has at least $n-1-k$ out-neighbors and there is at least one vertex in $\mathcal{G}$ with exactly $n-k-1$ out-neighbors

More recently, in [3], the above two approaches were merged and a combined linear program solution was proposed that we describe below in Theorem 1. There is another approach to combine the above two methods of local chromatic number and partial clique covering proposed in [16]. The approach in [16] uses the local chromatic number on partitions of the side information graph. The partitions are decided such that the sum of the local chromatic number of the complement graph of the partitions is minimized. However for the unicast index coding problem that we consider, Theorem 1 below can be shown to outperform this scheme of [16].

Arbabjolfaei and Kim [3] proposes the following result. Here $\mathcal{K}$ denotes the power set of vertices in $\mathcal{G}$ and $\mathcal{S} \in \mathcal{K}$ is a $k_{\mathcal{S}}$-partial clique.

Theorem 1 (Arbabjolfaei and Kim [3])**.** *The minimum broadcast rate of an index coding problem on the side information graph $\mathcal{G}$ is upper bounded by the optimal value of the following integer program,*

$$\begin{aligned} &\textit{minimize} && t \\ &\textit{subject to} && \textstyle\sum_{\mathcal{S}\in\mathcal{K}:\mathcal{S}\not\subseteq N(v,\mathcal{G})}(k_{\mathcal{S}}+1)\rho_{\mathcal{S}} \le t, \quad v \in V(\mathcal{G}) \\ &\textit{and} && \sum_{\mathcal{S}\in\mathcal{K}:v\in\mathcal{S}} \rho_{\mathcal{S}} \ge 1, \quad v \in V(\mathcal{G}) \\ & && \rho_{\mathcal{S}} \in \{0,1\}, \quad \mathcal{S}\in\mathcal{K}. \end{aligned} \tag{1}$$

In this paper, we give a modified definition of *local partial clique cover*, and provide a stronger bound than eq. (1) (see theorem 2 below). Our bound requires a significantly involved achievability scheme. This lead to better achievability bound for linear index codes than any previous approaches.

In addition, recently another approach to provide achievable index coding schemes is shown in [17], where the graph is covered by structures called *generalized interlinked cycles (GIC)*. We refrain from separately describing the scheme here. Instead we comment that the scheme can be generalized in spirit to that of k-partial clique covering from clique covering.

In the next section, we define the necessary graph parameters, and narrate our main theorems. We also give a concrete examples to show the better performance of our solution. All the proofs of our results are available in the full version [1].

II. Main result

We will need the following definitions.

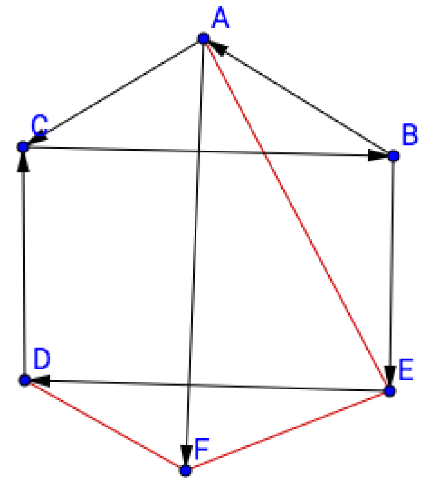

Fig. 2: Example of an Index Coding problem where the fractional local chromatic number and the fractional partial clique number based index codes are strictly larger than the proposed scheme for the fractional case. The undirected edges denote bidirectional edges.

Definition 1. *For a directed graph $\mathcal{G}$, a coloring of indices is proper if for every vertex v of $\mathcal{G}$ the color any of its out-neighbors, $N(v,\mathcal{G})$, is different from the color of vertex v. The minimum number of colors in a proper coloring is the chromatic number χ.*

Definition 2. *The local chromatic number χ_l of a directed graph $\mathcal{G}$ is the maximum number of colors in any out-neighborhood (including itself) of a vertex, minimized over all proper colorings of the undirected graph obtained from $\mathcal{G}$ by ignoring the orientation of edges in $\mathcal{G}$.*

The following definition of index coding and the optimal broadcast rate will be used throughout the paper. In the index coding problem, a *directed* side information graph $\mathcal{G}(V,E)$ is given. Each vertex $v_i \in V$ represents a receiver that is interested in knowing a uniform random vector $\boldsymbol{x}_i \in \mathbb{F}_q^{\ell}$. The receiver at v knows the values of the variables $\boldsymbol{x}_u, u \in N(v,\mathcal{G})$.

Definition 3 (Index coding)**.** *An* index code $\mathcal{C}$ *for* $\mathbb{F}_q^{\ell\times n}$ *with side information graph* $\mathcal{G}(V,E), V=\{v_1,v_2,\dots,v_n\}$*, is a set of codewords (vectors) in* $\mathbb{F}_q^{\ell\cdot n_0}$ *together with:*

1) *An encoding function f mapping inputs in $\mathbb{F}_q^{\ell\times n}$ to codewords, and*
2) *A set of deterministic decoding functions $g_1,\dots,g_n$ such that $g_i\Big(f(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n),\{\boldsymbol{x}_j : j\in N(v_i,\mathcal{G})\}\Big)=\boldsymbol{x}_i$ for every $i=1,\dots,n$.*

The encoding and decoding functions depend on $\mathcal{G}$. The number n_0 is called the broadcast rate of the index code $\mathcal{C}$ and ℓ is called the size *of a node. Note that n_0 can be a non-integer as long as $n_0\cdot\ell$ is integer. The minimum value of n_0 over all possible index codes and all possible ℓ is called the optimal broadcast rate.*

A. *Our main results*

Our main result is given by the following theorem.

Theorem 2. *The minimum broadcast rate of an index coding problem on the side information graph $\mathcal{G}$ is upper bounded by the optimum value of the following integer program.*

$$\begin{aligned}
&\textit{minimize} \quad t \\
&\textit{subject to} \sum_{\mathcal{S}\in\mathcal{K}} \min\{|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|,(k_{\mathcal{S}}+1)\}\rho_{\mathcal{S}} \le t, \quad v\in V(\mathcal{G}) && \text{(2a)}\\
&\textit{and} \sum_{\mathcal{S}\in\mathcal{K}:v\in\mathcal{S}} \rho_{\mathcal{S}} \ge 1, \quad v\in V(\mathcal{G}) && \text{(2b)}\\
&\rho_{\mathcal{S}}\in\{0,1\}, \quad \mathcal{S}\in\mathcal{K}, && \text{(2c)}
\end{aligned}$$

where $\overline{A}$ denotes the complement of set A.

The change in eq. (2) from eq. (1) is introduction of the term $|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|$ in the optimization under the summation. Although the set over which the summation is performed seemed to have expanded, it is not the case. Notice that, if $S\subseteq N(v,\mathcal{G})$ then $|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|=0$. This also shows that our scheme is at least as good as eq. (1). However the achievability scheme becomes significantly complicated because of the introduction of the term under the summation.

In comparison, for the linear program in eq. (1), a simpler achievability scheme is possible and we describe it in Appendix A for completeness. Note that, this simple achievability scheme was not mentioned in [3].

Any vertex subgraph of a k-partial clique is also a k-partial clique. Thus, the integer program in eq. (2) finds a partition of $\mathcal{G}$ into partial cliques i.e. the inequality in eq. (2b) can be replaced by equality. We call the optimal solution to eq. (2) the *local partial clique cover* number.

To construct a scheme to achieve the bound in theorem 2 we assign vectors to each vertex $u\in V(\mathcal{G})$, $\mathbf{v}_u$. And for each partial clique we use the column vectors of a generator matrix of an $(|\mathcal{S}|,k_{\mathcal{S}}+1)$-MDS code. The length of the index code is equal to the span of the vectors $\{\mathbf{v}_u\}_{u\in V(\mathcal{G})}$. The detailed scheme is described in [1].

The term $\min\{|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|,(k_{\mathcal{S}}+1)\}$ in eq. (2a) denotes the span of the non-neighbors of vertex v in the partition $\mathcal{S}$. Note that, we count at most $(k_{\mathcal{S}}+1)$ non-neighbors in a $k_{\mathcal{S}}$-partial clique because the span of the vectors corresponding to $\mathcal{S}$ is $k_{\mathcal{S}}+1$.

To find the local chromatic number, we partition the graph into cliques and minimize the maximum number of partitions in the neighborhood of any vertex in the complementary side information graph, $\bar{\mathcal{G}}$. In theorem 2, we replace the clique cover with a partial clique cover and minimize the maximum span of the vectors in the neighborhood of any vertex in $\overline{\mathcal{G}}$. This gives a generalization of the local chromatic number of $\overline{\mathcal{G}}$ for partial clique cover, and reduces to $\chi_l(\mathcal{G})$ when only 0-partial cliques are used. Since a clique cover is the special case of a k-partial cover (a clique is a 0-partial clique), eq. (2) trivially outperforms both the local chromatic number based-scheme and the partial clique cover scheme.

We can consider the fractional relaxation of the IP in eq. (2) and provide an even better upper-bound.

$$\begin{aligned}
&\text{minimize} \quad t \\
&\text{subject to} \quad \textstyle\sum_{\mathcal{S}\in\mathcal{K}} \min\{|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|,(k_{\mathcal{S}}+1)\}\rho_{\mathcal{S}} \le t, \quad v\in V(\mathcal{G})\\
&\text{and} \quad \sum_{\mathcal{S}\in\mathcal{K}:v\in\mathcal{S}} \rho_{\mathcal{S}} \ge 1, \quad v\in V(\mathcal{G})\\
&\rho_{\mathcal{S}}\in[0,1], \quad \mathcal{S}\in\mathcal{K}.
\end{aligned} \tag{3}$$

Theorem 3 (Linear programming relaxation)**.** *The minimum broadcast rate of an index coding problem on the side information graph $\mathcal{G}$ is at the most the optimum value of the linear programming of eq.* (3)*.*

The necessary adjustments required for the proof of this theorem is provided in [1].

Following the ideas of [3], we can further tighten the achievability bounds by considering recursive linear program.

Theorem 4 (Recursive LP)**.** *Let $IC_{FLP}(\mathcal{G})$ denote the optimizated value of the linear program below for graph $\mathcal{G}$:*

$$\begin{aligned}
&\textit{minimize} \quad t \\
&\textit{subject to} \quad \textstyle\sum_{\mathcal{S}\in\mathcal{K}} \min\{|\mathcal{S}\cap\overline{N(v,\mathcal{G})}|, IC_{FLP}(\mathcal{G}|_{\mathcal{S}})\}\rho_{\mathcal{S}} \le t, \quad v\in V(\mathcal{G})\\
&\textit{and} \quad \sum_{\mathcal{S}\in\mathcal{K}:v\in\mathcal{S}} \rho_{\mathcal{S}} \ge 1, \quad v\in V(\mathcal{G})\\
&\rho_{\mathcal{S}}\in[0,1], \quad \mathcal{S}\in\mathcal{K}.
\end{aligned} \tag{4}$$

Also define IC_{FLP} to be equal to 1 for singleton sets. The minimum broadcast rate of an index coding problem on the side information graph $\mathcal{G}$ is bounded from above by $IC_{FLP}(\mathcal{G})$. Here, for a graph $\mathcal{G}(V,E)$ and a set $S\subseteq V$, $\mathcal{G}|_S$ denotes the subgraph induced by S.

The proof of this theorem is given in [1].

1) Examples of better performance for our scheme: While it had been evident that our proposed bounds are at least as good as the previous bounds, we now give explicit example of graphs where our upper bounds are strictly smaller than all previous upper bounds. Consider the index coding problem described by the graph in fig. 2. For this graph, the index coding based on the fractional local chromatic number [15] has broadcast rate 4, the index coding based on the fractional partial clique clique covering [5] has broadcast rate $11/3$ and our proposed scheme (theorem 3) has broadcast rate $7/2$.

The proposed scheme of theorem 4 is also better than the corresponding recursive scheme proposed in [3, theorem 4]. For the graph in fig. 3 our scheme is strictly better than their scheme. The index coding broadcast rates for the graph are 3 and $7/2$ for our proposed scheme and the scheme in [3, theorem 4], respectively.

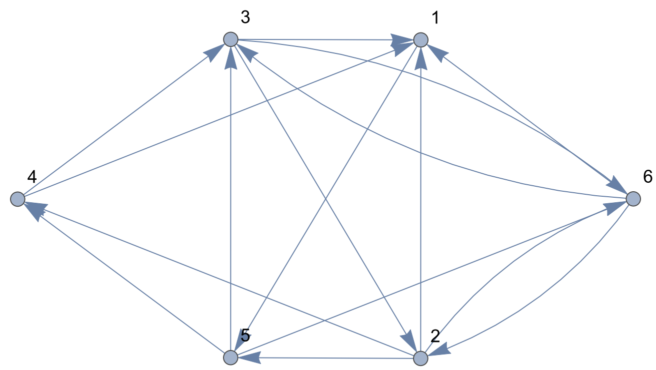

Fig. 3: Example of an Index Coding problem where the scheme in [3, theorem 4] are strictly larger than the proposed scheme for the fractional case.

Appendix

A. A Simple Achievability Scheme for [3, theorem 4]

Consider the following integer program proposed in [3],

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \textstyle\sum_{\mathcal{S}\in\mathcal{K}}(k_{\mathcal{S}}+1)\rho_{\mathcal{S}} \leq t, \\ & \qquad\qquad\qquad\qquad\qquad v \in V(\mathcal{G}) \\ \text{and} \quad & \sum_{\mathcal{S}\in\mathcal{K}:v\in\mathcal{S}} \rho_{\mathcal{S}} \geq 1, \quad v \in V(\mathcal{G}) \\ & \rho_{\mathcal{S}} \in [0,1], \quad \mathcal{S} \in \mathcal{K}. \end{aligned} \tag{5}$$

We propose a simple scheme corresponding to the integer program in eq. (5). Assume that the optima corresponding to the integer program is m, the partial cliques selected are $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_t$. Let $n_j := |\mathcal{S}_j|$ and $k_j := k_{\mathcal{S}_j}$. Let the cliques selected corresponding to the vertex v_i have indices $\{c_1^i, c_2^i, \ldots, c_{t_i}^i\} \subseteq [t]$.

We construct the following matrix G,

$$G = [\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \mathbf{u}_n] = [\Phi_1 G_1 \ \Phi_2 G_2 \ \ldots \Phi_t G_t]_{m\times n} \tag{6}$$

where $n = \sum_j n_j$ is the number of vertices in $\mathcal{G}$, $G_j, j \in [t]$ are $[n_j, k_j+1]$-MDS matrices, and Φ_j are $m \times (k_j+1)$ matrices defined as follows.

Let Φ denote a set of m linearly independent vectors of dimension m. We construct the matrix Φ_j such that its column vectors are vectors in Φ and the column vectors for the matrices $\Phi_{c_1^i}, \ldots, \Phi_{c_{t_i}^i}$ are full rank for all i. Note that such an assignment is possible because $m = \max_i \sum_{j=1}^{t_i}(k_{c_j^i}+1)$.

Note that, in this case, the upper bound on the alphabet size is $\max_{j\in[t]} n_j$ corresponding to the largest alphabet size needed for constructing G_j.

References

[1] Abhishek Agarwal and Arya Mazumdar. Local partial clique covers for index coding. *arXiv preprint arXiv:1603.02366*, 2016.

[2] Noga Alon, Eyal Lubetzky, Uri Stav, Amit Weinstein, and Avinatan Hassidim. Broadcasting with side information. In *Foundations of Computer Science, 2008 (FOCS'08), 49th Annual Symposium on*, pages 823–832. IEEE, 2008.

[3] Fatemeh Arbabjolfaei and Young-Han Kim. Local time sharing for index coding. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 286–290. IEEE, 2014.

[4] Ziv Bar-Yossef, Yitzhak Birk, TS Jayram, and Tomer Kol. Index coding with side information. *Information Theory, IEEE Transactions on*, 57(3):1479–1494, 2011. Preliminary version in FOCS 2006.

[5] Yitzhak Birk and Tomer Kol. Informed-source coding-on-demand (iscod) over broadcast channels. In *INFOCOM'98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1257–1264. IEEE, 1998.

[6] Anna Blasiak, Robert Kleinberg, and Eyal Lubetzky. Lexicographic products and the power of non-linear network coding. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 609–618. IEEE, 2011.

[7] Michelle Effros, Salim El Rouayheb, and Michael Langberg. An equivalence between network coding and index coding. *Information Theory, IEEE Transactions on*, 61(5):2478–2487, 2015.

[8] Salim El Rouayheb, Alex Sprintson, and Costas Georghiades. On the index coding problem and its relation to network coding and matroid theory. *Information Theory, IEEE Transactions on*, 56(7):3187–3195, 2010.

[9] Paul Erdös, Zoltán Füredi, András Hajnal, Péter Komjáth, Vojtech Rödl, and Ákos Seress. Coloring graphs with locally few colors. *Discrete mathematics*, 59(1):21–34, 1986.

[10] W. Haemers. An upper bound on the shannon capacity of a graph. *Algebraic methods in Graph Theory*, 25:267–272, 1978.

[11] Young-Han Kim. All index coding problems up to n=5 messages. http://circuit.ucsd.edu/~yhk/indexcoding.html. Accessed: 2016-02-17.

[12] Michael Langberg and Alex Sprintson. On the hardness of approximating the network coding capacity. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 315–319. IEEE, 2008.

[13] Eyal Lubetzky and Uri Stav. Nonlinear index coding outperforming the linear optimum. *Information Theory, IEEE Transactions on*, 55(8):3544–3551, 2009.

[14] Arya Mazumdar. On a duality between recoverable distributed storage and index coding. In *Information Theory, International Symposium on*, pages 1977–1981. IEEE, 2014.

[15] Karthikeyan Shanmugam, Alexandros G Dimakis, and Michael Langberg. Local graph coloring and index coding. In *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pages 1152–1156. IEEE, 2013.

[16] Karthikeyan Shanmugam, Alexandros G Dimakis, and Michael Langberg. Graph theory versus minimum rank for index coding. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 291–295. IEEE, 2014.

[17] Chandra Thapa, Lawrence Ong, and Sarah J Johnson. Generalized interlinked cycle cover for index coding. *arXiv preprint arXiv:1504.04806*, 2015.