2012 IEEE International Symposium on Information Theory Proceedings

Optimal Index Codes with Near-Extreme Rates

Son Hoang Dau
Division of Mathematical Sciences
School of Phys. and Math. Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371
Email: *shdau@ntu.edu.sg*

Vitaly Skachek[1]
Department of ECE
McGill University
3480 University Street
Montréal, QC H3A 2A7, Canada
Email: *vitaly.skachek@gmail.com*

Yeow Meng Chee
Division of Mathematical Sciences
School of Phys. and Math. Sciences
Nanyang Technological University
21 Nanyang Link, Singapore 637371
Email: *ymchee@ntu.edu.sg*

***Abstract*—The *min-rank* of a *digraph* was shown by Bar-Yossef *et al.* (2006) 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 work, the graphs and digraphs of near-extreme min-ranks are characterized. Those graphs and digraphs correspond to the ICSI instances having near-extreme transmission rates when using optimal scalar linear index codes. It is also shown that the decision problem of whether a *digraph* has min-rank two is NP-complete. By contrast, the same question for *graphs* can be answered in polynomial time.**

I. Introduction

Index Coding with Side Information (ICSI) [1] is a communication scheme dealing with broadcast channels in which receivers have prior side information about the messages to be broadcast. By using coding and by exploiting the knowledge about the side information, the sender may significantly reduce the number of required transmissions compared with the straightforward approach. Apart from being a special case of the well-known (non-multicast) Network Coding problem ([2], [3]), 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 oppoturnistic wireless networks ([1], [4], [5], [6]).

In the work of Bar-Yossef *et al.* [4], 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 corresponding to that instance. The concept of min-rank of a graph goes back to Haemers [7]. Min-rank serves as an upper bound for the celebrated Shannon capacity of a graph [8]. It was shown by Peeters [9] 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 interest in the Min-Rank problem has grown after the work of Bar-Yossef *et al.* [4]. Exact and heuristic algorithms for finding min-rank over the binary field of a digraph were developed in [10]. The min-rank of a random digraph was investigated in [11]. A dynamic programming approach was proposed in [12] to compute min-ranks of outerplanar graphs in polynomial time.

[1]The work of this author was done in part while he was with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

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 2, $n-2$, $n-1$, or n. Consequently, we show that the problem of deciding whether a digraph has min-rank two over $\mathbb{F}_2$ is NP-complete. By contrast, the same decision problem for graphs can be solved in polynomial time. The characterizations of graphs and digraphs with near-extreme min-ranks are summarized in the table below. The star mark indicates that the result is established in this paper.

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 ([9])	$\mathcal{D}$ 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 F (Fig. 4) 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*

The paper is organized as follows. Basic notations and definitions are presented in Section II. In Section III, we present characterization of graphs and digraphs of near-extreme min-ranks. In Section IV, we prove the hardness of the decision problem of whether a digraph has min-rank two. The full version of this paper is available at web.spms.ntu.edu.sg/~daus0001/ExtremeIC.pdf.

II. Notations and Definitions

Let $[n] \triangleq \{1, 2, \ldots, n\}$. Let $\mathbb{F}_q$ denote the finite field of q elements. The *support* of a vector $\boldsymbol{u} \in \mathbb{F}_q^n$ is defined to be the set $\mathrm{supp}(\boldsymbol{u}) = \{i \in [n] : u_i \neq 0\}$. 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 $\mathbb{F}_q$. We use $\boldsymbol{e}_i$ to denote the unit vector, which has a one at the ith position, and zeros elsewhere.

A *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 $\mathcal{E}(\mathcal{G})$, the edge set, is a set of *unordered* pairs of distinct vertices of $\mathcal{G}$. A typical edge of $\mathcal{G}$ is of the form $\{i, j\}$ where $i, j \in \mathcal{V}(\mathcal{G})$, and $i \neq j$. For a graph with n vertices, we often take $[n]$ as the vertex set. A *digraph*

$\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ is defined similarly, except that $\mathcal{E}(\mathcal{D})$ is called the arc set, which is a set of *ordered* pairs of distinct vertices of $\mathcal{D}$. A typical arc of $\mathcal{D}$ is of the form $e = (i, j)$ where $i, j \in \mathcal{V}(\mathcal{D})$, and $i \neq j$. We note that any graph can be viewed as a digraph with the same set of vertices, where each edge in $\mathcal{E}(\mathcal{G})$ is replaced with two antiparallel arcs in $\mathcal{E}(\mathcal{D})$.

If $(i, j) \in \mathcal{E}(\mathcal{D})$ then j is called an *out-neighbor* of i. The set of out-neighbors of a vertex i in the digraph $\mathcal{D}$ is denoted by $N_O^{\mathcal{D}}(i)$. We simply use $N_O(i)$ whenever there is no potential confusion. For a graph $\mathcal{G}$, we denote by $N^{\mathcal{G}}(i)$ the set of neighbors of i, namely, the set of vertices adjacent to i. The *order* of a digraph is its number of vertices. The *complement* of a digraph $\mathcal{D}$, denoted $\overline{\mathcal{D}}$, is a digraph with vertex set $\mathcal{V}(\mathcal{D})$ and arc set $\{(i, j) : i, j \in \mathcal{V}(\mathcal{D}), i \neq j, (i, j) \notin \mathcal{E}(\mathcal{D})\}$. The complement of a graph is defined analogously.

A circuit in a digraph $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ is a sequence of pairwise distinct vertices $\mathcal{C} = (i_1, i_2, \ldots, i_r)$, where $(i_s, i_{s+1}) \in \mathcal{E}(\mathcal{D})$ for all $s \in [r-1]$ and $(i_r, i_1) \in \mathcal{E}(\mathcal{D})$ as well. A digraph is called *acyclic* if it contains no circuits.

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 $x_i \in \mathbb{F}_q^t$ for all $i \in [n]$, 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 $\mathfrak{E}(\boldsymbol{x}) \in \mathbb{F}_q^{\kappa}$, $\kappa \in \mathbb{N}$, that enables each receiver R_i to recover x_i based on its side information. Such a mapping $\mathfrak{E}$ 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 $\mathfrak{E}$ is an $\mathbb{F}_q$-linear mapping, and *nonlinear* otherwise. The index code is called *scalar* if $t = 1$ and *vector* if $t > 1$. The length and the transmission rate of a scalar index code ($t = 1$) are identical.

Example II.1. Consider the following ICSI instance in which $n = 5$ and $X_1 = \{2\}$, $\mathcal{X}_2 = \{3\}$, $\mathcal{X}_3 = \{1, 4\}$, $\mathcal{X}_4 = \{5\}$, $\mathcal{X}_5 = \{2, 4\}$. Assume that $x_i \in \mathbb{F}_2$ ($i \in [5]$).

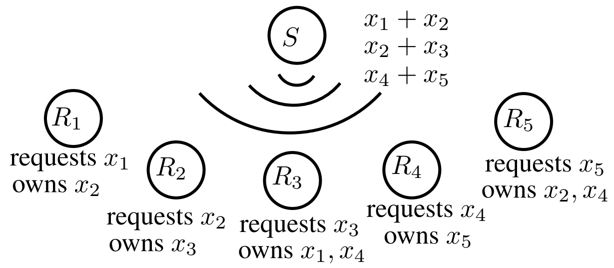
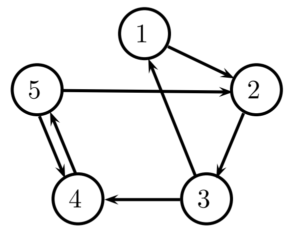


Fig. 1: Example of an ICSI instance and an index code

The sender broadcasts three packets $x_1 + x_2$, $x_2 + x_3$, and $x_4 + x_5$. This index code is of length three. The decoding process goes as follows. As R_1 knows x_2 and receives $x_1 + x_2$, it obtains $x_1 = x_2 + (x_1 + x_2)$; Similarly, R_2 obtains $x_2 = x_3 + (x_2 + x_3)$; R_3 obtains $x_3 = x_1 + (x_1 + x_2) + (x_2 + x_3)$; R_4 obtains $x_4 = x_5 + (x_4 + x_5)$; R_5 obtains $x_5 = x_4 + (x_4 + x_5)$. This index code saves *two* transmissions compared with the naïve approach when S broadcasts all five messages.

Each instance of the ICSI problem can be described by the so-called side information digraph [4]. Given n and $\mathcal{X}_i$, $i \in [n]$, the side information digraph $\mathcal{D}$ is defined as follows. The vertex set $\mathcal{V}(\mathcal{D}) = [n]$. The arc set $\mathcal{E}(\mathcal{D}) = \cup_{i \in [n]} \{(i, j) : j \in \mathcal{X}_i\}$. The side information digraph corresponding to the ICSI instance in Example II.1 is depicted in Fig. 2a.

Definition II.2 ([7], [4])**.** Let $\mathcal{D} = (\mathcal{V}(\mathcal{D}), \mathcal{E}(\mathcal{D}))$ be a digraph of order n. A matrix $\boldsymbol{M} = (m_{i,j}) \in \mathbb{F}_q^{n \times n}$ is said to *fit* $\mathcal{D}$ if $m_{i,i} \neq 0$ and $m_{i,j} = 0$ whenever $i \neq j$ and $(i, j) \notin \mathcal{E}(\mathcal{D})$. The *min-rank* of $\mathcal{D}$ over $\mathbb{F}_q$, denoted $\mathsf{minrk}_q(\mathcal{D})$, is defined to be the minimum rank of a matrix in $\mathbb{F}_q^{n \times n}$ that fits $\mathcal{D}$. Since a graph can be viewed as a special case of a digraph, the above definitions also apply to a graph.



(a) Side information digraph $\mathcal{D}$

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

(b) A rank-three matrix that fits $\mathcal{D}$

Fig. 2: Corresponding side information digraph (Example II.1)

Bar-Yossef *et al.* [4] showed that the length (transmission rate) of an optimal scalar linear index code for the ICSI instance described by $\mathcal{D}$ is precisely the min-rank of $\mathcal{D}$. Let $\beta_q(t, \mathcal{D})$ denotes the length of an optimal vector index code of block length t over $\mathbb{F}_q$ for an ICSI instance described by a digraph $\mathcal{D}$. Lubetzky *et al.* [13] defined the *broadcast rate* $\beta_q(\mathcal{D})$ of the corresponding ICSI instance to be $\lim_{t \to \infty} \beta_q(t, \mathcal{D})/t$ (see also [14]). Then $\mathsf{minrk}_q(\mathcal{D})$ is an upper bound on $\beta_q(\mathcal{D})$.

Theorem II.3 ([7], [4], [14])**.** *For any digraph $\mathcal{D}$ we have $\alpha(\mathcal{D}) \leq \beta_q(\mathcal{D}) \leq \mathsf{minrk}_q(\mathcal{D})$, where $\alpha(\mathcal{D})$ denotes the size of a maximum acyclic induced subgraph of $\mathcal{D}$. The same inequalities hold for a graph $\mathcal{G}$, where $\alpha(\mathcal{G})$ denotes the size of a maximum independent set of $\mathcal{G}$.*

III. Digraphs of Near-Extreme Min-Ranks

A. Digraphs of Min-Rank One

Proposition III.1. *A digraph has min-rank one over $\mathbb{F}_q$ if and only if it is a complete digraph.*

Due to Theorem II.3, it is also trivial to see that $\beta_q(\mathcal{D}) = 1$ if and only if $\mathcal{D}$ is a complete digraph.

B. Digraphs of Min-Rank Two

It was observed by Peeters [9] that a graph has min-rank two if and only if it is not a complete graph and its complement is a bipartite graph. We now focus only on 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 $\phi(i) \neq \phi(j)$ whenever $\{i,j\} \in \mathcal{E}(\mathcal{G})$. We often refer to $\phi(i)$ as the *color* of i. If there exists a k-coloring of $\mathcal{G}$, then we say that $\mathcal{G}$ is k-colorable.

Definition III.2. 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 $(i,j) \in \mathcal{E}(\mathcal{D})$ then $\phi(i) \neq \phi(j)$;

(C2) For each vertex i of $\mathcal{D}$, it holds that $\phi(j) = \phi(h)$ for all out-neighbors j and h of i.

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*.

Lemma III.3. *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 $i \in A$: either $N_O(i) \subseteq B$ or $N_O(i) \subseteq C$;*
2) *For every $i \in B$: either $N_O(i) \subseteq A$ or $N_O(i) \subseteq C$;*
3) *For every $i \in C$: either $N_O(i) \subseteq A$ or $N_O(i) \subseteq B$.*

Theorem III.4. *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: Suppose $\mathcal{V}(\mathcal{D}) = [n]$.

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 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 $B = \mathsf{supp}(\boldsymbol{M}_1) \cap \mathsf{supp}(\boldsymbol{M}_2)$, $A = \mathsf{supp}(\boldsymbol{M}_1) \setminus B$, and $C = \mathsf{supp}(\boldsymbol{M}_2) \setminus B$. Since the binary alphabet is considered and the matrix $\boldsymbol{M}$ has no zero rows, for every $i \in [n]$, one of the following must hold: (1) $\boldsymbol{M}_i = \boldsymbol{M}_1$; (2) $\boldsymbol{M}_i = \boldsymbol{M}_2$; (3) $\boldsymbol{M}_i = \boldsymbol{M}_1 + \boldsymbol{M}_2$. Hence $i \in \mathsf{supp}(\boldsymbol{M}_i) \subseteq A \cup B \cup C$ for every $i \in [n]$. This implies that $A \cup B \cup C = [n]$.

Suppose that $i \in A$. Then either $\boldsymbol{M}_i = \boldsymbol{M}_1$ or $\boldsymbol{M}_i = \boldsymbol{M}_1 + \boldsymbol{M}_2$. The former condition holds if and only if $\mathsf{supp}(\boldsymbol{M}_i) = A \cup B$, which in turns implies that $(i,j) \in \mathcal{E}(\mathcal{D})$ for all $j \in A \cup B \setminus \{i\}$. In other words, $(i,j) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $j \in A \cup B$. 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 $\mathsf{supp}(\boldsymbol{M}_i) = A \cup C$, which implies that $(i,j) \notin \mathcal{E}(\overline{\mathcal{D}})$ for all $j \in A \cup C$. In summary, for every $i \in A$ we have

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

In other words, for every $i \in A$, either $\mathcal{N}_O^{\overline{\mathcal{D}}}(i) \subseteq B$ or $\mathcal{N}_O^{\overline{\mathcal{D}}}(i) \subseteq C$. Analogous conditions hold for every $i \in B$ and for every $i \in C$ as well. Therefore, by Lemma III.3, $\overline{\mathcal{D}}$ 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 which fits $\mathcal{D}$. By Lemma III.3, 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 $i \in A$: either $N_O^{\overline{\mathcal{D}}}(i) \subseteq B$ or $N_O^{\overline{\mathcal{D}}}(i) \subseteq C$;
2) For every $i \in B$: either $N_O^{\overline{\mathcal{D}}}(i) \subseteq A$ or $N_O^{\overline{\mathcal{D}}}(i) \subseteq C$;
3) For every $i \in C$: either $N_O^{\overline{\mathcal{D}}}(i) \subseteq A$ or $N_O^{\overline{\mathcal{D}}}(i) \subseteq B$.

We construct an $n \times n$ matrix $\boldsymbol{M} = (m_{i,j})$ as follows. For each $i \in A$, if $N_O^{\overline{\mathcal{D}}}(i) \subseteq B$ then let $m_{i,j} = 1$ for $j \in A \cup C$, and $m_{i,j} = 0$ for $j \in B$. Otherwise, if $N_O^{\overline{\mathcal{D}}}(i) \subseteq C$ then let $m_{i,j} = 1$ for $j \in A \cup B$, and $m_{i,j} = 0$ for $j \in C$. For $i \in B$ and $i \in C$, $\boldsymbol{M}_i$ can be constructed analogously. It is obvious that $\boldsymbol{M}$ fits $\mathcal{D}$. Moreover, each row of $\boldsymbol{M}$ can be written as a linear combinations of the two binary vectors whose supports are $A \cup B$ and $B \cup C$, respectively. Therefore, $\mathsf{rank}_q(\boldsymbol{M}) \leq 2$. We complete the proof. ■

The following corollary characterizes the digraphs of min-rank two.

Corollary III.5. *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.* [15] that $\beta_2(\mathcal{G}) = 2$ if and only if $\overline{\mathcal{G}}$ is bipartite and $\mathcal{G}$ is not a complete graph. A characterization (by forbidden subgraphs) of digraphs $\mathcal{D}$ with $\beta_2(\mathcal{D}) = 2$ was also obtained therein.

C. Digraphs of Min-Ranks Equal to Their Orders

Proposition III.6. *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 III.7. *Let $\mathcal{D}$ be a digraph of order n. Then $\mathsf{minrk}_q(\mathcal{D}) = n$ if and only if $\mathcal{D}$ is acyclic.*

Proof: Equivalently, we show that $\mathsf{minrk}_q(\mathcal{D}) \leq n-1$ if and only if $\mathcal{D}$ has a circuit. Let $\mathcal{V}(\mathcal{D}) = [n]$.

Suppose $\mathcal{D}$ has a circuit $\mathcal{C} = (i_1, i_2, \ldots, i_r)$. We construct a matrix $\boldsymbol{M}$ fitting $\mathcal{D}$ as follows. Let $\mathcal{V}(\mathcal{C}) = \{i_1, \ldots, i_r\}$. For $j \notin \mathcal{V}(\mathcal{C})$, let $\boldsymbol{M}_j = \boldsymbol{e}_j$. For $s \in [r-1]$, let $\boldsymbol{M}_{i_s} = \boldsymbol{e}_{i_s} - \boldsymbol{e}_{i_{s+1}}$. Finally, let $\boldsymbol{M}_{i_r} = \boldsymbol{e}_{i_1} - \boldsymbol{e}_{i_r}$. Clearly, $\boldsymbol{M}$ fits $\mathcal{D}$. Moreover, as $\boldsymbol{M}_{i_r} = \sum_{s=1}^{r-1} \boldsymbol{M}_{i_s}$, we have $\mathsf{rank}_q(\boldsymbol{M}_{\mathcal{V}(\mathcal{C})}) \leq r-1$. Hence

$$\begin{aligned}\mathsf{rank}_q(\boldsymbol{M}) &= \mathsf{rank}_q\left(\boldsymbol{M}_{\mathcal{V}(\mathcal{C})}\right) + \mathsf{rank}_q\left(\boldsymbol{M}_{[n]\setminus\mathcal{V}(\mathcal{C})}\right) \\ &\leq (r-1) + (n-r) = n-1.\end{aligned}$$

Therefore, $\mathsf{minrk}_q(\mathcal{D}) \leq n-1$.

Conversely, suppose that $\mathsf{minrk}_q(\mathcal{D}) \leq n-1$. By Theorem II.3, $\alpha(\mathcal{D}) \leq n-1$. Therefore, $\mathcal{D}$ contains a circuit. ■

By Theorem II.3, it is not hard to see that Proposition III.6 and III.7 also hold if we replace $\mathsf{minrk}_q(\cdot)$ by $\beta_q(\cdot)$.

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

Definition III.8. A graph $\mathcal{G} = (\mathcal{V}(\mathcal{G}), \mathcal{E}(\mathcal{G}))$ is called a *star graph* if $|\mathcal{V}(\mathcal{G})| \geq 2$ and there exists a vertex $i \in \mathcal{V}(\mathcal{G})$ such that $\mathcal{E}(\mathcal{G}) = \big\{\{i,j\} : j \in \mathcal{V}(\mathcal{G}) \setminus \{i\}\big\}$.

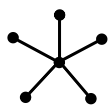


Fig. 3: A star graph

Theorem III.9. *Let $\mathcal{G}$ be a connected graph of order $n \geq 2$. Then $\mathsf{minrk}_q(\mathcal{G}) = n-1$ if and only if $\mathcal{G}$ is a star graph.*

Proof: We first suppose that $\mathsf{minrk}_q(\mathcal{G}) = n-1$. If $n = 2$ then $\mathcal{G}$ must be a complete graph, which is a star graph. We assume that $n \geq 3$. As $\mathcal{G}$ is connected, there exists a vertex i of degree at least two. Let i_1 and i_2 be any two distinct vertices adjacent to i. Our goal is to show that for every vertex $j \neq i$, we have $\{i, j\} \in \mathcal{E}(\mathcal{G})$, and those are all possible edges.

Firstly, suppose for a contradiction that $\{i, j\} \notin \mathcal{E}(\mathcal{G})$ for some $j \neq i$. Since $\mathcal{G}$ is connected, there exists k such that $\{j, k\} \in \mathcal{E}(\mathcal{G})$. Then either $k \neq i_1$ or $k \neq i_2$. Suppose that $k \neq i_2$. We create a matrix $\boldsymbol{M}$ as follows. Let $\boldsymbol{M}_j = \boldsymbol{M}_k = \boldsymbol{e}_j + \boldsymbol{e}_k$, $\boldsymbol{M}_i = \boldsymbol{M}_{i_2} = \boldsymbol{e}_i + \boldsymbol{e}_{i_2}$, and $\boldsymbol{M}_h = \boldsymbol{e}_h$ for $h \notin \{i, j, i_2, k\}$. Then $\boldsymbol{M}$ fits $\mathcal{G}$ and $\mathsf{rank}_q(\boldsymbol{M}) \leq n-2$. Therefore, $\mathsf{minrk}_q(\mathcal{G}) < n-1$. We obtain a contradiction. Thus, all other vertices are adjacent to i.

Secondly, suppose for a contradiction that there exist two adjacent vertices, namely j and k, both are different from i. As we just proved, both j and k must be adjacent to i. We create a matrix $\boldsymbol{M}$ as follows. We take $\boldsymbol{M}_i = \boldsymbol{M}_j = \boldsymbol{M}_k = \boldsymbol{e}_i + \boldsymbol{e}_j + \boldsymbol{e}_k$, and $\boldsymbol{M}_h = \boldsymbol{e}_h$ for $h \notin \{i, j, k\}$. Clearly $\boldsymbol{M}$ fits $\mathcal{G}$ and moreover, $\mathsf{rank}_q(\boldsymbol{M}) \leq n-2$, which implies that $\mathsf{minrk}_q(\mathcal{G}) < n-1$. We obtain a contradiction.

Conversely, assume that $\mathcal{G}$ is a star graph, where $\mathcal{E}(\mathcal{G}) = \big\{\{i, j_s\} : s \in [n-1]\big\}$, $j_s \in \mathcal{V}(\mathcal{G}) \setminus \{i\}$ for all $s \in [n-1]$. We create a matrix $\boldsymbol{M}$ fitting $\mathcal{G}$ by taking $\boldsymbol{M}_{j_s} = \boldsymbol{e}_i + \boldsymbol{e}_{j_s}$ for $s \in [n-1]$, and $\boldsymbol{M}_i = \boldsymbol{e}_i + \boldsymbol{e}_{j_1}$. Since $\boldsymbol{M}_i \equiv \boldsymbol{M}_{j_1}$, we deduce that $\mathsf{rank}_q(\boldsymbol{M}) \leq n-1$. Hence $\mathsf{minrk}_q(\mathcal{G}) \leq n-1$. On the other hand, since $\{j_s : s \in [n-1]\}$ is a maximum independent set in $\mathcal{G}$, we obtain that $\alpha(\mathcal{G}) = n-1$. By Theorem II.3, $\mathsf{minrk}_q(\mathcal{G}) \geq \alpha(\mathcal{G}) = n-1$. Thus, $\mathsf{minrk}_q(\mathcal{G}) = n-1$. ■

It is not hard to see that Theorem III.9 still holds with $\mathsf{minrk}_q(\cdot)$ being replaced by $\beta_q(\cdot)$.

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

A *matching* in a graph is a set of edges without common vertices. A *maximum matching* is a matching that contains the largest possible number of edges. The number of edges in a maximum matching in $\mathcal{G}$ is denoted by $\mathsf{mm}(\mathcal{G})$.

Theorem III.10. *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 F depicted in Fig. 4.*

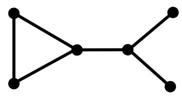


Fig. 4: The forbidden subgraph F

Sketch of the proof: For the ONLY IF direction, suppose that $\mathsf{minrk}_q(\mathcal{G}) = n-2$. It is not hard to see that $\mathsf{minrk}_q(\mathcal{G}) \leq n - \mathsf{mm}(\mathcal{G})$. Hence $\mathsf{mm}(\mathcal{G}) \leq 2$. As $\mathsf{mm}(\mathcal{G}) \in \{0, 1\}$ and $|\mathcal{V}(\mathcal{G})| \geq 6$ 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 F has min-rank *three* less than its order, $\mathcal{G}$ should not contain any subgraph isomorphic to F.

We now turn to the IF direction. Suppose that $\mathsf{mm}(\mathcal{G}) = 2$ and $\mathcal{G}$ does not contain any subgraph isomorphic to F. Then $\mathsf{minrk}_q(\mathcal{G}) \leq n - \mathsf{mm}(\mathcal{G}) = n-2$. As $\alpha(\mathcal{G}) \leq \mathsf{minrk}_q(\mathcal{G})$, it suffices to show that $\alpha(\mathcal{G}) = n-2$. Let $\{a, b\}$ and $\{c, d\}$ be the two edges in a maximum matching in $\mathcal{G}$. Let $U = \{a, b, c, d\}$ and $V = \mathcal{V}(\mathcal{G}) \setminus U$. The main idea is to show that we can always find two nonadjacent vertices in U which 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. ■

It is also not hard to see that Theorem III.10 still holds with $\mathsf{minrk}_q(\cdot)$ being replaced by $\beta_q(\cdot)$.

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

In this section, we first prove that the decision problem of whether a given digraph is fairly k-colorable (see Definition III.2) is NP-complete, for any given $k \geq 3$. The hardness of this problem, by Proposition III.1 and Corollary III.5, leads to the hardness of the decision problem of whether a given digraph has min-rank two over $\mathbb{F}_2$. The fair k-coloring problem is defined formally as follows.

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

Theorem IV.1. *The fair k-coloring problem is NP-complete for $k \geq 3$.*

Proof: This problem is obviously in NP. For NP-hardness, we reduce the k-coloring problem to the fair k-coloring problem. 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 $i \in [n]$, 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

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

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

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

Let $N^{\mathcal{G}}(i) = \{i_1, i_2, \ldots, i_{n_i}\}$. Then $\mathcal{D}_i$ is depicted in Fig. 5.

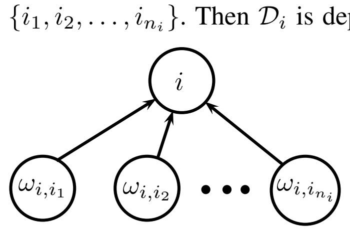


Fig. 5: Gadget $\mathcal{D}_i$ for each vertex i of $\mathcal{G}$

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

$$\mathcal{Q}_i = \big\{(p_i, i)\big\} \cup \big\{(p_i, \omega_{i',i'_j}) : i' \in [n],\ j \in [n_{i'}],\ i'_j = i\big\},$$

be the set consisting of (p_i, i) and the arcs that connect p_i and all the clones 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. 6, then $\mathcal{D}$ is the digraph in Fig. 7.

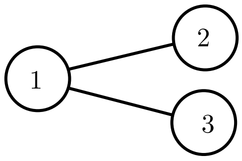


Fig. 6: An example of the graph $\mathcal{G}$

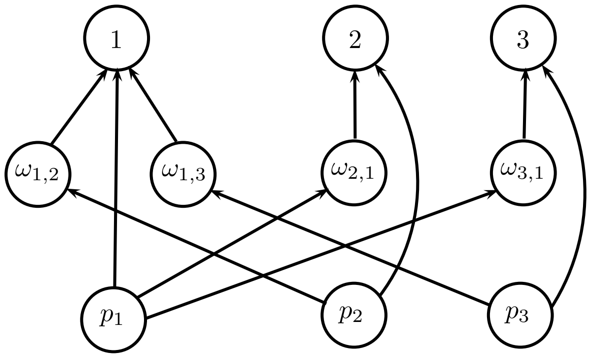


Fig. 7: The digraph $\mathcal{D}$ built from the graph $\mathcal{G}$ in Fig. 6

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 define $\phi_{\mathcal{D}} : \mathcal{V}(\mathcal{D}) \to [k]$ as follows

1) For every $i \in [n]$, $\phi_{\mathcal{D}}(i) \triangleq \phi_{\mathcal{G}}(i)$;
2) If $i'_j = i$, then $\phi_{\mathcal{D}}(\omega_{i',i'_j}) \triangleq \phi_{\mathcal{D}}(i) = \phi_{\mathcal{G}}(i)$, in other words, clones of i have the same color as i;
3) For every $i \in [n]$, $\phi_{\mathcal{D}}(p_i)$ is 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 III.2). 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, ω_{i',i'_j}) for $i'_j = i$, have different colors. It remains to verify that i and ω_{i,i_j} have different colors. On the one hand, ω_{i,i_j} is a clone of i_j, and hence has the same color as i_j. In other words,

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

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

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

Hence, $\phi_{\mathcal{D}}(\omega_{i,i_j}) \neq \phi_{\mathcal{D}}(i)$ ($i \in [n]$). Thus, (C1) is satisfied.

We now check if (C2) (see Definition III.2) is also satisfied. The out-neighbors of p_i are i and its clones, namely, ω_{i',i'_j} with $i'_j = i$. These vertices have the same color in $\mathcal{D}$ 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, $\phi_{\mathcal{D}}(\omega_{i',i'_j}) = \phi_{\mathcal{D}}(i)$ if $i'_j = i$. 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 in $|\mathcal{V}(\mathcal{G})|$. Since the k-coloring problem ($k \geq 3$) is NP-hard [16], we conclude that the fair k-coloring problem is also NP-hard. ∎

Corollary IV.2. *Given an arbitrary digraph $\mathcal{D}$, the decision problem of whether* $\mathsf{minrk}_2(\mathcal{D}) = 2$ *is NP-complete.*

Recall that for a graph $\mathcal{G}$, it was observed by Peeters [9] 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. Another related result, presented by Blasiak *et al.* [15], stated that there is a polynomial time algorithm to verify whether $\beta_2(\mathcal{D}) = 2$ for a general digraph $\mathcal{D}$.

References

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

[2] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1204–1216, 2000.

[3] R. Koetter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Trans. Netw.*, vol. 11, pp. 782–795, 2003.

[4] Z. Bar-Yossef, Z. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," in *Proc. 47th Annu. IEEE Symp. on Found. of Comput. Sci. (FOCS)*, 2006, pp. 197–206.

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

[6] 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*, 2006, pp. 243–254.

[7] W. Haemers, "An upper bound for the shannon capacity of a graph," *Algebr. Methods Graph Theory*, vol. 25, pp. 267–272, 1978.

[8] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 3, pp. 3–15, 1956.

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

[10] M. A. R. Chaudhry and A. Sprintson, "Efficient algorithms for index coding," in *Proc. IEEE Conf. on Comput. Commun. (INFOCOM)*, 2008, pp. 1–4.

[11] I. Haviv and M. Langberg, "On linear index coding for random graphs," 2011, preprint.

[12] Y. Berliner and M. Langberg, "Index coding with outerplanar side information," in *Proc. IEEE Symp. on Inform. Theory (ISIT)*, Saint Petersburg, Russia, 2011, pp. 869–873.

[13] E. Lubetzky and U. Stav, "Nonlinear index coding outperforming the linear optimum," *IEEE Trans. Inform. Theory*, vol. 55, no. 8, pp. 3544–3551, 2009.

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

[15] A. Blasiak, R. Kleinberg, and E. Lubetzky, "Index coding via linear programming," 2011, preprint.

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