

# Index coding with side information

Ehsan Ebrahimi Targhi  
University of Tartu

**Abstract.** The Index Coding problem has attracted a considerable amount of attention in the recent years. The problem is motivated by several applications in wireless networking and distributed computing. An instance of the index coding problem includes a sender that holds an input  $x = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n$  and communicates with  $n$  receivers  $R_1, \dots, R_n$  using wireless broadcast channel. Each receiver  $R_i$  needs to obtain bit  $x_i$  and has prior side information about  $x$ . In this report, we study the minimum length of linear index codes that is characterized by a measure on graphs, called the minrank.

## 1 Introduction

Source encoding is one of central areas of coding and information theory. Source coding involves changing the message source to a suitable code to be transmitted through the channel. According to Shannon's source coding theorem, the average number of bits necessary and sufficient to encode a source is equal (up to one bit) to the entropy of the source. In a situation that the receiver has some prior side information about the source message, source coding with side information deals with presenting encoding schemes that utilize the side information in order to reduce the length of the code.

Assume a situation where a server disseminates a set of data blocks over a broadcast channel to a set of caching clients. After transmission, each client possesses in its cache only a subset of the transmitted blocks, due to reception problem, limited storage space, etc. The client uses a backward channel to request blocks that he needs but it has not cached and to inform the server about the blocks that has in its cache. The server uses this information to minimize the amount of supplemental information that must be retransmitted in order to enable every client to acquire all its requested blockes.

In [1], it is assumed that every client either possesses an entire error-free block or does not possess it at all. We can formalize this as following. There is a server that has a bit string  $x = (x_1, x_2, \dots, x_n)$  of length  $n$  ( $x \in \{0, 1\}^n$ ) and wishes to transmit the requested bits of  $n$  receivers  $R_1, \dots, R_n$ . Each receiver  $R_i$  needs to obtain the bit (block)  $x_i$  and has prior side information about  $x$ . The side information is characterized by a simple directed graph  $G$  on  $\{1, 2, \dots, n\}$ , called side information graph, and the side information of client  $R_i$  is the projection of  $x$  on the set of all *outneighbours* of node  $i$  in graph  $G$ . In other words,

every node  $i \in \{1, 2, \dots, n\}$  is corresponded to the receiver  $R_i$  and the ordered pair  $(i, j)$  is an edge in  $G$  if and only if the receiver  $R_i$  knows the bit  $x_j$ .

The next example shows how the sender can reduce the length of transmitted data due to knowledge of side information of receivers.

**Example 1.**

Suppose that the side information graph is a directed cycle of length  $n$ . In other words, for every  $i = 2, \dots, n$  receiver  $R_i$  is interested in value  $x_i$  but knows value  $x_{i-1}$  as side information and receiver  $R_1$  knows  $x_n$ . The sender can transmit  $n - 1$ -tuple  $(x_1 \oplus x_2, x_2 \oplus x_3, \dots, x_{n-1} \oplus x_n)$  instead of broadcasting all the bits of  $x$ . Each receiver  $R_i$  for  $i > 1$  is able to recover bit  $x_i$  from  $x_{i-1} \oplus x_i$  due to knowledge of the bit  $x_{i-1}$ . Finally, receiver  $R_1$  XORs all the coordinates of  $n - 1$ -tuple broadcast by the sender together with  $x_n$  to recover  $x_1$ .

**Definition 1.** A deterministic INDEX code  $C$  for  $\{0, 1\}^n$  with side information graph  $G$  on  $n$  nodes, abbreviated as INDEX code for  $G$ , is a set of codewords in  $\{0, 1\}^l$  together with:

- An encoding function  $E$  mapping inputs in  $\{0, 1\}^n$  to codewords, and
- A set of decoding functions  $D_1, D_2, \dots, D_n$  such that  $D_i(E(x), x[N(i)]) = x_i$  for every  $i$ , where  $N(i)$  is the set of all outneighbours of node  $i$  in graph  $G$  and  $x[N(i)]$  is the projection of  $x$  on the set  $N(i)$ .

The rest of the paper is organized as follows. Section 2 gives a list of basic definitions and theorems related to graph theory. In section 3 we construct a linear INDEX code that its length is *minrank* of its side information graph and show that this code is optimal. Section 4 discusses the length of any INDEX code (not only linear) for some restricted graphs.

## 2 Graph Theory

In this section, we state some definitions and theorems related to graph theory needed for the paper, and refer the reader to [5] for more information.

**Graph:** A graph  $G$  is an ordered pair  $(V(G), E(G))$  consisting of a set  $V(G)$  of vertices and a set  $E(G)$ , disjoint from  $V(G)$ , of edges, together with an incidence function  $\Phi(G)$  that associates with each edge of  $G$  an unordered pair of (not necessarily distinct) vertices of  $G$ . A graph is a simple graph if for every  $u \in V(G)$ ,  $\{u, u\} \notin E(G)$ .

**Complement of a graph:** The complement of a simple graph  $G$ , denoted by  $\overline{G}$ , is the simple graph whose vertex set is  $V(G)$  and whose edges are the pairs of nonadjacent vertices of  $G$ .

**Subgraph:** A graph  $F$  is called a subgraph of a graph  $G$  if  $V(F) \subseteq V(G)$ ,  $E(F) \subseteq E(G)$ , and  $\Phi(F)$  is the restriction of  $\Phi(G)$  to  $E(F)$  (or  $\Phi(F)$  is a function with domain  $E(F)$  such that  $\Phi(F)(e) = \Phi(G)(e)$  for every  $e \in E(F)$ ).

**Induced subgraph:** A subgraph  $F = (V(F), E(F))$  of a graph  $G = (V(G), E(G))$  is an induced subgraph if for any pair of vertices  $u$  and  $v$  of  $F$ ,  $\{u, v\} \in E(F)$  if and only if  $\{u, v\} \in E(G)$ .

**Chromatic number:** The chromatic number of a graph  $G$ ,  $\chi(G)$ , is the smallest number of colors needed to color the vertices of  $G$  so that no two adjacent vertices share the same color.

**Clique:** A clique of a graph  $G$  is a set of mutually adjacent vertices, and the maximum size of a clique of a graph  $G$ , the clique number of  $G$ , is denoted  $\omega(G)$ .

**Perfect graph:** A graph  $G$  is perfect if  $\chi(H) = \omega(H)$  for every induced subgraph  $H$  of  $G$ .

**Vector cover:** A set of vertices such that each edge of the graph is incident to at least one vertex of the set.

**Independent set** in a graph  $G$  is a set of vertices no two of which are adjacent.  $\alpha(G) = \text{Max}\{|S| : S \text{ is independent set of } G\}$  is the independence number of  $G$ .

**Odd hole:** A cycle of odd length at least 5 is called an odd hole.

**Odd anti-hole** is the complement graph of an odd hole.

**Directed graph:** A directed graph  $D$  is an ordered pair  $(V(D), A(D))$  consisting of a set  $V(D)$  of vertices and a set  $A(D)$ , disjoint from  $V(D)$ , of arcs, together with an incidence function  $\Phi(D)$  that associates with each arc of  $D$  an ordered pair of (not necessarily distinct) vertices of  $D$ .

The set  $N(i) = \{j \in V(D) | (i, j) \text{ is an arc}\}$  is the set of all *outneighbours* of vertex  $i$ .

**Directed acyclic graph:** A directed graph with no directed cycles.

**MAIS(G)** is the size of the maximum acyclic (without directed cycle) induced subgraph of  $G$ .

**Theorem 1.** ([5]) *A graph  $G$  is perfect if and only if its complement is perfect.*

**Definition 2.** *Let  $G$  be a directed graph on  $n$  vertices without self-loops. A  $0-1$  matrix  $A = (a_{ij})$  fits  $G$  if for all  $i$  and  $j$ , the following two conditions hold:*

- $a_{ii} = 1$
- $a_{ij} = 0$  whenever  $(i, j)$  is not an arc of  $G$ .

**Definition 3.**  $\text{minrk}_2(G) = \min\{\text{rk}_2(A) : A \text{ fits } G\}$ , where  $\text{rk}_2(A)$  denotes the maximum number of linearly independent rows of matrix  $A$  over the field  $GF(2)$ .

**Proposition 1.** ([3], [4]) *For any (undirected) graph  $G$ ,  $\omega(\overline{G}) \leq \text{minrk}_2(G) \leq \chi(\overline{G})$ .*

### 3 Linear Codes

In this section we show how the length of a linear INDEX code with side information graph  $G$  is measured by  $\text{minrk}_2(G)$ . Specifically, we construct a linear

INDEX code for  $G$  whose length equals  $\text{minrk}_2(G)$  and show that  $\text{minrk}_2(G)$  is optimal bound for all linear INDEX codes for  $G$ .

**Theorem 2.** *For any side information graph  $G$ , there exists a linear INDEX code for  $G$  whose length equals  $\text{minrk}_2(G)$ . This bound is optimal for all linear INDEX codes for  $G$ .*

*Proof.* Let  $A$  be a matrix such that  $\text{minrk}_2(G) = \text{rk}_2(A)$ . Assume without loss of generality that the first  $k$  rows  $A_1, \dots, A_k$  are linearly independent. The sender encodes input  $x \in \{0, 1\}^n$  to  $k$  bits  $b_j = A_j \cdot x$  for  $1 \leq j \leq k$ .

For decoding, fix a receiver  $R_i$  for some  $i \in [n]$  and let  $A_i = \sum_{j=1}^k \lambda_j A_j$  for some choice of  $\lambda_j$ 's. The receiver can compute  $A_i \cdot x = \sum_{j=1}^k \lambda_j b_j$  using the  $k$ -bit encoding of  $x$ . Observe that the vector  $c_i = A_i - e_i$ , where  $e_i$  is the  $i$ th standard basis vector, has non-zero entries only in coordinates that are among the *outneighbours* of vertex  $i$  in  $G$ . This means that receiver  $R_i$  can compute  $c_i \cdot x$  using the side information and recover  $x_i$  via  $(A_i \cdot x) - (c_i \cdot x) = e_i \cdot x = x_i$ .

Next step is proving that this bound is optimal. Suppose  $C$  is an arbitrary linear INDEX code for  $G$  that encodes input  $x$  by the taking its inner product with each vector in  $S = \{u_1, u_2, \dots, u_k\}$ . We want to show that  $\text{minrk}_2(G) \leq k$ , but at first we need to prove following claim.

**Claim 1.** *For every  $i$ ,  $e_i$  belongs to the span of  $S \cup \{e_j : j \in N(i)\}$ .*

Before we prove the claim, we show how this claim yields the proof of Theorem 2. The claim shows that  $\forall i \in [n]; e_i = \sum_{j=1}^k \lambda_j u_j + \sum_{j \in N(i)} \mu_j e_j$ , for some choice of  $\lambda$  and  $\mu$ . Let  $A$  be the  $n \times n$  matrix that its rows are  $A_i := \sum_{j=1}^k \lambda_j u_j = e_i - \sum_{j \in N(i)} \mu_j e_j$ . Observe that  $A_i$  has value 1 in its  $i$ th coordinate and value 0 in coordinates outside  $N(i) \cup \{i\}$ . Hence, the matrix  $A$  fits  $G$  and it follows that  $\text{minrk}_2(G) \leq \text{rk}_2(A)$ . Also,  $A_i$  belongs to the span of  $S$  for every  $i$ . Therefore, the matrix  $A$  has rank at most  $k$ . We conclude that  $\text{minrk}_2(G) \leq k$ .

Now we prove the claim. Suppose by contrary that there is an  $i$  such that  $e_i$  is not in the subspace  $W$  spanned by the vectors in  $S \cup \{e_j : j \in N(i)\}$ . Recall that  $W^{\perp\perp} = W$ , where  $W^\perp = \{v : v \cdot w = 0 \text{ for all } w \in W\}$ . Therefore, the assumption that  $e_i \notin W$  implies that there is a vector  $x \in W^\perp$  such that  $x \cdot e_i \neq 0$ . On the other hand, an inner product of the vector  $x$  with every vector in  $S \cup \{e_j : j \in N(i)\}$  is equal to 0 because  $x \in W^\perp$ . It follows that the encoding of  $x$  is equal to  $0^k$  because  $x \cdot u_j = 0$  for  $1 \leq j \leq k$  and  $x_j = 0$  for every  $j \in N(i)$  because  $x \cdot e_j = 0$  for every  $j \in N(i)$ . The input  $0^n$  also has these two properties. This violates the correctness of the encoding because two different inputs with same side information, namely  $x$  and  $0^n$ , are encoded into one codeword.  $\square$

We make the encoding and decoding procedure of the above theorem more clear by applying it to an example.

**Example 2.**

Suppose the side information graph  $G$  is a directed cycle of length 5, Figure 1 (Recall Example 1).

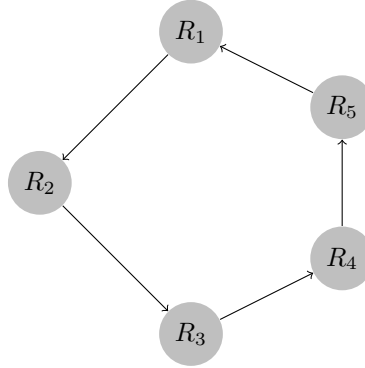


Figure 1: Directed cycle of length 5.

Observe that the matrix  $A$  fits  $G$ .

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

Now, we will apply elementary row operations to the matrix  $A$  to obtain  $rk_2(A)$ .

$$R_5 + R_1 \longrightarrow R_5 \implies A_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$R_5 + R_2 \longrightarrow R_5 \implies A_2 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

$$R_5 + R_3 \longrightarrow R_5 \implies A_3 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$R_5 + R_4 \longrightarrow R_5 \implies A_4 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$A_4$  has an upper triangular matrix  $A_{sub}$  as submatrix. Hence, the rank of  $A_4$  is 4. Let  $X = (x_1, x_2, \dots, x_5)$  be the input value of the sender. The sender transmits  $A_{sub}X^T = (x_1 \oplus x_2, x_2 \oplus x_3, x_3 \oplus x_4, x_4 \oplus x_5)$  to the receivers. It was shown in Example 1 that each receiver can obtain the requested bit.

## 4 Lower Bound for Restricted Graphs

In this section we investigate the length of any INDEX code when the side information graph is either a DAG, a perfect graph, an odd hole, or an odd anti-hole.

**Theorem 3.** ([1]) *The length of an INDEX code for graph  $G$  is at least  $MAIS(G)$ .*

### 4.1 Directed Acyclic Graphs

Nex preposition shows that the length of any INDEX code for  $G$  is at least  $minrk_2(G)$  when  $G$  is a directed acyclic graphs. We use DAG abbreviation for Directed Acyclic Graph.

**Proposition 2.** *Let  $G$  be a DAG on  $n$  nodes. Then, the length of any INDEX code for  $G$  is at least  $minrk_2(G)$ .*

*Proof.* Let  $C$  be any INDEX code for  $G$ . Since  $G$  is a DAG, then it is acyclic graph. It follows that  $MAIS(G) = n$ . Hence, by Theorem 3,  $len(C) \geq n$  and clearly  $n \geq minrk_2(G)$  that completes the proof.  $\square$

### 4.2 Perfect Graph

**Proposition 3.** *Let  $G$  be any perfect graph on  $n$  nodes. Then, the length of any INDEX code for  $G$  is at least  $minrk_2(G)$ .*

*Proof.* Let  $C$  be any INDEX code for  $G$ . By Theorem 3,  $len(C) \geq MAIS(G)$ . Since  $G$  is an undirected graph, then every edge in  $G$  is a directed cycle. Therefore, it follows that  $MAIS(G) = \alpha(G)$ . Clearly,  $\alpha(G) = \omega(\overline{G})$ , implying that  $len(C) \geq \omega(\overline{G})$ . Since  $G$  is perfect, the by Theorem 1 also  $\overline{G}$  is perfect and therefore  $\omega(\overline{G}) = \chi(\overline{G})$ . Hence  $len(C) \geq \chi(\overline{G})$ . By Proposition 1 we conclude  $len(C) \geq minrk_2(G)$ .  $\square$

### 4.3 Odd Holes

At first we characterize the minrank for odd holes. Specifically, we show that  $minrk_2(G) = n + 1$  where  $G$  is an odd hole and then prove that the length of any INDEX code for  $G$  is at least  $n + 1$ .

**Theorem 4.** *Let  $G$  be an odd hole of length  $2n+1$  ( $n \geq 2$ ). Then,  $\text{minrk}_2(G) = n + 1$ .*

*Proof.* It suffices to prove that  $\text{minrk}_2(G) \geq n + 1$  due to the facts that  $\chi(\overline{G}) = n + 1$  for an odd hole and  $\text{minrk}_2(G) \leq \chi(\overline{G})$  (proposition 1).

Suppose that the matrix  $A$  fits  $G$ . For convenience, we number the rows and the columns of  $A$  as  $0, 1, \dots, 2n$  and make all the index arithmetic below modulo  $2n$ . Let  $A_0, \dots, A_{2n}$  be the  $2n + 1$  rows of  $A$ . Observe that  $A$  has the following three properties for every  $i$  due to the assumption that  $A$  fits odd hole  $G$ . Following by notation  $A[i]$  we mean the  $i$ th coordinate of vector  $A$ .

1.  $A_i[i] = 1$ .
2.  $A_i[i - 1], A_i[i + 1] \in \{0, 1\}$ .
3.  $A_i[j] = 0$  for  $j \notin \{i - 1, i, i + 1\}$ .

For a row  $A_i$ , we call the rows  $A_0, \dots, A_{i-1}$  the "predecessors of  $A_i$ ". Note that  $A_0$  has no predecessors. We need the following two claim to conclude the proof.

**Claim 2.** *([1]) For  $i = 1, \dots, 2n - 2$ , either  $A_i$  is linearly independent of its predecessors or  $A_{i+1}$  is linearly independent of its predecessors.*

**Claim 3.** *([1]) At least one among  $A_1, A_{2n-1}, A_{2n}$  is linearly independent of its predecessors.*

Now, we use the above two claims to count the number of rows that are linearly independent. For each  $i$ , let  $Z_i = 1$  if the  $i$ th row of  $A$  is linearly independent of its predecessors and  $Z_i = 0$  otherwise. Note that the number of linearly independent rows of the matrix  $A$  is  $\sum_{i=0}^{2n} Z_i$ . The following three fact about the sequence  $Z_0, \dots, Z_{2n}$  are observable:

- 1)  $Z_0 = 1$ , because  $A_0$  simply does not have any predecessors.
- 2) For each  $i = 1, \dots, 2n - 3$ ,  $Z_i + Z_{i+1} \geq 1$  (Claim 2).
- 3)  $Z_1 + Z_{2n-1} + Z_{2n} \geq 1$  (Claim 3).

In order to obtain a lower bound for the sum  $\sum_{i=0}^{2n} Z_i$ , we write as follows:

$$2 \sum_{i=0}^{2n} Z_i = 2Z_0 + Z_1 + \sum_{i=0}^{2n-2} (Z_i + Z_{i+1}) + Z_{2n-1} + 2Z_{2n} \geq 2 + 2n - 2 + 1 + Z_{2n}.$$

Therefore,  $\sum_{i=0}^{2n} Z_i \geq (2n + 1)/2$ . However,  $\sum_{i=0}^{2n} Z_i$  is an integer, so  $\sum_{i=0}^{2n} Z_i \geq \lceil (2n + 1)/2 \rceil = n + 1$ . We showed that for every  $A$  that fits  $G$ ,  $\text{rk}_2(A) = \sum_{i=0}^{2n} Z_i \geq n + 1$ . Hence, it follows that  $\text{minrk}_2(G) \geq n + 1$ .  $\square$

In order to prove that the length of any INDEX code for an odd hole is at least  $\text{minrk}_2(G)$ , we need to study some properties of the *confusion graph* associated with INDEX coding.

**Definition 4.** The confusion graph  $C(G)$  associated with INDEX coding problem represented by a side information graph  $G$  is an (undirected) graph on  $\{0, 1\}^n$  such that  $x$  and  $x'$  are connected by an edge if for some  $i$ , we have  $x[N(i)] = x'[N(i)]$  but  $x_i \neq x'_i$ .

**Notation:** Let  $1_S$  denote the characteristic vector of a set  $S \subseteq [n]$ . In other words, the  $i$ th coordinate of vector  $1_S$  is 1 if  $i \in S$  and otherwise is 0.

**Lemma 1.** Let  $G$  be an undirected graph on  $n$  nodes and let  $C(G)$  be the confusion graph corresponding to INDEX coding for  $G$ . Then,

- 1) If  $S$  is a vertex cover of  $G$ , then any two inputs  $x, x' \in \{0, 1\}^n$  that agree on  $S$  (i.e.,  $x[S] = x'[S]$ ) are connected by an edge in  $C(G)$ .
- 2) If  $S$  is an independent set in  $G$ , then the set  $X_S = \{1_T | T \subseteq S\}$  forms a clique in  $C(G)$ .
- 3) If  $S, T$  are two disjoint and independent sets in  $G$ , and there exists some  $i \in S$  that has no neighbors in  $T$  or some  $j \in T$  that has no neighbors in  $S$ , then the inputs  $1_S$  and  $1_T$  are connected by an edge in  $C(G)$ .

Now we are able to prove the following theorem.

**Theorem 5.** Let  $G$  be an odd hole on  $2n + 1$  nodes ( $n \geq 2$ ). Then, the length of any INDEX code for  $G$  is at least  $\text{minrk}_2(G) = n + 1$ .

*Proof.* Let  $G$  be an odd hole and  $C$  be any INDEX code for  $G$ . If we prove that the number of codewords in  $C$  is greater than  $2^n$ , then we can conclude that  $\text{len}(C) \geq n + 1 = \text{minrk}_2(G)$ .

The three sets  $S_1 = \{1, 3, \dots, 2n - 1\}$ ,  $S_2 = \{2, 4, \dots, 2n\}$  and  $S_3 = \{2n + 1\}$  are independent in  $G$ . Hence, by Part 2 of Lemma 1,  $X_{S_i}$  is a clique in  $C(G)$  for each  $i \in \{1, 2, 3\}$ . It means that the receiver  $R_i$  has the same side information for all the elements of  $X_i$  and these elements differ in  $i$ th bit, so they should be encoded into different codewords. It follows that  $2^{|S_i|}$  codewords is needed to encode elements of  $X_{S_i}$ . This immediately results in  $|C| \geq 2^n$ .

Assume that  $|C| = 2^n$ . Observe that  $X_{S_1}, X_{S_2}, X_{S_3}$  have only all-zero vector  $((0, 0, \dots, 0))$  as a common input and are otherwise pairwise disjoint. since the number of codewords are  $2^n$  and every elements of  $X_{S_i}$  ( $i = 1, 2, 3$ ) is encoded to different codewords, then there must be at least three non-zero elements  $x_1 \in X_{S_1}$ ,  $x_2 \in X_{S_2}$  and  $x_3 \in X_{S_3}$  that are encoded into the same codeword. Therefore, there are three non-empty subset  $T_1 \subseteq S_1$ ,  $T_2 \subseteq S_2$  and  $T_3 \subseteq S_3$  such that  $x_1 = 1_{T_1}$ ,  $x_2 = 1_{T_2}$  and  $x_3 = 1_{T_3}$ . Consider any  $i \in T_1$ . By Part 3 of Lemma 1,  $i$  must have neighbor  $j \in T_2$  because  $T_1$  and  $T_2$  are independent sets and disjoint and also  $x_1$  and  $x_2$  are not connect by an edge in the confusion graph. Similarly, both  $i$  and  $j$  must have neighbors in  $T_3$ . The equality  $T_3 = \{2n + 1\}$  implies that  $i$  and  $j$  both are neighbors of  $2n + 1$ . This immediately results in the triangle  $(i, j, 2n + 1)$  in  $G$  that is a contradiction to the fact that all odd holes are free-triangle. Thus, we conclude  $|C| > 2^n$ .  $\square$



## 4.4 Odd Anti-Holes

**Theorem 6.** *Let  $G$  be an odd anti-hole on  $2n + 1$  nodes ( $n \geq 2$ ). Then, the length of any INDEX code for  $G$  is at least  $\text{minrk}_2(G) = 3$ .*

*Proof.* Let  $G$  be an odd anti-hole and  $C$  be any INDEX code for  $G$ . It suffices to prove  $|C| \geq 5$ .

Observe that sets  $S_1 = \{1\}$ ,  $S_2 = \{2, 3\}$ ,  $\dots$ ,  $S_{n+1} = \{2n, 2n + 1\}$  are independent and pairwise disjoint. Hence, the sets  $X_{S_1}, X_{S_2}, \dots, X_{S_{n+1}}$  have only the all-zero vector as a common inputs and are otherwise disjoint and also by Part 2 of Lemma 1 each of them forms a clique in  $C(G)$ . It follows that  $|X_{S_i}|$  codewords are needed to encode inputs of  $X_{S_i}$ , hence  $|C|$  should be greater than 4. Assume that  $|C| = 4$ . Therefore, there must be a single codeword that encodes a non-zero input of  $X_{S_i}$ , for every  $i = 1, \dots, n + 1$ . Suppose these inputs are  $x_1, \dots, x_{n+1}$ . Hence, there are non-empty sets  $T_1 \subseteq S_1, \dots, T_{n+1} \subseteq S_{n+1}$  such that  $1_{T_i} = x_i$ , for every  $1 \leq i \leq n + 1$ . The set  $\{x_1, x_2, \dots, x_{n+1}\}$  is an independent set in the confusion graph  $C(G)$ .

We will show that  $T_i = \{2i - 1\}$  for every  $i = 1, \dots, n + 1$ . Observe that  $T_1 = \{1\}$  because  $T_1 \subseteq S_1$  and  $1_{T_1} = x_1 \neq 0$ . Suppose by induction that  $T_i = \{2i - 1\}$ . We will show correctness for  $i + 1$ . Since  $x_i$  and  $x_{i+1}$  are not connected by an edge in the confusion graph  $C(G)$  and  $T_i$  and  $T_{i+1}$  are disjoint and independent sets in  $G$ , then by Part 3 of Lemma 1, every node in  $T_i$  must have a neighbor in  $T_{i+1}$  and vice versa.  $G$  is an anti hole and  $T_i = \{2i - 1\}$ , hence  $2i + 1$  is the only neighbor of  $2i - 1$  in the set  $S_{i+1}$ . It is concluded that  $T_{i+1} = \{2i + 1\}$ . Since  $x_1$  and  $x_{n+1}$  are not connected by an edge in the confusion graph  $C(G)$  and  $T_1$  and  $T_{n+1}$  are disjoint and independent sets in  $G$ , then by Part 3 of Lemma 1, the nodes 1 and  $2n + 1$  should be neighbors in graph  $G$  and this is a contradiction by the fact that  $G$  is an anti hole.  $\square$

## 5 Conclusions

In this paper we study how tools from the graph theory can be used in obtaining a lower bound for the length of INDEX codes. The idea is involving graph theory by considering the side information as a directed graph and identify a measure on graphs, the *minrank*, which characterizes the length of INDEX code for some classes of graphs.

## References

- [1] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, *Index coding with side information*, Proc. of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pp. 197-206.
- [2] Y. Birk, T. Kol, *Informed-Source Coding-on-Demand (ISCOD) over Broadcast Channels*, INFOCOM 1998: 1257-1264.

- [3] J.A. Bondy and U. S. R. Murty, *Graph Theory, Graduate texts in mathematics, Springer, New York, 2007.*
- [4] W. H. Haemers, *On some problems of Lovász concerning the Shannon capacity of a graph, IEEE Trans. Inf. Theory, vol. 25, pp. 231-232, 1979.*
- [5] W. H. Haemers, *An upper bound for the Shannon capacity of a graph, Algebr. Methods in Graph Theory, vol. 25, pp. 267-272, 1978.*
- [6] L. Lovász, *A characterization of perfect graphs, J. Combinator. Theory, Ser. B, vol. 13, pp. 95-98, 1972.*