Optimal Index Codes With Near-Extreme Rates

Son Hoang Dau, Vitaly Skachek, *Member, IEEE*, and Yeow Meng Chee, *Senior Member, IEEE*

***Abstract*— The min-rank of a digraph was shown to represent the length of an optimal scalar linear solution of the corresponding instance of the Index Coding with Side Information (ICSI) problem. In this paper, the graphs and digraphs of near-extreme min-ranks are studied. Those graphs and digraphs correspond to the ICSI instances having near-extreme transmission rates when using optimal scalar linear index codes. In particular, it is shown that the decision problem whether a digraph has min-rank two is NP-complete. By contrast, the same question for graphs can be answered in polynomial time. In addition, a circuit-packing bound is revisited, and several families of digraphs, optimal with respect to this bound, whose min-ranks can be found in polynomial time, are presented.**

***Index Terms*— Index coding, network coding, side information, broadcast.**

I. Introduction

Building communication schemes which allow participants to communicate efficiently has always been a challenging yet intriguing problem for information theorists. Index Coding with Side Information (ICSI) ([1], [2]) is a communication scheme dealing with broadcast channels in which receivers have prior side information about the messages to be transmitted. By using coding and exploiting the knowledge about the side information, the sender may significantly reduce the number of required transmissions compared with the straightforward approach. As a consequence, the efficiency of communication over this type of broadcast channels could be dramatically improved. Apart from being a special case of the well-known (non-multicast) Network Coding problem ([3], [4]), the ICSI problem has also found various potential applications on its owns, such as audio- and video-on-demand, daily newspaper delivery, data pushing, and opportunistic wireless networks ([1], [2], [5]–[8]).

Manuscript received March 15, 2012; revised August 28, 2013; accepted November 11, 2013. Date of publication December 18, 2013; date of current version February 12, 2014. This work was supported by the National Research Foundation of Singapore Research under Grant NRF-CRP2-2007-03. This paper was presented at the 2012 IEEE International Symposium on Information Theory.

S. H. Dau was with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371. He is now with the SUTD-MIT International Design Centre, Singapore University of Technology and Design, Singapore 138682 (e-mail: dausonhoang84@gmail.com).

V. Skachek was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Institute of Computer Science, Faculty of Mathematics and Computer Science, University of Tartu, Tartu 50409, Estonia (e-mail: vitaly.skachek@ut.ee).

Y. M. Chee is with the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371 (e-mail: ymchee@ntu.edu.sg).

Communicated by J. Körner, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2013.2295331

TABLE I
Characterizations of Graphs and Digraphs With Near-Extreme Min-Ranks

Min-Rank	Graph $\mathcal{G}$	Digraph $\mathcal{D}$
1	$\mathcal{G}$ is complete (trivial)	$\mathcal{D}$ is complete (trivial)
2	$\mathcal{G}$ is not complete and $\overline{\mathcal{G}}$ is 2-colorable ([12])	$\mathcal{D}$ is not complete and $\overline{\mathcal{D}}$ is fairly 3-colorable*†
$n-2$	$\mathcal{G}$ (connected) has a maximum matching of size two and does not contain $\mathfrak{F}$ (Fig. 1) as a subgraph*	unknown
$n-1$	$\mathcal{G}$ (connected) is a star graph*	unknown
n	$\mathcal{G}$ has no edges (trivial)	$\mathcal{D}$ has no circuits ([19])

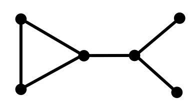

Fig. 1. The forbidden subgraph $\mathfrak{F}$.

In the work of Bar-Yossef *et al.* [5], the optimal transmission rate of scalar linear index codes for an ICSI instance was neatly characterized by the so-called *min-rank* of the side information digraph (i.e., directed graph, see Section II for definitions) corresponding to that instance. The concept of min-rank of a graph (i.e., undirected graph, see Section II for definitions) goes back to Haemers [9]. Min-rank serves as an upper bound for the celebrated Shannon capacity of a graph [10]. This upper bound, as pointed out by Haemers, although is usually not as good as the Lovász bound [11], is sometimes tighter and easier to compute. It was shown by Peeters [12] that computing the min-rank of a general graph (that is, the Min-Rank problem) is a hard task. More specifically, Peeters showed that deciding whether the min-rank of a graph is smaller than or equal to three is an NP-complete problem.

The work of Bar-Yossef *et al.* [5] has stimulated the interest in the Min-Rank problem. Exact and heuristic algorithms for finding min-ranks over the binary field of digraphs were developed in the work of Chaudhry and Sprintson [13]. The min-ranks of random digraphs were investigated by Haviv and Langberg [14]. A dynamic programming approach was proposed by Berliner and Langberg [15] to compute min-ranks of outerplanar graphs in polynomial time. Algorithms to approximate min-ranks of graphs with bounded min-ranks were studied by Chlamtac and Haviv [16].

In this paper, we study graphs and digraphs that have near-extreme min-ranks. In other words, we study ICSI instances with n receivers for which optimal *scalar linear* index codes have transmission rates 1, 2, $n-2$, $n-1$, or n. In particular, we show that deciding whether a digraph has min-rank two over the *binary* field is an NP-complete problem. Very recently, it was found by Maleki *et al.* [17] that the same problem for digraphs over sufficiently large field can be solved in polynomial time. By contrast, a graph has min-rank two over any finite field if and only it is not a complete graph and its complement is bipartite, a condition which can be verified in polynomial time (see, for instance, West [18, p. 495]).

The characterizations of graphs and digraphs with near-extreme min-ranks are summarized in Table I. The star mark "$*$" indicates that the result is established in this paper. The dagger mark "†" indicates that the result is proved only for the binary field.

The near-extreme cases are of significant interest from both theoretical and practical points of view. On the theoretical side, it is desirable to understand, which values of the min-rank in the range between 1 and n are easy to verify, and for which values it is hard. In particular, it is known that the Min-Rank problem is NP-hard [12] ($\mathsf{minrk}_q(\mathcal{G}) = 3$ is hard to verify). On the other hand, min-rank values of 1 and n are easy to verify in polynomial time for both graphs and digraphs. Therefore, there should exist threshold values of min-rank between 1 and n, for which the difficulty of the verification problem for the min-rank of a (di-)graph changes from easy to hard or vice versa. For graphs, such threshold values are 3 and some integer smaller than $n-2$ (not exactly known). By contrast, for digraphs, the thresholds are 2 (proved in this work) and $n-1$ (conjectured).

From the practical point of view, the use of length-one index codes in wireless communications has already been proposed (for instance, see COPE [7], [20], [21]), due to their simplicity and efficiency. However, the variety of scenarios where an index code of length one is applicable is limited (each client must know all except one message). An index code of length two is obviously the next potential candidate to be used. Thus, for index coding scenarios, where the corresponding min-rank is small, it is desirable that the sender could identify that efficiently, and employ an optimal index code if needed.

So far, families of graphs and digraphs whose min-ranks are either known or computable in polynomial time are the followings. For *graphs*, they are odd holes and odd anti-holes [19], perfect graphs [19], outerplanar graphs [15], and graphs with simple tree structure [22]. For *digraphs*, they are acyclic digraphs [19]. In this work, we point out several new families of *digraphs* for which the circuit-packing bound [23] is tight. For such families of digraphs, min-ranks can be found in polynomial time.

In the context of index coding, we only study min-ranks of digraphs over a *finite* field $\mathbb{F}_q$. However, all of our results, except Theorem 4.7, Corollary 4.8, and Theorem 5.2, still hold for an *arbitrary* field $\mathbb{F}$. This is because the characteristic of the field does not play any role in their proofs.

The paper is organized as follows. Basic notation and definitions are presented in Section II. The ICSI problem is formulated in Section III. Section IV is devoted to the characterizations of graphs and digraphs of near-extreme min-ranks. We prove the hardness of the Min-Rank problem for digraphs in Section V. Families of digraphs that attain the circuit-packing bound [23] are discussed in Section VI. We conclude the paper in Section VII.

II. Notation and Definitions

Let $[n]$ denote the set of integers $\{1, 2, \ldots, n\}$. Let $\mathbb{F}_q$ denote the finite field of q elements and $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{\mathbf{0}\}$. The *support* of a vector $\boldsymbol{u} \in \mathbb{F}_q^n$ is defined to be the set $\mathsf{supp}(\boldsymbol{u}) = \{i \in [n] : u_i \neq 0\}$. For an $n \times k$ matrix $\boldsymbol{M}$, let $\boldsymbol{M}_i$ denote the ith row of $\boldsymbol{M}$. For a set $E \subseteq [n]$, let $\boldsymbol{M}_E$ denote the $|E| \times k$ sub-matrix of $\boldsymbol{M}$ formed by rows of $\boldsymbol{M}$ which are indexed by the elements of E. For any matrix $\boldsymbol{M}$ over $\mathbb{F}_q$, we denote by $\mathsf{rank}_q(\boldsymbol{M})$ the rank of $\boldsymbol{M}$ over $\mathbb{F}_q$ (or the *q-rank* of $\boldsymbol{M}$). We use $\boldsymbol{e}_i$ to denote the unit vector, which has a one at the ith position, and zeros elsewhere.

A simple *graph* is a pair $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ where $\mathcal{V}(\mathcal{G})$ is the set of vertices of $\mathcal{G}$ and $\mathcal{E}(\mathcal{G})$ is a set of *unordered* pairs of distinct vertices of $\mathcal{G}$. We refer to $\mathcal{E}(\mathcal{G})$ as the set of *edges* of $\mathcal{G}$. A typical edge of $\mathcal{G}$ is of the form $\{u, v\}$ where $u \in \mathcal{V}(\mathcal{G})$, $v \in \mathcal{V}(\mathcal{G})$, and $u \neq v$. If $e = \{u, v\} \in \mathcal{E}(\mathcal{G})$ we say that u and v are adjacent. We also refer to u and v as the *endpoints* of e.

A simple *digraph* is a pair $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ where $\mathcal{V}(\mathcal{D})$ is the set of vertices of $\mathcal{D}$, and $\mathcal{E}(\mathcal{D})$ is a set of *ordered* pairs of distinct vertices of $\mathcal{D}$. We refer to $\mathcal{E}(\mathcal{D})$ as the set of arcs (or directed edges) of $\mathcal{D}$. A typical *arc* of $\mathcal{D}$ is of the form $e = (u, v)$ where $u \in \mathcal{V}(\mathcal{D})$, $v \in \mathcal{V}(\mathcal{D})$, and $u \neq v$. The vertices u and v are called the *endpoints* of the arc e.

Simple graphs and digraphs have no loops and no parallel edges and arcs, respectively. In the scope of this paper, only simple graphs and digraphs are considered. Therefore, we simply refer to them as graphs and digraphs for succinctness.

The number of vertices $|\mathcal{V}(\mathcal{D})|$ is called the *order* of $\mathcal{D}$, whereas the number of arcs $|\mathcal{E}(\mathcal{D})|$ is called the *size* of $\mathcal{D}$. The *complement* of a digraph $\mathcal{D}$, denoted by $\overline{\mathcal{D}}$, is defined as follows. The vertex set is $\mathcal{V}(\overline{\mathcal{D}}) = \mathcal{V}(\mathcal{D})$. The arc set is

$$\mathcal{E}(\overline{\mathcal{D}}) = \big\{(u, v) : \ u, v \in \mathcal{V}(\mathcal{D}), \ u \neq v, \ (u, v) \notin \mathcal{E}(\mathcal{D})\big\}.$$

Analogous concepts are also defined for graphs.

A digraph $\mathcal{D}$ is called *symmetric* if it satisfies the property that $(u, v) \in \mathcal{E}(\mathcal{D})$ if and only if $(v, u) \in \mathcal{E}(\mathcal{D})$. A symmetric digraph can be viewed as a graph, and vice versa. A *complete graph* is a graph that contains all possible edges. A *complete digraph* is a digraph that contains all possible arcs.

A collection of subsets $V_1, V_2, \ldots, V_k$ of a set V is said to *partition* V if $\cup_{i=1}^k V_i = V$ and $V_i \cap V_j = \varnothing$ for every $i \neq j$. In that case, $[V_1, V_2, \ldots, V_k]$ is referred to as a partition of V, and V_i's ($i \in [k]$) are called *parts* of the partition.

A graph $\mathcal{G}$ is called *bipartite* if $\mathcal{V}(\mathcal{G})$ can be partitioned into two subsets U and V such that for every edge $\{u, v\} \in \mathcal{E}(\mathcal{G})$, it holds that $u \in U$ and $v \in V$, or vice versa.

A *subgraph* of a graph $\mathcal{G}$ is a graph whose vertex set V is a subset of that of $\mathcal{G}$ and whose edge set is a subset of that of $\mathcal{G}$ restricted to the vertices in V. Let V be a subset of vertices in $\mathcal{V}(\mathcal{G})$. The subgraph of $\mathcal{G}$ *induced* by V is a graph whose vertex set is V, and edge set is $\{\{u, v\} : \ u \in V, \ v \in V, \{u, v\} \in \mathcal{E}(\mathcal{G})\}$. We refer to such a graph as an *induced subgraph* of $\mathcal{G}$. A subgraph and induced subgraph of a digraph can be defined in a similar manner.

A *path* in a graph $\mathcal{G}$ is a sequence of distinct vertices $(v_1, v_2, \ldots, v_r)$, such that $\{v_s, v_{s+1}\} \in \mathcal{E}(\mathcal{G})$ for all $s \in [r-1]$. A *directed path* in a digraph $\mathcal{D}$ is a sequence of distinct vertices $(v_1, v_2, \ldots, v_r)$, such that $(v_s, v_{s+1}) \in \mathcal{E}(\mathcal{D})$, for all $s \in [r-1]$.

A *circuit* in a digraph $\mathcal{D}$ is a sequence of pairwise distinct vertices

$$\mathcal{C} = (v_1, v_2, \ldots, v_r),$$

where $(v_s, v_{s+1}) \in \mathcal{E}(\mathcal{D})$ for all $s \in [r-1]$ and $(v_r, v_1) \in \mathcal{E}(\mathcal{D})$ as well. A digraph is called *acyclic* if it contains no circuits.

A graph is called *connected* if there is a path from each vertex in the graph to every other vertex. The *connected components* of a graph are its maximal connected subgraphs. Similarly, a digraph is called *strongly connected* if there is a directed path from each vertex in the graph to every other vertex. The *strongly connected components* of a digraph are its maximal strongly connected subgraphs.

If (u, v) is an arc in a digraph $\mathcal{D}$, then v is called an *out-neighbor* of u in $\mathcal{D}$. The set of out-neighbors of a vertex u in a digraph $\mathcal{D}$ is denoted by $N_O^{\mathcal{D}}(u)$. We simply use $N_O(u)$ whenever there is no potential confusion. We also denote by $N^{\mathcal{G}}(u)$ the set of neighbors of u in a graph $\mathcal{G}$, namely, the set of vertices adjacent to u in $\mathcal{G}$.

An *independent set* in a graph $\mathcal{G}$ is a set of vertices of $\mathcal{G}$ with no edges connecting any two of them. An independent set in $\mathcal{G}$ of largest cardinality is called a *maximum independent set* in $\mathcal{G}$. The cardinality of such a maximum independent set is referred to as the *independence number* of $\mathcal{G}$, denoted by $\alpha(\mathcal{G})$. We also use $\alpha(\mathcal{D})$ to denote the size of a maximum acyclic induced subgraph of a digraph $\mathcal{D}$ for the following reason. For a symmetric digraph $\mathcal{D}$, $\alpha(\mathcal{D})$ is equal to the size of a maximum independent set if $\mathcal{D}$ is regarded as a graph.

A clique of a graph is a set of vertices that induces a complete subgraph of that graph. A *clique cover* of a graph is a set of cliques that partition its vertex set. A *minimum clique cover* of a graph is a clique cover with the minimum number of cliques. The number of cliques in such a minimum clique cover of a graph is called the clique cover number of that graph. Similar concepts are defined for digraphs. We denote by $\mathsf{cc}(\mathcal{G})$ the clique cover number of a graph $\mathcal{G}$ and $\mathsf{cc}(\mathcal{D})$ the clique cover number of a digraph $\mathcal{D}$.

III. The Index Coding With Side Information Problem

The ICSI problem is formulated as follows. Suppose a sender S wants to send a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, where $x_i \in \Sigma^t$ for all $i \in [n]$, Σ is some alphabet, to n receivers $R_1, R_2, \ldots, R_n$. Each R_i possesses some prior side information, consisting of the blocks x_j, $j \in \mathcal{X}_i \subsetneq [n]$, and is interested in receiving a single block x_i. The sender S broadcasts a codeword $\mathfrak{E}(\boldsymbol{x}) \in \Sigma^\kappa$, where κ is some positive integer, that enables each receiver R_i to recover x_i based on its side information. Such a mapping $\mathfrak{E} : \Sigma^{nt} \to \Sigma^\kappa$ is called an *index code*. We refer to t as the *block length* and κ as the *length* of the index code. The ratio κ/t is called the *transmission rate* of the index code. The objective of S is to find an *optimal* index code, that is, an index code which has the minimum transmission rate. The index code is called *linear* if $\Sigma = \mathbb{F}_q$ for some prime power q and $\mathfrak{E}$ is a linear mapping. The index code is called *scalar* if $t = 1$ and *block* if $t > 1$. The length and the transmission rate of a scalar index code ($t = 1$) are identical.

Each instance of the ICSI problem can be described by the so-called *side information digraph* [5]. Given n and $\mathcal{X}_i$, $i \in [n]$, the *side information digraph* $\mathcal{D} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{D}))$ is defined as follows. The vertex set $\mathcal{V}(\mathcal{D}) = \{u_1, u_2, \ldots, u_n\}$. The edge set $\mathcal{E}(\mathcal{D}) = \cup_{i \in [n]} \big\{(u_i, u_j) : \ j \in \mathcal{X}_i\big\}$. Sometimes we simply take $\mathcal{V}(\mathcal{D}) = [n]$ and $\mathcal{E}(\mathcal{D}) = \cup_{i \in [n]} \big\{(i, j) : \ j \in \mathcal{X}_i\big\}$. If $\mathcal{D}$ is a symmetric digraph, we can regard $\mathcal{D}$ as a graph, and refer to $\mathcal{D}$ as the *side information graph*.

Definition 3.1 ([9]): Let $\mathcal{D} = \big(\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D})\big)$ be a digraph of order n, where $\mathcal{V}(\mathcal{D}) = \{u_1, u_2, \ldots, u_n\}$.

1) A matrix $\boldsymbol{M} = (m_{u_i, u_j}) \in \mathbb{F}_q^{n \times n}$ (whose rows and columns are labeled by the elements of $\mathcal{V}(\mathcal{D})$) is said to *fit* $\mathcal{D}$ if

$$\begin{cases} m_{u_i,u_j} \neq 0, & i = j, \\ m_{u_i,u_j} = 0, & i \neq j, \ (u_i, u_j) \notin \mathcal{E}(\mathcal{D}). \end{cases}$$

2) The *min-rank* of $\mathcal{D}$ over $\mathbb{F}_q$ is defined to be

$$\mathsf{minrk}_q(\mathcal{D}) \triangleq \min \big\{ \mathsf{rank}_q(\boldsymbol{M}) : \ \boldsymbol{M} \in \mathbb{F}_q^{n \times n} \text{ and } \boldsymbol{M} \text{ fits } \mathcal{D} \big\}.$$

Since a graph can be viewed as a symmetric digraph, the above definitions also apply to a graph.

Theorem 3.2 ([5], [24]): The length of an optimal scalar linear index code over $\mathbb{F}_q$ for the ICSI instance described by $\mathcal{D}$ is $\mathsf{minrk}_q(\mathcal{D})$.

Let $\beta_t(\mathcal{D})$ denote the length of an optimal *block* index code of block length t over $\Sigma = \{0, 1\}$ for an ICSI instance described by a digraph $\mathcal{D}$. Note that we do not require the index codes to be linear. Alon *et al.* [25] defined the *broadcast rate* $\beta(\mathcal{D})$ of the corresponding ICSI instance to be $\lim_{t \to \infty} \beta_t(\mathcal{D})/t$ (see also Blasiak *et al.* [26]). In words, the broadcast rate is the average minimum communication cost per symbol in each block x_i (for long blocks). The reciprocal of $\beta(\mathcal{D})$ is also referred to as the *capacity* of the ICSI instance described by $\mathcal{D}$ (see Langberg and Sprintson [27]). Theorem 3.3 demonstrates an intuitive fact that in terms of transmission rates, block index codes are at least as good as scalar index codes, which in turn are at least as good as scalar linear index codes.

Theorem 3.3 ([5], [9], [19], [25]): For any digraph $\mathcal{D}$ we have

$$\alpha(\mathcal{D}) \leq \beta(\mathcal{D}) \leq \beta_1(\mathcal{D}) \leq \mathsf{minrk}_2(\mathcal{D}) \leq \mathsf{cc}(D).$$

The same inequalities hold for graphs.

Similarly, we also have the following inequalities considering index codes over $\Sigma = \mathbb{F}_q$. The last inequality in this proposition is called the *clique-covering bound* for min-ranks.

Proposition 3.4: For any digraph $\mathcal{D}$ we have

$$\alpha(\mathcal{D}) \leq \mathsf{minrk}_q(\mathcal{D}) \leq \mathsf{cc}(D).$$

The same inequalities hold for graphs.

IV. Digraphs of Near-Extreme Min-Ranks

Some of the results presented below are folklore. However, we include them for the sake of completeness.

A. (Strongly) Connected Components and Min-Ranks

Lemma 4.1 (Folklore): Let $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ be a graph. Suppose that $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k$ are subgraphs of $\mathcal{G}$ that satisfy the following conditions

1) The sets $\mathcal{V}(\mathcal{G}_i)$, $i \in [k]$, partition $\mathcal{V}(\mathcal{G})$;
2) There is no edge of the form $\{u, v\}$ where $u \in \mathcal{V}(\mathcal{G}_i)$ and $v \in \mathcal{V}(\mathcal{G}_j)$ for $i \neq j$.

Then

$$\mathsf{minrk}_q(\mathcal{G}) = \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{G}_i).$$

In particular, the above equality holds if $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_k$ are all connected components of $\mathcal{G}$.

The proof follows from Definition 3.1.

Lemma 4.2 (Folklore): Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. If $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_k$ are all strongly connected components of $\mathcal{D}$, then

$$\mathsf{minrk}_q(\mathcal{D}) = \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i).$$

The proof of this lemma appears in the Appendix.

These two lemmas suggest that it is sufficient to study the min-ranks of connected graphs and strongly connected digraphs, respectively.

B. Digraphs of Min-Rank One

The following results are known for (di-)graphs of min-rank one.

Proposition 4.3 (Folklore): Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. Then $\mathsf{minrk}_q(\mathcal{D}) = 1$ if and only if $\mathcal{D}$ is a complete digraph. The same statement holds for a graph.

Corollary 4.4 follows by applying Theorem 3.3 and Proposition 4.3.

Corollary 4.4: Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. Then $\beta(\mathcal{D}) = 1$ if and only if $\mathcal{D}$ is a complete digraph. The same statement holds for a graph.

C. Digraphs of Min-Rank Two

In this section, only the *binary* alphabet is considered. We first introduce the following concept of a *fair coloring* of a digraph. Recall that a k-coloring of a graph $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ is a mapping $\phi : \mathcal{V}(\mathcal{G}) \to [k]$ which satisfies the condition that $\phi(u) \neq \phi(v)$ whenever $\{u, v\} \in \mathcal{E}(\mathcal{G})$. We often refer to $\phi(u)$ as the *color* of u. If there exists a k-coloring of $\mathcal{G}$, then we say that $\mathcal{G}$ is k-colorable.

Definition 4.5: Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. A *fair k-coloring* of $\mathcal{D}$ is a mapping $\phi : \mathcal{V}(\mathcal{D}) \to [k]$ that satisfies the following conditions:

(C1) If $(u, v) \in \mathcal{E}(\mathcal{D})$ then $\phi(u) \neq \phi(v)$;
(C2) For each vertex u of $\mathcal{D}$, it holds that $\phi(v) = \phi(\omega)$ for all out-neighbors v and ω of u.

If there exists a fair k-coloring of $\mathcal{D}$, we say that we can *color $\mathcal{D}$ fairly by k colors*, or, $\mathcal{D}$ is *fairly k-colorable*.

We refer to the condition (C2) as the *fairness* of the coloring, since this condition guarantees that all out-neighbors of each vertex share the same color.

Lemma 4.6: A digraph $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ is fairly 3-colorable if and only if there exists a partition of $\mathcal{V}(\mathcal{D})$ into three subsets A, B, and C that satisfy the following conditions

1) For every $u \in A$: either $N_O(u) \subseteq B$ or $N_O(u) \subseteq C$;
2) For every $u \in B$: either $N_O(u) \subseteq A$ or $N_O(u) \subseteq C$;
3) For every $u \in C$: either $N_O(u) \subseteq A$ or $N_O(u) \subseteq B$.

Proof: If $\mathcal{D}$ is fairly 3-colorable, let A, B, and C respectively be the sets of vertices of $\mathcal{D}$ that share the same color. Then clearly A, B, and C partition $\mathcal{V}(\mathcal{D})$. Moreover, since all out-neighbors of each vertex must have the same color, the three conditions above are obviously satisfied. Conversely, if those conditions are satisfied, then $\phi : \mathcal{V}(\mathcal{D}) \to [3]$, defined by

$$\phi(u) = \begin{cases} 1, & u \in A \\ 2, & u \in B \\ 3, & u \in C, \end{cases}$$

is a fair 3-coloring of $\mathcal{D}$. ∎

Theorem 4.7: Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. Then $\mathsf{minrk}_2(\mathcal{D}) \leq 2$ if and only if $\overline{\mathcal{D}}$, the complement of $\mathcal{D}$, is fairly 3-colorable.

Proof:

The ONLY IF direction:

By the definition of min-rank, $\mathsf{minrk}_2(\mathcal{D}) \leq 2$ implies the existence of an $n \times n$ binary matrix $\boldsymbol{M}$ of 2-rank at most two that fits $\mathcal{D}$. There must be some two rows of $\boldsymbol{M}$ that span its entire row space. Without loss of generality, suppose that they are the first two rows of $\boldsymbol{M}$, namely, $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ (these two rows might be linearly dependent if $\mathsf{minrk}_2(\mathcal{D}) < 2$). Let A, B, and C be disjoint subsets of $\mathcal{V}(\mathcal{D})$ such that

$$\mathsf{supp}(\boldsymbol{M}_1) = A \cup B, \ \mathsf{supp}(\boldsymbol{M}_2) = B \cup C.$$

Hence,

$$\mathsf{supp}(\boldsymbol{M}_1) \cap \mathsf{supp}(\boldsymbol{M}_2) = B.$$

Since the binary alphabet is considered and the matrix $\boldsymbol{M}$ has no zero rows, for every $u \in \mathcal{V}(\mathcal{D})$, one of the following must hold: (1) $\boldsymbol{M}_u = \boldsymbol{M}_1$; (2) $\boldsymbol{M}_u = \boldsymbol{M}_2$; (3) $\boldsymbol{M}_u = \boldsymbol{M}_1 + \boldsymbol{M}_2$. Hence for every $u \in \mathcal{V}(\mathcal{D})$

$$u \in \mathsf{supp}(\boldsymbol{M}_u) \subseteq A \cup B \cup C.$$

This implies that $A \cup B \cup C = \mathcal{V}(\mathcal{D})$.

Suppose that $u \in A$. Then either $\boldsymbol{M}_u = \boldsymbol{M}_1$ or $\boldsymbol{M}_u = \boldsymbol{M}_1 + \boldsymbol{M}_2$. The former condition holds if and only if $\mathsf{supp}(\boldsymbol{M}_u) = A \cup B$, which in turns implies that $(u, v) \in \mathcal{E}(\mathcal{D})$ for all $v \in A \cup B \setminus \{u\}$. In other words, $(u, v) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $v \in A \cup B$. Here $\overline{\mathcal{D}} = (\mathcal{V}(\overline{\mathcal{D}}), \mathcal{E}(\overline{\mathcal{D}}))$ is the complement of $\mathcal{D}$. The latter condition holds if and only if $\mathsf{supp}(\boldsymbol{M}_u) = A \cup C$, which implies that $(u, v) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $v \in A \cup C$. In summary, for every $u \in A$ we have

1) $(u,v) \notin \mathcal{E}(\overline{\mathcal{D}})$, for all $v \in A$;
2) Either $(u,v) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $v \in B$, or $(u,v) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $v \in C$;

In other words, for every $u \in A$, either $\mathcal{N}_O^{\overline{\mathcal{D}}}(u) \subseteq B$ or $\mathcal{N}_O^{\overline{\mathcal{D}}}(u) \subseteq C$. Analogous conditions hold for every $u \in B$ and for every $u \in C$ as well. Therefore, by Lemma 4.6, $\overline{\mathcal{D}}$ is fairly 3-colorable.

The IF direction:

Suppose now that $\overline{\mathcal{D}}$ is fairly 3-colorable. It suffices to find an $n \times n$ binary matrix $\boldsymbol{M}$ of rank at most two that fits $\mathcal{D}$. By Lemma 4.6, there exists a partition of $\mathcal{V}(\overline{\mathcal{D}})$ into three subsets A, B, and C that satisfy the following three conditions

1) For every $u \in A$: either $N_O^{\overline{\mathcal{D}}}(u) \subseteq B$ or $N_O^{\overline{\mathcal{D}}}(u) \subseteq C$;
2) For every $u \in B$: either $N_O^{\overline{\mathcal{D}}}(u) \subseteq A$ or $N_O^{\overline{\mathcal{D}}}(u) \subseteq C$;
3) For every $u \in C$: either $N_O^{\overline{\mathcal{D}}}(u) \subseteq A$ or $N_O^{\overline{\mathcal{D}}}(u) \subseteq B$.

We construct an $n \times n$ matrix $\boldsymbol{M} = (m_{u,v})$ as follows. For each $u \in A$, if $N_O^{\overline{\mathcal{D}}}(u) \subseteq B$ then let

$$m_{u,v} = \begin{cases} 1, & v \in A \cup C \\ 0, & v \in B. \end{cases}$$

Otherwise, if $N_O^{\overline{\mathcal{D}}}(u) \subseteq C$ then let

$$m_{u,v} = \begin{cases} 1, & v \in A \cup B \\ 0, & v \in C. \end{cases}$$

For $u \in B$ and $u \in C$, $\boldsymbol{M}_u$ can be constructed analogously. It is obvious that $\boldsymbol{M}$ fits $\mathcal{D}$. Moreover, each row of $\boldsymbol{M}$ can always be written as a linear combination of the two binary vectors whose supports are $A \cup B$ and $B \cup C$, respectively. Therefore, $\mathsf{rank}_2(\boldsymbol{M}) \le 2$. The proof is complete. ■

The following corollary characterizes the digraphs of min-rank two over $\mathbb{F}_2$.

Corollary 4.8: A digraph $\mathcal{D}$ has min-rank two over $\mathbb{F}_2$ if and only if $\overline{\mathcal{D}}$ is fairly 3-colorable and $\mathcal{D}$ is not a complete digraph.

For a graph $\mathcal{G}$, it was proved by Blasiak *et al.* [26] that $\beta(\mathcal{G}) = 2$ if and only if $\overline{\mathcal{G}}$ is bipartite and $\mathcal{G}$ is not a complete graph. A characterization of digraphs $\mathcal{D}$ with $\beta(\mathcal{D}) = 2$ was also obtained therein. More specifically, it was shown that $\beta(\mathcal{D}) = 2$ if and only if $\overline{\mathcal{D}}$ does not contain a subgraph isomorphic to an *almost alternating cycle*. The almost alternating $(2m+1)$-cycle $(m \ge 1)$ is defined as follows. Its vertex set consists of all integers between $-m$ and m, inclusive, and there is an edge from i to j if and only if $j - i \in \{m, m+1\}$. Based on this characterization, a polynomial time algorithm to recognize a digraph $\mathcal{D}$ with $\beta(\mathcal{D}) = 2$ was also derived in [26]. Hence, the question whether an optimal *block* index code of length *two* exists for an ICSI instance described by a digraph can be answered in polynomial time. For *scalar linear* index codes, the same question turns out to be hard. We prove later in Section V that the decision problem whether $\mathsf{minrk}_2(\mathcal{D}) = 2$ is NP-complete.

D. Digraphs of Min-Ranks Equal to Their Orders

We start with the following definition.

Definition 4.9: A *matching* in a graph is a set of edges without common vertices. A *maximum matching* is a matching

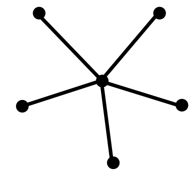

Fig. 2. A star graph.

that contains the largest possible number of edges. The number of edges in a maximum matching in $\mathcal{G}$ is denoted by $\mathsf{mm}(\mathcal{G})$.

Proposition 4.10 (Maximum-matching bound): For any graph $\mathcal{G}$ of order n, it holds that $\mathsf{minrk}_q(\mathcal{G}) \le n - \mathsf{mm}(\mathcal{G})$.

The proof follows from the clique-covering bound.

Proposition 4.11 (Folklore): Let $\mathcal{G}$ be a graph of order n. Then $\mathsf{minrk}_q(\mathcal{G}) = n$ if and only if $\mathcal{G}$ has no edges.

Proposition 4.12 (Follows from [19]): Let $\mathcal{D}$ be a digraph of order n. Then $\mathsf{minrk}_q(\mathcal{D}) = n$ if and only if $\mathcal{D}$ is acyclic.

It follows from Proposition 4.12 that the decision problem whether a digraph has min-rank equal to its order can be solved in polynomial time.

By Theorem 3.3, Proposition 4.11, and Proposition 4.12, the following corollary is straightforward.

Corollary 4.13. For a digraph $\mathcal{D}$, $\beta(\mathcal{D}) = |\mathcal{V}(\mathcal{D})|$ if and only if $\mathcal{D}$ is acyclic. For a graph $\mathcal{G}$, $\beta(\mathcal{G}) = |\mathcal{V}(\mathcal{G})|$ if and only if $\mathcal{G}$ has no edges.

E. Graphs of Min-Ranks One Less Than Their Orders

In this section, we consider (undirected) graphs. The corresponding case for digraphs is open. For a connected graph $\mathcal{G}$ of order at least two, it is easy to see that $\mathsf{mm}(\mathcal{G}) = 1$ if and only if it is a *star graph* (for an example, see Fig. 2), which is defined as follows.

Definition 4.14. A graph $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ is called a *star graph* if $|\mathcal{V}(\mathcal{G})| \ge 2$ and there exists a vertex $v \in \mathcal{V}(\mathcal{G})$ such that $\mathcal{E}(\mathcal{G}) = \big\{\{u,v\} : u \in \mathcal{V}(\mathcal{G}) \setminus \{v\}\big\}$.

It is straightforward to see that if $\mathsf{mm}(\mathcal{G}) = 1$ then $\alpha(\mathcal{G}) = n - 1$, as $\mathcal{G}$ is a star graph.

Proposition 4.15. Let $\mathcal{G}$ be a connected graph of order $n \ge 2$. Then $\mathsf{minrk}_q(\mathcal{G}) = n - 1$ if and only if $\mathsf{mm}(\mathcal{G}) = 1$ (or equivalently, $\mathcal{G}$ is a star graph).

Proof: We first suppose that $\mathsf{minrk}_q(\mathcal{G}) = n - 1$. By the maximum-matching bound, $n - 1 = \mathsf{minrk}_q(\mathcal{G}) \le n - \mathsf{mm}(\mathcal{G})$. Therefore, $\mathsf{mm}(\mathcal{G}) \le 1$. However, as $\mathsf{minrk}_q(\mathcal{G}) \ne n$, by Proposition 4.11 we have $\mathsf{mm}(\mathcal{G}) \ne 0$. Hence, $\mathsf{mm}(\mathcal{G}) = 1$.

Conversely, assume that $\mathsf{mm}(\mathcal{G}) = 1$. By the maximum-matching bound, $\mathsf{minrk}_q(\mathcal{G}) \le n - 1$. By Proposition 3.4, $\mathsf{minrk}_q(\mathcal{G}) \ge \alpha(\mathcal{G}) = n - 1$. Thus, $\mathsf{minrk}_q(\mathcal{G}) = n - 1$. ■

Corollary 4.16. Let $\mathcal{G}$ be a connected graph of order $n \ge 2$. Then $\beta(\mathcal{G}) = n - 1$ if and only if $\mathsf{mm}(\mathcal{G}) = 1$ ($\mathcal{G}$ is a star graph).

Proof: Suppose $\beta(\mathcal{G}) = n - 1$. Then either $\mathsf{minrk}_2(\mathcal{G}) = n - 1$ or $\mathsf{minrk}_2(\mathcal{G}) = n$. However, by Proposition 4.11, $\mathsf{minrk}_2(\mathcal{G}) = n$ implies that $\mathcal{G}$ has no edge. As a consequence, $\beta(\mathcal{G}) \ge \alpha(\mathcal{G}) = n$, which contradicts our assumption. Hence, $\mathsf{minrk}_2(\mathcal{G}) = n - 1$. According to Proposition 4.15, $\mathsf{mm}(\mathcal{G}) = 1$.

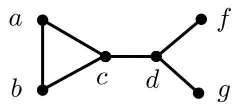

Fig. 3. The subgraph $\mathfrak{F}'$.

Conversely, suppose that $\mathsf{mm}(\mathcal{G}) = 1$. According to Proposition 4.15, we have

$$n - 1 = \alpha(\mathcal{G}) \leq \beta(\mathcal{G}) \leq \mathsf{minrk}_2(\mathcal{G}) = n - 1.$$

Hence, $\beta(\mathcal{G}) = n - 1$. ■

F. Graphs of Min-Ranks Two Less Than Their Orders

In this section, we consider (undirected) graphs. The corresponding case for digraphs is open. Here we also employ the matching language to characterize graphs of min-ranks two less than their orders.

Theorem 4.17. Suppose $\mathcal{G}$ is a connected graph of order $n \geq 6$. Then $\mathsf{minrk}_q(\mathcal{G}) = n - 2$ if and only if $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ does not contain a subgraph isomorphic to the graph $\mathfrak{F}$ depicted in Fig. 1.

The proof of this theorem appears in the Appendix.

Corollary 4.18. If $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ contains a subgraph isomorphic to $\mathfrak{F}$ (Fig. 1) then $\mathsf{minrk}_q(\mathcal{G}) = |\mathcal{V}(\mathcal{G})| - 3$.

Proof: Suppose $\mathfrak{F}'$ (Fig. 3) is a subgraph of $\mathcal{G}$ that is isomorphic to $\mathfrak{F}$.

As $\mathcal{G}$ does not have a matching of size three, each of the vertices c, f, and g is not adjacent to any vertex in $\mathcal{V}(\mathcal{G}) \setminus \mathcal{V}(\mathfrak{F}')$. Moreover, no pairs of vertices in $\mathcal{V}(\mathcal{G}) \setminus \mathcal{V}(\mathfrak{F}')$ are adjacent for the same reason. Therefore, $\{c, f, g\} \cup (\mathcal{V}(\mathcal{G}) \setminus \mathcal{V}(\mathfrak{F}'))$ is an independent set of size $|\mathcal{V}(\mathcal{G})| - 3$ in $\mathcal{G}$. Hence, $\mathsf{minrk}_q(\mathcal{G}) \geq \alpha(\mathcal{G}) \geq |\mathcal{V}(\mathcal{G})| - 3$. As $\mathsf{mm}(\mathcal{G}) = 2$, by the maximum-matching bound, $\mathsf{minrk}_q(\mathcal{G}) \leq |\mathcal{V}(\mathcal{G})| - 2$. As $\mathcal{G}$ contains $\mathfrak{F}'$, which is isomorphic to $\mathfrak{F}$, by Theorem 4.17, $\mathsf{minrk}_q(\mathcal{G}) \neq |\mathcal{V}(\mathcal{G})| - 2$. Thus, $\mathsf{minrk}_q(\mathcal{G}) = |\mathcal{V}(\mathcal{G})| - 3$. ■

Corollary 4.19. Theorem 4.17 holds verbatim if we replace $\mathsf{minrk}_q(\cdot)$ by $\beta(\cdot)$.

Proof: Suppose that $\beta(\mathcal{G}) = n-2$. Then $\mathsf{minrk}_2(\mathcal{G}) \in \{n-2, n-1, n\}$. By Proposition 4.11, Proposition 4.15, and their corollaries, for $\kappa \in \{n-1, n\}$, $\mathsf{minrk}_2(\mathcal{G}) = \kappa$ if and only if $\beta(\mathcal{G}) = \kappa$. Therefore, $\mathsf{minrk}_2(\mathcal{G}) = n - 2$. According to Theorem 4.17, $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ does not contain a subgraph isomorphic to $\mathfrak{F}$.

Conversely, as shown in the proof of Theorem 4.17 (the IF direction), $\alpha(\mathcal{G}) = \mathsf{minrk}_2(\mathcal{G}) = n - 2$. Therefore, $\beta(\mathcal{G}) = n - 2$ by Theorem 3.3. ■

V. The Hardness of the Min-Rank Problem for Digraphs

In this section, we first prove that it is an NP-complete problem to decide whether a given digraph is fairly k-colorable (see Definition 4.5), for any given $k \geq 3$. The hardness of this problem, by Lemma 4.3 and Corollary 4.8, leads to the hardness of the decision problem whether a given digraph has min-rank two over $\mathbb{F}_2$. The fair k-coloring problem is defined formally as follows.

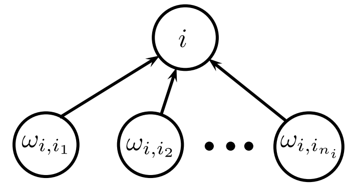

Fig. 4. Gadget $\mathcal{D}_i$ for each vertex i of $\mathcal{G}$.

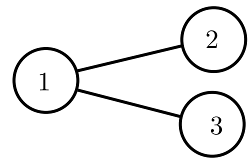

Fig. 5. An example of the graph $\mathcal{G}$.

Problem: **FAIR *k*-COLORING**
Input: A digraph $\mathcal{D}$, an integer k
Output: True if $\mathcal{D}$ is fairly k-colorable, False otherwise

Theorem 5.1. The fair k-coloring problem is NP-complete for $k \geq 3$.

Proof: This problem is obviously in NP, as the algorithm can guess a candidate for the fair coloring and verify that the candidate is indeed a fair coloring in polynomial time. For NP-hardness, we reduce the k-coloring problem to the fair k-coloring problem. Recall that the k-coloring problem is the decision problem whether a given graph is k-colorable. Suppose that $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ is an arbitrary graph. We aim to build a digraph $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ so that $\mathcal{G}$ is k-colorable if and only if $\mathcal{D}$ is fairly k-colorable. Suppose that $\mathcal{V}(\mathcal{G}) = [n]$. For each vertex $i \in [n]$, we build the following gadget, which is a digraph $\mathcal{D}_i = (\mathcal{V}_i, \mathcal{E}_i)$. The vertex set of $\mathcal{D}_i$ is

$$\mathcal{V}_i = \{i\} \cup \big\{\omega_{i,j} : \ j \in N^{\mathcal{G}}(i)\big\},$$

where $\omega_{i,j}$ are newly introduced vertices. We refer to $\omega_{i,j}$ as a *clone* (in $\mathcal{D}_i$) of the vertex $j \in [n]$. The arc set of $\mathcal{D}_i$ is

$$\mathcal{E}_i = \big\{(\omega_{i,j}, i) : \ j \in N^{\mathcal{G}}(i)\big\}.$$

Let $N^{\mathcal{G}}(i) = \{i_1, i_2, \ldots, i_{n_i}\}$. Then $\mathcal{D}_i$ can be drawn as in Fig. 4.

Additionally, we also introduce n new vertices $p_1, p_2, \ldots, p_n$. The digraph $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ is built as follows. The vertex set of $\mathcal{D}$ is

$$\mathcal{V}(\mathcal{D}) = \big(\cup_{i=1}^n \mathcal{V}_i\big) \cup \{p_1, p_2, \ldots, p_n\}.$$

Let

$$\mathcal{Q}_i = \big\{(p_i, i)\big\} \cup \big\{(p_i, \omega_{i',i}) : \ i' \in [n], \ i \in N^{\mathcal{G}}(i')\big\}$$

be the set consisting of (p_i, i) and the arcs that connect p_i and all the clones $\omega_{i',i}$ of i. The arc set of $\mathcal{D}$ is then defined to be

$$\mathcal{E}(\mathcal{D}) = \big(\cup_{i=1}^n \mathcal{E}_i\big) \cup \big(\cup_{i=1}^n \mathcal{Q}_i\big).$$

For example, if $\mathcal{G}$ is the graph in Fig. 5, then $\mathcal{D}$ is the digraph in Fig. 6.

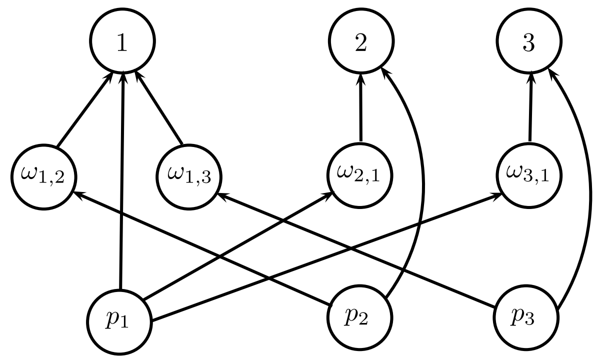

Fig. 6. The digraph $\mathcal{D}$ built from the graph $\mathcal{G}$ in Fig. 5.

Our goal now is to show that $\mathcal{G}$ is k-colorable if and only if $\mathcal{D}$ is fairly k-colorable.

Suppose that $\mathcal{G}$ is k-colorable and $\phi_{\mathcal{G}} : [n] \to [k]$ is a k-coloring of $\mathcal{G}$. We consider the mapping $\phi_{\mathcal{D}} : \mathcal{V}(\mathcal{D}) \to [k]$ defined as follows

1) For every $i \in [n]$, $\phi_{\mathcal{D}}(i) \triangleq \phi_{\mathcal{G}}(i)$;
2) If $i \in N^{\mathcal{G}}(i')$ then $\phi_{\mathcal{D}}(\omega_{i',i}) \triangleq \phi_{\mathcal{D}}(i) = \phi_{\mathcal{G}}(i)$, in other words, clones of i have the same color as i;
3) For every $i \in [n]$, $\phi_{\mathcal{D}}(p_i)$ can be chosen arbitrarily, as long as it is different from $\phi_{\mathcal{D}}(i)$.

We claim that $\phi_{\mathcal{D}}$ is a fair k-coloring for $\mathcal{D}$. We first verify the condition (C1) (see Definition 4.5). It is straightforward from the definition of $\phi_{\mathcal{D}}$ that the endpoints of each of the arcs of the forms (p_i, i) for $i \in [n]$, and $(p_i, \omega_{i',i})$ for $i \in N^{\mathcal{G}}(i')$, have different colors. It remains to check if i and $\omega_{i,j}$ for $j \in N^{\mathcal{G}}(i)$ have different colors. On the one hand, $\omega_{i,j}$ is a clone of j, and hence has the same color as j. In other words,

$$\phi_{\mathcal{D}}(\omega_{i,j}) = \phi_{\mathcal{D}}(j) = \phi_{\mathcal{G}}(j).$$

On the other hand, since $j \in N^{\mathcal{G}}(i)$, we obtain that

$$\phi_{\mathcal{G}}(j) \neq \phi_{\mathcal{G}}(i) = \phi_{D}(i).$$

Therefore, $\phi_{\mathcal{D}}(\omega_{i,j}) \neq \phi_{\mathcal{D}}(i)$ for all $i \in [n]$ and $j \in N^{\mathcal{G}}(i)$. Thus, (C1) is satisfied.

We now check if (C2) (see Definition 4.5) is also satisfied. The out-neighbors of p_i are i and its clones $\omega_{i',i}$ ($i \in N^{\mathcal{G}}(i')$). These vertices have the same color in $\mathcal{D}$, namely $\phi_{\mathcal{G}}(i)$, by the definition of $\phi_{\mathcal{D}}$. Thus (C2) is also satisfied. Therefore $\phi_{\mathcal{D}}$ is a fair k-coloring of $\mathcal{D}$.

Conversely, suppose that $\phi_{\mathcal{D}} : \mathcal{V}(\mathcal{D}) \to [k]$ is a fair k-coloring of $\mathcal{D}$. Condition (C2) guarantees that all clones of i have the same color as i, namely, $\phi_{\mathcal{D}}(\omega_{i',i}) = \phi_{\mathcal{D}}(i)$ if $i \in N^{\mathcal{G}}(i')$. Therefore, by (C1), if $\{i, j\} \in \mathcal{E}(\mathcal{G})$, that is, $j \in N^{\mathcal{G}}(i)$, then

$$\phi_{\mathcal{D}}(i) \neq \phi_{\mathcal{D}}(\omega_{i,j}) = \phi_{\mathcal{D}}(j).$$

Hence, if we define $\phi_{\mathcal{G}} : [n] \to [k]$ by $\phi_{\mathcal{G}}(i) = \phi_{\mathcal{D}}(i)$ for all $i \in [n]$, then it is a k-coloring of $\mathcal{G}$. Thus $\mathcal{G}$ is k-colorable.

Finally, notice that the order of $\mathcal{D}$ is a polynomial with respect to the order of $\mathcal{G}$. More specifically, $|\mathcal{V}(\mathcal{D})| = 2|\mathcal{V}(\mathcal{G})| + 2|\mathcal{E}(\mathcal{G})|$ and $|\mathcal{E}(\mathcal{D})| = |\mathcal{V}(\mathcal{G})| + 4|\mathcal{E}(\mathcal{G})|$. Moreover, building $\mathcal{D}$ from $\mathcal{G}$, and also obtaining a coloring of $\mathcal{G}$ from a coloring of $\mathcal{D}$, can be done in polynomial time with respect to the order of $\mathcal{G}$. Since the k-coloring problem ($k \geq 3$) is NP-hard [28], we conclude that the fair k-coloring problem is also NP-hard. ■

According to Theorem 5.1 and the work by Blasiak *et al.* [26] (see the discussion after Corollary 4.8), we obtain the following.

Theorem 5.2. Let $\mathcal{D}$ be an arbitrary digraph. Then the decision problem whether $\mathsf{minrk}_2(\mathcal{D}) = 2$ is NP-complete. However, the decision problem whether $\beta(\mathcal{D}) = 2$ can be solved in polynomial time.

Recall that by contrast, for a graph $\mathcal{G}$, it was observed by Peeters [12] that $\mathcal{G}$ has min-rank two if and only if $\overline{\mathcal{G}}$ is a bipartite graph and $\mathcal{G}$ is not a complete graph, which can be verified in polynomial time (see, for instance, West [18, p. 495]). Note that a graph is bipartite if and only if it is 2-colorable. This fact can also be derived by applying Theorem 4.7 to the digraph obtained from $\mathcal{G}$ by replacing each edge of $\mathcal{G}$ by two arcs of opposite directions.

VI. Circuit-Packing Bound Revisited

In this section, we discuss a circuit-packing bound [23] for the min-rank of a digraph. We investigate families of digraphs, whose min-ranks attain the bound and are computable in polynomial time.

A. The Bound

Let $\nu_0(\mathcal{D})$ be the *circuit packing number* of $\mathcal{D}$, namely, the maximum number of vertex-disjoint circuits in $\mathcal{D}$. Below, we reproduce an upper bound on min-ranks of digraphs, which uses the circuit packing number. This bound was first presented by Chaudhry *et al.* in [23].

Proposition 6.1 (Circuit-packing bound, [23]): The following holds for every digraph $\mathcal{D}$ of order n:

$$\mathsf{minrk}_q(\mathcal{D}) \leq n - \nu_0(\mathcal{D}).$$

Whereas for graphs the clique-cover bound is the best known bound, for digraphs that are not symmetric, this is not the case. The worst scenario for the clique-cover bound is when the digraph has no two arcs of opposite directions. For such a digraph, this bound becomes trivial, as the size of the smallest clique cover is equal to the order of the digraph. The following example emphasizes the fact that for certain digraphs, the circuit-packing bound can be *significantly tighter* than the clique-cover bound.

Example 6.2. Let $\mathcal{D}_k$ be the digraph of order $n = 3k$ depicted in Fig. 7. As there are no arcs of opposite directions, all cliques in $\mathcal{D}_k$ are of cardinality one. Therefore, the clique-cover bound gives $\mathsf{minrk}_q(\mathcal{D}_k) \leq 3k$. On the other hand, as $\mathcal{D}_k$ contains k vertex-disjoint circuits, namely $\mathcal{C}_i = (3i + 1, 3i + 2, 3i + 3)$ for $i = 0, 1, \ldots, k - 1$, the circuit-packing bound yields $\mathsf{minrk}_q(\mathcal{D}_k) \leq 2k = 3k - k$. The gap between the two bounds is one third of the order of the digraph.

B. Digraphs Attaining Circuit-Packing Bound

In this subsection, we present several new examples of families of digraphs that attain the circuit-packing bound.

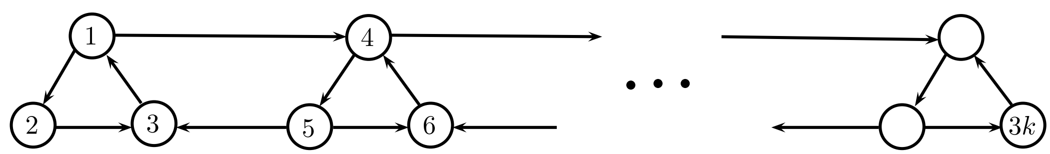

Fig. 7. Example where the circuit-packing bound is tighter than the clique-cover bound.

A *feedback vertex (arc, respectively) set* of $\mathcal{D}$ is a set of vertices (arcs, respectively) whose removal destroys all circuits in $\mathcal{D}$. Let $\tau_0(\mathcal{D})$ ($\tau_1(\mathcal{D})$, respectively) denote the *minimum size* of a feedback vertex (arc, respectively) set of $\mathcal{D}$. Then it is clear that $\alpha(\mathcal{D}) = n - \tau_0(\mathcal{D})$.

Corollary 6.3. If $\nu_0(\mathcal{D}) = \tau_0(\mathcal{D})$ then

$$\mathsf{minrk}_q(\mathcal{D}) = n - \nu_0(\mathcal{D}) = n - \tau_0(\mathcal{D}). \tag{1}$$

Proof: By Proposition 3.4 and Proposition 6.1 we have

$$n - \tau_0(\mathcal{D}) \le \mathsf{minrk}_q(\mathcal{D}) \le n - \nu_0(\mathcal{D}).$$

Hence, the proof follows. ∎

When $\mathcal{D}$ satisfies $\nu_0(\mathcal{D}) = \tau_0(\mathcal{D})$, we say that $\mathcal{D}$ satisfies the *min-max vertex equality*. In that case, the circuit-packing bound is tight. Similarly, let $\nu_1(\mathcal{D})$ denote the maximum number of arc-disjoint circuits in $\mathcal{D}$. We say that $\mathcal{D}$ satisfies the *min-max arc equality* if $\nu_1(\mathcal{D}) = \tau_1(\mathcal{D})$.

An example of digraphs that satisfy the min-max vertex equality is the *connectively reducible digraphs* [29]. This family of digraphs contains both the family of *fully reducible flow digraphs* [30] and the family of *cyclically reducible digraphs* [31] as special cases. A polynomial time algorithm was provided by Szwarcfiter [29] to recognize a member of this family and subsequently find a maximum set of vertex-disjoint circuits as well as a minimum feedback vertex set. Therefore, by Corollary 6.3, (1) holds for a connectively reducible digraph D. Moreover, $\mathsf{minrk}_q(\mathcal{D})$ can be found in polynomial time.

Another example of digraphs for which the circuit-packing bound is tight are the digraphs that *pack* [32]. A digraph packs if the min-max vertex equality holds for all of its subgraphs. The digraphs in this family are exactly ones that have no minor isomorphic to an odd double circuit or F_7, a special digraph of order 7 (interested readers may refer to [32] for more details, also for a structural characterization of this family of digraphs). For instance, *strongly planar* digraphs [32] belong to this family. As far as we know, there are no known polynomial time algorithms to find a minimum feedback vertex set of a digraph that packs.

Other examples of digraphs for which the circuit-packing bound is tight are the *line digraphs* of planar digraphs, of fully reducible flow digraphs, and of (special) Eulerian digraphs [33].

Definition 6.4. Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph. Then the digraph $\mathcal{L} = (\mathcal{V}(\mathcal{L}), \mathcal{E}(\mathcal{L}))$ with $\mathcal{V}(\mathcal{L}) = \mathcal{E}(\mathcal{D})$ and

$$\mathcal{E}(\mathcal{L}) = \big\{(e, e') : \; e = (u, v) \in \mathcal{E}(\mathcal{D}), \; e' = (v, w) \in \mathcal{E}(\mathcal{D})\big\},$$

is called the *line digraph* of $\mathcal{D}$. We denote the line digraph of $\mathcal{D}$ by $\mathcal{L}(\mathcal{D})$. The digraph $\mathcal{D}$ is called a *root digraph* of $\mathcal{L}(\mathcal{D})$.

Lemma 6.5. $\nu_0(\mathcal{L}(\mathcal{D})) = \nu_1(\mathcal{D})$.

The proof of this lemma appears in the Appendix.

Lemma 6.6. $\tau_0(\mathcal{L}(\mathcal{D})) = \tau_1(\mathcal{D})$.

The proof of this lemma appears in the Appendix.

Proposition 6.7. Let $\mathcal{D}$ be a digraph. If $\nu_1(\mathcal{D}) = \tau_1(\mathcal{D})$ then $\nu_0(\mathcal{L}(\mathcal{D})) = \tau_0(\mathcal{L}(\mathcal{D}))$ and

$$\mathsf{minrk}_q(\mathcal{L}(\mathcal{D})) = |\mathcal{E}(\mathcal{D})| - \nu_1(\mathcal{D}).$$

Proof: Suppose that $\nu_1(\mathcal{D}) = \tau_1(\mathcal{D})$. By Lemma 6.5 and Lemma 6.6, $\nu_0(\mathcal{L}(\mathcal{D})) = \tau_0(\mathcal{L}(\mathcal{D}))$. Therefore, by applying Corollary 6.3 to $\mathcal{L}(\mathcal{D})$ we obtain

$$\mathsf{minrk}_q(\mathcal{L}(\mathcal{D})) = |\mathcal{V}(\mathcal{L}(\mathcal{D}))| - \nu_0(\mathcal{L}(\mathcal{D})) = |\mathcal{E}(\mathcal{D})| - \nu_1(\mathcal{D}).$$

∎

Definition 6.8. A digraph that can be drawn on a plane in such a way that its (arcs) edges intersect only at their endpoints is called *planar*.

It is known that the min-max arc equality is satisfied for planar digraphs [34], for fully reducible flow digraphs [35], and for a special family of Eulerian digraphs [33]. Therefore, by Proposition 6.7, the min-max vertex equality is satisfied for the line digraphs of the members of these families. In summary, we have the following.

Corollary 6.9. The circuit-packing bound is tight for the following families of digraphs: connectively reducible digraphs, digraphs that pack, line digraphs of planar digraphs, line digraphs of fully reducible flow digraphs, and line digraphs of special Eulerian digraphs.

Consider the ICSI instances described by digraphs $\mathcal{D}$ with $\mathsf{minrk}_2(\mathcal{D}) = \alpha(\mathcal{D})$. By Theorem 3.3, $\mathsf{minrk}_2(\mathcal{D}) = \beta(\mathcal{D})$. Hence, for such instances, *scalar linear* index codes are as good as *block* index codes, in terms of transmission rates. Thus, for the ICSI instances described by families of digraphs listed in Corollary 6.9, scalar linear index codes achieve the best possible transmission rates. Previously, only perfect graphs and acyclic digraphs were known to have this property [19].

Definition 6.10. A digraph is called *partially planar* if all of its strongly connected components are planar.

Proposition 6.11. There is a polynomial time algorithm to recognize the line digraph of a partially planar digraph and subsequently determine its min-rank.

Proof: We present an algorithm as claimed. It consists of two phases.

1) **Recognition Phase:** To determine whether a given digraph $\mathcal{L}$ is the line digraph of a partially planar digraph, it suffices to determine whether each of its strongly connected components $\mathcal{L}_i$ ($i \in [k]$) is the line digraph of a planar digraph. All strongly connected

components of a digraph can be found in time linear in the number of edges [36].

For each $i \in [k]$, we can efficiently determine whether $\mathcal{L}_i$ is a line digraph of a digraph [37]. If yes, the algorithm outputs a digraph $\mathcal{D}'_i$, which is a root digraph of $\mathcal{L}_i$ and is strongly connected.

Suppose $\mathcal{L} = \mathcal{L}(\mathcal{D})$, where $\mathcal{D}$ is a digraph. Let $\mathcal{L}_i = \mathcal{L}(\mathcal{D}_i)$, where $\mathcal{D}_i$'s, $i \in [k]$, are all strongly connected components of $\mathcal{D}$ of order $\geqslant 2$. By [38, Theorem 3], $\mathcal{D}'_i$ and $\mathcal{D}_i$ are isomorphic, $i \in [k]$. To complete the Recognition Phase, one tests the planarity of $\mathcal{D}'_i$ for every $i \in [k]$. This can be done in time linear in the size of $\mathcal{D}$ [39].

2) **Min-Rank Computation Phase:** If $\mathcal{L}$ is the line digraph of a partially planar digraph, then $\mathsf{minrk}_q(\mathcal{L})$ can be computed efficiently. Indeed, by Lemma 4.2, it suffices to show that $\mathsf{minrk}_q(\mathcal{L}_i)$ for $i \in [k]$ can be found in polynomial time.

Since $\mathcal{D}'_i$ (which is isomorphic to $\mathcal{D}_i$) is planar, as it is shown in [34], $\nu_1(\mathcal{D}'_i) = \tau_1(\mathcal{D}'_i)$. Therefore, by Proposition 6.7,

$$\mathsf{minrk}_q(\mathcal{L}_i) = |\mathcal{E}(\mathcal{D}'_i)| - \nu_1(\mathcal{D}'_i),$$

where $\nu_1(\mathcal{D}'_i)$ can be computed efficiently ([40]). Therefore, $\mathsf{minrk}_q(\mathcal{L})$ can be computed efficiently. ■

In summary, we have the following result.

Corollary 6.12. There are polynomial time algorithms to recognize a member and subsequently determine the min-rank of that member of the following families of digraphs: connectively reducible digraphs (which includes fully reducible flow digraphs and cyclically reducible digraphs), and line digraphs of partially planar digraphs.

VII. Conclusion

We studied the ICSI instances whose optimal scalar linear index codes have near-extreme transmission rates. We presented new characterizations of side-information graphs with min-ranks $n-1$ and $n-2$ over a general finite field, and of digraphs with min-rank two over the binary field. We also showed that the decision problem whether a digraph has min-rank two (over a general finite field) is NP-complete. Finally, we presented several families of digraphs, whose min-ranks can be found efficiently.

Appendix

Proof of Lemma 4.2: Suppose that $\mathcal{V}_i$ is the set of vertices that induces $\mathcal{D}_i$, $i \in [k]$. Then $\{\mathcal{V}_i\}_{i\in[k]}$ forms a partition of $\mathcal{V}(\mathcal{D})$. By relabeling the vertices of $\mathcal{D}$ if necessary, we may assume without loss of generality that for every $i < j$

1) $u < v$ whenever $u \in \mathcal{V}_i$ and $v \in \mathcal{V}_j$;
2) There are no arcs of the form (v, u) where $u \in \mathcal{V}_i$ and $v \in \mathcal{V}_j$.

If $\boldsymbol{M}^{(i)}$ is a minimum-rank matrix that fits $\mathcal{D}_i$ ($i \in [k]$) then the diagonal block matrix $\boldsymbol{M}$ whose diagonal blocks are $\boldsymbol{M}^{(i)}$ clearly fits $\mathcal{D}$. Moreover,

$$\mathsf{rank}_q(\boldsymbol{M}) = \sum_{i=1}^{k} \mathsf{rank}_q(\boldsymbol{M}^{(i)}) = \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i).$$

Hence $\mathsf{minrk}_q(\mathcal{D}) \le \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i)$. It remains to show that $\mathsf{minrk}_q(\mathcal{D}) \ge \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i)$. Suppose that the matrix $\boldsymbol{M}$ fits $\mathcal{D}$. By the assumptions on $\mathcal{V}_i$'s ($i \in [k]$) stated at the beginning of the proof, $\boldsymbol{M}$ must be an upper-triangular block matrix. If we let $\boldsymbol{M}^{(i)}$ be the sub-matrix of $\boldsymbol{M}$ formed by the rows and columns indexed by the elements of $\mathcal{V}_i$, then $\boldsymbol{M}^{(i)}$ fits $\mathcal{D}_i$ and hence,

$$\mathsf{rank}_q(\boldsymbol{M}) \ge \sum_{i=1}^{k} \mathsf{rank}_q(\boldsymbol{M}^{(i)}) \ge \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i).$$

Thus, $\mathsf{minrk}_q(\mathcal{D}) \ge \sum_{i=1}^{k} \mathsf{minrk}_q(\mathcal{D}_i)$.

Proof of Theorem 4.17: For the ONLY IF direction, suppose that $\mathsf{minrk}_q(\mathcal{G}) = n-2$. By the maximum-matching bound, $n-2 \le n - \mathsf{mm}(\mathcal{G})$. Hence $\mathsf{mm}(\mathcal{G}) \le 2$. As $\mathsf{mm}(\mathcal{G}) \in \{0,1\}$ and $|\mathcal{V}(\mathcal{G})| \ge 6$ imply that either $\mathcal{G}$ has no edges ($\mathsf{minrk}_q(\mathcal{G}) = n > n-2$) or $\mathcal{G}$ is a star graph ($\mathsf{minrk}_q(\mathcal{G}) = n-1 > n-2$), we deduce that $\mathsf{mm}(\mathcal{G}) = 2$. Moreover, as the graph $\mathfrak{F}$ has min-rank *three* less than its order, $\mathcal{G}$ should not contain any subgraph isomorphic to $\mathfrak{F}$. Indeed, suppose for otherwise that $\mathfrak{F}'$ is a subgraph of $\mathcal{G}$ and $\mathfrak{F}'$ is isomorphic to $\mathfrak{F}$.

Consider the following block diagonal matrix $\boldsymbol{M}$ with two blocks $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$. The first block $\boldsymbol{B}_1$, a 6×6 matrix, corresponds to the rows and columns labeled by the vertices in $\mathfrak{F}'$. Moreover, we choose $\boldsymbol{B}_1$ so that it has q-rank three. This is possible since $\mathfrak{F}'$ is isomorphic to $\mathfrak{F}$ and $\mathsf{minrk}_q(\mathfrak{F}) = 3$. (Note that $3 = \alpha(\mathfrak{F}) \le \mathsf{minrk}_q(\mathfrak{F}) \le \mathsf{cc}(\mathfrak{F}) = 3$ implies that $\mathsf{minrk}_q(\mathfrak{F}) = 3$.) The second block $\boldsymbol{B}_2$ is chosen to be an $(n-6) \times (n-6)$ identity matrix. It corresponds to the rows and columns labeled by the vertices in $\mathcal{V}(\mathcal{G}) \setminus \mathcal{V}(\mathfrak{F}')$. Then $\boldsymbol{M}$ fits $\mathcal{G}$ and moreover,

$$\begin{aligned}\mathsf{rank}_q(\boldsymbol{M}) &= \mathsf{rank}_q(\boldsymbol{B}_1) + \mathsf{rank}_q(\boldsymbol{B}_2)\\ &= 3 + (n-6)\\ &= n-3.\end{aligned}$$

This implies that $\mathsf{minrk}_q(\mathcal{G}) \le n-3 < n-2$, which is impossible.

We now turn to the IF direction. Suppose that $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ does not contain any subgraph isomorphic to $\mathfrak{F}$. Then by the maximum-matching bound, $\mathsf{minrk}_q(\mathcal{G}) \le n-2$. As $\alpha(\mathcal{G}) \le \mathsf{minrk}_q(\mathcal{G})$, it suffices to show that $\alpha(\mathcal{G}) = n-2$.

Let $\{a,b\}$ and $\{c,d\}$ be the two edges of a maximum matching M in $\mathcal{G}$. Let $U = \{a,b,c,d\}$ and $V = \mathcal{V}(\mathcal{G}) \setminus U$. As $\mathcal{G}$ has at least six vertices, suppose that $V = \{f, g, \ldots\}$, where $f \ne g$. Since M is a maximum matching, V must be an independent set in $\mathcal{G}$. The idea is to show that we can always find two nonadjacent vertices in U that are not adjacent to any vertex in V. Such two vertices can be added to V to obtain an independent set of size $n-2$, which establishes the proof. We refer to such a pair of vertices as an *independent pair*.

For disjoint subsets I and J of $\mathcal{V}(\mathcal{G})$, let

$$s_{\mathcal{G}}(I,J) = \Big|\big\{\{i,j\}:\ i \in I,\ j \in J,\ \{i,j\} \in \mathcal{E}(\mathcal{G})\big\}\Big|.$$

Based on how the vertices in U are connected to each other, we consider the following five cases. Note that we only consider non-isomorphic configurations.

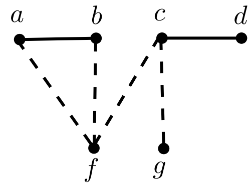

Fig. 8. Case 1.

Case 1: $s_{\mathcal{G}}(\{a,b\},\{c,d\}) = 0$.

There are four candidates for an independent pair, namely $\{a,c\}$, $\{a,d\}$, $\{b,c\}$, $\{b,d\}$. All of these pairs fail to be an independent pair if and only if either both a and b are adjacent to some vertices in V or both c and d are adjacent to some vertices in V. We show that either case never happens, by contradiction.

Suppose both a and b are adjacent to some vertices in V. (The case when both c and d are adjacent to some vertices in V is investigated analogously.) Without loss of generality, assume that a and f are adjacent. Then b must be adjacent to f but not to any other vertex in $\mathcal{V}$. Indeed, if b is adjacent to $h \in V$, $h \neq f$, then the set of three edges $\{a,f\}$, $\{b,h\}$, and $\{c,d\}$ form a matching of size three, which is impossible since $\mathsf{mm}(\mathcal{G}) = 2$. Similarly, a should not be adjacent to any other vertex in V rather than f.

As $\mathcal{G}$ is connected, f must be adjacent to either c or d. Without loss of generality, suppose f and c are adjacent. On the one hand, since $\mathcal{G}$ is connected, g must be adjacent to some vertex in U. On the other hand, g cannot be adjacent to any vertex in U, as

- if g and a are adjacent, then $\{a,g\}$, $\{b,f\}$, and $\{c,d\}$ form a matching of size three, which is impossible;
- if g and b are adjacent, then $\{a,f\}$, $\{b,g\}$, and $\{c,d\}$ form a matching of size three, which is impossible;
- if g and c are adjacent, then $\mathcal{G}$ has a subgraph isomorphic to $\mathfrak{F}$ (see Fig. 8), which is impossible;
- if g and d are adjacent, then $\{a,b\}$, $\{c,f\}$, and $\{d,g\}$ form a matching of size three, which is impossible.

We obtain a contradiction.

Case 2: $s_{\mathcal{G}}(\{a,b\},\{c,d\}) = 1$. Without loss of generality, suppose that $\{b,c\}$ is the only edge that connects $\{a,b\}$ and $\{c,d\}$.

There are three candidates for an independent pair, namely $\{a,c\}$, $\{a,d\}$, and $\{b,d\}$. All of these three pairs fail to be an independent pair only if at least one of the pairs $\{a,b\}$, $\{a,d\}$, and $\{c,d\}$ has both vertices adjacent to some vertices in V. We show below that this scenario cannot happen.

1) Assume that both a and b are adjacent to some vertices in V.
   Suppose without loss of generality that a and f are adjacent. Then the same argument as in Case 1 establishes that b must be adjacent to f but not to any other vertex in V. On the one hand, as $\mathcal{G}$ is connected, g must be adjacent to some vertex in U. On the other hand, as $\mathsf{mm}(\mathcal{G}) = 2$, g should not be adjacent to any vertex among a, b, and d. Moreover, g and c cannot be adjacent, for otherwise $\mathcal{G}$ would contain a subgraph isomorphic to $\mathfrak{F}$ (see Fig. 9). We obtain a contradiction.

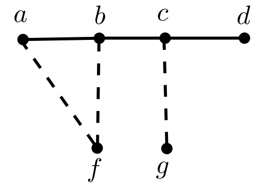

Fig. 9. Sub-case 1.

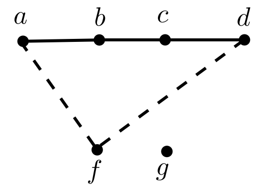

Fig. 10. Sub-case 2.

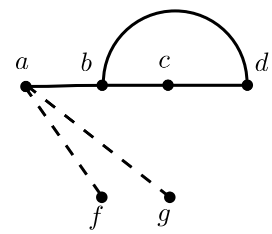

Fig. 11. Case 3, a possible arrangement of edges.

2) Assume that both a and d are adjacent to some vertices in V (Fig. 10). Suppose without loss of generality that a and f are adjacent. As there are no matchings of size three in $\mathcal{G}$, d is adjacent to f but not to any other vertex in V. Also, g is not adjacent to any vertex in U. However, this would imply that g is an isolated vertex of $\mathcal{G}$, which is impossible as $\mathcal{G}$ is connected.
3) Assume that both c and d are adjacent to some vertices in V. This sub-case is completely similar to the first sub-case.

Case 3: $s_{\mathcal{G}}(\{a,b\},\{c,d\}) = 2$ and the two edges that connect $\{a,b\}$ and $\{c,d\}$ share one common vertex. Without loss of generality suppose that these two edges are $\{b,c\}$ and $\{b,d\}$.

There are two candidates for an independent pair, namely $\{a,c\}$ and $\{a,d\}$. It suffices to show that a is not adjacent to any vertex in V and either c or d is not adjacent to any vertex in V.

Suppose that a is adjacent to a vertex, say f, in V. As $\mathsf{mm}(\mathcal{G}) = 2$, we deduce that g is not adjacent to any vertex among b, c, and d. Also, since $\mathcal{G}$ does not contain a subgraph isomorphic to $\mathfrak{F}$, we deduce that g cannot be adjacent to a (see Fig. 11). Hence g is an isolated vertex of $\mathcal{G}$, which is impossible as $\mathcal{G}$ is connected.

Now suppose that both c and d are adjacent to some vertices in V. Without loss of generality, suppose that c is adjacent to f. Then since $\mathsf{mm}(\mathcal{G}) = 2$, d must be adjacent to f but not to any other vertex in V. Also, g cannot be adjacent to any vertex among a, c, and d for the same reason. Moreover, as $\mathcal{G}$ does not contain a subgraph isomorphic to $\mathfrak{F}$, we deduce

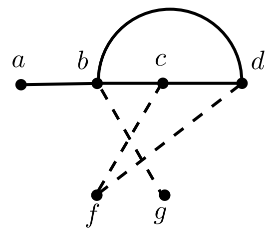

Fig. 12. Case 3, another possible arrangement of edges.

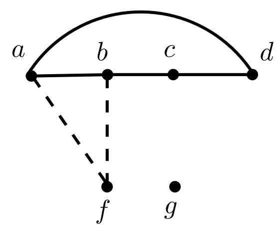

Fig. 13. Case 4.

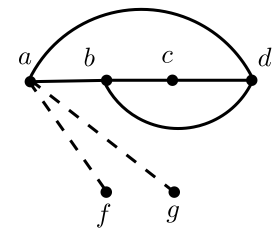

Fig. 14. Case 5.

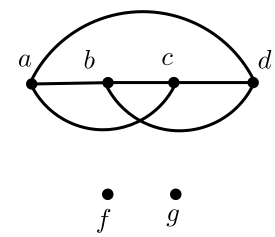

Fig. 15. Case 6.

that g is not adjacent to b (see Fig. 12). (Indeed, if g and b are adjacent, then the following subgraph of $\mathcal{G}$ is isomorphic to $\mathfrak{F}$: its vertex set is $\{a, b, c, d, f, g\}$, and its edge set is $\big\{\{c, d\}, \{d, f\}, \{c, f\}, \{c, b\}, \{b, a\}, \{b, g\}\big\}$.) Therefore, g is an isolated vertex of $\mathcal{G}$. We obtain a contradiction.

Case 4: $s_{\mathcal{G}}(\{a, b\}, \{c, d\}) = 2$ and the two edges that connect $\{a, b\}$ and $\{c, d\}$ share no common vertices. Suppose, without loss of generality, that these two edges are $\{a, d\}$ and $\{b, c\}$ (Fig. 13).

There are two candidates for an independent pair, namely $\{a, c\}$ and $\{b, d\}$. Both of these pairs fail to be an independent pair if and only if at least one of the four pairs $\{a, b\}$, $\{a, d\}$, $\{c, b\}$, and $\{c, d\}$ has both vertices adjacent to some vertices in V. By symmetry, it suffices to show that the scenario when both a and b are adjacent to some vertices in V never happens.

Suppose now that a and b are adjacent to some vertices in V.

Suppose that a and f are adjacent. The condition that $\mathsf{mm}(\mathcal{G}) = 2$ forces b to be adjacent to f but not to any other vertex in V. That condition also implies that g must be an isolated vertex in $\mathcal{G}$, which is impossible as $\mathcal{G}$ is connected.

Case 5: $s_{\mathcal{G}}(\{a, b\}, \{c, d\}) = 3$. Without loss of generality, suppose that $\{a, d\}$, $\{b, c\}$, and $\{b, d\}$ are the edges that connect $\{a, b\}$ and $\{c, d\}$. The only candidate for an independent pair is $\{a, c\}$. We prove by contradiction that both a and c are not adjacent to any vertex in V. By symmetry, it suffices to verify this property for only one of them.

Suppose that a is adjacent to some vertex in V. Let a be adjacent to f.

As $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ is connected, g must be adjacent to a. However, $\mathcal{G}$ now contains a subgraph whose edge set consists of $\{b, c\}, \{b, d\}, \{c, d\}, \{b, a\}, \{a, f\}, \{a, g\}$, which is isomorphic to $\mathfrak{F}$ (see Fig. 14). This contradicts our assumption.

Case 6: $s_{\mathcal{G}}(\{a, b\}, \{c, d\}) = 4$. In this case, the subgraph of $\mathcal{G}$ induced by $\{a, b, c, d\}$ is a complete graph (Fig. 15).

As $\mathcal{G}$ is connected, both f and g must be adjacent to some vertices in U. If f and g are adjacent to the same vertex in U, then $\mathcal{G}$ contains a subgraph isomorphic to $\mathfrak{F}$, which contradicts our assumption. For instance, if both f and g are adjacent to a, then this subgraph has vertex set $\{a, b, c, d, f, g\}$ and edge set consisting of the edges $\{b, c\}, \{c, d\}, \{b, d\}, \{b, a\}, \{a, f\}, \{a, g\}$. It is also easy to verify that if f and g are adjacent to different vertices in U, then $\mathcal{G}$ contains a matching of size three. This contradicts our assumption that $\mathsf{mm}(\mathcal{G}) = 2$. Thus, Case 6 never happens. ∎

Proof of Lemma 6.5:

1) $\nu_0(\mathcal{L}(\mathcal{D})) \geq \nu_1(\mathcal{D})$. It suffices to show that the existence of a set of arc-disjoint circuits in $\mathcal{D}$ implies the existence of a set of vertex-disjoint circuits of the same size in $\mathcal{L}(\mathcal{D})$. Let $\{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k\}$ be a set of arc-disjoint circuits in $\mathcal{D}$, where $\mathcal{C}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,r_i})$, $r_i \geq 2$, $i \in [k]$. Let $e_{i,j} = (v_{i,j}, v_{i,j+1})$, for $i \in [k]$ and $j \in [r_i - 1]$. Moreover, let $e_{i,r_i} = (v_{i,r_i}, v_{i,1})$ for $i \in [k]$. Let $\mathcal{C}'_i = (e_{i,1}, e_{i,2}, \dots, e_{i,r_i})$ for $i \in [k]$. Then $\mathcal{C}'_i$ is also a circuit in $\mathcal{L}(\mathcal{D})$ for every $i \in [k]$. Moreover, as the circuits $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_k$ share no common edges in $\mathcal{D}$, we deduce that $\mathcal{C}'_1, \mathcal{C}'_2, \dots, \mathcal{C}'_k$ share no common vertices in $\mathcal{L}(\mathcal{D})$. Therefore, they form a set of k vertex-disjoint circuits in $\mathcal{L}(\mathcal{D})$.
2) $\nu_0(\mathcal{L}(\mathcal{D})) \leq \nu_1(\mathcal{D})$. It suffices to show that the existence of a set of vertex-disjoint circuits in $\mathcal{L}(\mathcal{D})$ implies the existence of a set of arc-disjoint circuits of the same size in $\mathcal{D}$. Let $\{\mathcal{C}'_1, \mathcal{C}'_2, \dots, \mathcal{C}'_k\}$ be a set of vertex-disjoint circuits in $\mathcal{L}(\mathcal{D})$, where $\mathcal{C}'_i = \{e_{i,1}, e_{i,2}, \dots, e_{i,r_i}\}$ for $i \in [k]$. Suppose that $e_{i,j} = (v_{i,j}, v_{i,j+1}) \in \mathcal{E}(\mathcal{D})$ for $i \in [k]$ and $j \in [r_i]$, where $v_{i,j}$ and $v_{i,j+1}$ are vertices of $\mathcal{D}$. Then $v_{i,r_i+1} \equiv v_{i,1}$ for $i \in [k]$. For each $i \in [k]$, consider the sequence of (possibly repeated) vertices
$$v_{i,1}, v_{i,2}, \dots, v_{i,r_i+1}.$$
Since $v_{i,1} \equiv v_{i,r_i+1}$ and $(v_{i,j}, v_{i,j+1}) \in \mathcal{E}(\mathcal{D})$ for all $j \in [r_i]$, there exist j_0 and j_1 such that

- $1 \leq j_0 < j_1 \leq r_i$;
- $v_{i,j_0} \equiv v_{i,j_1+1}$;
- $v_{i,j_0}, v_{i,j_0+1}, \ldots, v_{i,j_1}$ are distinct.

Then $\mathcal{C}_i = (v_{i,j_0}, v_{i,j_0+1}, \ldots, v_{i,j_1})$ is a circuit in $\mathcal{D}$. Since the circuits $\mathcal{C}'_1, \mathcal{C}'_2, \ldots, \mathcal{C}'_k$ share no common vertices in $\mathcal{L}(\mathcal{D})$, we obtain that the circuits $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ share no common edges in $\mathcal{D}$. ■

Proof of Lemma 6.6: Let $F = \{e_1, e_2, \ldots, e_k\}$, where $e_i \in \mathcal{E}(\mathcal{D})$ for $i \in [k]$, be an arbitrary set of arcs of $\mathcal{D}$. We can also view F as a set of vertices of $\mathcal{L}(\mathcal{D})$. It suffices to show that F is a feedback arc set of $\mathcal{D}$ if and only if F is a feedback vertex set of $\mathcal{L}(\mathcal{D})$, for every such set F.

Let $\mathcal{D} - F$ be the digraph obtained from $\mathcal{D}$ by removing all arcs in F. Let $\mathcal{L}(\mathcal{D}) - F$ be the digraph obtained from $\mathcal{L}(\mathcal{D})$ by removing all vertices in F. Then $\mathcal{L}(\mathcal{D}) - F = \mathcal{L}(\mathcal{D} - F)$. As shown in the proof of Lemma 6.5, the existence of a circuit in $\mathcal{D} - F$ would result in the existence of a circuit in $\mathcal{L}(\mathcal{D} - F)$ and vice versa. Therefore, $\mathcal{D} - F$ is acyclic if and only if $\mathcal{L}(\mathcal{D}) - F$ is acyclic. Thus, F is a feedback arc set of $\mathcal{D}$ if and only if F is a feedback vertex set of $\mathcal{L}(\mathcal{D})$. ■

Acknowledgment

The authors wish to thank M. Langberg for providing the preprints [14], [15].

References

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

[2] Y. Birk and T. Kol, "Coding-on-demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2825–2830, Jun. 2006.

[3] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[4] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 782–795, Oct. 2003.

[5] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. 47th Annu. IEEE Symp. FOCS*, Oct. 2006, pp. 197–206.

[6] 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.

[7] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "XORs in the air: Practical wireless network coding," in *Proc. ACM SIGCOMM*, Sep. 2006, pp. 243–254.

[8] S. Katti, D. Katabi, H. Balakrishnan, and M. Médard, "Symbol-level network coding for wireless mesh networks," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 401–412, Oct. 2008.

[9] W. Haemers, "An upper bound for the Shannon capacity of a graph," *Colloq. Math. Soc. János Bolyai: Algebraic Methods in Graph Theory*, vol. 25. Szeged, Hungary, 1978, pp. 267–272.

[10] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 2, no. 3, pp. 3–15, Sep. 1956.

[11] L. Lovász, "On the Shannon capacity of a graph," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, pp. 1–7, Jan. 1979.

[12] R. Peeters, "Orthogonal representations over finite fields and the chromatic number of graphs," *Combinatorica*, vol. 16, no. 3, pp. 417–431, 1996.

[13] M. A. R. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proc. IEEE INFOCOM Workshops*, Apr. 2008, pp. 1–4.

[14] I. Haviv and M. Langberg, "On linear index coding for random graphs," in *Proc. IEEE ISIT*, Boston, MA, USA, Jul. 2012, pp. 2231–2235.

[15] Y. Berliner and M. Langberg, "Index coding with outerplanar side information," in *Proc. IEEE ISIT*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 869–873.

[16] E. Chlamtac and I. Haviv, "Linear index coding via semidefinite programming," in *Proc. 23rd Annu. ACM-SODA*, 2012, pp. 406–419.

[17] H. Maleki, V. Cadambe, and S. Jafar, "Index coding: An interference alignment perspective," in *Proc. IEEE ISIT*, Boston, MA, USA, Jul. 2012, pp. 2236–2240.

[18] D. B. West, *Introduction to Graph Theory*, 2nd ed. Prentice-Hall, 1996.

[19] 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.

[20] D. Traskov, M. Médard, P. Sadeghi, and R. Koetter, "Joint scheduling and instantaneously decodable network coding," in *Proc. 28th IEEE Conf. Global Telecommun.*, Nov./Dec. 2009, pp. 3835–3840.

[21] C. Zhan and Y. Xu, "Broadcast scheduling based on network coding in time critical wireless networks," in *Proc. IEEE Int. Symp. Netw. Coding*, Toronto, ON, Canada, Jun. 2010, pp. 1–6.

[22] S. H. Dau and Y. M. Chee. (2013). Polynomial time algorithm for min-ranks of graphs with simple tree structures. *Algorithmica*, [Online]. Available: http://dx.doi.org/10.1007/s00453-013-9789-9

[23] M. A. R. Chaudhry, Z. Asad, A. Sprintson, and M. Langberg, "On the complementary index coding problem," in *Proc. IEEE ISIT*, Saint Petersburg, Russia, Jul./Aug. 2011, pp. 306–310.

[24] E. Lubetzky and U. Stav, "Non-linear index coding outperforming the linear optimum," in *Proc. 48th Annu. IEEE Symp. FOCS*, Oct. 2007, pp. 161–168.

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

[26] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Index coding via linear programming," 2011 [Online]. Available: http://arxiv.org/abs/1004.1379

[27] M. Langberg and A. Sprintson, "On the hardness of approximating the network coding capacity," in *Proc. IEEE ISIT*, Toronto, ON, Canada, Jul. 2008, pp. 315–319.

[28] R. M. Karp, "Reducibility among combinatorial problems," *Complex. Comput. Comput.*, vol. 40, no. 4, pp. 85–103, 1972.

[29] J. L. Szwarcfiter, "On minimum cuts of cycles and maximum disjoint cycles," in *Graphs and Algorithms*, vol. 89. Providence, RI, USA: AMS, 1989, pp. 153–166.

[30] A. Frank and A. Gyarfas, "Directed graphs and computer programs," *Problèmes Combinatoires Théorie Graphes*, no. 260, pp. 157–158, 1976.

[31] C.-C. Wang, E. L. Lloyd, and M. L. Soffa, "Feedback vertex sets and cyclically reducible graphs," *J. ACM*, vol. 32, no. 2, pp. 296–313, 1985.

[32] B. Guenin and R. Thomas, "Packing directed circuits exactly," *Combinatorica*, vol. 31, no. 4, pp. 397–421, 2011.

[33] P. D. Seymour, "Packing circuits in Eulerian digraphs," *Combinatorica*, vol. 16, no. 2, pp. 223–231, 1996.

[34] C. Lucchesi and D. Younger, "A minimax theorem for directed graphs," *J. London Math. Soc.*, vol. 17, no. 2, pp. 369–374, 1978.

[35] V. Ramachandran, "A minimax arc theorem for reducible flow graphs," *SIAM J. Discrete Math.*, vol. 3, no. 4, pp. 554–560, 1990.

[36] R. E. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Comput.*, vol. 1, no. 2, pp. 146–160, 1972.

[37] M. Syslo, "A labeling algorithm to recognize a line digraph and output its root graph," *Inf. Process. Lett.*, vol. 15, no. 1, pp. 28–30, 1982.

[38] F. Harary and R. Norman, "Some properties of line digraphs," *Rendiconti Circolo Matematico Palermo*, vol. 9, no. 2, pp. 149–163, 1960.

[39] J. Hopcroft and R. Tarjan, "Efficient planarity testing," *J. ACM*, vol. 21, no. 4, pp. 549–568, 1974.

[40] C. Lucchesi, "A minimax equality for directed graphs," Ph.D. dissertation, Univ. Waterloo, Waterloo, ON, USA, 1976.

Son Hoang Dau received the B.S. degree in applied mathematics and informatics from the College of Science, Vietnam National University, Hanoi, Vietnam, in 2006. He received the M.S. degree (2009) and Ph.D. degree (2012) in mathematical sciences from the Division of Mathematical Sciences, Nanyang Technological University, Singapore. He is currently a postdoctoral research fellow in SUTD-MIT International Design Centre, Singapore University of Technology and Design.

His research interests include coding theory, network coding, distributed storage systems, and combinatorics.

Vitaly Skachek (M'08) received the B.A. (Cum Laude), M.Sc. and Ph.D. degrees in computer science from the Technion—Israel Institute of Technology, in 1994, 1998 and 2007, respectively.

In the summer of 2004, he visited the Mathematics of Communications Department at Bell Laboratories under the DIMACS Special Focus Program in Computational Information Theory and Coding. During 2007-2012, he held research positions with the Claude Shannon Institute, University College Dublin, with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, and with the Department of Electrical and Computer Engineering, McGill University, Montréal. He is now a Senior Lecturer with the Institute of Computer Science, University of Tartu.

Dr. Skachek is a recipient of the Permanent Excellent Faculty Instructor award, given by Technion.

Yeow Meng Chee (SM'08) received the B.Math. degree in computer science and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively.

Currently, he is a Professor at the Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. Prior to this, he was Program Director of Interactive Digital Media R&D in the Media Development Authority of Singapore, Postdoctoral Fellow at the University of Waterloo and IBM's Zürich Research Laboratory, General Manager of the Singapore Computer Emergency Response Team, and Deputy Director of Strategic Programs at the Infocomm Development Authority, Singapore.

His research interest lies in the interplay between combinatorics and computer science/engineering, particularly combinatorial design theory, coding theory, extremal set systems, and electronic design automation.