Locally Decodable Index Codes

Lakshmi Prasad Natarajan, *Member, IEEE*, Prasad Krishnan,
V. Lalitha, *Member, IEEE*, and Hoang Dau

Abstract—An index code for broadcast channel with receiver side information is *locally decodable* if each receiver can decode its demand by observing only a subset of the transmitted codeword symbols instead of the entire codeword. Local decodability in index coding is known to reduce receiver complexity, improve user privacy and decrease decoding error probability in wireless fading channels. Conventional index coding solutions assume that the receivers observe the entire codeword, and as a result, for these codes the number of codeword symbols queried by a user per decoded message symbol, which we refer to as *locality*, could be large. In this paper, we pose the index coding problem as that of minimizing the broadcast rate for a given value of locality (or vice versa) and designing codes that achieve the optimal trade-off between locality and rate. We identify the optimal broadcast rate corresponding to the minimum possible value of locality for all single unicast problems. We present new structural properties of index codes which allow us to characterize the optimal trade-off achieved by: vector linear codes when the side information graph is a directed cycle; and scalar linear codes when the minrank of the side information graph is one less than the order of the problem. We also identify the optimal trade-off among all codes, including non-linear codes, when the side information graph is a directed 3-cycle. Finally, we present techniques to design locally decodable index codes for arbitrary single unicast problems and arbitrary values of locality.

Index Terms—Broadcast rate, directed cycle, index coding, linear codes, locally decodable codes, minrank.

I. INTRODUCTION

THE fundamental communication problem in broadcast channels is to design an efficient coding scheme to satisfy the demands of multiple clients with minimal use of the shared communication medium. When the clients have *side-information* about the messages demanded by other users in

the network, *index coding* [1], [2] can achieve remarkable savings in channel use by broadcasting a coded version of the information symbols. All the receivers then simultaneously decode their demands using the broadcast codeword and their individual side-information. The objective of index coding is to minimize the number of uses of the broadcast channel, or equivalently, the *broadcast rate*. Index coding is a central problem in multi-user communication networks, not only because of its wide ranging applications, such as video-on-demand and daily newspaper delivery [1], but also because of its strong relation to other coding theoretic problems, such as network coding [3], [4], coded caching [5], codes for distributed data storage [6], [7] and distributed computation [8].

Of the several classes of index coding problems discussed in the literature since [1], the most widely studied is *single unicast index coding*, in which each message available at the server is demanded by a unique client. Several approaches have been taken, most popularly via graph theoretic ideas, to bound the optimal index coding rate from above and below; see, for example, [1], [2], [9]–[18]. The techniques used in these works naturally lead to converse results and constructions of index codes, most of the constructions being those of (scalar and vector) linear index codes.

Conventional index coding solutions in the literature assume that each receiver observes the entire transmitted codeword. If the network involves a large number of receivers, the length of the index code could also be large, and thus, the number of transmissions that each receiver has to observe could be significantly larger than the size of its demanded message. Thus index coding solutions which are optimal in terms of broadcast rate could be unfavorable in certain applications, such as when the power available at the wireless receivers is limited and the receivers can not afford to listen to radio transmissions for an extended period of time. In such scenarios it is desirable to use an index code that is *locally decodable* [19], i.e., an index code that requires each receiver to observe or query only a part of the transmitted codeword. We define the *locality* of an index code to be the ratio of the number of codeword symbols observed by a receiver to the number of message symbols it demands. The objective of designing locally decodable index codes is to minimize the broadcast rate and the locality simultaneously, and achieve the optimal trade-off between these two parameters.

Designing a locally decodable index code is a 'client aware' approach to the broadcast problem that takes into account the cost incurred by the clients or the receivers while participating in the communication protocol. If the index coded symbols are broadcast across different time or frequency resource blocks,

TABLE I

SUMMARY OF THE CONTENTS AND THE ORGANIZATION OF THIS PAPER. HERE β DENOTES THE INDEX CODING BROADCAST RATE, r IS THE LOCALITY AND N IS THE NUMBER OF MESSAGES OR RECEIVERS IN THE INDEX CODING PROBLEM

SECTION	SUMMARY OF CONTENTS	CLASS OF PROBLEMS CONSIDERED	MAIN RESULTS
Section II	Definitions and preliminaries	All	
Section III	Optimal β for $r = 1$ among all index codes	All	Theorem 1
Section IV-A	Properties of locally decodable vector linear index codes	All	Theorem 3
Section IV-B	Trade-off between β and r among vector linear index codes	Directed cycles	Theorem 4
Section V	Trade-off between β and r among all index codes	Directed 3-cycle	Theorem 5
Section VI-A	Properties of locally decodable scalar linear index codes	All	Corollary 2
Section VI-B	Trade-off between β and r among scalar linear codes	minrank $= N - 1$	Theorem 6
Section VI-C	Feasible receiver localities among scalar linear codes	Directed cycles	Theorem 7
Section VII	Constructions of locally decodable index codes	All	

a client might want to minimize the number of blocks it must listen to in order to decode its desired index coded message, so as to reduce its power consumption or utilize the remaining blocks to participate in other communication sessions. Locally decodable index codes allow us to reduce the number of coded packets that must be observed by each client. In contrast, conventional index coding (without locality considerations) is more 'channel centric' with its emphasis purely on minimizing the number of channel uses.

To the best of our knowledge, the idea of local decodability in index coding was introduced in [19] where the broadcast rates of random index coding problems, modeled as random graphs, were analyzed under a locality requirement. Locally decodable index codes were studied under the terminology 'k-limited access schemes' in [20] and were shown to improve user privacy in index coding. The authors of [20] provide constructions that modify any given binary scalar linear index code into a locally decodable scalar linear code at the cost of increased broadcast rate. The technique of allowing each receiver to observe only a fraction of the index coded symbols was also used in [21] and [22], to reduce the error rate in wireless fading channels and to reduce the receiver complexity, respectively.

While the notion of locality has been known in the literature, this parameter has not been rigorously formalized before. Also, a fundamental treatment of locally decodable index codes is not available to the best of our knowledge. In this work, we present a formal structure to the discussion regarding locally decodable index codes and present constructions of such codes along with their locality parameters for single unicast index coding problems. We derive some structural properties of index coding solutions that relate locality to broadcast rate. We use these properties to determine the optimal trade-off between broadcast rate and locality for some families of index coding problems.

Contributions

We now describe the contributions and organization of this paper. Table I provides a summary of the contributions and

the main results of this paper. In order to help the readers identify these main results in this paper we have highlighted them using boxes.

We formalize the ideas of locality and locality-broadcast rate trade-off in Section II, and show that the minimum possible locality for any single unicast index coding problem is one. In Section III, we identify the optimal broadcast rate corresponding to unit locality for any single unicast problem (considering non-linear index codes as well). This fully characterizes the minimum locality point in the locality-rate trade-off curve. In Section IV, we first derive some locality-related properties of vector linear index codes. We then utilize these results to derive the optimal locality-rate trade-off achieved by linear codes for the class of index coding problems whose side information graphs are directed cycles. For the specific case of directed 3-cycle we identify the optimal locality-rate trade-off achieved by any index code, including non-linear codes, in Section V. In Section VI, we derive some structural properties of scalar linear locally decodable index codes. We then characterize the locality-rate trade-off achieved by scalar linear codes for the family of problems whose minrank is one less than the number of messages. Note that directed cycles considered in Section IV are a subset of the family of side information graphs considered in Section VI. In Section VI, we also characterize the set of all possible receiver localities for directed cycles that can be achieved using scalar linear index codes. Sections III, IV, V, and VI consider specific families of index coding problems or a specific value of locality, and identify the optimal solutions for these cases. In Section VII we present techniques to design locally decodable index codes (which are not necessarily optimal) for arbitrary single unicast problems and arbitrary values of locality. We also show how the traditional *partition multicast* [1], [12] and *cycle covering* [11], [14] solutions to index coding can be modified to yield locally decodable index codes. We conclude the paper in Section VIII and discuss the relation between locally decodable index codes and overcomplete dictionaries for sparse representation of vectors.

The optimal index coding schemes identified in Sections III, IV, V, and VI are based on known index coding techniques. However, one of the main contributions of this paper is the development of new tools to derive good lower bounds on rate and locality that are vital in proving the optimality of these schemes.

Notation: For any positive integer N, we will denote the set $\{1, \ldots, N\}$ by $[N]$. Matrices and column vectors are denoted by bold upper and lower case letters, respectively, such as $\boldsymbol{A}$ and $\boldsymbol{x}$. The symbol wt denotes the Hamming weight of a vector. The finite field of size q is denoted as $\mathbb{F}_q$. The subspace spanned by vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ is denoted by $\mathsf{span}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N)$. The column space of a matrix $\boldsymbol{A}$ is denoted as $\mathcal{C}(A)$ and the null space of $\boldsymbol{A}$ is $\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$. The support set of a vector $\boldsymbol{x}$ is denoted as $\mathsf{supp}(\boldsymbol{x})$. For a set of column vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ and a subset $K \subset [N]$, we define $\boldsymbol{x}_K$ to be the column vector obtained by concatenating $\boldsymbol{x}_j, j \in K$. The empty set is denoted by ϕ. For a matrix $\boldsymbol{A}$, the component in j^{th} row and i^{th} column is denoted as $\boldsymbol{A}_{j,i}$. The transpose of $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^{\mathsf{T}}$.

II. SYSTEM MODEL & PRELIMINARIES

We consider single unicast index coding for a broadcast channel consisting of N receivers $\mathsf{Rx}_1, \ldots, \mathsf{Rx}_N$. The transmitter holds N messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ where the ith message $\boldsymbol{x}_i$ is demanded by Rx_i, and the messages $\boldsymbol{x}_j, j \in K_i$, are known at this receiver as side information, where $K_i \subseteq [N] \setminus \{i\}$. The side information graph $G = (\mathcal{V}, \mathcal{E})$ is the directed graph with vertex set $\mathcal{V} = [N]$ and edge set $\mathcal{E} = \{(i, j) | i \in [N], j \in K_i\}$. The transmitter encodes the messages into a codeword $\boldsymbol{c}$ and broadcasts the codeword to all the receivers through a noiseless broadcast channel. Each receiver decodes its demands using $\boldsymbol{c}$ and its own side information. Throughout this paper we will consider only single unicast index coding problems and denote a problem by its side information graph G.

We assume that the messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are vectors of length M over a finite alphabet $\mathcal{A}$, with $|\mathcal{A}| > 1$, and the codeword $\boldsymbol{c}$ is a vector of length ℓ over the same alphabet, i.e., $\boldsymbol{x}_i \in \mathcal{A}^M$, $i \in [N]$, and $\boldsymbol{c} \in \mathcal{A}^\ell$. We will assume that the alphabet $\mathcal{A}$ is arbitrary but fixed for a given index coding problem. The *broadcast rate* $\beta = \ell/M$ of an index code is the ratio of the codeword length to message length, and measures the bandwidth or time required by the source to broadcast the coded symbols to all the receivers. Please note that our formulation includes non-linear index codes also.

A. Locally Decodable Index Codes

Unlike the conventional index coding problem where each receiver is required to observe or query the entire codeword $\boldsymbol{c}$, we allow the receivers to observe only a part of the codeword in order to decode their demands. Index codes that satisfy this property are called *locally decodable* [19]. We will assume that Rx_i queries the subvector $\boldsymbol{c}_{R_i} = (c_j, j \in R_i)$, where $R_i \subseteq [\ell]$ is chosen in such a way that Rx_i can decode $\boldsymbol{x}_i$ using $\boldsymbol{c}_{R_i}$ and the available side information $\boldsymbol{x}_j, j \in K_i$. The *locality of the receiver* of Rx_i is $r_i = \frac{|R_i|}{M}$ is the ratio of the number of codeword symbols observed by Rx_i to the number of information symbols demanded.

Definition 1: The *overall locality* or simply the *locality* of an index code is defined as

$$r = \max_{i \in [N]} r_i = \max_{i \in [N]} \frac{|R_i|}{M} \tag{1}$$

and is equal to the maximum number of coded symbols queried by any of the N receivers per decoded information symbol. The *average locality* of an index code is

$$r_{\text{avg}} = \sum_{i \in [N]} \frac{r_i}{N} = \sum_{i \in [N]} \frac{|R_i|}{MN}. \tag{2}$$

■

Observe that the average locality is upper bounded by overall locality $r_{\text{avg}} \le r$. Without loss of generality we will consider only those index codes for which $R_1 \cup \cdots \cup R_N = [\ell]$ since the subset $\{c_j \mid j \in [\ell] \setminus (R_1 \cup \cdots \cup R_N)\}$ of codeword symbols need not be generated or transmitted by the encoder. For a given locality r, it is desirable to use an index code with as small a value of β as possible and vice versa, which leads us to the following definition.

Definition 2: Given an index coding problem G, the *optimal broadcast rate function* $\beta_G^*(r)$ is the infimum of the broadcast rates among all message lengths $M \ge 1$ and all valid index codes for G with locality at the most r. ■

The function $\beta_G^*(r)$ captures the trade-off between broadcast rate and locality, i.e., between the reduction in the number of channel uses possible through coding and the number of codeword symbols that a receiver has to observe to decode each message symbol.

Remark 1: Our system model assumes that the codeword vector and the message vectors are defined over the same alphabet $\mathcal{A}$. In some scenarios the underlying alphabet, say $\mathcal{Z}$, for the codeword $\boldsymbol{c}$ and the alphabet $\mathcal{A}$ for the messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ can be different. In Appendix I we generalize the definition of locality to the case where $\mathcal{Z}$ can be different from $\mathcal{A}$, and show that the optimal broadcast rate function is independent of the choice of $\mathcal{Z}$ and $\mathcal{A}$ (see Lemmas 15 and 16 in Appendix I). Hence, without loss of generality, we will only consider the case $\mathcal{Z} = \mathcal{A}$ in the rest of this paper. Also, from Lemma 16 of Appendix I, the value of $\beta_G^*(r)$ is independent of the choice of $\mathcal{A}$. ■

Remark 2: The cost incurred by a receiver in participating in an index coded broadcast can also be measured as the fraction of codeword symbols queried by the receiver. The fraction of coded symbols observed by Rx_i is $\delta_i = |R_i|/\ell$, and the worst-case among all receivers is $\delta = \max_i \delta_i$. Note that $\delta_i = r_i/\beta$, and hence, $\delta = r/\beta$. The design problem posed in this paper is to construct index codes that minimize β for a given value of r, or equivalently, minimize r for a given value of β. Since $\delta = r/\beta$, we observe that for a given value of β, minimizing δ is the same problem as minimizing r. Thus both choices of performance parameters, r and δ, capture the same underlying index coding problem. ■

We now prove some elementary properties of the function $\beta_G^*(r)$. We will rely on information-theoretic inequalities for this purpose. We will assume that the messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$

are random, independent of each other and are uniformly distributed in $\mathcal{A}^M$. The logarithms used in measuring mutual information and entropy will be calculated to the base $|\mathcal{A}|$. For example, the entropy of $\boldsymbol{x}_i$ is $H(\boldsymbol{x}_i) = M$ since $\boldsymbol{x}_i$ is uniformly distributed in $\mathcal{A}^M$.

Lemma 1: The locality r and the average locality r_{avg} of any valid index coding scheme satisfy $r, r_{\text{avg}} \geq 1$.

Proof: Considering the decoder at Rx_i, we have $I(\boldsymbol{x}_i; \boldsymbol{c}_{R_i}, \boldsymbol{x}_{K_i}) = H(\boldsymbol{x}_i) = M$. Since $\boldsymbol{x}_i$ is independent of $\boldsymbol{x}_{K_i}$ (because $i \notin K_i$), $I(\boldsymbol{x}_i; \boldsymbol{x}_K) = 0$, and

$$M = I(\boldsymbol{x}_i; \boldsymbol{c}_{R_i}, \boldsymbol{x}_{K_i}) = I(\boldsymbol{x}_i; \boldsymbol{x}_{K_i}) + I(\boldsymbol{x}_i; \boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i})$$
$$= I(\boldsymbol{x}_i; \boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i}) \leq H(\boldsymbol{c}_{R_i}) \leq |R_i|. \tag{3}$$

Hence, $r_i = |R_i|/M \geq 1$ for all $i \in [N]$, and thus, $r, r_{\text{avg}} \geq 1$. ∎

Note that uncoded transmission, i.e., $\boldsymbol{c} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, is a valid index code with $r = 1$. Hence, we will assume that the domain of the function β_G^* is the interval $1 \leq r < \infty$.

Lemma 2: The function $\beta_G^*(r)$ is convex and non-increasing.

Proof: The non-increasing property of β_G^* follows immediately from its definition. We use time-sharing to prove convexity, see Appendix II for details. ∎

For any valid index code, we have $|R_i| \leq \ell$, and hence, $r = \max_i |R_i|/M \leq \ell/M = \beta$. Hence, if there exists a valid index code with broadcast rate β, then its locality is at the most β, and hence,

$$\beta_G^*(\beta) \leq \beta. \tag{4}$$

We will denote by $\beta_{\text{opt}}(G)$ the infimum among the broadcast rates of all valid index codes for G (considering all possible message lengths $M \geq 1$ and all possible localities $r \geq 1$). Then it follows that $\beta_G^*(r) \geq \beta_{\text{opt}}$ for all $r \geq 1$. Together with (4), choosing $\beta = \beta_{\text{opt}}$, we deduce

$$\beta_G^*(\beta_{\text{opt}}) = \beta_{\text{opt}}. \tag{5}$$

Example 1 (Locality-Broadcast Rate Trade-Off of the Directed 3-Cycle): Let G be the directed 3-cycle, i.e., $N = 3$ and the side information at the three receivers are $K_1 = \{2\}$, $K_2 = \{3\}$ and $K_3 = \{1\}$. We show in Section V that for this index coding problem

$$\beta_G^*(r) = \max\{6 - 3r, 2\} \quad \text{for all } r \geq 1.$$

This function is shown in Fig. 1. We observe that in order to achieve any savings in rate compared to the uncoded transmission ($\beta = N = 3$), we necessarily require the locality to be strictly greater than 1, i.e., each receiver must necessarily query more codeword symbols than the message length to achieve savings in the broadcast channel uses. Also, the smallest locality required to achieve the minimum rate $\beta_{\text{opt}} = 2$ is $r = 4/3$. ∎

B. Locally Decodable Linear Index Codes

Some of the results presented in this paper are applicable to the class of linear index codes only. Linear index codes form the most widely studied family of index coding solutions, and enjoy low complexity encoding and decoding operations.
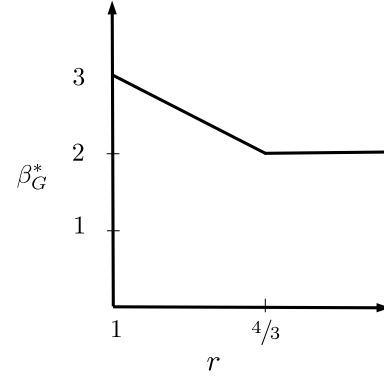

Fig. 1. The trade-off between the broadcast rate $\beta_G^*(r)$ and the locality r for the index coding problem represented by the 3-cycle G.

When considering (vector) linear index codes, we will assume that each message $\boldsymbol{x}_i$ is a vector of length M over some finite field $\mathbb{F}_q$, where q denotes the size of the field. The M components of the ith message vector $\boldsymbol{x}_i$ are denoted as $x_{i,1}, \ldots, x_{i,M}$. Encoding is performed by first concatenating the N messages into $\boldsymbol{x} = (\boldsymbol{x}_1^{\mathsf{T}}, \ldots, \boldsymbol{x}_N^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{F}_q^{MN}$ and multiplying this vector with a carefully designed encoding matrix $\boldsymbol{L} \in \mathbb{F}_q^{MN \times \ell}$ to generate a length ℓ the codeword $\boldsymbol{c}^{\mathsf{T}} = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}$. Note that the MN components of the concatenated vector $\boldsymbol{x} = (x_1, \ldots, x_{MN})^{\mathsf{T}}$ and the components of the individual message vectors are related as $x_{(i-1)M+m} = x_{i,m}$ for $i \in [N]$ and $m \in [M]$.

In the sequel we would like to view a vector linear index coding problem involving N vector messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{F}_q^M$ as a scalar linear problem defined over MN scalar messages $x_1, \ldots, x_{MN} \in \mathbb{F}_q$. In this case the ith receiver demands M scalar messages $x_{(i-1)M+1}, x_{(i-1)M+2}, \ldots, x_{iM}$, which correspond to the M components of the vector $\boldsymbol{x}_i$. This set of demands of Rx_i is represented by the index set $\mathcal{D}_i = \{(i-1)M + m \mid m \in [M]\}$. The scalar symbols available as side information at Rx_i correspond to the index set

$$\mathcal{K}_i = \left\{ (j-1)M + m \ \middle| \ j \in K_i, \, m \in [M] \right\}.$$

Thus the vector linear problem corresponding to the side information graph G with message length M is equivalent to a scalar linear problem with N receivers and MN messages, where the index set of demands of Rx_i is $\mathcal{D}_i$ and the set corresponding to the side information at Rx_i is $\mathcal{K}_i$. Note that $\mathcal{D}_i$ and $\mathcal{K}_i$ are subsets of $[MN]$.

We would like to characterize the trade-off between the broadcast rate β and the locality r of *linear* index codes over $\mathbb{F}_q$ for a given index coding problem G. We define the optimum locality-rate trade-off among all linear index codes for G over $\mathbb{F}_q$ as

$$\beta_{G,q}^*(r) = \inf \big\{ \beta \ \big| \ \exists \text{ a linear index code over } \mathbb{F}_q$$
$$\text{for } G \text{ with rate } \beta \text{ and locality} \leq r \big\},$$

where the infimum considers linear index codes over all possible message lengths $M \geq 1$ for the index coding problem G. Note the difference between $\beta_{G,q}^*(r)$ and $\beta_G^*(r)$; while the former is the optimal trade-off between rate and locality among

linear index codes (over $\mathbb{F}_q$), the latter considers all valid index codes (including linear and non-linear codes) for G.

Example 2 (A Simple Scalar Linear Code for Directed Cycles): Let the message length $M = 1$, $\mathbb{F}_q$ be any finite field, and G be the directed cycle of length N, i.e., $K_i = \{i + 1\}$ for $i \in [N - 1]$ and $K_N = \{1\}$. Consider the $N \times (N - 1)$ encoder matrix

$$\boldsymbol{L} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

that generates the codeword

$$\boldsymbol{c}^\mathsf{T} = (c_1, \ldots, c_{N-1}) = \boldsymbol{x}^\mathsf{T} \boldsymbol{L} = (x_1 + x_2, x_1 + x_3, \ldots, x_1 + x_N).$$

The receivers Rx_1 and Rx_N can decode their demands by querying c_1 and c_{N-1}, respectively, and hence, $|R_1| = |R_N| = 1$. For $1 < i < N$, receiver Rx_i queries the symbols $c_{i-1} = x_1 + x_i$ and $c_i = x_1 + x_{i+1}$, and uses its side information x_{i+1} to compute $c_{i-1} - c_i - x_{i+1}$, which equals its demand x_i. Hence, $|R_i| = 2$ for $1 < i < N$. Since $M = 1$, the locality of each receiver $r_i = 1$ if $i = 1$ or N, and $r_i = 2$ otherwise. The overall locality $r = 2$ and the average locality $r_{\mathrm{avg}} = 2(N - 1)/N$.

Since the graph G is symmetric, for any choice of $i \in [N - 1]$, the above coding scheme can be modified by an appropriate permutation of the rows of $\boldsymbol{L}$ to allow receiver localities $r_i = 1$ and $r_{i+1} = 1$ and locality $r_j = 2$ at all other receivers $j \neq i, i + 1$. ∎

III. MINIMUM-LOCALITY POINT IN THE LOCALITY-RATE TRADE-OFF

The minimum possible locality for any valid index coding solution is $r = 1$. Note that $r = 1$ immediately implies $r_{\mathrm{avg}} = 1$ as well. In this section we will show that the optimal index coding scheme with locality $r = 1$, considering non-linear schemes also, is the index code based on the fractional coloring of the *interference graph* corresponding to the problem. This explicitly identifies the value of the locality-rate trade-off function β_G^* at $r = 1$. In [19] it is remarked that the optimal broadcast rate with $r = 1$ among scalar linear index codes (i.e., $M = 1$, $\mathcal{A}$ is a finite field and the encoder is a linear transformation) is the chromatic number of the interference graph. Our results in this section generalize the result of [19] by considering all valid index codes, including non-linear codes. We first recall some graph-theoretic terminology related to index coding as used in [10], [15], state the main result of this section as Theorem 1 below and then prove the achievability and converse parts of this theorem in Sections III-A and III-B, respectively.

The *underlying undirected side information graph* $G_u = (\mathcal{V}, \mathcal{E}_u)$ of the side information graph $G = (\mathcal{V}, \mathcal{E})$ is the graph with vertex set $\mathcal{V} = [N]$ and an undirected edge set

$$\mathcal{E}_u = \{\,\{i, j\} \mid (i, j), (j, i) \in \mathcal{E}\,\},$$

that is, $\{i, j\} \in \mathcal{E}_u$ if and only if $i \in K_j$ and $j \in K_i$. The *interference graph* $\bar{G}_u = (\mathcal{V}, \bar{\mathcal{E}}_u)$ is the undirected complement of the graph G_u, i.e., $\bar{\mathcal{E}}_u = \{\,\{i, j\} \mid \{i, j\} \notin \mathcal{E}_u\}$. Note that

$$\{i, j\} \in \bar{\mathcal{E}}_u \text{ if and only if either } i \notin K_j \text{ or } j \notin K_i. \quad (6)$$

For positive integers a and b, an $a : b$ *coloring* of the undirected graph $\bar{G}_u = (\mathcal{V}, \bar{\mathcal{E}}_u)$ is a set $\{C_1, C_2, \ldots, C_N\}$ of N subsets $C_1, \ldots, C_N \subseteq [a]$, such that $|C_1| = \cdots = |C_N| = b$ and $C_i \cap C_j = \phi$ if $\{i, j\} \in \bar{\mathcal{E}}_u$. The elements of $[a]$ are *colors*, and each vertex of $\bar{G}_u$ is assigned b colors such that no two adjacent vertices have any colors in common. The *fractional chromatic number* χ_f of the undirected graph $\bar{G}_u$ is

$$\chi_f(\bar{G}_u) = \min\left\{ \frac{a}{b} \,\middle|\, \text{an } a : b \text{ coloring of } \bar{G}_u \text{ exists} \right\}.$$

The fractional chromatic number is a rational number and can be obtained as a solution to a linear program [23]. The *chromatic number* $\chi(\bar{G}_u)$ of the graph $\bar{G}_u$ is the smallest integer a such that an $a : 1$ coloring of $\bar{G}_u$ exists. In general, we have $\chi_f(\bar{G}_u) \leq \chi(\bar{G}_u)$. The main result of this section is

> *Theorem 1:* For any single unicast index coding problem G, the optimal broadcast rate for locality $r = 1$ is $\beta_G^*(1) = \chi_f(\bar{G}_u)$.

A. Proof of Theorem 1: Achievability Part

It is well known that there exists a coding scheme, which is known as the *fractional clique covering* or the *fractional coloring* solution, for any index coding problem G with broadcast rate $\beta = \chi_f(\bar{G}_u)$, for example, see [10]. It is straightforward to observe that $r = 1$ for this coding scheme. It then follows that $\beta_G^*(1) \leq \chi_f(\bar{G}_u)$. For completeness, we now recall the fractional coloring solution and deduce that $r = 1$ for this scheme.

Let $\chi_f(\bar{G}_u) = a/b$ for integers a and b, and let $C_1, \ldots, C_N \subseteq \{1, \ldots, a\}$ be an $a : b$ coloring of $\bar{G}_u$. Set the codeword length $\ell = a$ and message length $M = b$. Denote the components of the message vectors $\boldsymbol{x}_i \in \mathcal{A}^M$ using the variables $w_{i,t} \in \mathcal{A}$ as follows: $\boldsymbol{x}_i = (w_{i,t}, t \in C_i)$, i.e., one message symbol $w_{i,t}$ corresponding to each color t in the set C_i. Endow the set $\mathcal{A}$ with any abelian group structure $(\mathcal{A}, +)$. The symbols of the codeword $\boldsymbol{c} = (c_1, \ldots, c_\ell)$ are generated as $c_t = \sum_{i : t \in C_i} w_{i,t}$, for $t \in [\ell]$. Decoding at Rx_i is performed as follows. Note that $\boldsymbol{x}_i$ is composed of all symbols $w_{i,t}$ such that $t \in C_i$. In order to decode $w_{i,t}$, the receiver retrieves the code symbol c_t which is related to $w_{i,t}$ as

$$c_t = w_{i,t} + \sum_{\substack{j \neq i \\ j : t \in C_j}} w_{j,t}. \quad (7)$$

For any choice of the index j in the summation above, we have $i \neq j$ and $t \in C_i \cap C_j$. Since $C_1, \ldots, C_N$ is a coloring of $\bar{G}_u$ and $C_i \cap C_j \neq \phi$, we deduce that $\{i, j\} \notin \bar{\mathcal{E}}_u$, or equivalently, $i \in K_j$ and $j \in K_i$. Hence, for each $j \neq i$ such that $t \in C_j$, Rx_i knows the value of $w_{j,t}$, and thus, can recover $w_{i,t}$ from c_t using (7). Using a similar procedure Rx_i can decode all the M symbols in $\boldsymbol{x}_i$ from the M coded symbols $(c_t, t \in C_i)$.

This decoding method uses $R_i = C_i$, and hence, $r_i = |R_i|/M = 1$ for all $i \in [N]$, implying $r = 1$.

B. Proof of Theorem 1: Converse Part

The key result in the proof of the converse is Lemma 4, which states that $R_i \cap R_j = \phi$ for all i, j such that $\{i, j\} \in \bar{\mathcal{E}}_u$. Before providing the proof of the converse, we will give the intuition behind Lemma 4. The idea behind Lemma 4 is based on the analysis of scalar linear codes over $\mathbb{F}_2$ with locality 1; see also [19]. Note that each codeword symbol c_k of a binary scalar linear index code is a sum of a subset S_k of the scalars $x_1, \ldots, x_N \in \mathbb{F}_2$. Since the locality is 1, each receiver Rx_i observes exactly one codeword symbol, say $c_{f(i)}$. This implies that the set $S_{f(i)}$ must contain x_i, and any other symbol present in $S_{f(i)}$ must be available as side information at Rx_i. Now, if $j \notin K_i$, i.e., $\{i, j\} \in \bar{\mathcal{E}}_u$, then necessarily $x_j \notin S_{f(i)}$. On the contrary, the code symbol $c_{f(j)}$ observed by Rx_j will satisfy the property that $x_j \in S_{f(j)}$. We conclude that the code symbols $c_{f(i)}$ and $c_{f(j)}$ observed by Rx_i and Rx_j, respectively, are necessarily distinct. Thus $R_i \cap R_j = \{f(i)\} \cap \{f(j)\} = \phi$. Lemma 4 extends this result to all general index codes with locality 1.

Proof of the converse: Consider any valid index code, possibly non-linear, for G with locality $r = 1$, message length M and code length ℓ. We assume that $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ are independent and uniformly distributed over $\mathcal{A}^M$. Let $\boldsymbol{c}$ be the codeword generated by the index code. Since $r = 1$, from (1) and (3), we deduce that $|R_1| = \cdots = |R_N| = M$ for this valid index code.

Lemma 3: For any $i \in [N]$ and any $P \subset [N]$ such that $i \notin P$, we have (i) $I(\boldsymbol{c}_{R_i}; \boldsymbol{x}_i | \boldsymbol{x}_{K_i \cup P}) = M$; (ii) $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{K_i}) = 0$; and (iii) $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_P) = M$.

Proof: We first observe that $I(\boldsymbol{c}_{R_i}; \boldsymbol{x}_i | \boldsymbol{x}_{K_i \cup P}) = H(\boldsymbol{x}_i | \boldsymbol{x}_{K_i \cup P}) - H(\boldsymbol{x}_i | \boldsymbol{c}_{R_i}, \boldsymbol{x}_{K_i \cup P})$. Since $i \notin K_i \cup P$, $\boldsymbol{x}_i$ is independent of $\boldsymbol{x}_{K_i \cup P}$. Also, $\boldsymbol{x}_i$ can be decoded using $\boldsymbol{c}_{R_i}$ and $\boldsymbol{x}_{K_i}$. Hence, $H(\boldsymbol{x}_i | \boldsymbol{x}_{K_i \cup P}) = M$ and $H(\boldsymbol{x}_i | \boldsymbol{c}_{R_i}, \boldsymbol{x}_{K_i \cup P}) = 0$, thereby proving part *(i)*.

Using the result in part *(i)* and decomposing the mutual information term $I(\boldsymbol{c}_{R_i}; \boldsymbol{x}_i | \boldsymbol{x}_{K_i \cup P})$, we have

$$M = H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i \cup P}) - H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{K_i \cup P}). \qquad (8)$$

Since $\boldsymbol{c}_{R_i}$ is a length M vector, we have $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i \cup P}) \leq M$. Also, $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{K_i \cup P}) \geq 0$. Considering these facts together with (8), we deduce that

$$H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{K_i \cup P}) = 0 \text{ and } H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i \cup P}) = M. \quad (9)$$

Observe that (9) holds for any choice of P such that $i \notin P$. Choosing $P = \phi$ in the first equality in (9) proves part *(ii)* of this lemma. Now using the fact that $\boldsymbol{c}_{R_i}$ is of length M, and the second equality in (9), we have

$$M \geq H(\boldsymbol{c}_{R_i}) \geq H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_P) \geq H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{K_i \cup P}) = M.$$

This shows that $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_P) = M$, proving part *(iii)*. ∎

Lemma 4: For any $\{i, j\} \in \bar{\mathcal{E}}_u$, we have $R_i \cap R_j = \phi$.

Proof: Using (6), we will assume without loss of generality that $j \notin K_i$. We will now assume that $R_i \cap R_j \neq \phi$ and prove the lemma by contradiction. Let $t \in R_i \cap R_j$

and $P = \{i\} \cup K_i$. From part *(ii)* of Lemma 3, we have $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{K_i}) = 0$. In particular, since $t \in R_i$, we have

$$H(c_t | \boldsymbol{x}_i, \boldsymbol{x}_{K_i}) = H(c_t | \boldsymbol{x}_P) = 0. \qquad (10)$$

Note that $j \notin P$ since $j \neq i$ and $j \notin K_i$. From part *(iii)* of Lemma 3, we observe that $H(\boldsymbol{c}_{R_j} | \boldsymbol{x}_P) = M$. This implies that for any given realization of $\boldsymbol{x}_P$, the vector $\boldsymbol{c}_{R_j}$ is uniformly distributed over $\mathcal{A}^M$. Hence, the M coordinates of $\boldsymbol{c}_{R_j}$ are independent and uniformly distributed over $\mathcal{A}$. Since $t \in R_j$, we conclude that for any given realization of $\boldsymbol{x}_P$, c_t is uniformly distributed over $\mathcal{A}$, and hence, $H(c_t | \boldsymbol{x}_P) = 1$ which contradicts (10). ∎

Lemma 5: For any valid index coding scheme for G with $r = 1$, the broadcast rate $\beta \geq \chi_f(\bar{G}_u)$.

Proof: From Lemma 4, the subsets $R_1, \ldots, R_N \subseteq [\ell]$ are such that $R_i \cap R_j = \phi$ if $\{i, j\} \in \bar{\mathcal{E}}_u$ and $|R_i| = M$ for all $i \in [N]$. Hence, $\{R_1, \ldots, R_N\}$ is an $\ell : M$ coloring of $\bar{G}_u$. Consequently, the broadcast rate $\beta = \ell/M \geq \chi_f(\bar{G}_u)$. ∎

Combining the converse result in Lemma 5 with the achievability result in Section III-A, we arrive at Theorem 1. Theorem 1 can be easily generalized to the case where the message length M is fixed.

Corollary 1: The optimal broadcast rate for index coding problem G with locality $r = 1$ and message length M is

$$\min \left\{ \frac{a}{M} \,\middle|\, \text{an } a : M \text{ coloring of } \bar{G}_u \text{ exists} \right\}.$$

In particular, the optimum rate for $M = r = 1$ is $\chi(\bar{G}_u)$.

Proof: The achievability result is similar to the arguments used in Section III-A with the additional restriction that the subsets of colors $C_1, \ldots, C_N$ are all of size M. Converse follows by recognizing that the set of subsets $\{R_1, \ldots, R_N\}$ is an $\ell : M$ coloring of $\bar{G}_u$. When $M = 1$, the smallest ℓ for which an $\ell : 1$ coloring of $\bar{G}_u$ exists is the chromatic number of $\bar{G}_u$. ∎

IV. Vector Linear Codes for Directed Cycles

In this section, we will first derive some properties of locally decodable (vector) linear index codes (codes for any message length $M \geq 1$) in Section IV-A for arbitrary index coding problems. These structural properties of linear index codes will be useful in deriving the rate-locality trade-off of directed cycles presented in Section IV-B.

A. Structure of Locally Decodable Linear Index Codes

Following the notation from [24], for a vector $\boldsymbol{u} \in \mathbb{F}_q^{MN}$ and set $E \subset [MN]$, we write $\boldsymbol{u} \triangleleft E$ to denote $\mathrm{supp}(\boldsymbol{u}) \subseteq E$. Let us denote the columns of the encoder matrix $\boldsymbol{L}$ as $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_\ell \in \mathbb{F}_q^{MN}$. Then the kth symbol of the codeword is $c_k = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}_k$. Note that the ith receiver queries the subvector $\boldsymbol{c}_{R_i} = (c_k, k \in R_i)$, and utilizes the side information $\boldsymbol{x}_{\mathcal{K}_i} = (x_j, j \in \mathcal{K}_i)$, to decode the demand $\boldsymbol{x}_{\mathcal{D}_i} = (x_j, j \in \mathcal{D}_i)$.

We first consider a necessary and sufficient condition for a given encoder matrix $\boldsymbol{L}$ and receiver queries $R_1, \ldots, R_N$ to represent a valid index code for a single unicast problem G. We observe that the proofs of Lemmas 3.1 and 4.3 and Corollary 4.4 of [24] can be directly adapted to the scenario

of locally decodable index codes, immediately yielding the following constraints on the encoder matrix and receiver queries. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{MN}$ be the standard basis of $\mathbb{F}_q^{MN}$.

Theorem 2: For each $i \in [N]$, let Rx_i query the subvector $\boldsymbol{c}_{R_i}$ of the codeword $\boldsymbol{c}^{\mathsf{T}} = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}$ and have the side information $\boldsymbol{x}_{\mathcal{K}_i}$. Then Rx_i can decode its demand $\boldsymbol{x}_{\mathcal{D}_i}$ if and only if for each $j \in \mathcal{D}_i$ there exists a $\boldsymbol{u}_j \in \mathbb{F}_q^{MN}$ such that $\boldsymbol{u}_j \lhd \mathcal{K}_i$ and $\boldsymbol{u}_j + \boldsymbol{e}_j \in \mathrm{span}(\boldsymbol{L}_k, k \in R_i)$.

We will say that $\boldsymbol{L} \in \mathbb{F}_q^{MN \times \ell}$ is a valid encoder matrix corresponding to the queries $R_1, \ldots, R_N \subseteq [\ell]$ if it satisfies the criterion stated in Theorem 2 for decodability at all the receivers.

Observe that among the component symbols in the codeword $\boldsymbol{c} = (c_1, \ldots, c_\ell)^{\mathsf{T}}$, some are queried exactly once, i.e., queried by a single receiver, and the other symbols are queried by multiple receivers in the network. Let $\mathcal{S}_i = R_i \setminus (R_1 \cup \cdots \cup R_{i-1} \cup R_{i+1} \cdots \cup R_N)$ denote the index set of coded symbols that are queried only by Rx_i. Also, let $\mathcal{M}_i = R_i \cap (R_1 \cup \cdots \cup R_{i-1} \cup R_{i+1} \cdots \cup R_N)$ denote the index set of coded symbols that are queried by Rx_i and at least one other receiver. Note that $\mathcal{S}_i \cap \mathcal{M}_i = \phi$ and $\mathcal{S}_i \cup \mathcal{M}_i = R_i$ for each $i \in [N]$.

The following result shows that certain entries of the encoder matrix can be assumed to be equal to zero without affecting the locality or rate of the linear index code.

Theorem 3: Let $\boldsymbol{L} \in \mathbb{F}_q^{MN \times \ell}$ be a valid encoding matrix corresponding to the queries $R_1, \ldots, R_N$. Then there exists a valid encoding matrix $\boldsymbol{L}' \in \mathbb{F}_q^{MN \times \ell}$ for the queries $R_1, \ldots, R_N$ such that for each $i \in [N]$

$$\boldsymbol{L}'_k \lhd \mathcal{D}_i \text{ for all } k \in \mathcal{S}_i.$$

Proof: See Appendix III. ∎

The new index code guaranteed by Theorem 3 employs the same code length and the same set of queries as the given index code. Hence, the broadcast rate β, overall locality r and the average locality r_{avg} of the new code are identical to those of the given index code. Additionally, the new code guarantees that for any $i \in [N]$ any codeword symbol c_k, $k \in \mathcal{S}_i$, queried only by Rx_i, can be expressed as a linear combination of the demands of Rx_i. Since any valid encoder matrix can be modified to satisfy this property using Theorem 3, in the sequel, without loss of generality, we will only consider encoder matrices $\boldsymbol{L}$ that satisfy

$$\boldsymbol{L}_k \lhd \mathcal{D}_i \text{ for all } k \in \mathcal{S}_i \text{ and } i \in [N]. \tag{11}$$

Further, without loss of generality, we will assume that for each $i \in [N]$, the vectors $\boldsymbol{L}_k, k \in R_i$, are linearly independent. If this is not the case, then at least one of the codeword symbols c_k queried by Rx_i is a linear combination of the other queried symbols $\boldsymbol{c}_{R_i \setminus \{k\}}$. Reducing the index set of the queries of Rx_i from R_i to $R_i \setminus \{k\}$ does not affect decodability at Rx_i since c_k can be reconstructed from $\boldsymbol{c}_{R_i \setminus \{k\}}$. Note that this reduction in the queries does not increase the value of either the overall locality r or the average locality r_{avg} of

the index code. This process can be repeated till the columns of $\boldsymbol{L}$ corresponding to the queries of each of the receivers are linearly independent. Finally, any codeword symbol that is not queried by any of the receivers can be removed from the transmission since this symbol will not be used for decoding.

Let us denote the index set of codeword symbols that are queried exactly once by

$$\mathcal{S} = \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N. \tag{12}$$

Note that $\mathcal{S}_i \cap \mathcal{S}_j = \phi$ for any $i \neq j$. Hence, $|\mathcal{S}| = \sum_{i=1}^{N} |\mathcal{S}_i|$. Let

$$\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N \tag{13}$$

denote the index set corresponding to the codeword symbols that have been queried by more than one receiver. Observe that $\mathcal{S} \cap \mathcal{M} = \phi$ and $\mathcal{S} \cup \mathcal{M} = [\ell]$.

Lemma 6: For any valid index code for message length M, number of receivers N, rate β and average locality r_{avg},

$$|\mathcal{S}| \geq M(2\beta - N r_{\mathrm{avg}}),$$

where $|\mathcal{S}|$ is the number of codeword symbols that have been queried exactly once.

Proof: We will count the total number of queries made by all the receivers in two different ways and relate these expressions to arrive at the statement of this lemma. The number of queries made by Rx_i is $|R_i|$. Hence, the total number of queries made by all the receivers is $\sum_{i \in [N]} |R_i|$. From (2) this is equal to $MN r_{\mathrm{avg}}$. The number of times a codeword symbol c_k is queried is equal to 1 if $k \in \mathcal{S}$, and is at least 2 if $k \in \mathcal{M}$. Thus the total number of queries satisfies

$$MN r_{\mathrm{avg}} \geq \sum_{k \in \mathcal{S}} 1 + \sum_{k \in \mathcal{M}} 2 = |\mathcal{S}| + 2|\mathcal{M}|$$

$$= |\mathcal{S}| + 2(\ell - |\mathcal{S}|) = 2\ell - |\mathcal{S}|,$$

where we have used the fact $|\mathcal{S}| + |\mathcal{M}| = \ell$. Substituting $\ell = M\beta$ in the above inequality, we arrive at $MN r_{\mathrm{avg}} \geq 2M\beta - |\mathcal{S}|$ thereby proving the lemma. ∎

B. Optimal Linear Codes for Directed Cycles

We will now consider the index coding problem where G is a directed N-cycle, i.e., for $i = 1, \ldots, N-1$, $K_i = \{i+1\}$ and $K_N = \{1\}$. In this subsection we will provide the proof of the following result.

Theorem 4: For any $N \geq 3$, let G be a directed cycle of length N. For any finite field $\mathbb{F}_q$, the optimal trade-off between rate and locality among (vector) linear index codes for G over $\mathbb{F}_q$ is

$$\beta_{G,q}^*(r) = \max \left\{ \frac{N(N-1-r)}{N-2}, N-1 \right\}, \quad r \geq 1. \tag{14}$$

Note that locality $r \geq 1$ for any valid index coding scheme, and hence, $\beta_{G,q}^*(r)$ is defined for $r \geq 1$ only. The trade-off
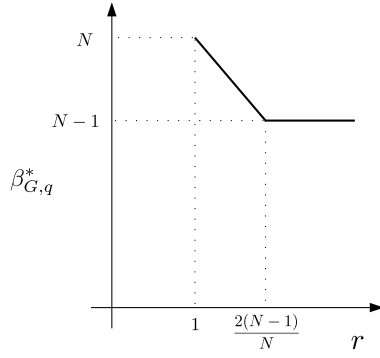
Fig. 2. The locality-rate trade-off of linear index codes for the directed N-cycle index coding problem.

between rate and locality is shown in Fig. 2. Sections IV-B1 and IV-B2 provide the proofs for the converse and achievability, respectively, of this rate-locality trade-off. It is known that for the directed N-cycle G, $N - 1$ is the minimum possible broadcast rate[1]. Our result shows that the smallest locality at which this rate is achievable using linear codes is $r = 2(N - 1)/N$. This locality is achievable if the message length M of the vector linear code is chosen carefully. In Section IV-B3, we provide a detailed analysis of the effect of the message length M on the locality r when the broadcast rate is $N - 1$.

1) Converse: In this subsection we will show that $\beta^*_{G,q}(r)$ is lower bounded by both $N-1$ and $N(N-1-r)/(N-2)$. It is clear that $\beta^*_{G,q}(r) \geq N - 1$, since even without any locality constraints the smallest possible broadcast rate for the directed N-cycle is $N - 1$. To complete the converse, we only need to show that $\beta^*_{G,q}(r) \geq N(N-1-r)/(N-2)$.

Suppose that the encoding matrix $\boldsymbol{L}$ is valid with respect to a set of queries $R_1, \ldots, R_N$. From Lemma 6, the number of codeword symbols queried exactly once $|\mathcal{S}| = \sum_{i \in [N]} |\mathcal{S}_i| \geq M(2\beta - Nr_{\mathrm{avg}})$. Hence there exists an $i \in [N]$ such that $|\mathcal{S}_i| \geq M(2\beta - Nr_{\mathrm{avg}})/N$. Without loss of generality, let us assume that

$$|\mathcal{S}_N| \geq M(2\beta - Nr_{\mathrm{avg}})/N. \tag{15}$$

We now relate this lower bound on $|\mathcal{S}_N|$ to the rank of $\boldsymbol{L}$ to complete the converse.

For $i = 1, \ldots, N - 1$, we have $\mathcal{K}_i = \{iM + 1, iM + 2, \ldots, (i+1)M\}$ and $\mathcal{D}_i = \{(i-1)M + 1, \ldots, iM\}$. From Theorem 2, for each $j \in \mathcal{D}_i$ there exists a $\boldsymbol{u}_j \lhd \mathcal{K}_i$ such that $\boldsymbol{e}_j + \boldsymbol{u}_j \in \mathcal{C}(\boldsymbol{L})$, where $\mathcal{C}(\boldsymbol{L})$ denotes the column span of the matrix $\boldsymbol{L}$. The non-zero entries of the $MN \times M$ matrix $[\boldsymbol{u}_j, \ j \in \mathcal{D}_i]$ are restricted to the M rows indexed by $\mathcal{K}_i$. Let $\boldsymbol{C}_i$ be the $M \times M$ submatrix of $[\boldsymbol{u}_j, \ j \in \mathcal{D}_i]$ corresponding to the rows $\mathcal{K}_i$. Considering the first $N - 1$ receivers $i = 1, \ldots, N - 1$ and each of their demands $j \in \mathcal{D}_i$, we obtain $M(N-1)$ vectors $\boldsymbol{e}_j + \boldsymbol{u}_j$, all which lie in $\mathcal{C}(\boldsymbol{L})$. Arranging

these column vectors into a matrix of size $MN \times M(N-1)$ we arrive at

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_1 & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 & \cdots & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{C}_{N-2} & \boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{C}_{N-1} \end{bmatrix}, \tag{16}$$

where each of the submatrices is of size $M \times M$. Note that the columns of this matrix are linearly independent.

Now considering Rx_N, we note that $\mathcal{D}_N = \{(N-1)M + 1, \ldots, MN\}$. The coded symbols $\boldsymbol{x}^\mathsf{T} \boldsymbol{L}_k$, $k \in \mathcal{S}_N$, are queried only by Rx_N. Using Theorem 3 and (11), we assume without loss of generality that $\boldsymbol{L}_k \lhd \mathcal{D}_N$ for all $k \in \mathcal{S}_N$, i.e., $\mathrm{supp}(\boldsymbol{L}_k) \subseteq \mathcal{D}_N$. Since the set of vectors $\{\boldsymbol{L}_k \,|\, k \in R_N\}$ is linearly independent and $\mathcal{S}_N \subseteq R_N$, we observe that the vectors $\boldsymbol{L}_k$, $k \in \mathcal{S}_N$, are linearly independent as well. Note that each of these vectors is a column of $\boldsymbol{L}$ and hence lies in $\mathcal{C}(\boldsymbol{L})$. Appending these $|\mathcal{S}_N|$ vectors as columns to the matrix in (16), we arrive at the block matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{C}_1 & \boldsymbol{I} & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}_2 & \cdots & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \vdots & \vdots & & \vdots & \vdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{C}_{N-2} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{C}_{N-1} & \boldsymbol{B}_N \end{bmatrix},$$

where $\boldsymbol{B}_N$ is an $M \times |\mathcal{S}_N|$ matrix with linearly independent columns. Note that each column of $\boldsymbol{A}$ lies in $\mathcal{C}(\boldsymbol{L})$, i.e., $\mathcal{C}(\boldsymbol{A}) \subseteq \mathcal{C}(\boldsymbol{L})$, and the columns of $\boldsymbol{A}$ are linearly independent, i.e., $\mathrm{rank}(\boldsymbol{A}) = M(N-1) + |\mathcal{S}_N|$. Thus,

$$\begin{aligned} \ell \geq \mathrm{rank}(\boldsymbol{L}) \geq \mathrm{rank}(\boldsymbol{A}) \qquad &(\text{since } \mathcal{C}(\boldsymbol{L}) \supseteq \mathcal{C}(\boldsymbol{A})) \\ = M(N-1) + |\mathcal{S}_N| & \\ \geq M(N-1) + M(2\beta - Nr_{\mathrm{avg}})/N \qquad &(\text{using (15)}) \end{aligned}$$

Since broadcast rate $\beta = \ell/M$, the above inequality yields $\beta \geq (N-1) + (2\beta - Nr_{\mathrm{avg}})/N$, which upon manipulation results in

$$\beta \geq N(N - 1 - r_{\mathrm{avg}})/(N - 2). \tag{17}$$

Since $r_{\mathrm{avg}} \leq r$, we arrive at $\beta \geq N(N - 1 - r)/(N - 2)$.

2) Achievability: In this subsection we show that the trade-off in (14) is achievable using linear index codes. We will show that the points $(r, \beta) = (1, N)$ and $(r, \beta) = (2(N - 1)/N, N - 1)$ are achievable. Then any point on the line segment $\beta = N(N-1-r)/(N-2), 1 \leq r \leq 2(N-1)/N$ can be achieved using time sharing between these two schemes. The achievability of the points $\beta = N-1$ and $r > 2(N-1)/N$ will follow immediately since the rate $N - 1$ is already achievable with $r = 2(N - 1)/N$.

[1] From [2, Theorem 3], we know that the broadcast rate is lower bounded by the size of the maximum acyclic induced subgraph of G, which is equal to $N - 1$ in this case. Further, Example 1 of [2] provides a scalar linear code that achieves rate $N - 1$ for directed N-cycles.

Achieving r = 1, β = N: The point $(r, \beta) = (1, N)$ can be achieved trivially using the uncoded scheme, i.e., the transmitted codeword equals the message vector $\boldsymbol{c} = \boldsymbol{x} \in \mathbb{F}_q^{MN}$. Each receiver Rx$_i$ queries $\boldsymbol{c}_{\mathcal{D}_i} = \boldsymbol{x}_{\mathcal{D}_i}$ to meet its demand.

Achieving r = 2(N−1)/N, β = N−1: Example 2 provides a family of N scalar linear codes for G, one for each choice of $i \in [N]$, with rate $\beta = N - 1$. The ith code provides localities $r_i = r_{i+1} = 1$ and $r_j = 2$ for all $j \neq i, i+1$, where we interpret $i + 1$ as 1 if $i = N$. Using $\boldsymbol{r} = (r_1, r_2, \ldots, r_N)$ to represent the tuple of receiver localities, we observe that rate $N - 1$ can be achieved using scalar linear codes for the following values of locality vector $\boldsymbol{r}$

$$\boldsymbol{r}_1 = (1, 1, 2, 2, \ldots, 2), \ \boldsymbol{r}_2 = (2, 1, 1, 2, \ldots, 2), \ldots,$$
$$\boldsymbol{r}_{N-1} = (2, 2, \ldots, 2, 1, 1), \ \boldsymbol{r}_N = (1, 2, \ldots, 2, 1). \quad (18)$$

If N is an odd integer, we time share the N scalar linear codes corresponding to $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N$. Observe that the overall scheme is a vector linear code for message length $M = N$, rate $N-1$ and locality $r = r_{\text{avg}} = 2(N-1)/N$. If N is an even integer, we time share $N/2$ scalar linear codes corresponding to $\boldsymbol{r}_1, \boldsymbol{r}_3, \ldots, \boldsymbol{r}_{N-1}$, that yields a vector linear code with $M = N/2$, rate $N - 1$ and $r = r_{\text{avg}} = 2(N - 1)/N$.

3) Dependence of Locality on Message Length: From the achievability scheme in Section IV-B2 we observe that for $\beta = N - 1$, the locality $r = r_{\text{avg}} = 2(N - 1)/N$ can be achieved using message length $M = N$ if N is odd, and $M = N/2$ if N is even. We will now show that the message length used by the proposed scheme is the minimum required to attain locality $2(N - 1)/N$ with rate $N - 1$.

Lemma 7: Let $N \geq 3$. The message length M of any index code with locality equal to $2(N - 1)/N$ for the directed N-cycle satisfies $M \geq N$ if N is odd, and $M \geq N/2$ if N is even.

Proof: Consider any valid coding scheme with locality $r = 2(N - 1)/N$. There exists an $i \in [N]$ such that

$$\frac{2(N - 1)}{N} = r = r_i = \frac{|R_i|}{M},$$

that is

$$M = \frac{|R_i| \, N}{2(N - 1)}.$$

If N is odd, N and $2(N - 1)$ have no common factors, and since M is an integer, we deduce that M must be a multiple of N, i.e., $M \geq N$. If N is even, using the fact $N/2$ and $N - 1$ have no common factors we arrive at $M \geq N/2$. ∎

From Lemma 7, it is clear that the minimum M required to attain $r = 2(N - 1)/N$ at rate $N - 1$ is $M = N$ if N is odd and $M = N/2$ if N is even. We will now derive the optimal locality when the message length is smaller than this quantity, i.e., $M < N$. We do so by analysing the two cases, $M < N/2$ and $N/2 \leq M < N$.

Locality When M < N/2: From (17) we deduce that for any vector linear scheme of rate $\beta = N - 1$, we have

$$r_{\text{avg}} \geq 2(N - 1)/N.$$

Thus, $\sum_{i=1}^{N} |R_i| = MNr_{\text{avg}} \geq 2M(N - 1)$. It follows that there exists an $i \in [N]$ such that

$$|R_i| \geq \frac{2M(N - 1)}{N} = 2M - \frac{M}{N/2}. \quad (19)$$

If $M < N/2$, considering the fact that $|R_i|$ is an integer, we deduce that $|R_i| \geq 2M$. Hence, $r_i = |R_i|/M \geq 2$, and thus, $r \geq 2$. This lower bound on r can be achieved by simply using the scalar linear code of Example 2 M times, leading to a vector linear code for message length M, rate $N - 1$ and $r = 2$. Note that this code still achieves the optimal value of average locality $r_{\text{avg}} = 2(N - 1)/N$.

Locality When N/2 ≤ M < N: If N is even, the message length $M = N/2$ is sufficient to attain $r = 2(N - 1)/N$. Thus it is enough to consider larger values of M, i.e., $N/2 \leq M < N$ only for N odd. From (19) and using the fact that $|R_i|$ is an integer, we arrive at $|R_i| \geq 2M - 1$. Thus,

$$r \geq r_i \geq 2 - \frac{1}{M}.$$

Assuming N is odd, this lower bound on r is achieved by time sharing the M scalar linear codes from Section IV-B2 corresponding to the tuples of localities

$$\boldsymbol{r}_1, \boldsymbol{r}_3, \ldots, \boldsymbol{r}_{N-3}, \boldsymbol{r}_N, \boldsymbol{r}_2, \boldsymbol{r}_4, \ldots, \boldsymbol{r}_{2M-(N+1)},$$

see (18). It is straightforward to show that this scheme has rate $N - 1$, $r = 2 - 1/M$ and $r_{\text{avg}} = 2(N - 1)/N$. Note that in the interval $N/2 \leq M < N$, the value of the optimal locality increases with M. Hence, the choice $M = (N+1)/2$ yields the smallest locality in this interval.

Example 3: Consider the problem of designing a vector linear index code for the directed 5-cycle, i.e., $N = 5$, with message length $M = 3$. Note that N is odd and $N/2 \leq M$. From Example 2 and (18), we know that there exist three scalar linear encoders, each of rate 4, for this index coding problem for which the tuple of receiver localities are

$$\boldsymbol{r}_1 = (1, 1, 2, 2, 2), \ \boldsymbol{r}_3 = (2, 2, 1, 1, 2) \text{ and } \boldsymbol{r}_5 = (1, 2, 2, 2, 1).$$

To arrive at a vector linear scheme of message length $M = 3$, we consider three generations of scalar messages and encode them using the above three scalar linear index coding schemes, respectively. The overall rate of this time-sharing scheme is $\beta = 4$. The total number of codeword symbols observed by each receiver for this vector linear code is $|R_1| = 4$, $|R_2| = 5$, $|R_3| = 5$, $|R_4| = 5$ and $|R_5| = 5$. Thus, this scheme has $r = 5/3$ and $r_{\text{avg}} = 8/5$, which are equal to $2 - 1/M$ and $2(N - 1)/N$, respectively. ∎

V. LOCALITY-BROADCAST RATE TRADE-OFF OF DIRECTED 3-CYCLE

Let G be the directed 3-cycle i.e., $N = 3$ and $K_1 = \{2\}$, $K_2 = \{3\}$ and $K_3 = \{1\}$. In this section we identify the optimal locality-rate trade-off function $\beta_G^*(r)$. The converse presented in this section applies to any valid index code (including non-linear codes), while the achievability proof uses the vector linear scheme of Section IV-B. Note that for this index coding problem the trade-off curves $\beta_G^*(r)$ and $\beta_{G,q}^*(r)$ are identical. The main result of this section is

Theorem 5: For the directed 3-cycle G, the optimal locality-broadcast rate trade-off function (considering non-linear codes also) is

$$\beta_G^*(r) = \max\{6 - 3r, 2\} \text{ for all } r \geq 1.$$

Using $N = 3$ in (14), we have $\beta_{G,q}^*(r) = \max\{6 - 3r, 2\}$ for $r \geq 1$. Achievability follows from observing that $\beta_G^*(r) \leq \beta_{G,q}^*(r)$, where the right hand side of the inequality considers only linear index codes.

In the rest of this section we prove the converse for Theorem 5. Towards proving the converse, we first present a result that exploits the symmetry in the side information graph. Let G be any single unicast index coding problem involving N messages such that the cyclic permutation σ on $[N]$ that maps $i \in [N]$ to $\sigma(i) = (i \mod N) + 1$ is an automorphism of G. Since we are considering non-linear codes too, an index code for G is represented by the encoding function $\mathfrak{E}$ and the decoding functions $\mathfrak{D}_1, \ldots, \mathfrak{D}_N$ defined over a finite alphabet $\mathcal{A}$. The encoder is a function $\mathfrak{E} : \mathcal{A}^{MN} \to \mathcal{A}^\ell$ that maps the message vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathcal{A}^M$ to the codeword $\boldsymbol{c} \in \mathcal{A}^\ell$. The decoder used at Rx_i is a function $\mathfrak{D}_i : \mathcal{A}^{|R_i|} \times \mathcal{A}^{M|K_i|} \to \mathcal{A}^M$ which maps the vectors $\boldsymbol{c}_{R_i}$ and $\boldsymbol{x}_j$, $j \in K_i$, to $\boldsymbol{x}_i$. We represent an index code by the tuple $(\mathfrak{E}, \mathfrak{D}_1, \ldots, \mathfrak{D}_N)$ of encoding and decoding functions.

Lemma 8: Let the cyclic permutation σ be an automorphism of G, and $(\mathfrak{E}, \mathfrak{D}_1, \ldots, \mathfrak{D}_N)$ be a valid index code for G with rate β and locality r. Then there exists a valid index code $(\mathfrak{E}', \mathfrak{D}_1', \ldots, \mathfrak{D}_N')$ for G with rate β and locality at the most r such that the index sets of codeword symbols observed by the N receivers $R_1', \ldots, R_N'$ for this code satisfy:

(i) $|R_1'| = |R_2'| \cdots = |R_N'|$; and
(ii) $|R_1' \cap R_2'| = |R_2' \cap R_3'| = \cdots = |R_N' \cap R_1'|$.

Proof: See Appendix V. ∎

Lemma 8 shows that, without affecting the broadcast rate and the locality, we can assume that the receiver queries in any valid index code for G satisfy the two symmetry properties listed above. In the rest of this section we will assume that G is a directed 3-cycle, and $(\mathfrak{E}, \mathfrak{D}_1, \mathfrak{D}_2, \mathfrak{D}_3)$ is a valid index code for G such that $|R_1| = |R_2| = |R_3|$ and $|R_1 \cap R_2| = |R_2 \cap R_3| = |R_3 \cap R_1|$.

For the sake of brevity, we abuse the notation mildly by using $i + 1$ to denote the receiver index $(i \mod 3) + 1$, and similarly we use $i + 2$ to denote $((i+1) \mod 3) + 1$. With this notation, for $i = 1, 2, 3$, the side information index set of the ith receiver is $K_i = \{i + 1\}$. Assume, as usual, that the messages $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are independently and uniformly distributed in $\mathcal{A}^M$.

Now considering the decoding operation at Rx_i, we have $I(\boldsymbol{x}_i; \boldsymbol{c}_{R_i} | \boldsymbol{x}_{i+1}) = H(\boldsymbol{x}_i) = M$. Expanding this term as a difference of conditional entropies, we have $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{i+1}) - H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) = M$. Using this with the upper bound $H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_{i+1}) \leq H(\boldsymbol{c}_{R_i}) \leq |R_i|$, we arrive at

$$H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \leq |R_i| - M.$$

Using the above inequality and the fact that $\boldsymbol{c}_{R_i}$ is a deterministic function of all three messages $\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \boldsymbol{x}_{i+2}$, we have

$I(\boldsymbol{c}_{R_i}; \boldsymbol{x}_{i+2} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) = H(\boldsymbol{c}_{R_i} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \leq |R_i| - M$. Hence,

$$H(\boldsymbol{x}_{i+2} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) - H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i}, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \leq |R_i| - M.$$

Since $H(\boldsymbol{x}_{i+2} | \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) = M$, we obtain the lower bound

$$H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i}, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \geq 2M - |R_i|. \tag{20}$$

Our objective now is to use the above inequality to obtain an upper bound on $|R_i \cap R_{i+2}|$, which can then be translated into a lower bound on ℓ, and hence, a lower bound on β. To do so, observe that $\boldsymbol{c}_{R_i}$ is composed of the following two sub-vectors $\boldsymbol{c}_{R_i \cap R_{i+2}}$ and $\boldsymbol{c}_{R_i \setminus R_{i+2}}$. Using (20), we obtain

$$H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i \cap R_{i+2}}, \boldsymbol{x}_i) \geq H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i}, \boldsymbol{x}_i, \boldsymbol{x}_{i+1}) \geq 2M - |R_i|.$$

Using this inequality, and the relation $H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_{i+2}}, \boldsymbol{x}_i) = 0$ (to satisfy the demands of Rx_{i+2}), we obtain the following

$$
\begin{aligned}
|R_{i+2} \setminus R_i| &\geq H(\boldsymbol{c}_{R_{i+2} \setminus R_i}) \geq I(\boldsymbol{x}_{i+2}; \boldsymbol{c}_{R_{i+2} \setminus R_i} | \boldsymbol{c}_{R_i \cap R_{i+2}}, \boldsymbol{x}_i) \\
&= H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i \cap R_{i+2}}, \boldsymbol{x}_i) \\
&\quad - H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_{i+2} \setminus R_i}, \boldsymbol{c}_{R_i \cap R_{i+2}}, \boldsymbol{x}_i) \\
&= H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_i \cap R_{i+2}}, \boldsymbol{x}_i) - H(\boldsymbol{x}_{i+2} | \boldsymbol{c}_{R_{i+2}}, \boldsymbol{x}_i) \\
&\geq 2M - |R_i|.
\end{aligned}
$$

Since $|R_1| = |R_2| = |R_3|$, we now have

$$
\begin{aligned}
|R_i \cap R_{i+2}| &= |R_{i+2}| - |R_{i+2} \setminus R_i| \\
&\leq |R_i| - (2M - |R_i|) \\
&= 2(|R_i| - M).
\end{aligned}
$$

Finally, since $|R_i \cap R_{i+2}|$ is independent of i, we have

$$
\begin{aligned}
\ell = |R_1 \cup R_2 \cup R_3| &\geq \sum_{j=1}^{3} |R_j| - \sum_{j=1}^{3} |R_j \cap R_{j+2}| \\
&= 3|R_i| - 3|R_i \cap R_{i+2}| \\
&\geq 3|R_i| - 3 \times 2(|R_i| - M) \\
&= 6M - 3|R_i|.
\end{aligned}
$$

Dividing both sides by the message length M, and remembering that all the receivers have the same locality $r = r_1 = r_2 = r_3$, we have $\beta = \ell/M \geq 6 - 3r$. Thus we have

$$\beta_G^*(r) \geq 6 - 3r \text{ for all } r \geq 1.$$

Further, the minimum possible broadcast rate $\beta_{\mathsf{opt}}(G) = 2$, and hence, $\beta_G^*(r) \geq 2$ for all $r \geq 1$. Combining this with the above inequality we have arrived at the converse $\beta_G^*(r) \geq \max\{6 - 3r, 2\}$ for all $r \geq 1$.

Remark 3: The main idea behind the derivation of $\beta_G^*(r)$ for the directed 3-cycle in this section is to lower bound the codelength using the identity $\ell = |R_1 \cup R_2 \cup R_3| \geq \sum_{i=1}^{3} |R_i| - |R_1 \cap R_2| - |R_2 \cap R_3| - |R_3 \cap R_1|$. The symmetry of the side information graph implies, through Lemma 8, that it is enough to find or bound the values of $|R_1|$ and $|R_1 \cap R_2|$. The other terms in this inequality are equal to either $|R_1|$ or $|R_1 \cap R_2|$. In contrast, for directed N-cycles with $N > 3$, we have

$$\ell = \left| \cup_{i=1}^{N} R_i \right| \geq \sum_{i=1}^{N} |R_i| - \sum_{i \neq j} |R_i \cap R_j|. \tag{21}$$

The symmetry of the side information graph and the results in Lemma 8 seem to be only partially helpful here. Although it is true that $|R_1| = \cdots = |R_N|$ (Lemma 8, property *(i)*), there is no guarantee that the terms $|R_i \cap R_j|$, $i \neq j$, are all equal. It is possible that $|R_1 \cap R_2| \neq |R_1 \cap R_3|$ when $N > 3$. In general, (21) contains a large number of terms that need to be individually analyzed and bounded. ∎

VI. SCALAR LINEAR CODES FOR LARGE MINRANK

In this section we analyze scalar linear codes that are locally decodable. If the minrank of the side-information graph G is N, then G is a directed acyclic graph and uncoded transmission is an optimal scalar linear index code. For this scheme, the locality of each receiver $r_i = 1$. Thus, we observe that when $\mathrm{minrk}(G) = N$, we can achieve the optimal broadcast rate $\beta = N$ and the minimum possible locality $r = r_{\mathrm{avg}} = 1$ simultaneously. Thus, the problem of designing locally decodable index codes is interesting only when $\mathrm{minrk}(G) \leq N - 1$.

In Section VI-A we identify some structural properties of scalar linear codes for arbitrary index coding problems G. These results are then specialized in Section VI-B to the family of index coding problems for which the number of messages is one greater than minrank, and the optimal trade-off between broadcast rate and locality is derived for this family of problems. Finally, in Section VI-C we consider directed N cycles (minrank is $N - 1$ for these problems) and characterize the set of all possible vectors of receiver localities $\boldsymbol{r} = (r_1, \ldots, r_N)$ that can be achieved through scalar linear coding when the broadcast rate is $N - 1$.

A. Preliminaries: Locally Decodable Scalar Linear Codes

We will now derive a few properties of scalar linear index codes that will be useful in relating broadcast rate to locality. Note that, in this case message length $M = 1$, demand set $\mathcal{D}_i = \{i\}$ and $\mathcal{K}_i = K_i$ for all $i \in [N]$. The locality of each receiver $r_i = |R_i|/M = |R_i|$ is an integer, and so is the overall locality $r = \max_i r_i$.

We say that a matrix $\boldsymbol{A} \in \mathbb{F}_q^{N \times N}$ *fits* $G = (\mathcal{V}, \mathcal{E})$ if the diagonal elements of $\boldsymbol{A}$ are all equal to 1, and the $(j, i)^{\mathrm{th}}$ entry of $\boldsymbol{A}$ is zero if $j \notin K_i$, i.e., $\boldsymbol{A}_{j,i} = 0$ if $(i, j) \notin \mathcal{E}$. The *minrank* of G over $\mathbb{F}_q$ is the minimum among the ranks of all possible matrices $\boldsymbol{A} \in \mathbb{F}_q^{N \times N}$ that fit G, and is denoted as $\mathrm{minrk}_q(G)$. It is known that the smallest possible scalar linear index coding rate is equal to $\mathrm{minrk}_q(G)$ [2], [24]. We also know from [2], [24] that a matrix $\boldsymbol{L}$ is a valid encoder matrix for G if and only if for each receiver $i \in [N]$, there exists a vector $\boldsymbol{u}_i \in \mathbb{F}_q^N$ such that $\boldsymbol{u}_i \lhd K_i$ and $\boldsymbol{u}_i + \boldsymbol{e}_i \in \mathcal{C}(\boldsymbol{L})$, where $\mathcal{C}$ denotes the column span of a matrix. If $\boldsymbol{L}$ is a valid encoder matrix, stacking these vectors we obtain the $N \times N$ matrix

$$\boldsymbol{A} = [\boldsymbol{u}_1 + \boldsymbol{e}_1 \quad \boldsymbol{u}_2 + \boldsymbol{e}_2 \quad \cdots \quad \boldsymbol{u}_N + \boldsymbol{e}_N].$$

Notice that $\boldsymbol{A}$ fits G and $\mathcal{C}(\boldsymbol{A}) \subseteq \mathcal{C}(\boldsymbol{L})$. We will say that $\boldsymbol{A}$ is a *fitting matrix corresponding to the encoder matrix* $\boldsymbol{L}$.

Suppose $\boldsymbol{L} \in \mathbb{F}_q^{N \times \ell}$ is a valid scalar linear encoder and the queries of the N receivers are $R_1, \ldots, R_N \subseteq [\ell]$. From Theorem 2, for each $i \in [N]$, there exists a vector $\boldsymbol{u}_i \lhd K_i$ such

that $\boldsymbol{u}_i + \boldsymbol{e}_i \in \mathrm{span}(\boldsymbol{L}_k, k \in R_i)$. Thus, there exist scalars $\alpha_{i,k}$, $k \in R_i$, such that $\boldsymbol{u}_i + \boldsymbol{e}_i = \sum_{k \in R_i} \alpha_{i,k} \boldsymbol{L}_k$. The ith receiver decodes its demand by computing $\sum_{k \in R_i} \alpha_{i,k} c_k - \boldsymbol{x}^{\mathsf{T}} \boldsymbol{u}_i$, which is equal to

$$\sum_{k \in R_i} \alpha_{i,k} \boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}_k - \boldsymbol{x}^{\mathsf{T}} \boldsymbol{u}_i = \boldsymbol{x}^{\mathsf{T}} (\boldsymbol{u}_i + \boldsymbol{e}_i) - \boldsymbol{x}^{\mathsf{T}} \boldsymbol{u}_i = x_i.$$

Notice that the receiver can compute $\boldsymbol{x}^{\mathsf{T}} \boldsymbol{u}_i$ using its side information since $\mathrm{supp}(\boldsymbol{u}_i) \subseteq K_i$. We will assume that each scalar $\alpha_{i,k}$ is non-zero since if $\alpha_{i,k} = 0$ the receiver does not need to query the coded symbol c_k. Finally, notice that stacking the vectors $\boldsymbol{u}_i + \boldsymbol{e}_i$, $i \in [N]$, we obtain a fitting matrix $\boldsymbol{A}$ corresponding to $\boldsymbol{L}$.

For certain choices of $S \subseteq [N]$, we will now relate the sizes of R_i, $i \in S$, and their union $\cup_{i \in S} R_i$. Let $\mathcal{N}(\boldsymbol{A})$ denote the null space of $\boldsymbol{A}$.

Lemma 9: Let $\boldsymbol{A}$ be any fitting matrix corresponding to a valid scalar linear encoder $\boldsymbol{L}$, $S \subseteq [N]$ be the support of a non-zero vector in $\mathcal{N}(\boldsymbol{A})$, and let the vectors $\boldsymbol{L}_k$, $k \in \cup_{i \in S} R_i$, be linearly independent. Then

$$\sum_{i \in S} |R_i| \geq 2 \left| \bigcup_{i \in S} R_i \right|.$$

Proof: Denote the columns of $\boldsymbol{A}$ by $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N$. Let $\boldsymbol{z} \in \mathcal{N}(\boldsymbol{A}) \setminus \{\boldsymbol{0}\}$ be such that $S = \mathrm{supp}(\boldsymbol{z})$. Since $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{0}$, we have $\sum_{i \in S} z_i \boldsymbol{A}_i = \boldsymbol{0}$, where the components z_i, $i \in S$, of the vector $\boldsymbol{z}$ are non-zero. Notice that there exist non-zero scalars $\alpha_{i,k}$ such that $\boldsymbol{A}_i = \sum_{k \in R_i} \alpha_{i,k} \boldsymbol{L}_k$. Hence, we have

$$\boldsymbol{0} = \sum_{i \in S} z_i \boldsymbol{A}_i = \sum_{i \in S} \sum_{k \in R_i} z_i \alpha_{i,k} \boldsymbol{L}_k.$$

All the scalars $z_i \alpha_{i,k}$ in the above linear combination are non-zero, and the set of vectors $\boldsymbol{L}_k$, $k \in \cup_{i \in S} R_i$, appearing in this linear combination are linearly independent. Hence, this linear combination is zero only if each $\boldsymbol{L}_k$, where $k \in \cup_{i \in S} R_i$, appears at least twice in the expansion $\sum_{i \in S} \sum_{k \in R_i} z_i \alpha_{i,k} \boldsymbol{L}_k$, i.e., only if each $k \in \cup_{i \in S} R_i$ is contained in at least two distinct sets R_i and R_j, $i \neq j$ and $i, j \in S$. Then a simple counting argument leads to the statement of this lemma. ∎

For any $S \subseteq [N]$, let G_S denote the subgraph of G induced by the vertices in S, i.e., the vertex set of G is S and edge set is $\{(i, j) \in \mathcal{E} \mid i, j \in S\}$. The subgraph G_S is the side information graph of the index coding problem obtained by restricting the index coding problem G to the messages x_i, $i \in S$. The following result can be used to manipulate the bound in Lemma 9.

Lemma 10: Let $\boldsymbol{L}$ be the encoder matrix of a valid scalar linear index code for G with receiver queries $R_1, \ldots, R_N$. For any $S \subseteq [N]$, we have

$$\left| \bigcup_{i \in S} R_i \right| \geq \mathrm{minrk}_q(G_S).$$

Proof: The submatrix $\boldsymbol{L}_S$ of $\boldsymbol{L}$ formed by the rows indexed by the set S is a valid encoder matrix for the index coding problem G_S. Since the receivers Rx_i, $i \in S$, query only the coded symbols with indices $k \in \cup_{i \in S} R_i$, the submatrix of $\boldsymbol{L}_S$ consisting of the columns with indices in $\cup_{i \in S} R_i$ is

also a valid scalar linear encoder for G_S. Hence, the code-length $|\cup_{i \in S} R_i|$ of this index code is lower bounded by $\mathrm{minrk}_q(G_S)$. ∎

The next result follows immediately from Lemmas 9 and 10.

Corollary 2: If L is an optimal scalar linear encoder for G, i.e., has codelength equal to $\mathrm{minrk}_q(G)$, A is a fitting matrix corresponding to L and $z \in \mathcal{N}(A) \setminus \{0\}$, then

$$\sum_{i \in S} r_i \geq 2 \, \mathrm{minrk}_q(G_S),$$

where $S = \mathsf{supp}(z)$.

Proof: The matrix L has linearly independent columns since the number of columns ℓ of L satisfies

$$\ell = \mathrm{minrk}_q(G) \leq \mathrm{rank}(A) \leq \mathrm{rank}(L) \leq \ell.$$

The corollary holds since $r_i = |R_i|$ for scalar linear codes and the vectors L_k, $k \in \cup_{i \in S} R_i$ satisfy the conditions of Lemma 9. ∎

B. Scalar Linear Coding When Minrank Is $N-1$

In the rest of this section we will assume $\mathrm{minrk}_q(G) = N-1$ and characterize the optimal localities r and r_{avg} among scalar linear index codes for such index coding problems. The main result of this section is Theorem 6.

Recall that since $M = 1$, the receiver localities r_i and the rate β are integers. Since the minimum scalar coding rate is equal to minrk_q, we are interested in the operating points corresponding to $\beta = N$ and $\beta = N - 1$. Note that the side information graph G contains at least one directed cycle, since otherwise, G is a directed acyclic graph and its minrank is equal to N [2], a contradiction. Let N_c denote the length of the smallest directed cycle contained in G.

If $N_c = 2$, there exist $i, j \in [N]$ such that $(i,j), (j,i) \in \mathcal{E}$, i.e., $i \in K_j$ and $j \in K_i$. The following scalar linear code attains the minimum possible locality $r = r_{\mathrm{avg}} = 1$ and the minimum possible rate $\beta = N - 1$ simultaneously. Transmit $x_i + x_j$ followed by transmitting the remaining $N-2$ information symbols uncoded. Rx$_i$ and Rx$_j$ can decode using $x_i + x_j$, and the remaining receivers query their demands directly from the codeword.

In the rest of this section we will assume that $N_c \geq 3$. Observe that rate $\beta = N$ can be achieved with smallest possible localities $r = r_{\mathrm{avg}} = 1$ using uncoded transmission. We will now consider the case $\beta = \mathrm{minrk}_q(G) = N-1$. Let L be any scalar encoder matrix with codelength $\ell = N - 1$, and A be a corresponding fitting matrix. Since $\ell = \mathrm{minrk}_q(G)$, we have $\ell \leq \mathrm{rank}(A) \leq \mathrm{rank}(L) \leq \ell$, and hence, $\mathrm{rank}(A) = \mathrm{rank}(L) = N - 1 = \ell$. Thus, the nullspace of A contains a non-zero vector.

Lemma 11: If $z \in \mathcal{N}(A) \setminus \{0\}$ and $S = \mathsf{supp}(z)$, the subgraph G_S of G induced by the vertices S contains at least one directed cycle.

Proof: Observe that the $|S| \times |S|$ submatrix A' of A composed of the rows and columns of A with indices in S fits G_S. Since $Az = 0$, the columns of A indexed by S are linearly dependent. This implies that

the columns of the submatrix A' are linearly dependent as well, and hence, $\mathrm{rank}(A') \leq |S| - 1$. It follows that $\mathrm{minrk}_q(G_S) \leq \mathrm{rank}(A') \leq |S| - 1$. Clearly G_S is not a directed acyclic graph since otherwise $\mathrm{minrk}_q(G_S) = |S|$. ∎

In order to use Corollary 2, we now derive a lower bound on $\mathrm{minrk}_q(G_S)$.

Lemma 12: If $\mathrm{minrk}_q(G) = N - 1$, then for any $S \subseteq [N]$, $\mathrm{minrk}_q(G_S) \geq |S| - 1$.

Proof: Consider the following valid scalar linear code for G. Encode the information symbols x_i, $i \in S$, using the optimal scalar linear code for G_S, and use uncoded transmission for the remaining symbols. The length of this code is lower bounded by $\mathrm{minrk}_q(G) = N - 1$, hence we obtain $\mathrm{minrk}_q(G_S) + N - |S| \geq N - 1$. ∎

We now prove the main result of this section.

Theorem 6: If $\mathrm{minrk}_q(G) = N - 1$ and the smallest directed cycle in G is of length $N_c \geq 3$, the optimal locality for scalar linear coding for G with rate $N - 1$ is

$$r = 2 \text{ and } r_{\mathrm{avg}} = \frac{N + N_c - 2}{N}.$$

Proof: Converse: Let A be a fitting matrix corresponding to any valid scalar linear code for G with rate $N - 1$. Let $z \in \mathcal{N}(A) \setminus \{0\}$ and $S = \mathsf{supp}(z)$. From Corollary 2 and Lemma 12, $\sum_{i \in S} r_i \geq 2 \, \mathrm{minrk}_q(G_S) \geq 2(|S| - 1)$. Using the trivial bound $r_i \geq 1$ for $i \notin S$, we have

$$\sum_{i \in [N]} r_i = \sum_{i \in S} r_i + \sum_{i \notin S} r_i$$
$$\geq 2(|S| - 1) + N - |S| = N + |S| - 2.$$

From Lemma 11, we know that G_S contains a cycle, and hence, the number of vertices $|S|$ in G_S is at least N_c. Thus,

$$r_{\mathrm{avg}} = \frac{\sum_{i \in [N]} r_i}{N} \geq \frac{N + |S| - 2}{N} \geq \frac{N + N_c - 2}{N}.$$

Since $N_c \geq 3$, we have $r \geq r_{\mathrm{avg}} \geq (N + 1)/N$, and since r is an integer we conclude that $r \geq 2$.

Achievability: Let $C \subseteq [N]$ be the set of vertices that form the smallest directed cycle in G. Note that the subgraph G_C is a directed cycle of length $|C| = N_c$. We encode the symbols x_i, $i \in C$, using the scalar linear code given in Example 2 and send the remaining $N - N_c$ symbols uncoded. This achieves the codelength $N-1$. From Example 2, the sum locality within the cycle $\sum_{i \in C} r_i = 2(N_c - 1)$ and the maximum locality within the cycle $\max_{i \in C} r_i = 2$. The locality of the remaining receivers is $r_i = 1$, $i \notin C$. This scheme achieves the optimal values of r and r_{avg} for rate $N - 1$. ∎

Remark 4: Corollary V.2 of [20] shows that, over the binary field $q = 2$, if $\mathrm{minrk}_q(G) = N-1$, then a scalar linear coding rate of $N - 1$ is achievable for any choice of locality $r \geq 2$. Our results show that $r = 2$ is optimal (if $N_c \geq 3$) and also provide the optimal value of the average locality r_{avg} for scalar linear codes over an arbitrary finite field $\mathbb{F}_q$. ∎

C. Feasible Receiver Localities for Directed Cycles

We observe from the results in Section VI-B that the interaction between rate and locality for scalar linear codes for graphs

with minrank equal to $N-1$ depends on the rate-locality trade-off of directed cycles. Further, one of the elementary index coding schemes (for arbitrary single unicast problems), known as *cycle covering* [11], [14], partitions the side information graph into disjoint cycles and applies a scalar linear code for each cycle. In view of the important role directed cycles play in index coding, we now determine the possible values of the tuple of receiver localities $\boldsymbol{r} = (r_1, \ldots, r_N)$ for scalar linear coding when G is a directed N-cycle with the desired broadcast rate being $N-1$.

Definition 3: A vector $\boldsymbol{r} = (r_1, \ldots, r_N)$ of receiver localities is *feasible* for G at rate ℓ through scalar linear coding if there exists a scalar linear code for G with codelength at the most ℓ with the localities of $\mathsf{Rx}_1, \ldots, \mathsf{Rx}_N$ at the most $r_1, \ldots, r_N$, respectively. ∎

In the rest of this section we will assume that G is a directed N-cycle. We have $N_c = N$ in this case, and Theorem 6 implies that for any valid scalar coding scheme with rate $N-1$, $r_{\text{avg}} \geq 2(N-1)/N$. This implies that if the vector of receiver localities $\boldsymbol{r}$ is feasible for G using scalar linear codes at rate $N-1$, then $r_1 + \cdots + r_N \geq 2(N-1)$. We also know that each $r_i \geq 1$, $i \in [N]$. We now show that these two necessary conditions are also sufficient for the feasibility of $\boldsymbol{r}$. The rest of this section is the proof of the sufficiency part of

---

Theorem 7: Let G be a directed N-cycle and $\mathbb{F}_q$ be any finite field. A vector of receiver localities $\boldsymbol{r} = (r_1, \ldots, r_N)$ consisting of strictly positive integers is feasible for G at rate $N-1$ through scalar linear coding over $\mathbb{F}_q$ if and only if $\sum_{i \in [N]} r_i \geq 2(N-1)$.

We will assume without loss of generality that $\sum_{i \in [N]} r_i = 2(N-1)$ and that the components of $\boldsymbol{r}$ are in ascending order. We might have to relabel the vertices of G for this latter assumption to hold. In general, we will assume $r_1 \leq \cdots \leq r_N$, and that Rx_i demands x_i and knows the message $x_{\pi(i)}$ as side information, where π is a permutation on $[N]$ and $\pi(i) \neq i$ for any $i \in [N]$.

Our proof for Theorem 7 is constructive. Given a vector $\boldsymbol{r}$ satisfying the conditions in the theorem, we construct a scalar linear code of rate $N-1$ that achieves receiver localities $r_1, \ldots, r_N$. Towards this, we first construct a bipartite graph $\mathcal{B}$ satisfying certain adjacency properties, and then design the index code using $\mathcal{B}$.

The Bipartite Graph $\mathcal{B}$: The bipartite graph $\mathcal{B}$ consists of the vertex sets $\{u_1, \ldots, u_N\}$ representing the receivers and $\{v_1, \ldots, v_{N-1}\}$ representing the codeword symbols of the length $N-1$ index code. The edges between the two vertex sets denote the queries made by the receivers to decode their demands. Thus, designing $\mathcal{B}$ is identical to designing the sets $R_1, \ldots, R_N$. An edge $\{u_i, v_k\}$ exists if and only if $k \in R_i$, i.e., the kth codeword symbol is observed by Rx_i. The degree of u_i is $r_i = |R_i|$, and the degree of v_k is the number of receivers querying the kth coded symbol.

Lemma 13: Let $N \geq 2$ and $\boldsymbol{r} = (r_1, \ldots, r_N)$ be a vector of strictly positive integers satisfying $\sum_i r_i = 2(N-1)$ and $r_1 \leq r_2 \leq \cdots \leq r_N$. There exists a bipartite graph $\mathcal{B}$ with vertex set $\{u_1, \ldots, u_N\} \cup \{v_1, \ldots, v_{N-1}\}$ such that
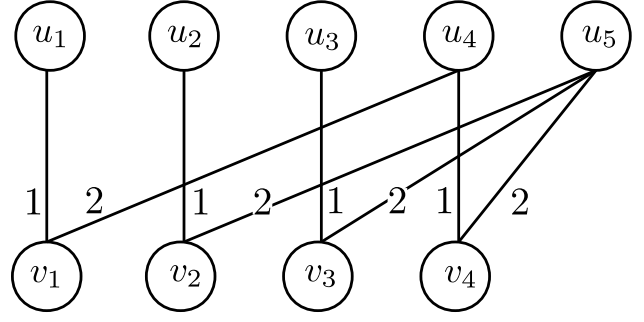

Fig. 3. A bipartite graph $\mathcal{B}$ for the index coding problem of Example 4. The edges of $\mathcal{B}$ are weighted using scalars from $\mathbb{F}_3$, with weights equal to either 1 or $2 = -1$.

(i) the degree of u_i is r_i, for $i \in [N]$;
(ii) the degrees of $v_1, \ldots, v_{N-1}$ are all equal to 2; and
(iii) for $i \in [N-1]$, $\{u_i, v_i\}$ is an edge and the neighbors of u_i are a subset of $\{v_1, \ldots, v_i\}$.

Proof: We provide a constructive proof of this lemma in Appendix IV. Note that property *(iii)* in this lemma is equivalent to $i \in R_i$ and $R_i \subseteq [i]$ for all $i \in [N-1]$. ∎

The neighbors of u_i correspond to the set $R_i \subseteq [N-1]$. We will represent the neighbors of v_k by the set $S_k \subset [N]$, i.e., $S_k = \{i \in [N] \mid k \in R_i\}$. Note that $|S_k| = 2$ for $k \in [N-1]$. For each edge $\{u_i, v_k\}$, we associate a weight $w_{i,k} \in \mathbb{F}_q$ as follows. If the characteristic of $\mathbb{F}_q$ is two, all the weights are equal to 1. Otherwise, for each $k \in [N-1]$, the two edges incident on v_k are assigned the two distinct weights 1 and -1, respectively, in an arbitrary fashion. Such an assignment of weights $w_{i,k}$ implies, over any field $\mathbb{F}_q$,

$$\sum_{i \in S_k} w_{i,k} = 1 - 1 = 0, \quad \text{for all } k \in [N-1]. \quad (22)$$

Example 4: Consider a directed cycle of length $N = 5$, where the side-informations of the receivers Rx_1, Rx_2, Rx_3, Rx_4, Rx_5 are the messages x_3, x_1, x_5, x_2, x_4, respectively. Note that $(\pi(1), \pi(2), \ldots, \pi(5)) = (3, 1, 5, 2, 4)$. Here, G is the cycle $(1, 3, 5, 4, 2, 1)$. Let the desired locality vector be $\boldsymbol{r} = (r_1, r_2, r_3, r_4, r_5) = (1, 1, 1, 2, 3)$. Let us consider code construction over the field $\mathbb{F}_3 = \{0, 1, 2\}$, where $-1 = 2$. A bipartite graph $\mathcal{B}$ corresponding to this problem is shown in Fig. 3. Note that the degrees of $u_1, \ldots, u_5$ are $1, 1, 1, 2, 3$, respectively. The degrees of $v_1, \ldots, v_4$ are equal to 2. For each $i \leq 4$, $\{u_i, v_i\}$ is an edge and the neighbors of u_i are a subset of $\{1, 2, \ldots, i\}$. For each $k \leq 4$, the weights of the two edges incident on v_k are distinct, and are equal to 1 and 2, respectively. ∎

Designing the Scalar Linear Index Code: Let $\mathbb{F}_q$ be any finite field. We will use the following fitting matrix $\boldsymbol{A}$ to design the index code over $\mathbb{F}_q$. The ith column $\boldsymbol{A}_i$ of the matrix $\boldsymbol{A}$ contains exactly two non-zero components: the entry in position i is 1, and the entry in position $\pi(i)$ is -1. Note that each row of $\boldsymbol{A}$ has exactly two non-zero entries with corresponding values 1 and -1. Hence we have

$$\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_N = \boldsymbol{0}. \quad (23)$$

We now construct an encoding matrix $\boldsymbol{L}$ with $N-1$ columns $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{N-1}$ corresponding to the fitting matrix $\boldsymbol{A}$ and for

the receiver queries $R_1, \ldots, R_N$ dictated by $\mathcal{B}$. A sufficient condition for L to satisfy these properties is

$$A_i = \sum_{k \in R_i} w_{i,k} L_k, \text{ for } i \in [N], \tag{24}$$

since this implies $x_i - x_{\pi(i)} = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{A}_i$ can be obtained from the coded symbols $\boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}_k$, $k \in R_i$. We prove the existence of $\boldsymbol{L}$ by viewing (24) as a set of N linear equations in unknowns $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_{N-1}$, and showing that this system has a unique solution. Towards this, we claim that the last equation in this system is linearly dependent on the first $N-1$ equations, and hence, is redundant. The proof for this claim follows from using (22) and (23) as below

$$\boldsymbol{0} = \boldsymbol{A}_1 + \cdots + \boldsymbol{A}_N - \sum_{k \in [N-1]} \sum_{i \in S_k} w_{i,k} \boldsymbol{L}_k$$

(last term corresponds to all the edges in $\mathcal{B}$)

$$= \boldsymbol{A}_1 + \cdots + \boldsymbol{A}_N - \sum_{i \in [N]} \sum_{k \in R_i} w_{i,k} \boldsymbol{L}_k$$

(alternative way to enumerate the edges in $\mathcal{B}$)

$$= \sum_{i \in [N-1]} \left(\boldsymbol{A}_i - \sum_{k \in R_i} w_{i,k} \boldsymbol{L}_k\right) + \left(\boldsymbol{A}_N - \sum_{k \in R_N} w_{N,k} \boldsymbol{L}_k\right).$$

We conclude that (24) is equivalent to the following system consisting of $N-1$ equations with an equal number of unknowns

$$A_i = \sum_{k \in R_i} w_{i,k} L_k, \text{ for } i \in [N-1]. \tag{25}$$

From Lemma 13 we know that $i \in R_i$ and $R_i \subseteq [i]$. Thus, the unknowns involved in the ith equation of the linear system (25) are a subset of $\{\boldsymbol{L}_1, \ldots, \boldsymbol{L}_i\}$ and necessarily include the unknown $\boldsymbol{L}_i$. Hence, (25) is a triangular system of linear equations, and therefore, has a unique solution. We conclude that the locality vector $\boldsymbol{r}$ is feasible.

Example 5: We will continue with the index coding problem of Example 4. The fitting matrix for this problem is

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 0 & 1 \end{bmatrix}.$$

We now determine the encoder matrix $\boldsymbol{L}$, that consists of 4 columns $\boldsymbol{L}_1, \ldots, \boldsymbol{L}_4$, using the bipartite graph $\mathcal{B}$ in Fig 3. The system of linear equations (25) for this problem is

$$\boldsymbol{A}_1 = \boldsymbol{L}_1, \ \boldsymbol{A}_2 = \boldsymbol{L}_2, \ \boldsymbol{A}_3 = \boldsymbol{L}_3, \text{ and } \boldsymbol{A}_4 = 2\boldsymbol{L}_1 + \boldsymbol{L}_4.$$

The unique solution to this system is

$$\boldsymbol{L} = [\boldsymbol{L}_1 \cdots \boldsymbol{L}_4] = [\boldsymbol{A}_1 \ \boldsymbol{A}_2 \ \boldsymbol{A}_3 \ \boldsymbol{A}_4 + \boldsymbol{A}_1] = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 2 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{bmatrix}.$$

With this encoding matrix, the transmitted codeword is

$$\boldsymbol{c} = (c_1, c_2, c_3, c_4)$$
$$= (x_1 - x_3, x_2 - x_1, x_3 - x_5, x_1 - x_2 - x_3 + x_4).$$

The receivers $\mathsf{Rx}_1, \mathsf{Rx}_2, \ldots, \mathsf{Rx}_5$ decode their demanded messages using a subset of codeword symbols and their own side information as follows

$$x_1 = c_1 + x_3, \qquad x_2 = c_2 + x_1, \qquad x_3 = c_3 + x_5,$$
$$x_4 = c_4 - c_1 + x_2 \quad \text{and} \quad x_5 = -c_2 - c_3 - c_4 + x_4.$$

The localities at the five receivers are $(r_1, r_2, \ldots, r_5) = (1, 1, 1, 2, 3)$, which are as stipulated in Example 4. ∎

VII. CONSTRUCTING LOCALLY DECODABLE INDEX CODES

In the previous sections we considered index codes for specific cases: locality is equal to one, side information graph is a directed cycle, and minrank is one less than the order of the problem. We developed matching converses for each of these cases and identified optimal index codes. In this section we propose techniques to construct linear locally decodable index codes for arbitrary index coding problems. We also show that some of the traditional index code constructions can be modified to yield locally decodable codes.

Recall the decoding technique for scalar linear codes from Section VI-A. Let the encoding matrix be $\boldsymbol{L}$, the corresponding fitting matrix be $\boldsymbol{A}$ and the ith column of $\boldsymbol{A}$ be $\boldsymbol{A}_i = \boldsymbol{e}_i + \boldsymbol{u}_i$ where $\boldsymbol{u}_i \lhd K_i$. Receiver Rx_i decodes its demand by computing a specific linear combination of codeword symbols and its side information. The linear combination $\boldsymbol{c}^{\mathsf{T}} \boldsymbol{d}_i$, $\boldsymbol{d}_i \in \mathbb{F}_q^\ell$, of the codeword symbols is chosen such $\boldsymbol{L} \boldsymbol{d}_i = \boldsymbol{A}_i$, and hence,

$$\boldsymbol{c}^{\mathsf{T}} \boldsymbol{d}_i = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{L} \boldsymbol{d}_i = \boldsymbol{x}^{\mathsf{T}} \boldsymbol{A}_i = \boldsymbol{x}^{\mathsf{T}} (\boldsymbol{e}_i + \boldsymbol{u}_i). \tag{26}$$

With this choice of linear combination, the demand x_i can be decoded as $x_i = \boldsymbol{c}^{\mathsf{T}} \boldsymbol{d}_i - \boldsymbol{x}^{\mathsf{T}} \boldsymbol{u}_i$, Note that $R_i = \mathsf{supp}(\boldsymbol{d}_i)$ since the receiver needs to observe only those components of $\boldsymbol{c}$ that correspond to the non-zero entries of $\boldsymbol{d}_i$. Thus, the locality of the ith receiver equals the Hamming weight of $\boldsymbol{d}_i$, i.e., $r_i = \mathsf{wt}(\boldsymbol{d}_i)$. If $\boldsymbol{A}_i$ is one of the columns of the encoding matrix $\boldsymbol{L}$, then $\boldsymbol{d}_i$ can be chosen such that $\mathsf{wt}(\boldsymbol{d}_i) = 1$ resulting in $r_i = 1$. If $\boldsymbol{L}$ does not contain $\boldsymbol{A}_i$ as one of its columns, then we have the naive upper bound $r_i = \mathsf{wt}(\boldsymbol{d}_i) \le \ell$.

A. Separation Based Scheme Using Covering Codes

We can arrive at a locally decodable code using a separation based technique where the encoder matrix $\boldsymbol{L} = \boldsymbol{L}'\boldsymbol{H}$ is the product of two matrices: an optimal index coding matrix $\boldsymbol{L}'$ with number of columns equal to $\text{minrk}_q(G)$, and the parity-check matrix $\boldsymbol{H}$ of a *covering code* $\mathscr{C}$ with length ℓ and dimension $\ell - \text{minrk}_q(G)$. The linear code $\mathscr{C}$ is chosen such that its covering radius is equal to the desired locality r, i.e., Hamming spheres of radius r centered around the codewords of $\mathscr{C}$ cover the entire space $\mathbb{F}_q^\ell$. Among all covering codes over $\mathbb{F}_q$ with covering radius r and co-dimension $\text{minrk}_q(G)$ we choose the one with the smallest possible blocklength ℓ. Note that the number of rows of $\boldsymbol{H}$ is $\text{minrk}_q(G)$. Since the covering radius of $\mathscr{C}$ is r, it is well known that, any vector of length $\text{minrk}_q(G)$ is a linear combination of at the most r columns of $\boldsymbol{H}$. If $\boldsymbol{A}$ is a fitting matrix corresponding to $\boldsymbol{L}'$, every column of $\boldsymbol{A}$ is a linear combination of columns of $\boldsymbol{L}'$. Thus, $\boldsymbol{A}_i$ can be expressed as a linear combination of at the most r columns

of the matrix $L = L'H$. We conclude that L is a valid scalar linear encoder matrix for G with locality r.

B. Codes Using Acyclic Induced Subgraph Covers of G

We will design vector linear codes using the acyclic induced subgraphs of G. First, we require the following result.

Note that for any subset $S \subset [N]$ of vertices of G, G_S denotes the subgraph of G induced by S.

Lemma 14: Let $\ell \geq \operatorname{minrk}_q(G)$ and the subgraph G_S of G induced by the subset $S \subseteq [N]$ be a directed acyclic graph. There exists a scalar linear index code of length ℓ for G such that $r_i = 1$ for every $i \in S$.

Proof: Let $L \in \mathbb{F}_q^{N \times \ell}$ be the encoding matrix of any valid scalar linear code for G, and $A = [A_1 \cdots A_N]$ be the fitting matrix corresponding to L. Since G_S is acyclic, there exists a *topological ordering* $i_1, \ldots, i_{|S|}$ of its vertex set $S = \{i_1, \ldots, i_{|S|}\}$, i.e., for any $1 \leq a < b \leq |S|$, there exists no directed edge (i_b, i_a) in G_S, and hence, G does not contain the edge (i_b, i_a). It follows that for any choice of $1 \leq a < b \leq |S|$, $i_a \notin K_{i_b}$, and hence, $A_{i_a, i_b} = 0$. Further, for any $1 \leq a \leq |S|$, $A_{i_a, i_a} = 1$ since the diagonal entries of A are equal to 1. Let E be the $|S| \times |S|$ square submatrix of A composed of the rows and columns indexed by S. It follows that if the rows and columns of E are permuted according to the topological ordering $i_1, \ldots, i_{|S|}$, then E is lower triangular and all the entries on its main diagonal are equal to 1. Thus E is a full-rank matrix, and hence, the columns $A_{i_1}, \ldots, A_{i_{|S|}}$ of A are linearly independent.

Consider the matrix $L' \in \mathbb{F}_q^{N \times \ell}$ constructed as follows. Let the first $|S|$ columns of L' be $A_{i_1}, \ldots, A_{i_{|S|}}$. The remaining $\ell - |S|$ columns of L' are chosen from among the columns of L such that the column spaces of L' and L are identical. This is possible since $A_{i_1}, \ldots, A_{i_{|S|}}$ are linearly independent and are contained in the column space of L. By construction, the column space of L' contains the column space of A, and hence, L' is a valid scalar linear index coding matrix for G. Also, for any $i \in S$, A_i is a column of L', and hence, $r_i = 1$. ∎

Definition 4: A set of P subsets $S_1, \ldots, S_P \subseteq [N]$ of the vertex set of the side information graph G is a *Q-fold acyclic induced subgraph (AIS) cover* of G if *(i)* $S_1 \cup \cdots \cup S_P = [N]$, *(ii)* each $i \in [N]$ is an element of at least Q of the P subsets $S_1, \ldots, S_P$, and *(iii)* all the P induced subgraphs $G_{S_1}, \ldots, G_{S_M}$ are acyclic. ∎

Given an AIS cover $S_1, \ldots, S_P$ of G and any $\ell \geq \operatorname{minrk}_q(G)$, we construct a vector linear code as follows. From Lemma 14, we know that for each $j \in [P]$, there exists a valid scalar linear encoding matrix, say $L^{(j)}$, with codelength ℓ such that the locality of every receiver $i \in S_j$ is 1. Consider a vector linear index code that encodes P independent instances of the scalar messages $x_1, \ldots, x_N \in \mathbb{F}_q$ using the encoding matrices $L^{(1)}, \ldots, L^{(P)}$, respectively. The broadcast rate of this scheme is ℓ. If $S_1, \ldots, S_P$ is a Q-fold AIS cover of G, for each $i \in [N]$, there exist Q scalar linear encoders among $L^{(1)}, \ldots, L^{(P)}$ that provide locality 1 at the ith receiver. The locality provided by the remaining $(P-Q)$ encoders is at most ℓ at this receiver. Thus the number of encoded symbols queried by any receiver in this vector linear coding scheme is at the most $Q + (P-Q)\ell$. Normalizing this by the number of message

instances P, we observe that the locality of this scheme is at the most $(Q + (P-Q)\ell)/P$. Thus, we have proved

Theorem 8: If there exists a Q-fold AIS cover of G consisting of P subsets of its vertex set, and if $\ell \geq \operatorname{minrk}_q(G)$, there exists a vector linear code for G with broadcast rate ℓ, message length $M = P$, and locality at the most $(Q + (P-Q)\ell)/P$, and hence,

$$\beta_{G,q}^* \left(\frac{Q + (P-Q)\ell}{P} \right) \leq \ell.$$

As an application of Theorem 8, consider the following coding scheme. Let the parameter $t \geq 1$ be such that the side information graph G contains no cycles of length t or less. With $P = \binom{N}{t}$, let $S_1, \ldots, S_P$ be the set of all subsets of $[N]$ of size t. The subgraph of G induced by S_j, for any $j \in [P]$, is acyclic since $|S_j| = t$. Further, each $i \in [P]$ is an element of $Q = \binom{N-1}{t-1}$ of the P subsets $S_1, \ldots, S_P$. Hence the resulting locality is $(Q + (P-Q)\ell)/P$ which can easily be shown to be equal to $(t + (N-t)\ell)/N$. Hence, we have

Corollary 3: If G contains no cycles of length t or less and $\ell \geq \operatorname{minrk}_q(G)$, then we can achieve broadcast rate ℓ with locality at the most $(t + (N-t)\ell)/N$.

Using $t = 1$ in Corollary 3 we immediately arrive at the following result.

Corollary 4: Let G be any index coding problem and $\ell \geq \operatorname{minrk}_q(G)$. There exists a vector linear index code for G with message length $M = N$, broadcast rate $\beta = \ell$ and locality $r \leq (1 + (N-1)\ell)/N$.

C. Side Information Graphs With Circular Symmetry

We will now construct vector linear index codes for side information graphs G that satisfy a symmetry property. Consider the permutation σ on the set $[N]$ that maps $i \in [N]$ to $\sigma(i) = (i \mod N) + 1$, i.e., $\sigma(1) = 2, \sigma(2) = 3, \ldots, \sigma(N) = 1$. In this subsection we will assume G to be any directed graph with vertex set $[N]$ such that σ is an automorphism of G. Such unicast index coding problems have been considered before in the literature, see [25], and are related to topological interference management [26]. The following theorem shows that any rate $\ell \geq \operatorname{minrk}_q(G)$ can be achieved using a vector linear code with r at the most $\ell(N - \ell + 1)/N$.

Theorem 9: If the cyclic permutation σ is an automorphism of G and if $\ell \geq \operatorname{minrk}_q(G)$, then

$$\beta_{G,q}^* \left(\frac{\ell(N - \ell + 1)}{N} \right) \leq \ell.$$

Proof: Since $\ell \geq \operatorname{minrk}_q(G)$, there exists an $A \in \mathbb{F}_q^{N \times N}$ that fits G and which is of rank ℓ. Let L be an $N \times \ell$ matrix composed of a set of ℓ linearly independent columns of A. Note that when L is used as a scalar linear index code there exist ℓ receivers with locality 1 since exactly ℓ columns of A appear as columns of L.

Let P be the permutation matrix obtained by cyclically shifting down the rows of the $N \times N$ identity matrix by one position. It is straightforward to verify that PAP^T fits the graph $\sigma(G)$. Since σ is an automorphism of G, we have $\sigma(G) = G$, and hence, PAP^T fits G. Since the column space of PL is identical to that of PAP^T, PL represents a valid

scalar linear index code for G. Extending this argument we deduce that for any $j \in [N]$, the matrix $\boldsymbol{A}^{(j)} = \boldsymbol{P}^j \boldsymbol{A}(\boldsymbol{P}^j)^{\mathsf{T}}$ fits G, and the column space of $\boldsymbol{L}^{(j)} = \boldsymbol{P}^j \boldsymbol{L}$ is identical to that of $\boldsymbol{A}^{(j)}$. The matrices $\boldsymbol{L}^{(j)}$ and $\boldsymbol{A}^{(j)}$, $j \in [N]$, represent N valid scalar linear codes and the corresponding fitting matrices for G. The receiver Rx_i has locality 1 when using the jth code $\boldsymbol{L}^{(j)}$ if the ith column of $\boldsymbol{A}^{(j)}$ appears as a column of $\boldsymbol{L}^{(j)}$, else its locality is at the most ℓ. Using the fact that the group $\{\sigma, \sigma^2, \ldots, \sigma^N = 1\}$ acts transitively on the vertices of G, we observe that for any $i \in [N]$, there exist ℓ distinct values of j such that the ith column of $\boldsymbol{A}^{(j)}$ is a column of $\boldsymbol{L}^{(j)}$. Thus, for any $i \in [N]$, the locality of Rx_i is 1 for ℓ of the N codes $\boldsymbol{L}^{(1)}, \ldots, \boldsymbol{L}^{(N)}$, and is at the most ℓ for the remaining $N - \ell$ codes. Time-sharing these N codes we obtain a vector linear code with message length $M = N$, rate ℓ and locality at the most $(\ell + \ell(N - \ell))/N = \ell(N - \ell + 1)/N$ for each of the N receivers. ∎

D. Codes From Optimal Coverings of G

Several index coding schemes in the literature partition the given problem (side information graph G) into subproblems (subgraphs of G), and apply a pre-defined coding technique on each of these subproblems independently. The overall codelength is the sum of the codelengths of the individual subproblems. The broadcast rate is then reduced by optimizing over all possible partitions of G. We will now quickly recall a few such *covering-based* schemes, and then show that they can be modified to guarantee locality.

Partition Multicast or Partial Clique Covering [1], [12], [13]: Let G_S be the subgraph of G induced by the subset of vertices $S \subseteq [N]$. The number of information symbols in the index coding problem G_S is $|S|$, and the side information of receiver $i \in S$ in G_S is $K_i \cap S$. The partition multicast scheme uses a scalar linear encoder for G_S where the encoding matrix is the transpose of the parity-check matrix of an MDS code of length $|S|$ and dimension $\min_{i \in S} |K_i \cap S|$. This code for G_S encodes messages of length $m_S = 1$ and has codelength $\ell_S = |S| - \min_{i \in S} |K_i \cap S|$. We will use the trivial value of overall locality $r_S = \ell_S$ for this coding scheme. The finite field $\mathbb{F}_q$ must be sufficiently large for the said MDS codes to exist.

Cycle Covering [11], [14]: If G_S is a directed cycle of length $|S|$, then it is encoded using the scalar linear coding scheme described in Example 2. resulting in message length $m_S = 1$, index codelength $\ell_S = |S| - 1$, and locality $r_S = 2$. If G_S is not a directed cycle, then its information symbols are transmitted uncoded resulting in $m_S = 1$, $\ell_S = |S|$ and locality $r_S = 1$.

In similar vein to partition multicast and cycle covering schemes, consider the following proposed technique that applies the optimal scalar linear code over each subgraph G_S.

Minrank Covering: Encode each subgraph G_S using its own optimal scalar linear index code. The message length $m_S = 1$, codelength ℓ_S equals $\operatorname{minrk}_q(G_S)$, and use trivial value for locality $r_S = \operatorname{minrk}_q(G_S)$. By partitioning G into subgraphs G_S of small minrank we can achieve a small locality for the overall scheme.

Now consider any covering-based index coding technique (such as partition multicast, cycle covering or minrank covering) for G. Let the scalar linear index code associated with the subgraph G_S, $S \subseteq [N]$, have codelength ℓ_S and locality r_S. The overall index code uses a partition of the vertex set $[N]$, which is represented by the tuple $(a_S, S \subset [N])$, where each $a_S \in \{0, 1\}$ is such that the partition of $[N]$ consists of all subsets S with $a_S = 1$. Note that $(a_S, S \subset [N])$ represents a partition of $[N]$ if and only if $\sum_{S:i \in S} a_S = 1$ for every $i \in [N]$, i.e., every vertex i is contained in exactly one of the subsets in the partition. The covering-based index coding technique applies an index code of length ℓ_S and locality r_S to each subgraph G_S with $a_S = 1$ independently. Thus the codelength of the overall index code is $\ell = \sum_{S:a_S=1} \ell_S = \sum_{S \subseteq [N]} a_S \ell_S$ and locality is $r = \max_{S:a_S=1} r_S = \max_{S \subseteq [N]} a_S r_S$. By optimizing over all possible partitions of G, we have

Theorem 10 (Covering With Locality): Consider a family of scalar linear index codes, one for each G_S, $S \subseteq [N]$, with length ℓ_S and locality r_S. Given any $r \geq 1$, there exists a scalar linear code for G with locality at the most r and rate equal to the solution of the following integer program: minimize $\sum_{S \subseteq [N]} a_S \ell_S$ subject to the following constraints, $\sum_{S:i \in S} a_S = 1$, for all $i \in [N]$, and $a_S r_S \leq r$, where $a_S \in \{0, 1\}$ for all $S \subseteq [N]$.

The second constraint $a_S r_S \leq r$ in Theorem 10 ensures that the locality of the resulting coding scheme is at the most r. Since $a_S \in \{0, 1\}$, this implies that when solving for the optimal partition, the integer program considers only those subsets S with $r_S \leq r$, i.e., locality r is achieved by partitioning G into subproblems of small locality.

The above technique can be extended to vector linear codes by using a linear programming relaxation to allow $0 \leq a_S \leq 1$. The subproblems G_S can themselves be encoded using vector linear codes of small locality, as in Theorems 8 and 9, say with rate β_S and locality r_S. Then the overall achievable rate by time-sharing over these subproblems, for a given locality r, is the solution to $\min \sum_{S \subseteq [N]} a_S \beta_S$ subject to $\sum_{S:i \in S} a_S r_S \leq r$ and $\sum_{S:i \in S} a_S = 1$ for each $i \in [N]$, where $a_S \in [0, 1]$ for all $S \subseteq [N]$.

VIII. Discussion

We introduced the problem of designing index codes that are locally decodable and have identified several techniques to construct such codes. We identified the optimal trade-off between rate and locality for the following cases: when locality $r = 1$, side information graph is the directed 3-cycle, vector linear codes for directed cycles of any length, and scalar linear codes when minrank is one less than the number of messages.

A. Relation to Overcomplete Dictionaries

Locally decodable index codes seem to be related to the problem of sparse representation of sets of vectors. A scalar linear index code with locality r is characterized by a valid encoder matrix $\boldsymbol{L}$ with a corresponding fitting matrix $\boldsymbol{A}$ such that any column of $\boldsymbol{A}$ is some linear combination of at the most r columns of $\boldsymbol{L}$. Thus the columns of $\boldsymbol{L}$ serve as an

overcomplete basis for a sparse representation of the columns of $\boldsymbol{A}$. Given an index coding problem, we must design a fitting matrix $\boldsymbol{A}$ and a corresponding overcomplete basis $\boldsymbol{L}$ for $\boldsymbol{A}$ such that both the locality r and the codelength ℓ are small.

B. On Generalization of Results on Locality-Rate Trade-Off

Stronger achievability and/or converse results may be required to gain deeper insights into the locality-rate trade-off of a general index coding problem. In Section VI we identified optimal scalar linear codes for side information graphs G with $\mathrm{minrk}_q(G) = N - 1$. The key property used in Section VI is that any such directed graph G contains at least one directed cycle (see Lemma 11 and Theorem 6). A natural direction to generalize this result is to consider side information graphs G with smaller values of minrank, such as $\mathrm{minrk}_q(G) \leq N - 2$. However, the structural properties of directed graphs with $\mathrm{minrk}_q(G) = N - 2$ are not as extensively understood in the index coding community as the more extreme (and simpler) case of $\mathrm{minrk}_q(G) = N - 1$. For instance, reference [27] identifies a subset of index coding problems whose minrank equals $N - 2$.

Another avenue of work is the generalization of the results on directed 3-cycle (in Section V) to directed cycles of length $N > 3$. The case $N = 3$ is highly structured and this facilitated our analysis in Section V. However, some of this symmetry is lost when N takes values larger than 3 and this imposes analytical difficulties (see Remark 3).

One of the extreme operating points in the trade-off between locality r and rate β is the point where the broadcast rate is optimal, i.e., minimum possible with no restriction imposed by locality. Thus, when deriving the optimal trade-off between r and β we are naturally restricted to the family of index coding problems for which the minimum broadcast rate or capacity is explicitly known. However, determining the capacity of index coding problems is difficult in general, and the capacity is known for only some families of side information graphs.

APPENDIX I
INVARIANCE OF LOCALITY-RATE TRADE-OFF WITH RESPECT TO CHANNEL AND MESSAGE ALPHABETS

The channel model considered in this paper assumes that the message alphabet and the channel alphabet are identical, i.e., all the messages and the codeword are vectors over the same alphabet $\mathcal{A}$. The definition of rate β, locality r and the optimal locality-rate trade-off $\beta_G^*(r)$ given in Section II are specific to this model. In this section, we consider the general scenario where the channel alphabet, denoted by $\mathcal{Z}$, and the message alphabet $\mathcal{A}$ can be different. We will define locality when $\mathcal{A} \neq \mathcal{Z}$, and show that the optimal locality-rate trade-off (that captures non-linear codes as well) is independent of the choice of $\mathcal{A}$ and $\mathcal{Z}$ as long as these sets are finite and of size at least 2. The results of this appendix are summarized in Appendix I-D.

A. The General System Model

As in Section II, we will assume that G is a given side information graph with N vertices, and the messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$

are elements of $\mathcal{A}^M$, where $\mathcal{A}$ is the message alphabet and M is the message length. The codeword $\boldsymbol{c}$ is a vector of length ℓ over the channel alphabet $\mathcal{Z}$, i.e., $\boldsymbol{c} = (c_1, \ldots, c_\ell) \in \mathcal{Z}^\ell$. Both $\mathcal{A}$ and $\mathcal{Z}$ are finite sets. The receiver Rx_i observes $\boldsymbol{c}_{R_i} = (c_k, k \in R_i) \in \mathcal{Z}^{|R_i|}$, where $R_i \subseteq [\ell]$, and decodes its demand $\boldsymbol{x}_i$ using $\boldsymbol{c}_{R_i}$ and its side information $\boldsymbol{x}_j$, $j \in K_i$. The rate of this code is defined as

$$\beta = \frac{\log_2 |\mathcal{Z}|^\ell}{\log_2 |\mathcal{A}|^M} = \frac{\ell}{M} \times \frac{\log_2 |\mathcal{Z}|}{\log_2 |\mathcal{A}|}.$$

This is consistent with the definitions of rate used in the literature; for instance, see [13]. We define the locality of receiver Rx_i as

$$r_i = \frac{\log_2 |\mathcal{Z}|^{|R_i|}}{\log_2 |\mathcal{A}|^M} = \frac{|R_i|}{M} \times \frac{\log_2 |\mathcal{Z}|}{\log_2 |\mathcal{A}|}.$$

Here, $\log_2 |\mathcal{Z}|^{|R_i|}$ measures the number of bits observed from the channel and $\log_2 |\mathcal{A}|^M$ is the information content of the desired message (in bits). Thus r_i is the number of transmission bits queried per decoded message bit at Rx_i. The average locality r_{avg} and overall locality r are the arithmetic average and the maximum of $r_1, \ldots, r_N$, respectively. When $\mathcal{Z} = \mathcal{A}$ these definitions of β, r_i, r_{avg} and r coincide with the ones used in the main text of this paper.

Definition 5: The *optimal broadcast rate* $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ is the infimum among the broadcast rates β of all valid index codes (including non-linear codes) with message alphabet $\mathcal{A}$, channel alphabet $\mathcal{Z}$, and with locality at the most r, considering all possible message lengths $M \geq 1$. ∎

The definition of $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ captures the possible dependence (if any) of the locality-rate trade-off on the choice of message and channel alphabets. The function $\beta_G^*(r)$ in the main text is identical to $\beta_G^*(r, \mathcal{A}, \mathcal{A})$.

B. Invariance With Respect to Channel Alphabet

We now show that $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ is independent of $\mathcal{Z}$. We will do so by showing that $\beta_G^*(r, \mathcal{A}, \mathcal{Z}') \leq \beta_G^*(r, \mathcal{A}, \mathcal{Z})$ for any choice of alphabets $\mathcal{Z}$ and $\mathcal{Z}'$. Reversing the roles of $\mathcal{Z}$ and $\mathcal{Z}'$ in this inequality shows that $\beta_G^*(r, \mathcal{A}, \mathcal{Z}) \leq \beta_G^*(r, \mathcal{A}, \mathcal{Z}')$, and hence, $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ is independent of $\mathcal{Z}$.

Lemma 15: Consider any valid index code for G with message and channel alphabets $\mathcal{A}$ and $\mathcal{Z}$ with rate β and overall locality r, with $|\mathcal{A}|, |\mathcal{Z}| \geq 2$. For any finite set $\mathcal{Z}'$, with $|\mathcal{Z}'| \geq 2$, and any $\epsilon > 0$, there exists a valid index code for G with message alphabet $\mathcal{A}$, channel alphabet $\mathcal{Z}'$, rate at the most $(1 + \epsilon)\beta$ and locality at the most $(1 + \epsilon)r$.

Proof: Let $\mathscr{C}$ be the given valid index code over the channel alphabet $\mathcal{Z}$, with codelength ℓ, message length M, receiver queries $R_1, \ldots, R_N$. Denote the decoders at the N receivers by $\mathscr{D}_1, \ldots, \mathscr{D}_N$. We will use this index code to build a valid code over the channel alphabet $\mathcal{Z}'$. To do so, we choose positive integers t and s such that

$$\frac{\log_2 |\mathcal{Z}|}{\log_2 |\mathcal{Z}'|} \leq \frac{t}{s} \leq (1 + \epsilon) \frac{\log_2 |\mathcal{Z}|}{\log_2 |\mathcal{Z}'|}. \tag{27}$$

We consider s generations of messages and encode each of them, independently, using $\mathscr{C}$. For $j \in [s]$, we encode the jth generation $\boldsymbol{x}_1^{(j)}, \ldots, \boldsymbol{x}_N^{(j)} \in \mathcal{A}^M$ using $\mathscr{C}$ to obtain the

codeword $\boldsymbol{c}^{(j)} = (c_1^{(j)}, \ldots, c_\ell^{(j)}) \in \mathcal{Z}^\ell$. Note that the number of symbols encoded per message, considering all generations, is $M' = sM$. Now, to design an index code $\mathscr{C}'$ for the channel alphabet $\mathcal{Z}'$, we map the vectors $\boldsymbol{c}^{(1)}, \ldots, \boldsymbol{c}^{(s)}$ to a vector of length $\ell' = t\ell$ over $\mathcal{Z}'$ as follows. For each $k \in [\ell]$, we use any one-to-one map from $\mathcal{Z}^s$ to $\mathcal{Z}'^t$ to map the vector $(c_k^{(1)}, \ldots, c_k^{(s)})$ to $(d_k^{(1)}, \ldots, d_k^{(t)})$. This is possible since $|\mathcal{Z}|^s \leq |\mathcal{Z}'|^t$, see (27). The codeword transmitted for the index code $\mathscr{C}'$ is the vector $(d_k^{(j)}, k \in [\ell], j \in [t])$.

To decode the ith message, receiver Rx_i observes the symbols $(d_k^{(j)}, k \in R_i, j \in [t])$. The number of transmissions queried by Rx_i is $|R_i'| = t|R_i|$. For each $k \in R_i$, the receiver uses $(d_k^{(1)}, \ldots, d_k^{(t)})$ to retrieve $(c_k^{(1)}, \ldots, c_k^{(s)})$. Thus, Rx_i can compute $(c_k^{(j)}, k \in R_i)$, for every $j \in [s]$. For each $j \in [s]$, the vector $(c_k^{(j)}, k \in R_i)$ can be used with the side information available at Rx_i to decode $\boldsymbol{x}_i^{(j)}$ using the decoder $\mathscr{D}_i$ of the original index code $\mathscr{C}$. This shows that $\mathscr{C}'$ is a valid index code with channel alphabet $\mathcal{Z}'$.

Using (27) we observe that the rate β' of $\mathscr{C}'$ satisfies

$$\beta' = \frac{\ell' \log_2 |\mathcal{Z}'|}{M' \log_2 |\mathcal{A}|} = \frac{t\ell \log_2 |\mathcal{Z}'|}{sM \log_2 |\mathcal{A}|} \leq (1+\epsilon) \frac{\ell \log_2 |\mathcal{Z}|}{M \log_2 |\mathcal{A}|}$$
$$= (1+\epsilon)\beta.$$

Similarly, the locality r_i' of the ith receiver for the index code $\mathscr{C}'$ is

$$r_i' = \frac{|R_i'| \log_2 |\mathcal{Z}'|}{M' \log_2 |\mathcal{A}|} = \frac{t|R_i| \log_2 |\mathcal{Z}'|}{sM \log_2 |\mathcal{A}|} \leq (1+\epsilon) \frac{|R_i| \log_2 |\mathcal{Z}|}{M \log_2 |\mathcal{A}|}$$
$$= (1+\epsilon)r_i,$$

where r_i is the receiver locality of Rx_i for the code $\mathscr{C}$. Hence, the overall locality of $\mathscr{C}'$ is at the most $(1+\epsilon)r$. ∎

Considering vanishingly small values of ϵ, and choosing β arbitrarily close to $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$, Lemma 15 implies that $\beta_G^*(r, \mathcal{A}, \mathcal{Z}') \leq \beta_G^*(r, \mathcal{A}, \mathcal{Z})$.

C. Invariance With Respect to Message Alphabet

We will use arguments similar to Appendix I-B to show that $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ is independent of $\mathcal{A}$. In particular, we show that for any choice of finite sets $\mathcal{A}$ and $\mathcal{A}'$, with $|\mathcal{A}|, |\mathcal{A}'| \geq 2$, $\beta_G^*(r, \mathcal{A}', \mathcal{Z}) \leq \beta_G^*(r, \mathcal{A}, \mathcal{Z})$. From the symmetry of this result, it follows that the inequality holds in the other direction as well, implying equality.

Lemma 16: Suppose there exists a valid index code for G with message alphabet $\mathcal{A}$ and channel alphabet $\mathcal{Z}$ of rate β and locality r, with $|\mathcal{A}|, |\mathcal{Z}| \geq 2$. Let $\mathcal{A}'$ be any finite set with size at least two. For any $\epsilon > 0$, there exists a valid index code for G with message alphabet $\mathcal{A}'$, channel alphabet $\mathcal{Z}$, with rate at the most $(1+\epsilon)\beta$ and locality at the most $(1+\epsilon)r$.

Proof: Let $\mathscr{C}$ denote a valid index code with codelength ℓ, message length M, receiver localities $r_1, \ldots, r_N$, and receiver queries $R_1, \ldots, R_N$. Note that its rate is $\beta = \ell/M$ and locality of Rx_i is $r_i = |R_i|/M$. We now choose positive integers M' and s such that

$$\frac{\log_2 |\mathcal{A}'|}{M \log_2 |\mathcal{A}|} \leq \frac{s}{M'} \leq (1+\epsilon) \frac{\log_2 |\mathcal{A}'|}{M \log_2 |\mathcal{A}|} \quad (28)$$

We use $\mathscr{C}$ to design a valid index code $\mathscr{C}'$ for message alphabet $\mathcal{A}'$ with message length M'. For $\mathscr{C}'$, let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N \in \mathcal{A}'^{M'}$ denote the N messages. From (28), an injective function exists from $\mathcal{A}'^{M'}$ to $\mathcal{A}^{sM}$. For each $i \in [N]$, we use such a function to map $\boldsymbol{y}_i$ to a vector $(\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(s)}) \in \mathcal{A}^{sM}$, where $\boldsymbol{x}_i^{(j)} \in \mathcal{A}^M$ for all $j \in [s]$. With this, we represent a message $\boldsymbol{y}_i$ over $\mathcal{A}'$ using s generations of M-length vectors $\boldsymbol{x}_i^{(j)}$ over $\mathcal{A}$. The side information at Rx_i is $\boldsymbol{y}_n, n \in K_i$, which is equivalent to $\boldsymbol{x}_n^{(j)}, n \in K_i$ and $j \in [s]$.

We transmit the s generations of messages independently using the index code $\mathscr{C}$. The resulting index code $\mathscr{C}'$ has codelength $\ell' = s\ell$, and message length M' over the message alphabet $\mathcal{A}'$. The rate of this code is

$$\beta' = \frac{\ell' \log_2 |\mathcal{Z}|}{M' \log_2 |\mathcal{A}'|} = \frac{s\ell \log_2 |\mathcal{Z}|}{M' \log_2 |\mathcal{A}'|} \leq (1+\epsilon) \frac{\ell \log_2 |\mathcal{Z}|}{M \log_2 |\mathcal{A}|}$$
$$= (1+\epsilon)\beta,$$

where we have used (28). The number of codeword symbols observed by Rx_i for the code $\mathscr{C}'$ is $s|R_i|$ since s instances of $\mathscr{C}$ are used to encode $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$. Thus the locality r_i' of Rx_i under the code $\mathscr{C}'$ is

$$r_i' = \frac{s|R_i| \log_2 |\mathcal{Z}|}{M' \log_2 |\mathcal{A}'|} \leq (1+\epsilon) \frac{|R_i| \log_2 |\mathcal{Z}|}{M \log_2 |\mathcal{A}|} = (1+\epsilon)r_i.$$

We conclude that $\mathscr{C}'$ has rate at the most $(1+\epsilon)\beta$ and locality at the most $(1+\epsilon)r$. ∎

If we choose ϵ and β arbitrarily close to 0 and $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$, respectively, we observe from Lemma 16 that $\beta_G^*(r, \mathcal{A}', \mathcal{Z}) \leq \beta_G^*(r, \mathcal{A}, \mathcal{Z})$.

D. Summary

From the results of this appendix we observe that the value of $\beta_G^*(r, \mathcal{A}, \mathcal{Z})$ is independent of $\mathcal{A}$ and $\mathcal{Z}$. Hence, for any choice of alphabet $\mathcal{A}$

$$\beta_G^*(r, \mathcal{A}, \mathcal{A}) = \inf_{\mathcal{A}', \mathcal{Z}'} \beta_G^*(r, \mathcal{A}', \mathcal{Z}'),$$

i.e., $\beta_G^*(r, \mathcal{A}, \mathcal{A})$ captures the ultimate locality-rate trade-off even when optimization over message and channel alphabets is allowed.

APPENDIX II
PROOF OF LEMMA 2

Assume $r_1, r_2 \geq 1$ and let $\epsilon > 0$. For each $j = 1, 2$, there exists an index code with broadcast rate $\beta_j \leq \beta_G^*(r_j) + \epsilon$ and locality at the most r_j. We will denote the blocklength of this code by ℓ_j, message length by M_j and the subsets of the indices used by the N receivers as $R_{1,j}, \ldots, R_{N,j}$, where $j = 1, 2$. For some choice of non-negative integers k_1 and k_2, consider a time-sharing scheme where the first index code is used $k_1 M_2$ times, and the second index code is used $k_2 M_1$ times. For this composite scheme, the overall message length is $M = k_1 M_2 M_1 + k_2 M_1 M_2 = M_1 M_2(k_1 + k_2)$. The blocklength is $\ell = k_1 M_2 \ell_1 + k_2 M_1 \ell_2$, and the number of codeword symbols utilized by the ith receiver to decode

its desired message is $|R_i| = k_1 M_2 |R_{i,1}| + k_2 M_1 |R_{i,2}|$. The locality of this time-sharing scheme can be upper bounded as

$$\max_i \frac{|R_i|}{M} = \max_i \frac{k_1 |R_{i,1}|}{(k_1 + k_2) M_1} + \frac{k_2 |R_{i,2}|}{(k_1 + k_2) M_2}$$
$$\leq \frac{k_1}{k_1 + k_2} r_1 + \frac{k_2}{k_1 + k_2} r_2.$$

Similarly, the broadcast rate β of this time-sharing scheme can be shown to be equal to $k_1 \beta_1 / (k_1 + k_2) + k_2 \beta_2 / (k_1 + k_2)$, which is upper bounded as

$$\beta = \frac{k_1}{k_1 + k_2} \beta_1 + \frac{k_2}{k_1 + k_2} \beta_2$$
$$\leq \frac{k_1}{k_1 + k_2} \beta_G^*(r_1) + \frac{k_2}{k_1 + k_2} \beta_G^*(r_2) + \epsilon.$$

Denoting $k_1 r_1 / (k_1 + k_2) + k_2 r_2 / (k_1 + k_2)$ by r, and by letting $\epsilon \to 0$, we observe that

$$\beta_G^*(r) \leq \frac{k_1}{k_1 + k_2} \beta_G^*(r_1) + \frac{k_2}{k_1 + k_2} \beta_G^*(r_2).$$

Convexity follows by approximating any real number in the interval $(0, 1)$ by a rational number $k_1 / (k_1 + k_2)$ to any desired accuracy by using sufficiently large k_1 and k_2.

APPENDIX III
PROOF OF THEOREM 3

We will design a new encoding matrix $\boldsymbol{L}'$ by modifying the subset of the columns of the given matrix $\boldsymbol{L}$ corresponding to the column indices $\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$. For the remaining indices $k \in \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_N$, the kth columns of $\boldsymbol{L}$ and $\boldsymbol{L}'$ are equal, i.e., $\boldsymbol{L}_k = \boldsymbol{L}'_k$. For an arbitrary $i \in [N]$, we will now explain the construction of the column vectors $\boldsymbol{L}'_k$, $k \in \mathcal{S}_i$. Since the symbols $\boldsymbol{x}^{\mathsf{T}} \boldsymbol{L}'_k$, $k \in \mathcal{S}_i$, are queried only by Rx_i and are unused by other receivers, we only need to consider the constraints that are imposed by the demands of Rx_i while designing the column vectors $\boldsymbol{L}'_k$, $k \in \mathcal{S}_i$.

We will introduce the notation which will be used in the rest of the proof. For any $E \subseteq [MN]$, let $U_E = \mathsf{span}(\boldsymbol{e}_k, k \in E)$, i.e., U_E is the subspace of all vectors whose support is a subset of E. For any $F \subseteq [\ell]$, let $V_F = \mathsf{span}(\boldsymbol{L}_k, k \in F)$ and $V'_F = \mathsf{span}(\boldsymbol{L}'_k, k \in F)$. Since $R_i = \mathcal{S}_i \cup \mathcal{M}_i$, we have $V_{R_i} = V_{\mathcal{S}_i} + V_{\mathcal{M}_i}$ and $V'_{R_i} = V'_{\mathcal{S}_i} + V'_{\mathcal{M}_i}$, where the addition corresponds to sum of subspaces. From Theorem 2 and using the fact that $\boldsymbol{L}$ is a valid encoder matrix, we have $\boldsymbol{e}_j \in V_{R_i} + U_{\mathcal{K}_i}$, for all $j \in \mathcal{D}_i$, i.e., we have

$$U_{\mathcal{D}_i} = \mathsf{span}(\boldsymbol{e}_j, j \in \mathcal{D}_i) \subseteq V_{R_i} + U_{\mathcal{K}_i} = V_{\mathcal{S}_i} + V_{\mathcal{M}_i} + U_{\mathcal{K}_i}. \tag{29}$$

Again using Theorem 2, we observe that $\boldsymbol{L}'$ allows Rx_i to decode its demand if and only if

$$U_{\mathcal{D}_i} \subseteq V'_{R_i} + U_{\mathcal{K}_i} = V'_{\mathcal{S}_i} + V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}. \tag{30}$$

Using the validity of the encoder matrix $\boldsymbol{L}$, we will first lower bound $|\mathcal{S}_i|$ which is the number of coded symbols queried uniquely by Rx_i. From (29), we obtain

$$U_{\mathcal{D}_i} = (V_{\mathcal{S}_i} + V_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}. \tag{31}$$

Let $V_{\mathsf{in}} = (V_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}$ denote the subspace of $V_{\mathcal{M}_i} + U_{\mathcal{K}_i}$ contained in $U_{\mathcal{D}_i}$, and let V_{out} be any subspace such that $V_{\mathsf{out}} \cap U_{\mathcal{D}_i} = \{\boldsymbol{0}\}$ and $V_{\mathsf{in}} + V_{\mathsf{out}} = V_{\mathcal{M}_i} + U_{\mathcal{K}_i}$. Continuing from (31), we claim that

$$U_{\mathcal{D}_i} = (V_{\mathcal{S}_i} + V_{\mathsf{in}} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i} \tag{32}$$
$$= ((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}) + V_{\mathsf{in}}. \tag{33}$$

It is clear that the subspace in (32) contains the subspace in (33) since $V_{\mathsf{in}} \subseteq U_{\mathcal{D}_i}$. To prove that (32) is contained in (33), assume that $\boldsymbol{y} \in V_{\mathcal{S}_i} + V_{\mathsf{out}}$ and $\boldsymbol{z} \in V_{\mathsf{in}}$ are such that $\boldsymbol{y} + \boldsymbol{z} \in U_{\mathcal{D}_i}$. Since $\boldsymbol{z} \in V_{\mathsf{in}} \subseteq U_{\mathcal{D}_i}$ and $\boldsymbol{y} + \boldsymbol{z} \in U_{\mathcal{D}_i}$, we conclude that $\boldsymbol{y} \in U_{\mathcal{D}_i}$ as well. Thus $\boldsymbol{y} \in (V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}$, and hence, $\boldsymbol{y} + \boldsymbol{z} \in ((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}) + V_{\mathsf{in}}$.

Considering the dimensions of the subspaces in (33), we have

$$\dim(U_{\mathcal{D}_i}) \leq \dim((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}) + \dim(V_{\mathsf{in}}). \tag{34}$$

In order to proceed with the proof of the theorem, we will now show that $\dim((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}) \leq |\mathcal{S}_i|$. To do so, assume that $\boldsymbol{y}_j + \boldsymbol{z}_j$, $j = 1, \ldots, n$, form a basis for $(V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}$ where $\boldsymbol{y}_j \in V_{\mathcal{S}_i}$ and $\boldsymbol{z}_j \in V_{\mathsf{out}}$ and $n = \dim((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i})$. If the scalars $\alpha_1, \ldots, \alpha_n$ are such that $\sum_j \alpha_j \boldsymbol{y}_j = \boldsymbol{0}$, then

$$\sum_{j=1}^{n} \alpha_j (\boldsymbol{y}_j + \boldsymbol{z}_j) = \sum_{j=1}^{n} \alpha_j \boldsymbol{z}_j \in V_{\mathsf{out}}.$$

Since $\boldsymbol{y}_j + \boldsymbol{z}_j \in U_{\mathcal{D}_i}$, we also observe that $\sum_j \alpha_j (\boldsymbol{y}_j + \boldsymbol{z}_j) \in U_{\mathcal{D}_i}$. Using the fact $V_{\mathsf{out}} \cap U_{\mathcal{D}_i} = \{\boldsymbol{0}\}$, we deduce that $\sum_j \alpha_j (\boldsymbol{y}_j + \boldsymbol{z}_j) = \boldsymbol{0}$, and hence, $\alpha_1 = \cdots = \alpha_n = 0$. We conclude that $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \in V_{\mathcal{S}_i}$ are linearly independent, and therefore

$$\dim((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i}) = n \leq \dim(V_{\mathcal{S}_i}) \leq |\mathcal{S}_i|. \tag{35}$$

Note that the columns of $\boldsymbol{L}$ and $\boldsymbol{L}'$ corresponding to the column indices $\mathcal{M}_i$ are equal, and hence $V'_{\mathcal{M}_i} = V_{\mathcal{M}_i}$. Using this fact together with (34) and (35), we have

$$|\mathcal{S}_i| \geq \dim((V_{\mathcal{S}_i} + V_{\mathsf{out}}) \cap U_{\mathcal{D}_i})$$
$$\geq \dim(U_{\mathcal{D}_i}) - \dim(V_{\mathsf{in}})$$
$$= \dim(U_{\mathcal{D}_i}) - \dim((V_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i})$$
$$= \dim(U_{\mathcal{D}_i}) - \dim((V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}).$$

From (30), in order to satisfy the claim of this theorem, it is sufficient to chose the $|\mathcal{S}_i|$ vectors $\boldsymbol{L}'_k$, $k \in \mathcal{S}_i$, such that $\boldsymbol{L}'_k \lhd \mathcal{D}_i$, i.e., $\boldsymbol{L}'_k \in U_{\mathcal{D}_i}$ and

$$V'_{\mathcal{S}_i} + ((V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}) \supseteq U_{\mathcal{D}_i}.$$

This is always possible since the difference in the dimensions of $U_{\mathcal{D}_i}$ and $(V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}$ is at the most $|\mathcal{S}_i|$. One way to construct $\boldsymbol{L}'_k$, $k \in \mathcal{S}_i$, is as follows. We begin with a basis for $(V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i}$. These vectors form a linearly independent set in $U_{\mathcal{D}_i}$. We extend this set to a basis for $U_{\mathcal{D}_i}$. The number of additional vectors in this basis is $\dim(U_{\mathcal{D}_i}) - \dim((V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i})$. These additional vectors together with $|\mathcal{S}_i| - \dim(U_{\mathcal{D}_i}) + \dim((V'_{\mathcal{M}_i} + U_{\mathcal{K}_i}) \cap U_{\mathcal{D}_i})$ all-zero vectors are chosen as the columns $\boldsymbol{L}'_k$, $k \in \mathcal{S}_i$.

APPENDIX IV
CONSTRUCTION OF THE BIPARTITE GRAPH $\mathcal{B}$

We require the following lemma to prove the correctness of the proposed construction. Note that $r_1 = |R_1|, \ldots, r_N = |R_N|$ are the degrees of the vertices $u_1, \ldots, u_N$, respectively, in $\mathcal{B}$, and are the localities of the N receivers in the index coding problem.

Lemma 17: For any $i \in [N]$, $r_1 + \cdots + r_i \leq 2i - 1$.

Proof: Let $r_1 + \cdots + r_i = y$. The average of the first i entries of $\boldsymbol{r}$ is y/i. Since the components of $\boldsymbol{r}$ are in the ascending order, we have $r_{i+1}, r_{i+2}, \ldots, r_N \geq y/i$. Using this in

$$y + r_{i+1} + r_{i+2} + \cdots + r_N = r_1 + \cdots + r_N = 2(N-1),$$

we have $y + (N-i)y/i \leq 2(N-1)$. This implies $y \leq 2(N-1)i/N < 2i$, and since y is an integer, we deduce that $y \leq 2i - 1$. ∎

Proof of Lemma 13

The construction of $\mathcal{B}$ starts with the vertex sets $\{u_1, \ldots, u_N\}$ and $\{v_1, \ldots, v_{N-1}\}$, and an empty set of edges. Edges are appended sequentially by considering the vertices $u_1, \ldots, u_N$, in that order. At step i, we determine all the neighbors of u_i, while making sure that the degrees of $v_1, \ldots, v_{N-1}$ do not exceed 2. Using $i = 1$ in Lemma 17, we deduce that r_1 is necessarily equal to 1. We append $\{u_1, v_1\}$ to the edge set, and this completes the first step. For $2 \leq i \leq N-1$, at step i, we append $\{u_i, v_i\}$ to the edge set of $\mathcal{B}$, and identify $r_i - 1$ vertices among $\{v_1, \ldots, v_{i-1}\}$ with degrees equal to 1, and include edges between u_i and these $r_i - 1$ vertices. We will shortly show that such a set of $r_i - 1$ vertices indeed exist. This procedure ensures that degrees of $v_1, \ldots, v_{N-1}$ do not exceed 2, and for any $i \leq N-1$, we have $i \in R_i$ and $R_i \subseteq [i]$. Finally, at step N, we append edges between u_N and all the vertices among $\{v_1, \ldots, v_{N-1}\}$ that have degree 1. Since $\sum_{i \in [N]} r_i = 2(N-1)$, at the end of step N, it is guaranteed that the degrees of $u_1, \ldots, u_N$ are $r_1, \ldots, r_N$, respectively, and that the degrees of $v_1, \ldots, v_{N-1}$ are all equal to 2.

It only remains to show that at step i, for $2 \leq i \leq N-1$, there exist $r_i - 1$ vertices among $v_1, \ldots, v_{i-1}$ with degree equal to 1. We prove this by induction. Let us first consider $i = 2$. We know from Lemma 17 that $r_1 = 1$ and $r_1 + r_2 \leq 3$, i.e, $r_2 \leq 2$ and $r_2 - 1 \leq 1$. At the end of step 1, v_1 has degree 1, which is because of the edge $\{u_1, v_1\}$. Since at the most 1 vertex with degree 1 is required, step 2 can be completed. Now consider $3 \leq i \leq N-1$, and assume that steps $1, \ldots, (i-1)$ have been completed. Now consider the end of step $(i-1)$. At this point, the edges $\{u_1, v_1\}, \ldots, \{u_{i-1}, v_{i-1}\}$ are present in $\mathcal{B}$. Thus the degrees of each of $v_1, \ldots, v_{i-1}$ is either 1 or 2. Let the number of vertices with degree 1 among $v_1, \ldots, v_{i-1}$ at the end of step $(i-1)$ be z. Thus the sum of the degrees of $v_1, \ldots, v_{i-1}$ is $z + 2(i-1-z) = 2i - z - 2$. Since the neighbors of each of $v_1, \ldots, v_{i-1}$ are subsets of $u_1, \ldots, u_{i-1}$ at the end of step $(i-1)$, and since the degrees of $u_1, \ldots, u_{i-1}$ are $r_1, \ldots, r_{i-1}$ at this stage, we have $2i - z - 2 = r_1 + \cdots + r_{i-1}$. Using Lemma 17, we have $2i - 1 \geq r_1 + \cdots + r_i = 2i - z - 2 + r_i$

which implies $z \geq r_i - 1$. This proves that step i can be completed.

APPENDIX V
PROOF OF LEMMA 8

Since σ is an automorphism of G, so is σ^n for any $n \in [N]$. Note that the group $\{\sigma, \sigma^2, \ldots, \sigma^N = 1\}$ acts transitively on the vertex set of G. Let $(\mathfrak{E}, \mathfrak{D}_1, \ldots, \mathfrak{D}_N)$ be an index code for G with rate β, message length M, receiver localities $r_1, \ldots, r_N$, and receiver queries $R_1, \ldots, R_N$. Note that $r \geq r_{\mathrm{avg}} = \sum_{i \in [N]} r_i / N$. We will consider N index coding schemes $(\mathfrak{E}^{(n)}, \mathfrak{D}_1^{(n)}, \ldots, \mathfrak{D}_N^{(n)})$, $n \in [N]$, each of which is derived from $(\mathfrak{E}, \mathfrak{D}_1, \ldots, \mathfrak{D}_N)$ by permuting the roles of the messages $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$. Specifically, the nth encoder $\mathfrak{E}^{(n)}$ is the encoder $\mathfrak{E}$ applied to the nth left cyclic shift of the message tuple $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, i.e.,

$$\mathfrak{E}^{(n)}\left((\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)\right) = \mathfrak{E}((\boldsymbol{x}_{n+1}, \boldsymbol{x}_{n+2}, \ldots, \boldsymbol{x}_N, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n))$$
$$= \mathfrak{E}\left((\boldsymbol{x}_{\sigma^n(1)}, \ldots, \boldsymbol{x}_{\sigma^n(N)})\right).$$

For any $i \in [N]$, in the above expression of $\mathfrak{E}^{(n)}$, the message $\boldsymbol{x}_i$ is the $(i-n)_N^{\mathrm{th}}$ argument of $\mathfrak{E}$ where $(i-n)_N = (i-n)$ if $(i-n) \geq 1$ and $(i-n)_N = i - n + N$ otherwise. Hence, the encoding function $\mathfrak{E}^{(n)}$ operates on the message $\boldsymbol{x}_i$ in the same manner as $\mathfrak{E}$ operates on the message $\boldsymbol{x}_{(i-n)_N}$. Using the fact that σ is an automorphism of G, it is easy to see that, when the nth code is used, the ith receiver can decode $\boldsymbol{x}_i$ as $\mathfrak{D}_{(i-n)_N}(\boldsymbol{c}_{R_{(i-n)_N}}, \boldsymbol{x}_{K_i})$. Thus, the number of codeword symbols queried by the ith receiver in the nth code is $|R_{(i-n)_N}| = M r_{(i-n)_N}$, where M is the message length.

Now consider a time sharing scheme that utilizes each of the N encoders $\mathfrak{E}^{(1)}, \ldots, \mathfrak{E}^{(N)}$ exactly once. The overall message length for this scheme is MN, the broadcast rate is β, and the number of codeword symbols queried by the ith receiver is

$$|R_i'| = \sum_{n \in [N]} |R_{(i-n)_N}| = M \sum_{n \in [N]} r_n = MN r_{\mathrm{avg}},$$

which is independent of i. Also, the overall locality of this time sharing scheme is $r' = \max_i |R_i'|/MN = r_{\mathrm{avg}} \leq r$. To complete the proof observe that for any $i \in [N]$ and $j = i \bmod N + 1$, we have

$$|R_i' \cap R_j'| = \sum_{n \in [N]} |R_{(i-n)_N} \cap R_{(i-n+1)_N}|$$
$$= \sum_{n \in [N]} |R_n \cap R_{(n+1)_N}|,$$

which is independent of i.

ACKNOWLEDGMENT

References

[1] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," in *Proc. IEEE INFOCOM Conf. Comput. Commun. 17th Annu. Joint Conf. IEEE Comput. Commun. Soc. Gateway 21st Century*, vol. 3, Mar. 1998, pp. 1257–1264.

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

[3] S. El Rouayheb, A. Sprintson, and C. Georghiades, "On the index coding problem and its relation to network coding and matroid theory," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3187–3195, Jul. 2010.

[4] 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.

[5] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

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

[7] K. Shanmugam and A. G. Dimakis, "Bounding multiple unicasts through index coding and locally repairable codes," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 296–300.

[8] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos, and K. Ramchandran, "Speeding up distributed machine learning using codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1514–1529, Mar. 2018.

[9] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *Proc. 49th IEEE Symp. Found. Comput. Sci. (FOCS)*, Oct. 2008, pp. 823–832.

[10] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Index coding via linear programming," 2010, *arXiv:1004.1379*. [Online]. Available: http://arxiv.org/abs/1004.1379

[11] M. A. R. Chaudhry, Z. Asad, A. Sprintson, and M. Langberg, "On the complementary index coding problem," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2011, pp. 244–248.

[12] A. S. Tehrani, A. G. Dimakis, and M. J. Neely, "Bipartite index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2246–2250.

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

[14] M. J. Neely, A. S. Tehrani, and Z. Zhang, "Dynamic index coding for wireless broadcast networks," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7525–7540, Nov. 2013.

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

[16] K. Shanmugam, A. G. Dimakis, and M. Langberg, "Graph theory versus minimum rank for index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2014, pp. 291–295.

[17] F. Arbabjolfaei and Y.-H. Kim, "Local time sharing for index coding," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 286–290.

[18] A. Agarwal and A. Mazumdar, "Local partial clique and cycle covers for index coding," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.

[19] I. Haviv and M. Langberg, "On linear index coding for random graphs," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 2231–2235.

[20] M. Karmoose, L. Song, M. Cardone, and C. Fragouli, "Privacy in index coding: k-limited-access schemes," *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 2625–2641, May 2020.

[21] A. Thomas, K. Radhakumar, C. Attada, and B. S. Rajan, "Single uniprior index coding with min–max probability of error over fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6050–6059, Jul. 2017.

[22] M. B. Vaddi and B. S. Rajan, "Reduced complexity index codes and improved upperbound on broadcast rate for neighboring interference problems," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 1247–1251.

[23] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, vol. 207. Cham, Switzerland: Springer, 2013.

[24] S. H. Dau, V. Skachek, and Y. M. Chee, "On the security of index coding with side information," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3975–3988, Jun. 2012.

[25] H. Maleki, V. R. Cadambe, and S. A. Jafar, "Index coding—An interference alignment perspective," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5402–5432, Sep. 2014.

[26] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, Jan. 2014.

[27] L. Ong, "A new class of index coding instances where linear coding is optimal," in *Proc. Int. Symp. Netw. Coding (NetCod)*, Jun. 2014, pp. 1–6.

Lakshmi Prasad Natarajan (Member, IEEE) received the B.E. degree in electronics and communication from the College of Engineering, Guindy, in 2008, and the Ph.D. degree from the Indian Institute of Science, Bengaluru, in 2013. From 2014 to 2016, he held a post-doctoral position at the Department of Electrical and Computer Systems Engineering, Monash University, Australia. He is currently an Assistant Professor with the Department of Electrical Engineering, IIT Hyderabad. His primary research interests are coding techniques and information theory for communication systems.

Prasad Krishnan received the B.E degree from the College of Engineering, Anna University, Guindy, in 2007, and the Ph.D. degree from the Indian Institute of Science, Bengaluru, in 2014. Since 2014, he has been a Faculty Member with the International Institute of Information Technology Hyderabad (IIIT-H). His research interests include applications of coding theory in network coding, index coding, codes for distributed storage, and multi-user communication.

V. Lalitha (Member, IEEE) received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, in 2003, and the M.E. and Ph.D. degrees from the Indian Institute of Science (IISc), Bengaluru, in 2005 and 2015, respectively. She worked as an Engineer at Qualcomm, Hyderabad, in 2006, and a Design Engineer at Conexant Systems, Noida, from 2006 to 2008. She worked as a Research Intern at Microsoft Research, Bengaluru. Since May 2015, she has been working as an Assistant Professor at the International Institute of Information Technology Hyderabad (IIIT Hyderabad), where she is currently with the Signal Processing and Communications Research Center. Her research interests include coding for distributed storage and computing, index coding, polar codes, and learning-based codes. She was a recipient of Prof. I.S.N. Murthy Medal from IISc in 2005 and the TCS Research Scholarship in 2011. Her paper received the runner-up best paper award at NCC 2019. She is the present Chapter Chair of the Joint Chapter of Communications and Signal Processing Societies, IEEE Hyderabad Section.

Hoang Dau received the B.S. degree in applied mathematics and informatics from Vietnam National University, Hanoi, Vietnam, in 2006, and the M.S. and Ph.D. degrees in mathematics from Nanyang Technological University, Singapore, in 2009 and 2012, respectively. He is currently a Lecturer with the Discipline of Computer Science and Information Technology, School of Science, RMIT University, Australia. His research interests include coding theory, combinatorics, and blockchains.