

Minrank of Embedded Index Coding Problems and its Relation to Connectedness of a Bipartite Graph

Anjana Ambika Mahesh, and B. Sundar Rajan

Department of Electrical Communication Engineering, Indian Institute of Science, Bengaluru 560012, KA, India

E-mail: {anjanamahesh,bsrajan}@iisc.ac.in

Abstract—This paper deals with embedded index coding problem (EICP), introduced by A. Potter and M. Wootters, which is a decentralized communication problem among users with side information. An alternate definition of the parameter minrank of an EICP, which has reduced computational complexity compared to the existing definition, is presented. A graphical representation for an EICP is given using directed bipartite graphs, called bipartite problem graph, and the side information alone is represented using an undirected bipartite graph called the side information bipartite graph. Inspired by the well-studied single unicast index coding problem, graphical structures, similar to cycles and cliques, are identified in the side information bipartite graph of a single unicast embedded index coding problem (SUEICP). Transmission schemes based on these graphical structures, called tree cover scheme and bi-clique cover scheme are also presented. For a class of SUEICPs, scalar linear optimal solution is given using bi-clique cover. A relation between connectedness of the side information bipartite graph and the number of transmissions required in a scalar linear solution of an EICP is established.

Index Terms—Embedded index coding, graphical representation, minrank, covering schemes.

I. INTRODUCTION

Motivated by applications in device to device multicast and distributed computing, embedded index coding problem (EICP) was introduced by Potter and Wootters in [1], which is a decentralized version of the well-studied index coding problem (ICP) [2]. In an EICP, the user nodes themselves act as both senders and receivers and hence the only messages which can be requested by a user are those that are present with at least one other user. The solution to an EICP is called an embedded index code.

Minrank of an EICP was defined in [3] using a matrix representation of the EICP, called the side information matrix, and was shown to be equal to the length of an optimal scalar linear embedded index code. In this paper, we give an alternate definition of minrank of an EICP, consistent with the definition in [3] but along the lines of the definition of minrank of a general ICP given in [4], which has a reduced computational complexity than that required by the approach in [3].

A graphical representation for an EICP was given in [1] using a directed graph called problem graph which is merely an alternate representation of the general ICP. In this paper, we represent an EICP using directed bipartite graphs, as in [5], called the bipartite problem graph, which is a more intuitive and insightful representation as explained in [6]. With a given side information at the user nodes, there exist multiple EICPs corresponding to different demand vectors. The side

information at the users is represented using an undirected bipartite graph called the side information bipartite graph.

For an SUEICP [7], graphical structures like cycles and cliques were identified to give advantage in reducing the number of transmissions required and hence simple transmission schemes which makes use of these graphical structures [2], [8], [9] were also presented. We identify similar graphical structures in the side information bipartite graph of single unicast EICPs which will help in reducing the number of transmissions required and also propose transmission schemes utilizing these graph structures. For a class of SUEICPs, we construct optimal scalar linear embedded index codes using one such graph structure.

We also establish a relation between the connectedness of a side-information bipartite graph and the minrank of the EICPs with the same side information as that represented by the graph. This has been motivated by the application of EIC in the V2V phase of collaborative message dissemination protocol of Vehicular Adhoc Networks (VANETs) [10], [11]. VANETs are formed by vehicles that are within hundred meters of each other moving at low speeds in the same direction. The side information at these vehicles is obtained from an earlier R2V communication phase where a road-side unit transmits to vehicles in its range. Since the vehicles are close-by and moving very slowly, the possibility of them receiving a lot of packets in common is quite high and hence the side information bipartite graph will be heavily connected.

The technical contributions in this paper are as follows.

- We give an alternate definition for minrank of an EICP which is computationally more efficient and prove that the length of an optimal linear solution to an EICP is equal to the minrank.
- A graphical representation of an EICP is given using directed bipartite graphs which we call as the bipartite problem graph. The side information alone is represented using an undirected bipartite graph, called the side information bipartite graph.
- Two transmission schemes, called tree cover scheme and bi-clique cover scheme, are introduced for an SUEICP, which make use of certain graphical structures called regular trees and bi-cliques.
- For a class of SUEICPs, optimal scalar linear embedded index codes are constructed based on bi-clique covers.
- A theorem which establishes a relation between the connectedness of the side information bipartite graph and

the minrank of the corresponding EICPs is presented.

The rest of this paper is organized as follows. Embedded index coding problem is formally set up in section II. This is followed by a definition of minrank of an EICP and a relation between minrank of an EICP and the length of an optimal linear solution to the EICP in section III. A graphical representation for an EICP is given, and graph structures called regular trees and bi-cliques identified in the side information bipartite graph for an SUEICP and transmission schemes utilizing these graph structures are presented in section IV. A theorem which establishes a relation between the connectedness of a side information bipartite graph and the minrank of the EICPS which have their side information as that represented by the side information bipartite graph is given in section VI. Finally the paper is concluded in section VII by identifying directions for future research.

Notations: For a prime power q , \mathbb{F}_q denotes the finite field with q elements. For a positive integer n , $[n]$ denotes the set $\{1, 2, \dots, n\}$. A t -subset of $[n]$ is a subset of $[n]$ of size t . The n -length vector \mathbf{e}_i is the i^{th} standard basis vector for \mathbb{F}_q^n . For an n -length vector \mathbf{x} and a set $S \subset [n]$, $\mathbf{x} \triangleleft S$ indicates that the vector \mathbf{x} has its support in S . The transpose of a vector \mathbf{v} is denoted as \mathbf{v}^T and that of a matrix A is denoted as A^T .

II. PROBLEM SETUP

Consider an EICP with N users, $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ and a set of M messages $\mathcal{X} = \{x_1, x_2, \dots, x_M\}$, $x_i \in \mathbb{F}_q$, with user u_i demanding a subset of messages $\mathcal{W}_i \subset \mathcal{X}$ and possessing a non-intersecting subset of messages, indexed by an ordered set \mathcal{K}_i , as side-information. The goal of the EICP is to satisfy the message requests of all the users with minimum number of transmissions by the users themselves. There is no central server which possesses all the messages in \mathcal{X} and hence the side-information possessed by the users is such that $\bigcup_{i=1}^K \mathcal{X}_{\mathcal{K}_i} = \mathcal{X}$, i.e., every message is present with at least one user. No user possesses all the messages, i.e., $\mathcal{X}_{\mathcal{K}_i} \subsetneq \mathcal{X}$, $\forall i$, since then that user can act as a central server and the EICP reduces to the centralized ICP and any solution of the centralized ICP can be transmitted by this user which possesses all the messages. Further, it is also assumed that no message is available at all users as then that message won't be demanded by any user and can as well be removed from the system.

Since a user demanding k messages can be split into k users each demanding a single message and all the k users having the same side information as the original user, in the rest of this paper, we consider that each user demands a single message in an EICP. Let the message demanded by a user u_i be denoted as x_{d_i} , where $d_i \in [M]$ and let the vector \mathbf{d} denote the vector of indices of messages demanded by all the N users, i.e., $\mathbf{d} = (d_1, d_2, \dots, d_N)$. Further, let the side information possessed by all the users be denoted by the set $\mathcal{K} = \{\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_N\}$. An EICP with N users, M messages, $M \leq N$, side information set \mathcal{K} and demand vector \mathbf{d} is denoted as $\mathcal{E}(N, M, \mathcal{K}, \mathbf{d})$.

A solution to an EICP, called an embedded index code, is a set of transmissions made by all or a subset of the users such that the demands of all the users are met. An embedded index code is called linear if all the transmissions involved are linear combinations of the messages. An optimal embedded index code is one with minimum number of transmissions.

III. MINRANK OF AN EICP

In [3], a matrix representation of an EICP was given and a parameter called minrank was derived from this matrix representation which characterized the length of an optimal scalar linear embedded index code. In this section, an alternate definition of minrank is proposed and a proof that the minrank characterizes the length of an optimal embedded index code is given along the lines of the proof in [4].

Definition 1 (Minrank of an EICP). *For an EICP $\mathcal{E}(N, M, \mathcal{K}, \mathbf{d})$, the minrank of \mathcal{E} over \mathbb{F}_q is defined as*

$$\kappa_q(\mathcal{E}) \triangleq \min \{ \text{rank}_{\mathbb{F}_q}(\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]}) : \mathbf{v}_i \in \mathbb{F}_q^M, \mathbf{v}_i \triangleleft \mathcal{K}_i \text{ s.t. } \forall i \in [N], \exists j \in [N], j \neq i, \text{ s.t. } (\mathbf{e}_{d_i} + \mathbf{v}_i) \triangleleft \mathcal{K}_j \} \quad (1)$$

Remark 1. *The definition of minrank is similar to that in [4] except for the extra condition that for each of the vectors in the set $\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]}$, there should be a user who has all the messages in the support set of that vector in its side information set. So, the minimization of rank is only over those sets $\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]}$ where for each of the vectors $\mathbf{e}_{d_i} + \mathbf{v}_i$ in the set, there exists at least one user which could transmit the corresponding coded message.*

Theorem 1. *For a given EICP $\mathcal{E}(K, N, \mathcal{K}, \mathbf{d})$, the length of an optimal linear embedded index code is equal to the minrank $\kappa_q(\mathcal{E})$.*

Proof. The proof follows along the lines of the proof in [4]. Consider the message vector $\mathbf{x} = (x_1, x_2, \dots, x_M)$. From a transmission of the form $T_i = \mathbf{x}(\mathbf{e}_{d_i} + \mathbf{v}_i)^T$, $\mathbf{v}_i \triangleleft \mathcal{K}_i$, user u_i can decode its demanded message x_{d_i} as $T_i - \mathbf{x}\mathbf{v}_i^T$. Thus, if there are N transmissions $\{T_i\}_{i=1}^N$, all the users in \mathcal{U} can decode their demanded messages. It is, in fact, sufficient to have $\text{rank}_{\mathbb{F}_q}(\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]})$ transmissions. However, for each $i \in [N]$, there should be a user u_j , $j \neq i$ which can transmit T_i , i.e., u_j should have all the messages involved in T_i in its side information. If we only consider sets $\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]}$ where each of the elements $(\mathbf{e}_{d_i} + \mathbf{v}_i)$ satisfies the condition that there exists some user which could transmit $T_i = \mathbf{x}(\mathbf{e}_{d_i} + \mathbf{v}_i)^T$, and perform minimization of the rank over \mathbb{F}_q of these sets, the minimum rank obtained will be equal to $\kappa_q(\mathcal{E})$. Since any linear embedded index code consists of a set of transmissions of the form T_i , the optimal length of a linear embedded index code for the given EICP \mathcal{E} is equal to $\kappa_q(\mathcal{E})$. \square

A comparative analysis of the number of rank computations required in using the definition of minrank in [3] and the definition in this paper is given in [6] and the same is illustrated using the following numerical example.

Example 1. Consider an EICP with $N = 4$ users, $M = 4$ messages over \mathbb{F}_2 , the side information at the users given by $\mathcal{K}_1 = \{2, 3\}$, $\mathcal{K}_2 = \{1, 3\}$, $\mathcal{K}_3 = \{4\}$, $\mathcal{K}_4 = \{1, 2\}$ and the demand vector given by $\mathbf{d} = \{1, 2, 3, 4\}$. Consider user u_1 which demands the message x_1 . The contribution of u_1 to the set $\{\mathbf{e}_{d_i} + \mathbf{v}_i\}_{i \in [N]}$ are of the form $\mathbf{e}_1 + \mathbf{v}_1$, where \mathbf{v}_1 can take values from the set $\{\mathbf{e}_2, \mathbf{e}_3, \mathbf{0}\}$ since $x_1 + x_2$ can be transmitted by u_4 , $x_1 + x_3$ by u_2 and x_1 independently by either u_2 or u_4 . Similarly \mathbf{v}_2 can take values from $\{\mathbf{e}_1, \mathbf{e}_3, \mathbf{0}\}$ whereas \mathbf{v}_3 and \mathbf{v}_4 can only take the value $\mathbf{0}$. Thus, the rank minimization is performed only over 9 matrices. However, if we followed the approach in [3], it would be required to compute the ranks over \mathbb{F}_2 of $2^{\sum_{i=1}^N |\mathcal{K}_i|^2} = 2^{13} = 8192$ matrices of size 4×7 and $2^{\sum_{i=1}^N |\mathcal{K}_i|^2} \times 2^{\sum_{i=1}^N |\mathcal{K}_i|} = 2^{20} = 1048576$ matrices of size 4×11 .

IV. A GRAPHICAL REPRESENTATION OF AN EICP

In this section, we propose a graphical representation of an EICP using directed bipartite graphs and present transmission schemes for SUEICPs based on coverings using special graph structures. An explanation of why the representation in this paper is better than that in [1] as well as the proofs of all the results in the rest of this paper are given in [6].

A. Bipartite Problem Graph and Side Information Bipartite Graph

Definition 2 (Bipartite Problem Graph). Let $\mathcal{E}(N, M, \mathcal{K}, \mathbf{d})$ be an instance of the EICP. Its graphical representation called the bipartite problem graph is given by a directed bipartite graph \mathcal{G} on the vertex set $V = (\mathcal{U}, \mathcal{X})$ and the directed edge set $E = \{(u_i, x_j) : j \in \mathcal{K}_i\} \cup \{(x_i, u_j) : d_j = i\}$.

The edges directed from the vertex set \mathcal{U} to the vertex set \mathcal{X} represent side information and the edges in the opposite direction represent the demanded messages. We use the notation $\mathcal{E}(\mathcal{G})$ to refer to the EICP corresponding to a given bipartite problem graph \mathcal{G} . For a system with N users $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$, M messages $\mathcal{X} = \{x_1, x_2, \dots, x_M\}$ and side information set \mathcal{K} at the users, there could be $\prod_{i=1}^N (M - |\mathcal{K}_i|)$ possible demand vectors and corresponding to each of these demands, there is an EICP. Let the set of this $\prod_{i=1}^N (M - |\mathcal{K}_i|)$ arising from a side information set \mathcal{K} be denoted as $\mathcal{E}_{\mathcal{K}}$. The side information set \mathcal{K} which is common to all these EICPs is represented using an undirected bipartite graph as explained below.

Definition 3 (Side Information Bipartite Graph). A graphical representation of the side information set \mathcal{K} is given by an undirected bipartite graph \mathcal{G}_S on the vertex set $V = (\mathcal{U}, \mathcal{X})$ and the edge set $E = \{(u_i, x_j) : j \in \mathcal{K}_i\}$.

Given a side information bipartite graph \mathcal{G}_S , the set of possible demand vectors is denoted as $D_{\mathcal{G}_S}$. Since a side information set \mathcal{K} is analogous to the side information bipartite graph \mathcal{G}_S , the set of EICPs $\mathcal{E}_{\mathcal{K}}$ is also denoted as $\mathcal{E}_{\mathcal{G}_S}$. The EICP corresponding to a demand vector $\mathbf{d} \in D_{\mathcal{G}_S}$ is denoted

as $\mathcal{E}(\mathcal{G}_S, \mathbf{d})$. The bipartite problem graph of the EICP in Example 1 is given in Fig. 1(a) and the corresponding side information bipartite graph is given in Fig. 1(b).

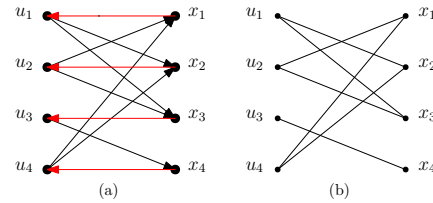


Fig. 1: Bipartite Problem Graph and Side information Bipartite Graph of the EICP in Example 1.

In the following subsection, we identify certain graph structures which, if present in the side information bipartite graph, can result in savings in the number of transmissions required to solve the corresponding EICPs and present a couple of transmission schemes utilizing these structures.

B. Graphical Structures in Single Unicast EICP

A single unicast EICP (SUEICP) is defined as follows.

Definition 4 (Single Unicast EICP). An EICP $\mathcal{E}(N, M, \mathcal{K}, \mathbf{d})$ is said to be single unicast if

- 1) $M = N$, and
- 2) $d_i \neq d_j$, for $i \neq j$.

Since in a SUEICP, the number of messages is equal to the number of users and each of the users demand a distinct message, without loss of generality, let us consider that user u_i demands the message x_i , i.e., $d_i = i$. Let an SUEICP with N users and N messages and demand vector \mathbf{d} such that $d_i = i$ be represented using its side information bipartite graph \mathcal{G}_S and denoted as $\mathcal{E}(N, \mathcal{G}_S)$.

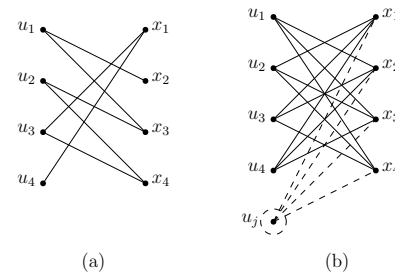


Fig. 2: Regular Tree and Bi-clique structures in SUEICP

Definition 5 (Regular Tree). Consider a bipartite graph on the partite sets $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$, $n \geq 3$, with the edge set, $E = \{(a_i, b_{i+1}), i \in [n]\} \cup \{(a_i, b_{i+2}), i \in [n-1]\}$, where $b_{n+1} = b_1$. Such a bipartite graph on $2n$ vertices, denoted as $T_{n,n}$, is called a regular tree.

Remark 2. In the bipartite graph $T_{n,n}$, every node in A except a_n has degree two and similarly every node in B except b_2 has degree two. The nodes a_n and b_2 have degree one each. Thus, the bipartite graph $T_{n,n}$ on $2n$ vertices has a total of $2n - 1$ edges and hence is a tree.

Fig. 2(a) shows a regular tree $T_{4,4}$. Suppose $T_{4,4}$ represents the side information bipartite graph of an SUEICP with $N = M = 4$ and the demand vector $\mathbf{d} = (1, 2, 3, 4)$. If u_1 transmits $x_2 + x_3$, u_2 transmits $x_3 + x_4$, and u_3 transmits $x_4 + x_1$, the demands of all the four users are met. Further, it can be verified that any set of two transmissions are not sufficient to satisfy the demands of all the 4 users and hence the transmission scheme with 3 transmissions is scalar linear optimal. We generalize this transmission scheme for an SUEICP on 4 users to a scheme for an SUEICP on N users in the following lemma.

Lemma 1. *If the side information bipartite graph \mathcal{G}_S of an SUEICP with N users and N messages, $N \geq 3$, is a regular tree $T_{N,N}$, then $N - 1$ transmissions are necessary and sufficient to satisfy the demands of all the users.*

Definition 6 (Bi-clique). *Consider a bipartite graph on the partite sets $A = \{a_1, a_2, \dots, a_n\}$ and $B = \{b_1, b_2, \dots, b_n\}$ where each of the node $a_i \in A$ is connected to all the nodes in $B \setminus \{b_i\}$ which implies that each node $b_i \in B$ is connected to all the nodes $A \setminus \{a_i\}$. This $n - 1$ regular bipartite graph is called a “bi-clique” and is denoted as $B_{n,n}$.*

Definition 7 (Covered Bi-clique). *For a bi-clique $B_{n,n}$ on the partite sets A and B , which is a subgraph of a bipartite graph G on the partite sets $A_G \supset A$ and $B_G \supseteq B$, if there exists a node $a \in A_G \setminus A$ such that a is connected to all nodes in B , then the bi-clique is called a “covered” bi-clique, denoted as $B_{n,n}^c$ and the node a is called the covering node.*

Fig. 2(b) shows a bi-clique $B_{4,4}$. Suppose it is a sub-graph on 4 user nodes and 4 message nodes of the side information bipartite graph \mathcal{G}_S of an SUEICP. Assume that the user vertex u_j and the edges coming from it shown in dotted lines are absent. Then, for the demand vector \mathbf{d} such that $d_i = i$, $i \in [4]$, with two transmissions given by $T_1 = x_2 + x_3 + x_4$ transmitted by u_1 and $T_2 = x_1$ transmitted by u_2 , the demands of all four users can be satisfied. However, if the user u_j and the dotted edges incident on it are present in \mathcal{G}_S , then for the demand vector $\mathbf{d} = \{1, 2, 3, 4\}$ corresponding to the users $\{u_1, u_2, u_3, u_4\}$, a single transmission $T_1 = x_1 + x_2 + x_3 + x_4$ by user u_j is sufficient to satisfy the demands of all the 4 users. In this scenario, the user u_j is called the “covering user”.

Lemma 2. *Consider a bi-clique $B_{n,n} \subseteq \mathcal{G}_S$ formed by the user nodes $\mathcal{U}_B = \{u_{i_1}, u_{i_2}, \dots, u_{i_n}\}$ and message nodes $\mathcal{X}_B = \{x_{i_1}, x_{i_2}, \dots, x_{i_n}\}$. For the SUEICP $\mathcal{E}(\mathcal{G}_S, \mathbf{d})$ corresponding to the demand vector \mathbf{d} such that $d_i = i$, to satisfy the demands of the users in \mathcal{U}_B we need $2 - \mathcal{I}(B_{n,n}^c)$ transmissions, where,*

$$\mathcal{I}(B_{n,n}^c) = \begin{cases} 1 & \text{if } B_{n,n} \text{ is covered,} \\ 0 & \text{otherwise} \end{cases},$$
 is the indicator function on whether the bi-clique $B_{n,n}$ is covered or not.

Remark 3. *A cycle on $n \geq 3$ vertices in the side information graph of an SUEICP is equivalent to a regular tree on n user vertices and n message vertices, $T_{n,n} \subseteq \mathcal{G}_S$ w.r.t number of transmissions required and a clique on $n \geq 3$ vertices in the side information graph of an SUEICP is equivalent to a covered*

bi-clique $B_{n,n}^c$ on n user vertices and n message vertices.

To describe covering schemes based on the above graph structures, we identify graph structures equivalent to regular trees and bi-cliques for $n = 1$ and $n = 2$. For $n = 1$, both regular tree as well as a covered bi-clique is a single edge as shown in Fig. 3(a) where the message is transmitted independently. For $n = 2$, a tree $T_{2,2}$, will not represent an SUEICP as for the graph $T_{2,2}$ to be connected, one of the two user nodes must know both the messages. The graph structure in an SUEICP which requires one transmission to satisfy the demands of two users is a minimally connected graph on 3 user nodes and 2 message nodes as shown in Fig. 3(b) which is also the covered bi-clique $B_{2,2}^c$.

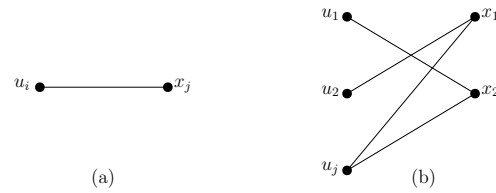


Fig. 3: Graph structures in \mathcal{G}_S corresponding to both regular trees and covered bi-cliques for $n = 1$ and $n = 2$.

Lemma 3 (Tree Cover Scheme). *The tree-cover scheme for the SUEICP $\mathcal{E}(N, \mathcal{G}_S)$ identifies a maximal set of message-disjoint regular trees, say $\{T_{n_i, n_i}\}_{i=1}^K$, in \mathcal{G}_S such that the union of these trees covers the message vertices of \mathcal{G}_S . The total number of transmissions required to satisfy the demands of all the N users in the SUEICP $\mathcal{E}(N, \mathcal{G}_S)$ using the tree cover scheme with K regular trees, is $N - K + K_e$, where K_e is the number of single edge trees.*

Lemma 4 (Bi-Clique Cover Scheme). *For an SUEICP $\mathcal{E}(N, \mathcal{G}_S)$, a bi-clique cover scheme identifies a minimal set of message-disjoint bi-cliques such that the union of these bi-cliques covers the message vertex set of \mathcal{G}_S . The total number of transmissions to solve the SUEICP $\mathcal{E}(N, \mathcal{G}_S)$ using the bi-clique cover scheme with K message-disjoint bi-cliques $\{B_{n_i, n_i}\}_{i=1}^K$ is equal to $\sum_{i=1}^K (2 - \mathcal{I}(B_{n_i, n_i}^c))$ which is bounded between K and $2K$.*

Remark 4. *If all the bi-cliques identified in the bi-clique covering scheme are covered bi-cliques, then it is the same as the covering scheme given by Algorithm 2 in [1]. A bi-clique which is not covered will be identified as two cliques in the problem graph by Algorithm 2 in [1] and hence requires two transmissions same as that required by the bi-clique covering scheme in this paper.*

Remark 5. *While the solution obtained by the bi-clique covering scheme is a task-based solution as defined in [1], the solution given by the tree covering scheme is not task-based, in general. Each bi-clique represents a co-operative data exchange problem [12] and hence is of practical interest. Further, since [13] showed that the length of a task based*

solution is at most equal to the square of the minrank of the corresponding centralized ICP, the gap between the length of an optimal solution of the EICP and that of the clique cover is at most quadratic.

V. OPTIMAL SOLUTION FOR A CLASS OF SUEICP

The class of SUEICPs with symmetric neighboring consecutive side information, abbreviated as SNCS-SUEICP, has a user u_i requesting the message x_i and knowing U consecutive messages before x_i and D consecutive messages after x_i , i.e., $\mathcal{K}_i = \{i-U, i-U+1, \dots, i-1\} \cup \{i+1, i+2, \dots, i+D\}$, where the calculations are performed modulo N .

Theorem 2. For the class of SNCS-SUEICPs with $D = U + 1, U \geq 1$, minrank κ is equal to $\left\lceil \frac{N-1}{U+1} \right\rceil$.

Proof. Converse : Consider the class of SNCS-SUEICPs which is the centralized version of SNCS-SUEICP. It was shown in [14] that the length of an optimal scalar linear index code for such problems is given by $\left\lceil \frac{N - \max(U, D) + \min(U, D)}{\min(U, D) + 1} \right\rceil$ which when restricted to the class of SNCS-SUEICPs with $D = U + 1, U \geq 1$, gives the minrank to be $\left\lceil \frac{N-1}{U+1} \right\rceil$. Since we know that the minrank of an EICP is at least as much as the minrank of the corresponding ICP, $\kappa \geq \left\lceil \frac{N-1}{U+1} \right\rceil$.

Achievability Scheme : Let $p \triangleq \left\lceil \frac{N-1}{U+1} \right\rceil$. Consider the following set of p transmissions,

$$T_i = \sum_{j \in [D]} x_{(i-1)D+j}, \quad i \in \{1, 2, \dots, p-1\}, \text{ and}$$

$$T_p = x_{(p-1)D+1} + x_{(p-1)D+2} + \dots + x_N.$$

$T_i, 1 < i \leq p$ is transmitted by user $u_{(i-1)D}$ and T_1 is transmitted by u_N . This is a feasible scheme as user $u_{(i-1)D}, 1 < i \leq p$ has the messages $\{x_{(i-1)D+1}, x_{(i-1)D+2}, \dots, x_{iD}\}$ in its side information which are summed up to form the transmission T_i and user u_N has x_1, x_2, \dots, x_D which are added to form the transmission T_1 . Consider a user u_j . The transmission containing its requested message x_j is formed by x_j added to messages with indices in the set $\{j - (D-1), j - (D-2), \dots, j-1, j+1, \dots, j + (D-1)\}$, each of which is present with user u_j and hence it can decode the message x_j . Thus, from the above set of p transmissions, each of the N users can decode their requested message and hence is a valid embedded index code. \square

The set of transmissions in the achievability scheme above can be obtained as a bi-clique cover with all the p bi-cliques being covered bi-cliques. In the following section, we present a relation between the connectedness of the side information bipartite graph and the minrank of an EICP.

VI. CONNECTEDNESS OF THE BIPARTITE GRAPH AND MINRANK

The motivation to look for a relation between the connectedness of the side information bipartite graph and the minrank of the corresponding EICPs has been applications where the users have a lot of common messages in their side information.

Further, for EICPs with $N = M = 3$, the minrank is strictly less than the number of unique messages demanded only for problems with connected side information graphs and when all three messages are demanded. In this section, we prove a theorem connecting minrank $\kappa_q(\mathcal{G})$ and connectedness of \mathcal{G}_S .

For a given side information bipartite graph \mathcal{G}_S with the partite sets \mathcal{U} and \mathcal{X} , let \mathcal{X}' denote the subset of \mathcal{X} obtained by removing vertices of degree 1 in \mathcal{X} , i.e., $\mathcal{X}' = \mathcal{X} \setminus \{x_j : \deg(x_j) = 1\}$ and the induced sub-graph on the vertex set $(\mathcal{U}, \mathcal{X}')$ be denoted by \mathcal{G}'_S . The set \mathcal{X}' is the set of messages which could be possibly coded in the embedded index code as message nodes with degree 1 are present at only one user and cannot be coded and there are no degree zero message nodes as each message is assumed to be present at least at one user. For a demand vector \mathbf{d} , the unique messages demanded in \mathbf{d} from the message set in \mathcal{G} is denoted by $\text{uniq}(\mathbf{d}_{\mathcal{G}})$.

Theorem 3. For a side-information bipartite graph \mathcal{G}_S , for every demand vector $\mathbf{d} \in D_{\mathcal{G}_S}$ such that $\text{uniq}(\mathbf{d}_{\mathcal{G}'_S}) = \mathcal{X}'$, the minrank of the EICP $\mathcal{E}(\mathcal{G}_S, \mathbf{d})$ is strictly less than the number of distinct messages demanded, i.e., $\kappa_q(\mathcal{E}) \leq |\text{uniq}(\mathbf{d}_{\mathcal{G}_S})| - 1$, if \mathcal{G}_S is connected.

Corollary 1. For a connected side information graph \mathcal{G}_S , for EICPs where all the M messages are demanded, the number of transmissions required is strictly less than M , i.e., $\forall \mathbf{d} \in D_{\mathcal{G}_S}$ such that $|\text{uniq}(\mathbf{d}_{\mathcal{G}_S})| = M, \kappa_q(\mathcal{E}(\mathcal{G}_S, \mathbf{d})) < M$.

Remark 6. While Lemma 1 showed that for an SUEICP whose side information bipartite graph is a regular tree $T_{N,N}$, the number of transmissions required is $N - 1$, by applying the corollary above, it can be seen that for any tree on N user nodes and N message nodes, the number of transmissions required for an SUEICP is less than or equal to $N - 1$.

VII. CONCLUSION

This paper looked at embedded index coding problems for which an alternate definition of minrank, which is computationally more efficient, and a graphical representation using directed bipartite graphs were given. The side information alone was represented using an undirected bipartite graph, which was called the side information bipartite graph. For the single unicast class of EICPs, a couple of transmission schemes were presented, one of which was shown to give scalar linear optimal solutions to a sub-class of single unicast EICPs. Further, connectedness of the side information bipartite graph was shown to be a sufficient condition for reducing the minrank below the number of distinct messages demanded. It will be interesting to explore whether stricter requirements on the connectivity of the side information bipartite graph will give better savings in transmission.

ACKNOWLEDGMENT

This work was supported partly by the Science and Engineering Research Board (SERB) of Department of Science and Technology (DST), Government of India, through J.C. Bose National Fellowship to Prof. B. Sundar Rajan.

REFERENCES

- [1] A. Porter and M. Wootters, "Embedded Index Coding," in *IEEE Transactions on Information Theory*, vol. 67, no. 3, pp. 1461-1477, March 2021.
- [2] Y. Birk and T. Kol, "Informed-source coding-on-demand (ISCOD) over broadcast channels," Proceedings. IEEE INFOCOM '98, the Conference on Computer Communications. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Gateway to the 21st Century (Cat. No.98, San Francisco, CA, USA, 1998, pp. 1257-1264 vol.3.
- [3] A. A. Mahesh, N. Sageer Karat and B. S. Rajan, "Min-rank of Embedded Index Coding Problems," 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, CA, USA, 2020, pp. 1723-1728.
- [4] S. H. Dau, V. Skachek, and Y. M. Chee, "Error correction for index coding with side information," in *IEEE Transactions on Information Theory*, vol. 59, no. 3, pp. 1517-1531, Mar. 2013.
- [5] A. S. Tehrani, A. G. Dimakis and M. J. Neely, "Bipartite index coding," in *2012 IEEE International Symposium on Information Theory Proceedings*, Cambridge, MA, USA, 2012, pp. 2246-2250.
- [6] A. A. Mahesh, B. S. Rajan, "Minrank of Embedded Index Coding Problems and its Relation to Connectedness of a Bipartite Graph", arXiv:2201.08680 [cs.IT].
- [7] L. Ong and C. K. Ho, "Optimal index codes for a class of multicast networks with receiver side information," in *Proc. 2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, 2012, pp. 2213-2218.
- [8] M. A. R. Chaudhry, Z. Asad, A. Sprintson, and M. Langberg, "On the complementary index coding problem," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2011, pp. 224-248.
- [9] A. Blasiak, R. D. Kleinberg, and E. Lubetzky, "Broadcasting with side information: Bounding and approximating the broadcast rate," in *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5811-5823, Sept. 2013.
- [10] W. Zhu, D. Li and W. Saad, "Multiple Vehicles Collaborative Data Download Protocol via Network Coding," in *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1607-1619, April 2015.
- [11] Jesy P., N. S. Karat, Deepthi P.P., and B. S. Rajan, "Index Coding in Vehicle to Vehicle Communication," in *IEEE Transactions on Vehicular Technology*, Vol.69, No. 10, Oct. 2020, pp.11926-11936.
- [12] S. El Rouayheb, A. Sprintson and P. Sadeghi, "On coding for cooperative data exchange," in *Proc. IEEE Information Theory Workshop on Information Theory (ITW)*, Cairo, 2010, pp. 1-5.
- [13] I. Haviv, "Task-Based Solutions to Embedded Index Coding," in *IEEE Transactions on Information Theory*, vol. 66, no. 10, pp. 6144-6149, Oct. 2020.
- [14] H. Maleki, V. R. Cadambe and S. A. Jafar, "Index Coding An Interference Alignment Perspective," in *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5402-5432, Sept. 2014.
- [15] S. El Rouayheb, A. Sprintson and C. Georghiades, "On the Index Coding Problem and Its Relation to Network Coding and Matroid Theory," in *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3187-3195, July 2010.
- [16] M. B. Vaddi and B. S. Rajan, "Optimal vector linear index codes for some symmetric side information problems," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, 2016, pp. 125-129.