Hardness of Linear Index Coding on Perturbed Instances*

Dror Chawin† Ishay Haviv†

Abstract

The index coding problem is concerned with the amount of information that a sender has to transmit to multiple receivers in a way that enables each of them to retrieve its requested data relying on prior side information. For linear index coding, the problem is characterized by the minrank parameter of a graph that represents the side information map of the receivers. Previous work has shown that it is NP-hard to determine the minrank parameter of graphs. In this work, we study the computational complexity of the minrank parameter on perturbed instances, obtained from worst-case instances by a random extension of the side information available to the receivers. This setting is motivated by applications of index coding, in which the side information is accumulated via repeated transmissions that suffer from loss of data due to noisy communication or storage capacity. We prove that determining the minrank parameter remains computationally hard on perturbed instances. Our contribution includes an extension of several hardness results of the minrank parameter to the perturbed setting as well as a general technique for deriving the hardness of the minrank parameter on perturbed instances from its hardness on worst-case instances.

1 Introduction

Index coding is a canonical problem in the area of network information theory. It is concerned with the design of coding schemes for broadcasting information to multiple receivers relying on their prior side information. Since its introduction in 1998 by Birk and Kol [6], it provides a simple and yet rich framework for research on coding-on-demand communication problems motivated by various applications, e.g., network and satellite communication, distributed storage and caching, device-to-device relaying, and interference management (see, e.g., [2]).

The index coding problem involves n receivers $R_1, \ldots, R_n$ and a sender that holds an n-symbol message $x \in \Sigma^n$ over some alphabet Σ. Every receiver R_i is interested in the ith symbol x_i of x and has some side information on x, comprising a subset of the symbols x_j with $j \in [n] \setminus \{i\}$. The side information map is naturally represented by a digraph $G = ([n], E)$, which contains a directed edge (i, j) if the side information of the receiver R_i includes x_j. When the side information map is symmetric, we view G as an undirected graph. The sender wishes to broadcast a short message to the receivers in a way that enables each of them to retrieve its symbol. An index code for G over Σ of length ℓ is an encoding function $E : \Sigma^n \to \Sigma^\ell$ such that for every message $x \in \Sigma^n$ and for every

*A preliminary version appeared in Proceedings of the 58th Annual Allerton Conference on Communication, Control, and Computing, 2022.

†School of Computer Science, The Academic College of Tel Aviv-Yaffo, Tel Aviv 61083, Israel. Research supported by the Israel Science Foundation (grant No. 1218/20).

$i \in [n]$, the receiver R_i is able to decode x_i given $E(x)$ and the side information available to it. Given a digraph G that represents the side information of the receivers and given an alphabet Σ, the index coding problem asks to design an index code for G over Σ of as short as possible length. For example, if G is a complete graph and the alphabet Σ is binary, it suffices to broadcast to the receivers the xor of the bits of x, hence the minimal length of an index code in this case is 1.

To demonstrate the applicability of the index coding problem, let us consider the following scenario described in [4]. Suppose that a server disseminates a sequence of data blocks over a broadcast channel to a set of caching clients. After the main transmission, each client holds some subset of the transmitted blocks, whereas the other blocks are lost due to intermittent reception or limited storage capacity. Assuming that the data blocks are large and that the amount of metadata per block is independent of its size, each client can use a slow backward channel to inform the server about the indices of the blocks available in its cache. Then, the goal is to minimize the length of the additional information needed to be transmitted in order to enable the clients to discover their required blocks. In certain cases, several rounds of transmissions are made until all clients receive the required data. During these rounds, the clients might extend their available information (see, e.g., [20]).

A special case of the index coding problem that has received a considerable amount of attention in the literature is that of linear index coding, where the alphabet Σ is a field $\mathbb{F}$ and the encoding function E is linear over $\mathbb{F}$. It was shown by Bar-Yossef, Birk, Jayram, and Kol [4] that the minimal length of a linear index code for a digraph G over a field $\mathbb{F}$ is precisely characterized by a quantity called minrank, denoted by $\mathrm{minrk}_{\mathbb{F}}(G)$ (see Definition 2.1). The latter was originally introduced by Haemers [13] in the study of the Shannon capacity of graphs and has applications for various areas of theoretical computer science, e.g., circuit complexity [21], communication complexity [19], and randomized computation [15]. The behavior of the minrank of random graphs was studied over various fields in [14, 12, 1]. Note that for certain instances of the index coding problem, a non-linear index code can be much shorter than any linear index code [7]. Nevertheless, the study of linear index codes is of great interest due to the simplicity and the efficiency of their encoding and decoding procedures.

The characterization of the linear index coding problem via the minrank parameter motivates its study from a computational perspective. A result of Peeters [18] asserts that for every field $\mathbb{F}$, it is NP-hard to decide whether the minrank of a given graph over $\mathbb{F}$ is at most 3. For the non-symmetric case, it was shown by Dau, Skachek, and Chee [9] that it is NP-hard to decide whether the minrank of a given digraph over the binary field $\mathbb{F}_2$ is at most 2. This is in contrast to (undirected) graphs, whose minrank over any field is at most 2 if and only if their complement is bipartite, a property that can be checked in polynomial time. Assuming certain variants of Khot's unique games conjecture [16], it was shown by Langberg and Sprintson [17] that, for all integers $k_2 > k_1 \geq 3$ and over every finite field, it is NP-hard to distinguish graphs with minrank at most k_1 from graphs with minrank at least k_2. In fact, by combining the approach of [17] with a result of [10], it follows that this hardness result holds even for a super-constant k_2 (namely, for $k_2 = \Omega(\log \log n)$, where n stands for the number of vertices). In the recent paper [8], it was shown that for every sufficiently large integer k and for every finite field $\mathbb{F}$, it is NP-hard to distinguish graphs with minrank at most k from graphs with minrank at least $2^{(1-o(1))\cdot k/2}$. Note that the latter hardness result relies only on the assumption $\mathsf{P} \neq \mathsf{NP}$.

1.1 Hardness of Linear Index Coding on Perturbed Instances

The hardness results stated above imply that it is unlikely that there exists a polynomial-time algorithm for determining the minrank of a given graph over a given field. However, those hardness results refer to the worst-case complexity of the problem. In this paper, we study the stability of the hardness of the minrank parameter to random perturbations. Namely, we consider the question of whether the minrank parameter remains computationally hard when a worst-case instance is randomly perturbed. This question is motivated by applications of the index coding problem where the side information of the receivers is extended in a random manner. For example, suppose that a sender and a collection of receivers are given a side information map, and due to the computational hardness of the minrank parameter, they are unable to efficiently design an appropriate economical linear index code. To handle this situation, the sender broadcasts the whole information to the receivers, where due to noisy communication or storage limitation, only a subset of the transmitted information becomes available to each receiver. Then, the objective is to design a short linear index code with respect to the side information accumulated by the receivers.

For concreteness, suppose that we are given a digraph $G = ([n], E)$ that represents the side information currently available to the n receivers. Suppose further that the sender broadcasts the entire message $x \in \mathbb{F}^n$ to the n receivers and that the ith receiver gets every symbol x_j with $j \in [n] \setminus \{i\}$ independently with some probability $p \in (0,1)$. After this transmission, the updated side information map is represented by a digraph $\widetilde{G}$ obtained from G by adding every non-edge of G with probability p.[1] We refer to the digraph $\widetilde{G}$ as a perturbed instance of G and study the complexity of determining the minrank over a field $\mathbb{F}$ of such perturbed instances. To do so, we employ an approach for proving hardness results on perturbed instances developed by Bennett, Reichman, and Shinkar [5]. This approach was used in [5] for several classical NP-hard problems (e.g., independence number, chromatic number, vertex cover, Hamilton cycle, subset sum, and constraint satisfaction).

We prove several hardness results on the minrank parameter of perturbed instances. Our first result shows that the hardness result of [9] on the minrank parameter of digraphs over $\mathbb{F}_2$ holds for perturbed instances, as stated below.

Theorem 1.1. *Unless* NP = RP, *for every constant* $p \in (0,1)$, *there is no polynomial-time algorithm that given a digraph* G *and a digraph* $\widetilde{G}$ *obtained from* G *by adding every non-edge of* G *independently with probability* p, *decides whether* $\mathrm{minrk}_{\mathbb{F}_2}(\widetilde{G}) \leq 2$ *with high success probability over the choice of* $\widetilde{G}$.

Our next result shows that the hardness result of [18] on the minrank parameter of graphs holds for perturbed instances, as stated below.

Theorem 1.2. *Unless* NP = RP, *for every constant* $p \in (0,1)$ *and for every field* $\mathbb{F}$, *there is no polynomial-time algorithm that given a graph* G *and a graph* $\widetilde{G}$ *obtained from* G *by adding every non-edge of* G *in-*

[1] A noisy transmission of the message x to the receivers allows the ith receiver to get its required symbol x_i. In this case, the ith receiver is already satisfied, hence its vertex can be omitted from the updated side information graph. It would thus be natural to consider the random graph $\widetilde{G}$ obtained from G by first removing every vertex with probability p and then adding every non-edge between remaining vertices with probability p. For simplicity of presentation, and to be more aligned with the framework of [5], we consider throughout this work the model in which the ith receiver learns from the transmission only symbols x_j with $j \in [n] \setminus \{i\}$. Note, however, that the statements of Theorems 1.1 and 1.2 hold with respect to the model that allows removal of vertices as well.

dependently with probability p, decides whether $\mathrm{minrk}_{\mathbb{F}}(\widetilde{G}) \leq 3$ *with high success probability over the choice of* $\widetilde{G}$.

Note that the hardness result of Theorem 1.2 further holds when the bound on the minrank is required to hold over all fields $\mathbb{F}$ simultaneously.

The proofs of Theorems 1.1 and 1.2 rely on reductions that are robust to random perturbations (see Section 2.3). Following the approach applied in [5] in the context of the 3-colorability problem, a main component of these reductions is the notion of blowup of graphs (or digraphs), defined as follows. For a graph G and for an integer R, the R-blowup of G is the graph obtained from G by replacing every vertex of G by an independent set of size R and every edge of G by the R^2 edges connecting the vertices associated with its endpoints. Our analysis shows, roughly speaking, that if R is sufficiently large then the R-blowup operation makes the minrank parameter of the complement stable to random perturbations. The proof of Theorem 1.1 uses the characterization of [9] for digraphs with minrank at most 2 over $\mathbb{F}_2$ (see Lemma 3.3). The proof of Theorem 1.2 involves a gadget graph that was used in the hardness proof of [18] (see Section 3.2.1) along with a concentration result on the clique number of random graphs (see Theorem 2.4)

Finally, we provide a general technique for deriving hardness results on the minrank parameter of perturbed instances with perturbation probability $1/2$ from standard worst-case hardness results. For this method to be applicable, the assumed worst-case hardness should hold for a polynomially large gap, namely, for distinguishing between graphs with minrank at most k from graphs whose minrank is $\Omega(k^3)$ (in contrast to the limited gaps given in [18] and in [9]). For a precise statement, see Proposition 3.18. In the analysis, we study the typical effect of a random perturbation on the minrank of graphs, borrowing a technique of [5] (see Lemma 3.17). As applications, we show that the worst-case hardness results on the minrank parameter given in [17] and in [8] yield hardness results on perturbed instances.

2 Preliminaries

Throughout the paper, we omit all floor and ceiling signs whenever they are not crucial. All logarithms are in base 2, unless otherwise specified. Undirected graphs are referred to as graphs, and directed graphs are referred to as digraphs. All the considered graphs and digraphs are simple.

2.1 Minrank

The minrank parameter, introduced in [13], is defined as follows.

Definition 2.1 (Minrank)**.** *Let $G = (V, E)$ be a digraph on the vertex set $V = [n]$, and let $\mathbb{F}$ be a field. We say that a matrix $M \in \mathbb{F}^{n \times n}$* represents *$G$ if $M_{i,i} \neq 0$ for every $i \in V$, and $M_{i,j} = 0$ for every distinct vertices $i, j \in V$ such that $(i, j) \notin E$. The* minrank *of G over $\mathbb{F}$ is defined as*

$$\mathrm{minrk}_{\mathbb{F}}(G) = \min\{\mathrm{rank}_{\mathbb{F}}(M) \mid M \text{ represents } G \text{ over } \mathbb{F}\}.$$

The definition is naturally extended to graphs by replacing every edge with two oppositely directed edges.

It was shown in [4] that for every side information digraph G and for every field $\mathbb{F}$, the smallest length of a linear index code for G over $\mathbb{F}$ is $\mathrm{minrk}_{\mathbb{F}}(G)$.

The chromatic number of a graph G, denoted by $\chi(G)$, is the smallest integer k needed for a proper k-coloring of G, i.e., a coloring of its vertices with k colors such that every two adjacent vertices are assigned distinct colors. For every graph G and for every finite field $\mathbb{F}$, it holds that

$$\log_{|\mathbb{F}|} \chi(G) \leq \text{minrk}_{\mathbb{F}}(\overline{G}) \leq \chi(G). \tag{1}$$

Indeed, for the first inequality, let G be a graph on the vertex set $V = [n]$, set $k = \text{minrk}_{\mathbb{F}}(\overline{G})$, and let $M \in \mathbb{F}^{n \times n}$ be a matrix that represents $\overline{G}$ and satisfies $\text{rank}_{\mathbb{F}}(M) = k$. The matrix M can be written as $M = A^T \cdot B$ for two matrices $A, B \in \mathbb{F}^{k \times n}$. Observe that for every two adjacent vertices $i, j \in V$ in G, the ith and jth columns of A are distinct, because the inner product of the ith column of B with the former is nonzero whereas its inner product with the latter is zero. This implies that by assigning to every vertex $i \in V$ the ith column of A, we get a proper coloring of G, hence $\chi(G) \leq |\mathbb{F}|^k$, implying the required bound. For the second inequality, consider a proper k-coloring of a graph G on the vertex set $V = [n]$, and let $M \in \mathbb{F}^{n \times n}$ be the $0,1$ matrix whose value in the (i, j) entry is 1 if and only if the vertices i and j belong to the same color class. Observe that M represents $\overline{G}$ and that $\text{rank}_{\mathbb{F}}(M) \leq k$, hence $\text{minrk}_{\mathbb{F}}(\overline{G}) \leq k$.

2.2 Computational Complexity

In what follows, we describe the notions from the area of computational complexity used throughout this paper. For more details, we refer the reader to [3, Chapters 2 and 7].

Let Σ be some alphabet, and let $A = (A_{\mathsf{YES}}, A_{\mathsf{NO}})$ be a pair of disjoint subsets of Σ^*. The computational problem associated with A (in short, the problem A) is that of distinguishing A_{YES} from A_{NO}, that is, deciding whether a given instance $x \in A_{\mathsf{YES}} \cup A_{\mathsf{NO}}$ lies in A_{YES} or in A_{NO}. In the former case, x is referred to as a YES instance, and in the latter as a NO instance. When the sets A_{YES} and A_{NO} do not cover Σ^*, the problem A is referred to as a promise problem to express the fact that an instance of the problem is promised to lie in $A_{\mathsf{YES}} \cup A_{\mathsf{NO}}$. A reduction from a problem $A = (A_{\mathsf{YES}}, A_{\mathsf{NO}})$ to a problem $B = (B_{\mathsf{YES}}, B_{\mathsf{NO}})$ is a mapping f from the instances of A to instances of B, such that for every instance x of A, it holds that if $x \in A_{\mathsf{YES}}$ then $f(x) \in B_{\mathsf{YES}}$, and if $x \in A_{\mathsf{NO}}$ then $f(x) \in B_{\mathsf{NO}}$.

The complexity class NP is the class of all problems that can be verified by a polynomial-time algorithm. A problem B is NP-hard if for every $A \in \mathsf{NP}$, there exists a polynomial-time reduction from A to B. For any $0 \leq \alpha < \beta \leq 1$, let $\mathsf{BPP}(\alpha, \beta)$ denote the complexity class that consists of all problems $A = (A_{\mathsf{YES}}, A_{\mathsf{NO}})$ for which there exists a randomized algorithm with polynomial time that accepts every $x \in A_{\mathsf{YES}}$ with probability at least β and accepts every $x \in A_{\mathsf{NO}}$ with probability at most α. The complexity classes BPP and RP are defined by $\mathsf{BPP} = \mathsf{BPP}(\frac{1}{3}, \frac{2}{3})$ and $\mathsf{RP} = \mathsf{BPP}(0, \frac{1}{2})$. It is well known that if $\mathsf{NP} \subseteq \mathsf{BPP}$ then $\mathsf{NP} = \mathsf{RP}$.

2.3 Robustness of NP-hardness

We turn to describe the notion of robust reductions, introduced in [5], which is used to prove hardness results on perturbed instances. Let $A = (A_{\mathsf{YES}}, A_{\mathsf{NO}})$ and $B = (B_{\mathsf{YES}}, B_{\mathsf{NO}})$ be two promise problems. For every instance y of B, consider a distribution $\mathsf{noise}(y)$, and suppose that given an instance y it is possible to sample from $\mathsf{noise}(y)$ in polynomial time. Let f be a mapping

from the instances of A to instances of B, and let $\varepsilon \geq 0$. We say that f is a noise-*robust reduction with error* ε from A to B if for every instance x of A, it holds that

1. if $x \in A_{\mathsf{YES}}$, then $f(x) \in B_{\mathsf{YES}}$ and $\mathsf{noise}(f(x)) \in B_{\mathsf{YES}}$ with probability at least $1 - \varepsilon$, and
2. if $x \in A_{\mathsf{NO}}$, then $f(x) \in B_{\mathsf{NO}}$ and $\mathsf{noise}(f(x)) \in B_{\mathsf{NO}}$ with probability at least $1 - \varepsilon$.

The promise problem B is said to be NP-*hard under a* noise-*robust reduction with error* ε if there exists a polynomial-time noise-robust reduction with error ε from an NP-hard problem to B. The notion of hardness under noise-robust reductions is useful for proving hardness results on perturbed instances, as described by the following proposition (see [5, Proposition 1.3]).

Proposition 2.2. *Let $B = (B_{\mathsf{YES}}, B_{\mathsf{NO}})$ be a promise problem, and let $\varepsilon, \alpha \geq 0$ be fixed constants satisfying $\varepsilon + \alpha < 1/2$. For every instance y of B, consider a distribution $\mathsf{noise}(y)$, such that given y it is possible to sample from $\mathsf{noise}(y)$ in polynomial time. Suppose that B is NP-hard under a noise-robust reduction with error ε. Then, unless $\mathsf{NP} = \mathsf{RP}$, there is no polynomial-time algorithm that given an instance y of B and a sample $y' \sim \mathsf{noise}(y)$, decides whether $y' \in B_{\mathsf{YES}}$ or $y' \in B_{\mathsf{NO}}$ with success probability at least $1 - \alpha$ over the choice of y'.*

Proof: Suppose that there exists a polynomial-time noise-robust reduction f with error ε from an NP-hard problem A to B. For the contrapositive, suppose that there exists a polynomial-time algorithm, denoted by Alg, that given an instance y of B and a sample $y' \sim \mathsf{noise}(y)$, decides whether $y' \in B_{\mathsf{YES}}$ or $y' \in B_{\mathsf{NO}}$ with success probability at least $1 - \alpha$ over the choice of y'. Consider the randomized algorithm that given an instance x of A applies the reduction f to obtain an instance $y = f(x)$ of B, picks at random an instance $y' \sim \mathsf{noise}(y)$, and runs Alg on the pair (y, y'). The running time of this algorithm is clearly polynomial.

We turn to analyze the success probability. For every input x, with probability at least $1 - \varepsilon$ over the choice of y', it holds that if $x \in A_{\mathsf{YES}}$ then $y' \in B_{\mathsf{YES}}$ and if $x \in A_{\mathsf{NO}}$ then $y' \in B_{\mathsf{NO}}$, and Alg succeeds on (y, y') with probability at least $1 - \alpha$. By the union bound, the probability that the algorithm fails on x is at most $\varepsilon + \alpha < 1/2$, hence the algorithm succeeds with the complement probability, which is at least some constant larger than $1/2$. By standard amplification, it follows that A can be solved by a randomized polynomial-time algorithm with high success probability, hence $\mathsf{NP} \subseteq \mathsf{BPP}$, which yields that $\mathsf{NP} = \mathsf{RP}$. ∎

2.4 Random Graphs and Subgraphs

Definition 2.3. *For a graph or a digraph $G = (V, E)$ and for $p \in [0, 1]$, let $G_{p,e} = (V, E')$ denote a random subgraph of G on the vertex set V where each edge of E is included in E' independently with probability p.*

For the complete graph K_n on the vertex set $[n]$, we denote its random subgraph $(K_n)_{p,e}$ by the standard notation $G(n, p)$. We need the following result on the distribution of the clique number of $G(n, p)$ (for a proof, see, e.g., [11, Theorem 7.3 and Equation (7.7)]).

Theorem 2.4. *For every fixed $p \in (0, 1)$, there exists a constant $a_p > 0$, such that the probability that there exists a clique of size at least $a_p \cdot \log n$ in the random graph $G(n, p)$ is at least $1 - e^{-\Omega(n^2/\log^5 n)}$.*

3 Proofs of Results

3.1 Digraphs with Minrank at Most Two

In this section we prove the following hardness result. Here, for a digraph $D = (V, E)$, the complement $\overline{D}$ of D is the digraph on V with edge set $\{(u, v) \in V \times V \mid u \neq v,\ (u, v) \notin E\}$.

Theorem 3.1. *Unless* NP = RP, *for all constants $p \in (0, 1)$ and $\beta > \frac{1}{2}$, there is no polynomial-time algorithm that given a digraph D and a random subgraph $D' \sim D_{1-p,e}$ decides whether* $\mathrm{minrk}_{\mathbb{F}_2}(\overline{D'}) \leq 2$ *with success probability at least β over the choice of D'.*

Notice that if $D' \sim D_{1-p,e}$ then the digraph $\overline{D'}$ is obtained from $\overline{D}$ by adding each of its non-edges independently with probability p, hence Theorem 3.1 confirms Theorem 1.1. The proof employs a characterization of Dau et al. [9] for digraphs with minrank at most 2 over $\mathbb{F}_2$, which is based on the notion of fair colorings of digraphs.

3.1.1 Fair Colorability

The notion of fair 3-colorability is defined as follows (see [9, Definition IV.5 and Lemma IV.6]).

Definition 3.2. *A digraph $D = (V, E)$ is* fairly 3-colorable *if there exists a partition of its vertex set into three sets $V = V_1 \cup V_2 \cup V_3$, such that for every vertex $u \in V_i$ with $i \in [3]$, it holds that the set $N_D(u) = \{v \in V \mid (u, v) \in E\}$ of the out-neighbors of u in D is contained in V_j for some $j \in [3] \setminus \{i\}$. We refer to such a partition as a* fair 3-coloring *of D with V_1, V_2, V_3 as color classes.*

The following result of [9] relates the minrank of a digraph over $\mathbb{F}_2$ to the fair 3-colorability of its complement.

Lemma 3.3 ([9, Theorem IV.7])**.** *For every digraph D, it holds that D is fairly 3-colorable if and only if* $\mathrm{minrk}_{\mathbb{F}_2}(\overline{D}) \leq 2$.

We also need the following special case of a result of [9], which was combined there with Lemma 3.3 above to obtain the hardness of the minrank of digraphs over $\mathbb{F}_2$.

Theorem 3.4 ([9, Theorem V.1])**.** *The problem of deciding whether a given digraph is fairly 3-colorable is* NP*-hard.*

3.1.2 Proof of Theorem 3.1

Fix a constant $p \in (0, 1)$. By Proposition 2.2, it suffices to show that the problem of deciding whether a given digraph D satisfies $\mathrm{minrk}_{\mathbb{F}_2}(\overline{D}) \leq 2$ is NP-hard under a noise-robust reduction with error $o(1)$, where $\mathsf{noise}(D)$ refers to the distribution $D_{1-p,e}$ and the $o(1)$ term tends to zero as the input size grows. In fact, by considering the complement digraph, it follows from Lemma 3.3 that it suffices to show such a reduction for the problem of deciding whether a given digraph is fairly 3-colorable, whose NP-hardness is given by Theorem 3.4. We present a noise-robust reduction from this problem to itself.

The reduction. For a given digraph $G = (V, E)$ with n vertices, the reduction outputs the R-blowup D of the digraph G, where $R = c \cdot n$ for some constant $c = c(p)$ to be determined later. Namely, every vertex $v \in V$ of G is replaced in D by a set I_v of R vertices, and every directed edge $(u, v) \in E$ of G is replaced in D by all the R^2 possible directed edges from the vertices of I_u to those of I_v. Notice that the reduction can be implemented in polynomial time.

Correctness. Let $G = (V, E)$ be an input digraph with n vertices, and let D be its R-blowup for the R defined by the reduction. Let D' be a random subgraph of D distributed like $D_{1-p,e}$.

We first claim that if G is fairly 3-colorable then D' is fairly 3-colorable with probability 1. To see this, consider a fair 3-coloring of G and observe that it induces a fair 3-coloring of D by assigning the color of every vertex $v \in V$ in G to the vertices of I_v in D. This implies that every subgraph of D is fairly 3-colorable, hence the random subgraph D' is fairly 3-colorable with probability 1.

For the soundness proof, let us consider the event $\mathcal{E}$ defined as follows.

Definition 3.5. *Let $\mathcal{E}$ denote the event that for every choice of $\frac{R}{3}$-subsets $I'_u \subseteq I_u$ for the vertices $u \in V$, the following holds. For every vertex $u \in V$, there exists a vertex $w \in I'_u$ such that for every $v \in N_G(u)$, there exists an edge in D' from w to some vertex in I'_v.*

We turn to show that the event $\mathcal{E}$ occurs with high probability over the choice of D'. Note that the $o(1)$ term tends to zero as n tends to infinity.

Lemma 3.6. *The event $\mathcal{E}$ occurs with probability $1 - o(1)$.*

Proof: Let $u \in V$ be a vertex of G, and set $d = |N_G(u)|$. Let A_u denote the event that there exists a choice of $\frac{R}{3}$-subsets $I'_v \subseteq I_v$ for $v \in \{u\} \cup N_G(u)$, for which no vertex $w \in I'_u$ satisfies that for every $v \in N_G(u)$, there exists an edge in D' from w to some vertex in I'_v. Fix a choice of $\frac{R}{3}$-subsets $I'_v \subseteq I_v$ for the vertices $v \in \{u\} \cup N_G(u)$. Observe that for every fixed pair of vertices $w \in I'_u$ and $v \in N_G(u)$, the probability that D' does not include an edge from w to any vertex in I'_v is $p^{R/3}$. This implies that the probability that D' includes for every $v \in N_G(u)$ an edge from w to some vertex in I'_v is

$$(1 - p^{R/3})^d \geq 1 - d \cdot p^{R/3}.$$

Hence, the probability that no vertex $w \in I'_u$ satisfies this condition does not exceed

$$(d \cdot p^{R/3})^{R/3}.$$

By the union bound over the choices of the sets I'_v for $v \in \{u\} \cup N_G(u)$, it follows that the probability of the event A_u is at most

$$\begin{aligned} \binom{R}{R/3}^{d+1} \cdot (d \cdot p^{R/3})^{R/3} &\leq 2^{R \cdot (d+1)} \cdot 2^{(\log d) \cdot R/3} \cdot p^{R^2/9} \\ &\leq 2^{4Rn/3} \cdot 2^{(\log p) \cdot R^2/9} \leq 2^{-\Omega(n^2)}, \end{aligned}$$

where the second inequality follows by $d < n$ and the third by the fact that $R = c \cdot n$ for a sufficiently large constant c, say, $c = -24/\log p$. Using again the union bound, it follows that the probability that there exists a vertex $u \in V$ for which the event A_u occurs does not exceed $n \cdot 2^{-\Omega(n^2)} \leq 2^{-\Omega(n^2)}$. This implies that the event $\mathcal{E}$ occurs with probability $1 - o(1)$, and we are done. ∎

The following lemma shows that whenever the event $\mathcal{E}$ holds, so does the soundness of the reduction.

Lemma 3.7. *If the digraph D' is fairly 3-colorable and the event $\mathcal{E}$ occurs, then G is fairly 3-colorable.*

Proof: Suppose that D' is fairly 3-colorable and that the event $\mathcal{E}$ occurs. Consider a fair 3-coloring of D', and recall that $|I_u| = R$ for every $u \in V$. Since the given coloring uses three colors, for every vertex $u \in V$, there exists an $\frac{R}{3}$-subset $I'_u \subseteq I_u$ such that the vertices of I'_u share a common color. Since the event $\mathcal{E}$ occurs, for every vertex $u \in V$, there exists a vertex in I'_u that is connected by an edge in D' to some vertex in I'_v for every $v \in N_G(u)$. Since the given 3-coloring of D' is fair, this implies that all the vertices of the sets I'_v with $v \in N_G(u)$ share a common color, and this color is different from the color of the vertices of I'_u. Therefore, by assigning to every vertex $u \in V$ the color of the vertices of I'_u in D', we obtain that G is fairly 3-colorable, as required. ■

Equipped with the above lemmas, we complete the soundness proof of the reduction. Suppose that G is not fairly 3-colorable. By Lemma 3.7, the digraph D' is not fairly 3-colorable unless the event $\mathcal{E}$ does not occur. By Lemma 3.6, the probability that $\mathcal{E}$ does not occur is $o(1)$, so with probability $1 - o(1)$ over the choice of D', it holds that D' is not fairly 3-colorable, as desired. In addition, the digraph D itself forms a sample from D' for which the event $\mathcal{E}$ occurs, hence it is not fairly 3-colorable. This completes the proof of Theorem 3.1.

3.2 Graphs with Minrank at Most Three

In this section we prove the following hardness result.

Theorem 3.8. *Unless* NP = RP, *for all constants $p \in (0,1)$ and $\beta > \frac{1}{2}$ and for every field $\mathbb{F}$, there is no polynomial-time algorithm that given a graph G and a random subgraph $G' \sim G_{1-p,e}$ decides whether* $\mathrm{minrk}_{\mathbb{F}}(\overline{G'}) \leq 3$ *with probability at least β over the choice of G'.*

Notice that if $G' \sim G_{1-p,e}$ then the graph $\overline{G'}$ is obtained from $\overline{G}$ by adding each of its non-edges independently with probability p, hence Theorem 3.8 confirms Theorem 1.2. The proof borrows a gadget graph that was used by Peeters [18] in his hardness proof for the minrank parameter. We start by presenting this gadget and some of its properties.

3.2.1 A Gadget Graph

For two given vertices u and v, the gadget graph $P_{u,v}$ consists of six vertices: the vertices u, v and four additional vertices, denoted a, b, c, d. It includes two triangles whose vertices are $\{a, b, u\}$ and $\{c, d, v\}$ and a matching that connects the vertices a, b, u to the vertices v, c, d respectively (see Figure 1).

The following lemma summarizes two simple properties of $P_{u,v}$ given in [18].

Lemma 3.9 ([18])**.** *The gadget graph $P_{u,v}$, given in Figure 1, satisfies the following properties.*

1. *There exists a proper 3-coloring of $P_{u,v}$ that assigns to u and v the same color, and there exists a proper 3-coloring of $P_{u,v}$ that assigns to u and v distinct colors.*

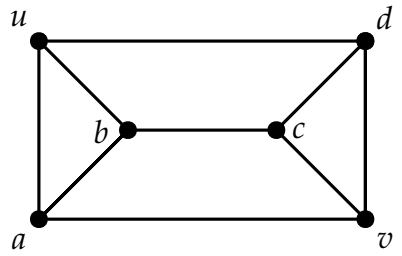

Figure 1: The gadget graph $P_{u,v}$

2. *Let M be a matrix that represents the complement of the graph $P_{u,v}$ over a field $\mathbb{F}$ with $\mathrm{rank}_{\mathbb{F}}(M) \leq 3$. Then, M has one of the following two patterns, where each $\star$ stands for a nonzero element of $\mathbb{F}$.*

$$\begin{array}{c} \\ u \\ a \\ b \\ c \\ d \\ v \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} u & a & b & c & d & v \end{array} \\ \begin{pmatrix} \star & 0 & 0 & \star & 0 & 0 \\ 0 & \star & 0 & 0 & \star & 0 \\ 0 & 0 & \star & 0 & 0 & \star \\ \star & 0 & 0 & \star & 0 & 0 \\ 0 & \star & 0 & 0 & \star & 0 \\ 0 & 0 & \star & 0 & 0 & \star \end{pmatrix} \end{array} \qquad \begin{array}{c} \\ u \\ a \\ b \\ c \\ d \\ v \end{array}\!\!\begin{array}{c} \begin{array}{cccccc} u & a & b & c & d & v \end{array} \\ \begin{pmatrix} \star & 0 & 0 & 0 & 0 & \star \\ 0 & \star & 0 & \star & 0 & 0 \\ 0 & 0 & \star & 0 & \star & 0 \\ 0 & \star & 0 & \star & 0 & 0 \\ 0 & 0 & \star & 0 & \star & 0 \\ \star & 0 & 0 & 0 & 0 & \star \end{pmatrix} \end{array}.$$

We next prove the following lemma that extends an argument of [18].

Lemma 3.10. *Let $H = (V, E)$ be a graph, and let M be a matrix that represents its complement $\overline{H}$ over a field $\mathbb{F}$ such that $\mathrm{rank}_{\mathbb{F}}(M) \leq 3$. Let $A_1, \ldots, A_t \subseteq V$ be t sets of vertices, such that*

1. *for every $i \in [t]$, the rows of M that correspond to the vertices of A_i are pairwise proportional (i.e., equal up to a scalar multiplication over $\mathbb{F}$),*
2. *for every $i \in [t]$ and for every distinct $u, v \in A_i$, the graph H contains a $P_{u,v}$ gadget, and*
3. *for every distinct $i, j \in [t]$, there exist vertices $u \in A_i$ and $v \in A_j$ for which the graph H contains a $P_{u,v}$ gadget.*

Then, there exist three (not necessarily distinct) rows in M such that for every $i \in [t]$, the rows of M that correspond to the vertices of A_i are proportional to one of them.

Proof: For every $i \in [t]$, let V_i denote the subspace spanned by the rows of M that correspond to the vertices of A_i. By assumption, we have $\dim(V_i) = 1$.

We first claim that for every $i \in [t]$, the values of every row of M in the coordinates that correspond to vertices of A_i are either all zero or all nonzero. If $|A_i| = 1$, then this trivially holds. Otherwise, let u and v be two vertices of A_i. By assumption, the graph H contains a $P_{u,v}$ gadget. Using $\mathrm{rank}_{\mathbb{F}}(M) \leq 3$, Item 2 of Lemma 3.9 implies that the restriction of the matrix M to the vertices of this gadget has one of the two patterns given by the lemma. In fact, since the rows of M that correspond to u and v are proportional, this gadget must fit the right-hand pattern. Now, observe that the rows of the vertices u, a, b of this gadget span the entire row space of M, and that this subspace does not include a vector whose value in the coordinate of u is zero and whose value

in the coordinate of v is not (or vice versa). Hence, the values of every row of M in the coordinates of A_i are either all zero or all nonzero.

Now, for every $i \in [t]$, let U_i denote the subspace spanned by the rows of M whose values in the coordinates that correspond to the vertices of A_i are zeros. Note that the subspace V_i is nonzero on all the coordinates that correspond to vertices of A_i, hence V_i is not contained in U_i.

We claim that for every $i, j \in [t]$, either $V_i \subseteq U_j$ or $V_i = V_j$. Indeed, for distinct $i, j \in [t]$, the assumption of the lemma implies that the graph H contains a $P_{u,v}$ gadget for some vertices $u \in A_i$ and $v \in A_j$. By combining the assumption $\text{rank}_{\mathbb{F}}(M) \leq 3$ with Item 2 of Lemma 3.9, the restriction of the matrix M to the vertices of this gadget has one of the two patterns given by the lemma. For each of them, it follows that the rows of M whose value in the coordinate of u is zero span a subspace of dimension at least 2, hence $\dim(U_i) \geq 2$. Since V_i is not contained in U_i, using again the fact that $\text{rank}_{\mathbb{F}}(M) \leq 3$, it follows that $\dim(U_i) = 2$. Furthermore, if the restriction of M to the vertices of $P_{u,v}$ corresponds to the left-hand pattern given in Lemma 3.9, then the value of the row of v in the coordinate of the vertex u is zero, hence all its values in the coordinates that correspond to vertices of A_i are zeros, and thus $V_j \subseteq U_i$. If, however, it corresponds to the right-hand pattern, then it follows that $V_i = V_j$.

Finally, suppose that V_i and V_j are distinct subspaces for some $i, j \in [t]$. The above discussion implies that every subspace V_k which is different from both V_i and V_j satisfies $V_k \subseteq U_i \cap U_j$. Since V_i and V_j are distinct, it follows that U_i and U_j are also distinct (because V_i is contained in U_j but not in U_i), so by combining $\dim(U_i) = \dim(U_j) = 2$ with $\text{rank}_{\mathbb{F}}(M) \leq 3$, it follows that $\dim(U_i \cap U_j) = 1$, hence $V_k = U_i \cap U_j$. This implies that there are no more than three distinct subspaces V_i for $i \in [t]$, hence there exist three rows in M as required by the lemma. ■

3.2.2 Proof of Theorem 3.8

Fix a constant $p \in (0, 1)$ and a field $\mathbb{F}$. By Proposition 2.2, it suffices to show that the problem of deciding whether a given graph G satisfies $\text{minrk}_{\mathbb{F}}(\overline{G}) \leq 3$ is NP-hard under a noise-robust reduction with error $o(1)$, where $\text{noise}(G)$ refers to the distribution $G_{1-p,e}$ and the $o(1)$ term tends to zero as the input size grows.

The reduction. We reduce from the 3-colorability problem. Let $G = (V, E)$ be an instance of this problem. Set $n = |V|$ and $R = 2^{c \cdot (\log n)^{1/2}}$ for some constant $c = c(p)$ to be determined later. The reduction outputs a graph H defined as follows. For every vertex $u \in V$, define a set

$$C_u = \{(u, i) \mid i \in [R]\}$$

of R vertices. We refer to C_u as the cloud of the vertex u and define $C = \cup_{u \in V} C_u$. For every two vertices $u, v \in V$ that are adjacent in G, we add to H all the R^2 possible edges between the vertices of C_u and those of C_v. Finally, for every two distinct vertices $(u, i), (v, j) \in C$, including the case $u = v$, we add to H the gadget graph $P_{(u,i),(v,j)}$ (see Section 3.2.1). Note that for every such gadget, we add to the graph four new vertices and nine edges. Notice that the definition of R implies that the reduction can be implemented in polynomial time.

Correctness. Let $G = (V, E)$ be an input graph with $n = |V|$, let H be the graph defined by the above reduction, and let H' be a random subgraph of H distributed like $H_{1-p,e}$.

We first claim that if G is 3-colorable then $\text{minrk}_{\mathbb{F}}(\overline{H'}) \leq 3$ with probability 1. To see this, consider a proper 3-coloring of G, and for every $u \in V$, assign the color of u to all the vertices of the cloud C_u in H. Since every two adjacent vertices in G receive distinct colors, it follows that the endpoints of the edges of H that connect vertices of distinct clouds receive distinct colors as well. Further, by Item 1 of Lemma 3.9, for every two distinct vertices $(u,i),(v,j) \in C$, the 3-coloring can be properly extended to the other four vertices of the gadget $P_{(u,i),(v,j)}$. It thus follows that H is 3-colorable, implying that its subgraph H' is 3-colorable with probability 1. This yields, using (1), that $\text{minrk}_{\mathbb{F}}(\overline{H'}) \leq \chi(H') \leq 3$ with probability 1, as desired.

For the soundness proof, let us consider the event $\mathcal{E}$ defined as follows.

Definition 3.11. *Let $\mathcal{E}$ denote the event that for some integer $r \geq 3$, there exists a choice of r-subsets $C'_u \subseteq C_u$ for the vertices $u \in V$ such that*

1. *for every $u \in V$ and for every distinct $(u,i),(u,j) \in C'_u$, all the edges of the gadget $P_{(u,i),(u,j)}$ of H are included in H', and*

2. *for every choice of $\frac{r}{3}$-subsets $C''_u \subseteq C'_u$ for the vertices $u \in V$, it holds that for every distinct $u,v \in V$, there exist vertices $(u,i) \in C''_u$ and $(v,j) \in C''_v$ for which all the edges of the gadget $P_{(u,i),(v,j)}$ of H are included in H', and in addition, if (u,i) and (v,j) are adjacent in H, then the edge that connects them is included in H'.*

We turn to show that the event $\mathcal{E}$ occurs with high probability over the choice of H'.

Lemma 3.12. *The event $\mathcal{E}$ occurs with probability $1 - o(1)$.*

Proof: It will be convenient to assume here that the random subgraph H' of H is chosen in two stages. In the first stage, every edge of the gadgets $P_{(u,i),(u,j)}$ with $u \in V$ and $i,j \in [R]$, i.e., gadgets that correspond to pairs of vertices in the same cloud, is removed independently with probability p. In the second stage, every other edge of H, that is, every edge that connects vertices from distinct clouds in H or from gadgets $P_{(u,i),(v,j)}$ with distinct $u,v \in V$, is removed independently with probability p. Note that the graph obtained after these two stages is distributed like $H_{1-p,e}$

Consider first the subgraph of H obtained after the first stage. For every $u \in V$, let F_u denote the graph on the vertex set $[R]$, in which every two distinct vertices $i,j \in [R]$ are adjacent if all the edges of the gadget $P_{(u,i),(u,j)}$ of H survive in the first stage. Since every such gadget has nine edges, it follows that the graph F_u is distributed like $G(R,q)$ where $q = (1-p)^9$. By Theorem 2.4, for some constant $a = a(p) > 0$ and for $r = a \cdot \log R = a \cdot c \cdot \sqrt{\log n}$, the probability that there is no clique of size r in F_u does not exceed $e^{-\Omega(R^2/\log^5 R)}$. By the union bound, the probability that there exists $u \in V$ for which there is no such clique in F_u is at most $n \cdot e^{-\Omega(R^2/\log^5 R)} = o(1)$, where the $o(1)$ bound follows from the definition of R (recall that $R = 2^{\Theta((\log n)^{1/2})}$). Hence, with probability $1 - o(1)$, there exists a choice of r-subsets $C'_u \subseteq C_u$ for $u \in V$ satisfying Item 1 of Definition 3.11.

Consider now the subgraph H' obtained after the second stage. We turn to show that with high probability the above sets C'_u for $u \in V$, whose choice is independent of the removal of edges in the second stage, satisfy Item 2 of Definition 3.11. To see this, fix two distinct vertices $u,v \in V$, and let $A_{u,v}$ denote the event that there exist sets $C''_u \subseteq C'_u$ and $C''_v \subseteq C'_v$ of size $|C''_u| = |C''_v| = \frac{r}{3}$ for which no pair of vertices $(u,i) \in C''_u$ and $(v,j) \in C''_v$ satisfies that all the edges of $P_{(u,i),(v,j)}$ in H as well as the edge that connects them, if exists in H, are included in H'. Fix a choice of such sets

C''_u and C''_v of size $\frac{r}{3}$, and observe that for every $(u,i) \in C''_u$ and $(v,j) \in C''_v$, the probability that all of the aforementioned edges are included in H' is either $(1-p)^9$ or $(1-p)^{10}$. It thus follows that the probability that no such pair of vertices satisfies this condition is at most $(1-(1-p)^{10})^{r^2/9}$. By the union bound, the probability of the event $A_{u,v}$ does not exceed

$$\binom{r}{r/3}^2 \cdot (1-(1-p)^{10})^{r^2/9}.$$

Using again the union bound, the probability that for some pair of distinct vertices $u, v \in V$ the event $A_{u,v}$ occurs is at most

$$\binom{n}{2} \cdot \binom{r}{r/3}^2 \cdot (1-(1-p)^{10})^{r^2/9} \leq n^2 \cdot 2^{2r} \cdot e^{-(1-p)^{10} \cdot r^2/9} = o(1),$$

where the $o(1)$ bound follows from the definition of $r = a \cdot c \cdot \sqrt{\log n}$, assuming that the constant $c = c(p)$ in the definition of R is sufficiently large. By applying again the union bound, we obtain that with probability $1 - o(1)$, there exists a choice of r-subsets $C'_u \subseteq C_u$ for $u \in V$ satisfying Items 1 and 2 of Definition 3.11, hence with such probability, the event $\mathcal{E}$ occurs. ■

The following lemma shows that whenever the event $\mathcal{E}$ holds, so does the soundness of the reduction.

Lemma 3.13. *If the graph H' satisfies* $\text{minrk}_{\mathbb{F}}(\overline{H'}) \leq 3$ *and the event $\mathcal{E}$ occurs, then G is 3-colorable.*

Proof: Suppose that the random subgraph H' of H satisfies $\text{minrk}_{\mathbb{F}}(\overline{H'}) \leq 3$, and let M be a matrix that represents $\overline{H'}$ over $\mathbb{F}$ and satisfies $\text{rank}_{\mathbb{F}}(M) \leq 3$. Suppose further that the event $\mathcal{E}$ occurs, and let $C'_u \subseteq C_u$ be the choice of r-subsets for the vertices $u \in V$ guaranteed by Definition 3.11. By Item 1 of this definition, for every $u \in V$, the graph H' includes a gadget $P_{(u,i),(u,j)}$ for every distinct vertices (u,i) and (u,j) in C'_u. Hence, for every $u \in V$ we can apply Lemma 3.10 to the r sets $\{(u,i)\}$ with $(u,i) \in C'_u$, to obtain that there are three rows in M such that every row of M associated with a vertex of C'_u is proportional to one of them. In particular, there exists a set $C''_u \subseteq C'_u$ of size $|C''_u| = \frac{r}{3}$ such that the rows of M that correspond to the vertices of C''_u are pairwise proportional.

Now, by Item 2 of Definition 3.11, for every distinct $u, v \in V$ there exist vertices $(u,i) \in C''_u$ and $(v,j) \in C''_v$ such that H' includes a gadget $P_{(u,i),(v,j)}$. Hence, we can apply Lemma 3.10 to the n sets C''_u with $u \in V$ to obtain that there are three rows in M, denoted w_1, w_2, w_3, such that every row of M associated with a vertex of C''_u for some $u \in V$ is proportional to one of them. Consider a partition of V into three sets V_1, V_2, V_3, where V_i is a set of vertices $u \in V$ for which the rows of M that correspond to the vertices of C''_u are proportional to w_i. We claim that no edge of G connects two vertices from the same part of this partition, hence G is 3-colorable. Indeed, let u and v be adjacent vertices in G. Then, by Item 2 of Definition 3.11, there exists an edge in H' that connects some vertex $(u,i) \in C''_u$ and some vertex $(v,j) \in C''_v$. This implies that the (u,i) entry of the row of M that corresponds to the vertex (v,j) is zero, whereas its (v,j) entry is nonzero. Hence, the rows of M that correspond to the vertices of C''_u and to those of C''_v are not proportional, yielding that u and v lie in different parts of the partition. ■

Equipped with the above lemmas, we complete the soundness proof of the reduction. Suppose that G is not 3-colorable. By Lemma 3.13, the graph H' satisfies $\text{minrk}_{\mathbb{F}}(\overline{H'}) \geq 4$ unless the event $\mathcal{E}$ does not occur. By Lemma 3.12, the probability that $\mathcal{E}$ does not occur is $o(1)$, so with probability $1 - o(1)$ over the choice of H', it holds that $\text{minrk}_{\mathbb{F}}(\overline{H'}) \geq 4$. In addition, the graph H itself forms a sample from H' for which the event $\mathcal{E}$ occurs, hence it satisfies $\text{minrk}_{\mathbb{F}}(\overline{H}) \geq 4$. This completes the proof of Theorem 3.8.

3.3 The Minrank of Perturbed Graphs

In this section we study the behavior of the minrank parameter on perturbed graphs with perturbation probability $1/2$. The proofs follow ideas applied in [5] in the context of the chromatic number of graphs. Let us start with the following lemma.

Lemma 3.14. *For a graph $G = (V, E)$, let $E = E_1 \cup E_2$ be a partition of its edge set into two sets, and consider the graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$. Then, for every field $\mathbb{F}$, it holds that*

$$\text{minrk}_{\mathbb{F}}(\overline{G}) \leq \text{minrk}_{\mathbb{F}}(\overline{G_1}) \cdot \text{minrk}_{\mathbb{F}}(\overline{G_2}).$$

Proof: Set $k_1 = \text{minrk}_{\mathbb{F}}(\overline{G_1})$ and $k_2 = \text{minrk}_{\mathbb{F}}(\overline{G_2})$, and let M_1 and M_2 be matrices of rank k_1 and k_2 over $\mathbb{F}$ that represent the graphs $\overline{G_1}$ and $\overline{G_2}$ respectively. Let M be the matrix defined by

$$M_{i,j} = (M_1)_{i,j} \cdot (M_2)_{i,j}$$

for all vertices i and j. The matrix M represents the graph $\overline{G}$ over $\mathbb{F}$. Indeed, the values on the diagonal are all nonzero. Further, every distinct non-adjacent vertices i, j in $\overline{G}$ are adjacent in G, and are thus adjacent in G_1 or in G_2. It follows that the (i, j) entry is zero in at least one of the matrices M_1 and M_2, hence it is zero in M as well. Finally, it is well known and easy to check that M is a principal sub-matrix of the Kronecker product $M_1 \otimes M_2$, hence

$$\text{rank}_{\mathbb{F}}(M) \leq \text{rank}_{\mathbb{F}}(M_1 \otimes M_2) = k_1 \cdot k_2.$$

This implies that $\text{minrk}_{\mathbb{F}}(\overline{G}) \leq k_1 \cdot k_2$, as required. ∎

As a consequence, we obtain the following bound on the expectation of the minrank parameter of perturbed graphs with perturbation probability $1/2$.

Lemma 3.15. *Let $\mathbb{F}$ be a field, and let G be a graph with $k = \text{minrk}_{\mathbb{F}}(\overline{G})$. Then, the random subgraph $G' \sim G_{1/2,e}$ satisfies*

$$\mathbf{E}\left[\text{minrk}_{\mathbb{F}}(\overline{G'})\right] \geq \sqrt{k}.$$

Proof: For a graph $G = (V, E)$, consider the two random subgraphs G_1 and G_2 on V, such that every edge of G is chosen uniformly and independently to be included either in G_1 or in G_2. Observe that both G_1 and G_2 are distributed like $G_{1/2,e}$ and thus like G'. This implies that

$$\begin{aligned}
\mathbf{E}\left[\text{minrk}_{\mathbb{F}}(\overline{G'})\right] &= \frac{1}{2} \cdot \mathbf{E}\left[\text{minrk}_{\mathbb{F}}(\overline{G_1}) + \text{minrk}_{\mathbb{F}}(\overline{G_2})\right] \\
&\geq \mathbf{E}\left[\sqrt{\text{minrk}_{\mathbb{F}}(\overline{G_1}) \cdot \text{minrk}_{\mathbb{F}}(\overline{G_2})}\right] \\
&\geq \mathbf{E}\left[\sqrt{\text{minrk}_{\mathbb{F}}(\overline{G})}\right] = \sqrt{k},
\end{aligned}$$

where the second inequality follows from Lemma 3.14. This completes the proof. ∎

We next obtain a lower bound on the typical minrank parameter of perturbed graphs (rather than on their expected minrank). We use the following lemma that was derived in [5] from a result in additive combinatorics.

Lemma 3.16 ([5, Lemma 3.1])**.** *For an integer m and for $\alpha \in (0,1]$, let $\mathcal{A}$ be a collection of subsets of $[m]$ such that $|\mathcal{A}| = \alpha \cdot 2^m$. Then, there exist three sets $A_1, A_2, A_3 \in \mathcal{A}$ such that $|A_1 \cup A_2 \cup A_3| \geq m - 4/\alpha^3$.*

Equipped with Lemma 3.16, we prove the following.

Lemma 3.17. *For a field $\mathbb{F}$, an integer k, and $\alpha \in (0,1)$, let $G = (V, E)$ be a graph that satisfies*

$$\text{minrk}_{\mathbb{F}}(\overline{G}) > k^3 + 8/(1-\alpha)^3.$$

Then, the random subgraph $G' \sim G_{1/2,e}$ satisfies

$$\Pr\left[\text{minrk}_{\mathbb{F}}(\overline{G'}) > k\right] > \alpha.$$

Proof: We prove the contrapositive. Suppose that the random subgraph $G' \sim G_{1/2,e}$ satisfies $\text{minrk}_{\mathbb{F}}(\overline{G'}) > k$ with probability at most α. Put $m = |E|$, and let $\mathcal{A}$ denote the collection of all subsets $E' \subseteq E$ for which the minrank over $\mathbb{F}$ of the complement of the graph (V, E') is at most k. Since the edge set of G' is chosen uniformly over all the subsets of E, our assumption implies that $|\mathcal{A}| \geq (1-\alpha) \cdot 2^m$. By Lemma 3.16, it follows that there exist three sets $E_1, E_2, E_3 \in \mathcal{A}$ for which it holds that $|E_1 \cup E_2 \cup E_3| \geq m - 4/(1-\alpha)^3$.

Let V' be the set of vertices $v \in V$ for which all the edges of G that are incident with v belong to $E_1 \cup E_2 \cup E_3$. For each $i \in [3]$, put $G_i = (V, E_i)$, and let $G_i' = (V', E_i')$ denote its induced subgraph on the vertex set V'. Since $E_i \in \mathcal{A}$, it follows that $\text{minrk}_{\mathbb{F}}(\overline{G_i'}) \leq \text{minrk}_{\mathbb{F}}(\overline{G_i}) \leq k$. By Lemma 3.14, the minrank over $\mathbb{F}$ of the complement of the graph $(V', E_1' \cup E_2' \cup E_3')$ does not exceed k^3. Further, in the induced subgraph of G on $V \setminus V'$, every vertex lies on an edge of $E \setminus (E_1 \cup E_2 \cup E_3)$, hence $|V \setminus V'| \leq 8/(1-\alpha)^3$. This implies that $\text{minrk}_{\mathbb{F}}(\overline{G}) \leq k^3 + 8/(1-\alpha)^3$, and we are done. ■

By combining Lemma 3.17 with Proposition 2.2 applied with the identity reduction, one can derive hardness results on the minrank parameter of perturbed instances with perturbation probability $1/2$ from worst-case hardness with a sufficiently large gap between YES and NO instances. This is stated in the following proposition.

Proposition 3.18. *Suppose that for some integers $k_2 > k_1$ and for some field $\mathbb{F}$, it is* NP*-hard to decide whether a given graph G satisfies $\text{minrk}_{\mathbb{F}}(G) \leq k_1$ or $\text{minrk}_{\mathbb{F}}(G) \geq k_2$. Then, unless* NP $=$ RP*, there is no polynomial-time algorithm that given a graph G and a random subgraph $G' \sim G_{1/2,e}$ decides whether $\text{minrk}_{\mathbb{F}}(\overline{G'}) \leq k_1$ or $\text{minrk}_{\mathbb{F}}(\overline{G'}) \geq \Omega(k_2^{1/3})$, with high success probability over the choice of G'.*

We finally combine Proposition 3.18 with known worst-case hardness results on the minrank parameter to obtain the following consequences.

- Let $\mathbb{F}$ be a fixed finite field, and let k be a sufficiently large integer. It was shown in [8] that it is NP-hard to decide whether a graph G satisfies $\text{minrk}_{\mathbb{F}}(G) \leq k$ or $\text{minrk}_{\mathbb{F}}(G) \geq 2^{(1-o(1))\cdot k/2}$. By Proposition 3.18, we obtain that unless NP $=$ RP, there is no polynomial-time algorithm that given a graph G and a random subgraph $G' \sim G_{1/2,e}$ decides whether $\text{minrk}_{\mathbb{F}}(\overline{G'}) \leq k$ or $\text{minrk}_{\mathbb{F}}(\overline{G'}) \geq 2^{(1-o(1))\cdot k/6}$, with high success probability over the choice of G'.

- Let $\mathbb{F}$ be a fixed finite field. It was shown in [17] that assuming certain variants of the unique games conjecture, for all integers $k_2 > k_1 \geq 3$, it is NP-hard to decide whether a given graph G satisfies $\mathrm{minrk}_{\mathbb{F}}(G) \leq k_1$ or $\mathrm{minrk}_{\mathbb{F}}(G) \geq k_2$. By Proposition 3.18, we obtain that under the same conjectures and assuming that NP $\neq$ RP, for all integers $k_2 > k_1 \geq 3$, there is no polynomial-time algorithm that given a graph G and a random subgraph $G' \sim G_{1/2,e}$ decides whether $\mathrm{minrk}_{\mathbb{F}}(\overline{G'}) \leq k_1$ or $\mathrm{minrk}_{\mathbb{F}}(\overline{G'}) \geq k_2$, with high success probability over the choice of G'. We remark that this statement can also be derived by combining a hardness result on the chromatic number of perturbed instances, which follows from [5, Lemma 3.9], with the relations between these graph parameters given in (1).

Acknowledgments

We thank the anonymous referees for their insightful comments and suggestions that improved the presentation of this paper.

References

[1] N. Alon, I. Balla, L. Gishboliner, A. Mond, and F. Mousset. The minrank of random graphs over arbitrary fields. *Isr. J. Math.*, 235:63–77, 2020.

[2] F. Arbabjolfaei and Y. Kim. Fundamentals of index coding. *Found. Trends Commun. Inf. Theory*, 14(3–4):163–346, 2018.

[3] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2006.

[4] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol. Index coding with side information. *IEEE Trans. Inform. Theory*, 57(3):1479–1494, 2011. Preliminary vesrion in FOCS'06.

[5] H. Bennett, D. Reichman, and I. Shinkar. On percolation and NP-hardness. *Random Struct. Algorithms*, 54(2):228–257, 2019. Preliminary version in ICALP'16.

[6] 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. Inform. Theory*, 52(6):2825–2830, 2006. Preliminary version in INFOCOM'98.

[7] A. Blasiak, R. Kleinberg, and E. Lubetzky. Lexicographic products and the power of non-linear network coding. In *Proc. of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS'11)*, pages 609–618, 2011.

[8] D. Chawin and I. Haviv. Improved NP-hardness of approximation for orthogonality dimension and minrank. *SIAM J. Discret. Math.*, 37(4):2670–2688, 2023. Preliminary version in STACS'23.

[9] S. H. Dau, V. Skachek, and Y. M. Chee. Optimal index codes with near-extreme rates. *IEEE Trans. Inform. Theory*, 60(3):1515–1527, 2014. Preliminary version in ISIT'12.

[10] I. Dinur and I. Shinkar. On the conditional hardness of coloring a 4-colorable graph with super-constant number of colors. In *Proc. of the 13th Intl. Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX'10)*, pages 138–151, 2010.

[11] A. Frieze and M. Karoński. *Introduction to Random Graphs*. Cambridge University Press, 2015.

[12] A. Golovnev, O. Regev, and O. Weinstein. The minrank of random graphs. *IEEE Trans. Inform. Theory*, 64(11):6990–6995, 2018. Preliminary version in RANDOM'17.

[13] W. H. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(2):231–232, 1979.

[14] I. Haviv and M. Langberg. On linear index coding for random graphs. In *Proc. of the IEEE Int. Symposium on Information Theory (ISIT'12)*, pages 2231–2235, 2012.

[15] I. Haviv and M. Langberg. H-wise independence. *Chic. J. Theor. Comput. Sci.*, 2019, 2019. Preliminary version in ITCS'13.

[16] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. of the 34th Annual ACM Symposium on Theory of Computing (STOC'02)*, pages 767–775, 2002.

[17] M. Langberg and A. Sprintson. On the hardness of approximating the network coding capacity. *IEEE Trans. Inform. Theory*, 57(2):1008–1014, 2011. Preliminary version in ISIT'08.

[18] R. Peeters. Orthogonal representations over finite fields and the chromatic number of graphs. *Combinatorica*, 16(3):417–431, 1996.

[19] P. Pudlák, V. Rödl, and J. Sgall. Boolean circuits, tensor ranks, and communication complexity. *SIAM J. Comput.*, 26(3):605–633, 1997.

[20] A. Sharififar, N. Aboutorab, and P. Sadeghi. An update-based maximum column distance coding scheme for index coding. *IEEE J. Sel. Areas Inf. Theory*, 2(4):1282–1299, 2021. Preliminary version in ISIT'21.

[21] L. G. Valiant. Graph-theoretic arguments in low-level complexity. In *Proc. of the 6th International Symposium on Mathematical Foundations of Computer Science (MFCS'77)*, pages 162–176, 1977.