Index Coding with Side Information

Ziv Bar-Yossef[*] Yitzhak Birk[†] T. S. Jayram[‡] Tomer Kol[§]

Abstract

Motivated by a problem of transmitting data over broadcast channels (Birk and Kol, INFOCOM 1998), we study the following coding problem: a sender communicates with n receivers $R_1, \ldots, R_n$. He holds an input $x \in \{0,1\}^n$ and wishes to broadcast a single message so that each receiver R_i can recover the bit x_i. Each R_i has prior side information *about x, induced by a directed graph G on n nodes; R_i knows the bits of x in the positions $\{j \mid (i,j)$ is an edge of $G\}$. We call encoding schemes that achieve this goal* INDEX codes for $\{0,1\}^n$ with side information graph G.

In this paper we identify a measure on graphs, the minrank*, which we conjecture to exactly characterize the minimum length of* INDEX *codes. We resolve the conjecture for certain natural classes of graphs. For arbitrary graphs, we show that the minrank bound is tight for both linear codes and certain classes of non-linear codes. For the general problem, we obtain a (weaker) lower bound that the length of an* INDEX *code for any graph G is at least the size of the maximum acyclic induced subgraph of G.*

1. Introduction

Source coding is one of the central areas of coding and information theory. Shannon's famous source coding theorem states that 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 many distributed applications, though, the receiver may have some prior *side information* about x, before it is sent. Source coding with side information addresses encoding schemes that exploit the side information in order to reduce the length of the code. Classical results in this area [16, 19, 18] describe how to achieve optimal rates with respect to the joint entropy of the source and the side information.

Witsenhausen [17] initiated the study of the zero-error side information problem. For every source input $x \in \mathcal{X}$, the receiver gets an input $y \in \mathcal{Y}$ that gives some information about x. This is captured by restricting the pairs (x, y) to belong to a fixed set $\mathcal{L} \subseteq \mathcal{X} \times \mathcal{Y}$. Both the sender and the receiver know $\mathcal{L}$, and thus each of them, given his own input, has information about the other's input. Witsenhausen showed that fixed-length side information codes were equivalent to colorings of a related object called the *confusion graph*, and thus the logarithm of the chromatic number of this graph tightly characterizes the minimum number of bits needed to encode the source. Further results by Alon and Orlitsky [2] and Koulgi *et al.* [12] showed that graph-theoretic information measures could be used to characterize both the average length of variable-length codes, as well as asymptotic rates of codes that simultaneously encode multiple inputs drawn from the same source.

In this paper, we study a new variant of source coding with side information, first proposed by Birk and Kol [6] in the context of a server that disseminates a set of data blocks (e.g., the daily newspaper) over a broadcast channel (e.g., satellite or coaxial cable) to a set of caching clients. At the end of the main transmission, each client possesses some subset of the transmitted blocks, be it due to intermittent reception, "interest filters" or limited storage capacity. Also, any given client is only interested in some subset of the blocks, and requests retransmission of those blocks that it needs but does not possess. There is no communication among clients, but a (slow) "backward" channel can be used by a client to send requests and metadata to the server. Each client *requests* a subset of the data blocks, and advises the server of the data blocks already available in its cache. Assuming large blocks and in view of the fact that the amount

[*] Department of Electrical Engineering, Technion, Haifa 32000, Israel. Email: `zivby@ee.technion.ac.il`. Supported by the European Commission Marie Curie International Re-integration Grant.

[†] Department of Electrical Engineering, Technion, Haifa 32000, Israel. Email: `birk@ee.technion.ac.il`.

[‡] IBM Almaden Research Center, 650 Harry Road, San Jose 95120, CA, USA. Email: `jayram@almaden.ibm.com`.

[§] Department of Electrical Engineering, Technion, Haifa 32000, Israel. Email: `tomer@tx.technion.ac.il`.

of metadata per block is independent of block size, the challenge is to minimize the amount of supplemental information that must be broadcast by the server in order to enable every client to derive all its *requested* blocks.

Birk and Kol [6] suggested the idea of *coding on demand by an informed source (ISCOD)*. With ISCOD, the server uses its knowledge of the cache contents and *requested* blocks of each client along with a systematic erasure correcting code (e.g., Reed-Solomon) to derive a set of supplemental data blocks that would jointly enable every client to derive its requested blocks. The supplemental blocks are then transmitted. Each client uses a subset of the received supplemental blocks along with some of its cached blocks to derive its requested block(s). Instance-specific upper bounds on the amount of data that must be transmitted are presented, along with some heuristic algorithms. The bounds are nevertheless shown not to be tight. No lower bounds are presented. Finally, [6] presents a two-way protocol for exchanging control information between the server and the clients.

A client may request multiple blocks. With a broadcast channel, however, this is equivalent to multiple single-request clients, each with the same cache content as the original one, and is so represented. In [6], it is pointed out that when a given block is requested by multiple clients, the main communication savings is through only transmitting it once. Both [6] and the current paper only address the case of unique requests.

The above scenario is formalized as a source coding with side information problem as follows (cf. [6]). There is a sender who has an input x from a source alphabet $\mathcal{X}$ (in this paper we confine ourselves to the alphabet $\mathcal{X} = \{0,1\}^n$). There are n receivers $R_1, \dots, R_n$, where for each i, R_i is interested in the bit x_i. The side information is characterized by a simple directed graph G (no self loops or parallel edges) on $\{1, 2, \dots, n\}$. For a subset $S \subseteq [n]$, $x[S]$ denotes the projection of x on the coordinates in S. The side information of R_i equals $x[N(i)]$ where $N(i) \triangleq \{j \in V \mid (i,j) \text{ is an edge}\}$ denotes the set of out-neighbors of i in the graph G.

Example 1. Let $R_1, R_2, \dots, R_n$ be the n receivers (clients) over a broadcast channel whose source alphabet is $\mathcal{X} = \{0,1\}^n$. For an input (data) $x \in \mathcal{X}$, each receiver R_i is interested in the value x_i (requested block) but knows x_{i-1} as side information (cached block). (Abusing notation slightly, receiver R_1 knows x_n.) The side information graph is thus a directed cycle of length n. Since x_{i-1} is "independent" of x_i, it may not be clear at first how the sender (server) can take advantage of the side information of the receivers to shorten the broadcast. However, there is a strategy in which the sender can save one bit: rather than send all the bits of x, the sender broadcasts the $n-1$ parities $x_1 \oplus x_2, x_2 \oplus x_3, \dots, x_{n-1} \oplus x_n$. Now, each receiver R_i for $i > 1$ can recover x_i by taking the parity of $x_{i-1} \oplus x_i$ with x_{i-1}. The receiver R_1 on the other hand just xors the $n-1$ parities broadcast by the sender together with x_n to recover x_1.

Definition 2 (INDEX codes)**.** A deterministic INDEX code $\mathcal{C}$ for $\{0,1\}^n$ with side information graph G on n nodes, abbreviated as "INDEX code for G", is a set of codewords in $\{0,1\}^\ell$ together with:

1. An encoding function E mapping inputs in $\{0,1\}^n$ to codewords, and
2. 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.

The graph G is known in advance to the sender and the receivers; thus the encoding and decoding functions typically depend on G. The *length* of $\mathcal{C}$, denoted by $\operatorname{len}(\mathcal{C})$, is defined to be ℓ.

The above problem can also be cast in an equivalent setting with a single receiver: The receiver is given an index i and the side information $x[N(i)]$ as inputs and wants to recover the value x_i. (The equivalence follows from the fact the sender does not know the index i given to the receiver, and thus has to use an encoding that allows recovering x_i, for *any* i.) Using this equivalent form, we can contrast our side information problem with Witsenhausen's zero-error side information problem. A first notable difference is that while in Witsenhausen's setting the *entire* input x has to be recovered, in our setting only a single bit x_i is needed. This allows significant savings in the encoding length, as the following example demonstrates: Suppose the side information graph is a perfect matching on n nodes. Since the receiver has only a single bit of side information, then $n-1$ bits are necessary to recover the entire input. On the other hand, if only a single bit is needed, then the sender can encode his input by the $n/2$ parities of pairs of matched bits. A second difference from Witsenhausen's setting is that the type of side information addressed in our problem is restricted to side information graphs. This natural restriction emanates from the broadcast application mentioned above and also imposes more structure that enables us to obtain an interesting combinatorial characterization of the minimum length of INDEX codes in terms of the side information graphs.

We also consider in this paper randomized INDEX codes, in which the encoding and decoding functions are allowed to be randomized and are even allowed to use a common public random string. Decoding needs to succeed only with *high probability*, taken over the random choices made by the encoding and decoding functions.

Our contributions. In this paper we identify a graph functional, called *minrank*, which we show to characterize the minimum length of INDEX codes, for natural types of codes and for wide classes of side information graphs. Let G be a directed graph on n vertices without self-loops. We say that a 0-1 matrix $A = (a_{ij})$ *fits* G if for all i and j: (i) $a_{ii} = 1$, and (ii) $a_{ij} = 0$ whenever (i, j) is *not* an edge of G. Thus, $A - I$ is the adjacency matrix of an *edge subgraph* of G, where I denotes the identity matrix. Let $\mathrm{rk}_2(\cdot)$ denote the 2-rank of a 0-1 matrix, namely, its rank over the field $GF(2)$.

Definition 3. $\mathrm{minrk}_2(G) \triangleq \min\{\mathrm{rk}_2(A) : A \text{ fits } G\}$

The above measure for *undirected* graphs was considered by Haemers [11] in the context of proving bounds for the Shannon capacity Θ of undirected graphs. For an undirected graph G whose adjacency matrix is M, the 2-rank of $M + I$ (which fits G) has also been studied in the algebraic graph theory community. For example, Brouwer and van Eijl [7] and Peeters [15] study this quantity for strongly regular and distance-regular graphs, respectively. It has been shown by Peeters [14] that computing $\mathrm{minrk}_2(G)$ is NP-hard. Finally, it is known that minrk_2 has the “sandwich property”, similar to other natural quantities such as the Lovász Theta function:

Proposition 4 ([10, 11])**.** *For any undirected graph G,* $\omega(\overline{G}) \leq \Theta(G) \leq \mathrm{minrk}_2(G) \leq \chi(\overline{G})$*. Moreover, each of these inequalities is strict.*

Our first result (see Section 3) shows that $\mathrm{minrk}_2(G)$ completely characterizes the minimum length of *linear* INDEX codes (i.e., ones whose encoding function is linear), for *arbitrary* directed side information graphs G:

Theorem 5. *The optimal length of a linear* INDEX *code for a side information graph G equals* $\mathrm{minrk}_2(G)$.

The upper bound in the above theorem strictly improves a previous result of Birk and Kol [6]. Birk and Kol showed a construction of a linear INDEX code, whose length is the “cover cost” of the side information graph (and showed that the construction is suboptimal). For undirected graphs, the cover cost is the same as the chromatic number of the complement graph. Since the minrank can be strictly smaller than this chromatic number, it immediately follows that this bound beats the Birk and Kol bound. The lower bound for linear codes is of interest, since linear codes are possibly the most natural type of codes. In fact, all the existing INDEX codes (with or without side information) we are aware of are linear.

Our second contribution is a lower bound which holds for general INDEX codes including deterministic and randomized INDEX codes. This result is presented in Section 4.

Theorem 6. *The length of any δ-error randomized* INDEX *code for G is at least* $\mathrm{MAIS}(G) \cdot (1 - H_2(\delta))$*, where* $\mathrm{MAIS}(G)$ *is the size of the maximum acyclic induced subgraph of G and* $H_2(\cdot)$ *is the binary entropy function.*

If G is undirected, then $\mathrm{MAIS}(G)$ equals the size of the largest independent set in G, i.e., $\omega(\overline{G})$. Given the gap between $\omega(\overline{G})$ and $\mathrm{minrk}_2(G)$ mentioned above, a natural question is whether $\mathrm{minrk}_2(G)$ characterizes the optimal length of general INDEX codes for general graphs G.

Conjecture 7. *The optimal length of a general* INDEX *code for G equals* $\mathrm{minrk}_2(G)$*, i.e. linear codes achieve the optimal length over all codes for G.*

In Section 5 we give supporting evidence for this conjecture by proving that $\mathrm{minrk}_2(G)$ is a lower bound on the minimum length of a wide class of non-linear codes. An INDEX code is called *linearly-decodable*, if all its n decoding functions are linear. A linearly-decodable code need not be linearly encodable. A simple argument shows that the length of a linearly-decodable INDEX code for any graph G is at least $\mathrm{minrk}_2(G)$. We relax the notion of linearly-decodable codes to “semi-linearly-decodable” codes. An INDEX code is *k-linearly-decodable*, if at least k of its decoding functions are linear. Note that n-linearly-decodable codes are simply linearly-decodable, while 0-linearly-decodable codes are unrestricted. We are able to prove the conjecture for k-linearly-decodable codes when $k \geq n - 2$:

Theorem 8. *For any graph G, and for any* $k \geq n - 2$*, the length of any k-linearly-decodable* INDEX *code for G is at least* $\mathrm{minrk}_2(G)$.

Our lower bound for general codes (Theorem 6) immediately gives tight bounds for directed acyclic graphs and undirected graphs G that satisfy $\omega(\overline{G}) = \mathrm{minrk}_2(G) = \chi(\overline{G})$. In particular, they hold for perfect graphs[1]. In Section 6, we are able to prove that minrank characterizes the minimum length of INDEX codes,

[1] Recall that an undirected graph G is called *perfect*, if for every induced subgraph G' of G, $\omega(\overline{G'}) = \chi(\overline{G'})$. Perfect graphs include a wide class of graphs such as trees, bipartite graphs, interval graphs, chordal graphs, etc.

even for non-perfect graphs, namely *odd holes* (undirected odd-length cycles of length at least 5) and *odd anti-holes* (complements of odd holes).

Theorem 9. *Let G be any graph, which is either a DAG, a perfect graph, an odd hole, or an odd anti-hole. Then, the length of any* INDEX *code for G is at least* $\mathrm{minrk}_2(G)$.

Finally, we consider the following natural direct sum-type problem: If a graph G has k connected components $G_1, \ldots, G_k$, then is the length of the best INDEX code for G equal to the sum of the lengths of the best codes for $G_1, \ldots, G_k$? The answer should intuitively be affirmative, but a direct proof seems to be elusive. In fact, using the techniques of Feder *et al.* [9], one can show a connection between the two, but incurring a loss of an additive term that depends linearly on k. After lower bounding the length of a code by its *information cost*, we are able to prove a tight direct sum theorem w.r.t. the information cost measure. We note that almost all our lower bounds hold not only for the length of INDEX codes but also for their information cost. This result is presented in Section 4.

Techniques. The many results presented in this paper required us to resort to a multitude of techniques from linear algebra, information theory, Fourier analysis, and combinatorics.

The lower bounds for linearly-encodable and linearly-decodable codes are based on dimension arguments from linear algebra. To extend the lower bound for linearly-decodable codes to semi-linearly-decodable codes, we used an intriguing "balance property" of Boolean functions: If all linear Boolean functions are "balanced" on some set U (i.e., get the same number of 0's and 1's on the set), then all Boolean functions (whether linear or not) are balanced on U. To prove this property, we use Fourier analysis to represent arbitrary Boolean functions as linear combinations of linear functions. We then introduce the notion of "conditional minrank" of a Boolean matrix and explore its properties using the balance property. This in turn allows us to extend the lower bound for linearly-decodable codes to $(n-2)$-linearly-decodable codes. Extension of the proof technique to hold for k-linearly-decodable codes, for $k < n-2$, would require better understanding of the conditional minrank measure.

The lower bound for general (randomized) codes and the direct sum theorem are proved via information theory arguments. We extend previous arguments from [5, 4] to obtain a direct sum theorem for the *information cost* of codes.

Finally, our lower bounds for odd holes and odd anti-holes are purely combinatorial. We employ a connection between vertex covers of a graph G and the structure of the confusion graph corresponding to the INDEX coding for G. We note that dealing with odd holes, and with the pentagon in particular, turned out to be very challenging, because the standard technique of lower bounding the chromatic number of the corresponding confusion graph via its independence number does not work.

Related work. There are settings other than source coding in which INDEX codes have been addressed. Ambainis *et al.* [3] considered the so called "random access codes"[2], which are identical to randomized INDEX codes without side information. Their main thrust was proving tight bounds on the length of the codes in the quantum setting, where inputs can be encoded by qubits rather than classical bits; their result applied to the classical setting is a special case of our Theorem 6 for the case when G is the empty graph.

The problem of INDEX coding with side information can also be cast as a *one-way communication complexity* problem of the *indexing* function [13] (from which the term INDEX codes was coined) with the additional twist of side information. Alice (the sender) is given an input x, sends a single message to Bob. Bob is given an index i and the side information $x[N(i)]$, and wants to learn x_i. Another formulation of INDEX coding is in terms of *network coding* [20, 1]. As such, it represents a restricted case of a single source, a single encoder and a single channel, but with the important addition of a special flavor of side information. Parts of this information are known to different decoders, and the encoder is fully aware of this knowledge.

Notation. Throughout the paper, we use the following notations. Let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Let e_i denote the i-th standard basis vector. The dimensions of these vectors are understood from the context. For a subset $S \subseteq [n]$, we denote by $x[S]$ the projection of a vector x on the coordinates in S.

2. Sandwich property of minrank

We start with an observation relating minrank to other well-known graph measures.

[2] We chose the term INDEX codes to avoid confusion since the term "random access" denotes a different concept in the information theory community.

Proposition 4 (restated) *For any undirected graph G, $\omega(\overline{G}) \leq \Theta(G) \leq \mathrm{minrk}_2(G) \leq \chi(\overline{G})$. Moreover, each of these inequalities is strict.*

Proof. Fix an optimal coloring of $\overline{G}$. Define the 0-1 matrix A by $A_{ij} = 1$ if i and j get the same color in $\overline{G}$, and 0, otherwise. The matrix A fits G, and $\mathrm{rk}_2(A) = \chi(\overline{G})$. Hence, $\mathrm{minrk}_2(G) \leq \mathrm{rk}_2(A) = \chi(\overline{G})$.

Recall that the Shannon capacity $\Theta(G)$ of a graph G is defined as $\lim_{k\to\infty} \alpha(G^k)^{1/k}$. Here G^k denotes the (strong) k-th power of G, where there is an edge between distinct $(u_1, u_2, \ldots, u_k)$ and $(v_1, v_2, \ldots, v_k)$ if and only if for all j either $u_j = v_j$ or u_j is connected to v_j in G. It can be verified that G^k has an independent set of size $\alpha(G)^k$, so $\Theta(G) \geq \alpha(G) = \omega(\overline{G})$.

Suppose A fits G such that $\mathrm{rk}_2(A) = \mathrm{minrk}_2(G)$. It can be verified that the k-th matrix tensor power of A, denoted by $A^{\otimes k}$, fits G^k. Since $A^{\otimes k}$ has a square identity sub-matrix corresponding to a largest independent set in G^k, we have $\alpha(G^k) \leq \mathrm{rk}_2(A^{\otimes k})$. It is well known that $\mathrm{rk}_2(B^{\otimes k}) = \mathrm{rk}_2(B)^k$ for any matrix B, so $\mathrm{rk}_2(A^{\otimes k}) = \mathrm{rk}_2(A)^k = \mathrm{minrk}_2(G)^k$. Taking the k-th root on both sides and letting $k \to \infty$ proves the required bound.

From the results in [10], it is known that the family of Symplectic graphs G_n with parameter n satisfies $\mathrm{minrk}_2(G_n) = 2n+1$ whereas $\chi(\overline{G}_n) = 2^n + 1$, exhibiting a large gap between these two measures. In contrast, gap between $\mathrm{minrk}_2(G)$ and $\omega(\overline{G})$, to the best of our knowledge, is via odd cycles: for a cycle of length $2n+1$ its minrank equals $n+1$ whereas its independence number equals n. Lovász's classic paper which introduced the θ-function shows that the Shannon capacity of the 5-cycle equals $\sqrt{5}$, which is strictly smaller than its minrank. □

3. Linear codes

In this section we obtain a tight characterization of the length of linear INDEX codes for all side information graphs G.

Theorem 5 (restated) *The optimal length of a linear* INDEX *code for a side information graph G equals* $\mathrm{minrk}_2(G)$.

Proof. Let A be the matrix that fits G whose rank equals $\mathrm{minrk}_2(G) \triangleq k$. Assume without loss of generality that the span of the first k rows $A_1, \ldots, A_k$ equals the span of all the rows of A. The encoding function is simply the k bits $b_j \triangleq A_j \cdot x$ for $1 \leq j \leq k$.

Decoding proceeds as follows. 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 λ_j's. The receiver first computes $A_i \cdot x = \sum_{j=1}^k \lambda_j b_j$ using the k-bit encoding of x. Now, consider the vector $c_i = A_i - e_i$, where e_i is the i-th standard basis vector. Observe that the only non-zero entries in c_i correspond to coordinates which are among the neighbors of i in G. This means that the receiver can compute $c_i \cdot x$ using the side information. Receiver R_i can now recover x_i via $(A_i \cdot x) - (c_i \cdot x) = e_i \cdot x = x_i$.

For the lower bound, suppose $\mathcal{C}$ is an arbitrary linear INDEX code for G defined by the set $S = \{u_1, u_2, \ldots, u_k\}$, i.e. x is encoded by the taking its inner product with each vector in S.

Claim 10. *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 to finish the proof of the lower bound. Fix an $i \in [n]$; the claim shows that $e_i = \sum_{j=1}^k \lambda_j u_j + \sum_{j \in N(i)} \mu_j e_j$, for some choice of λ and μ. Rearranging, we have $\sum_j \lambda_j u_j = e_i - \sum_{j\in N(i)} \mu_j e_j \triangleq A_i$. It follows that A_i has value 0 in coordinates outside $N(i)$ and that A_i belongs to the span of S. Therefore, the matrix A whose rows are given by $A_1, A_2, \ldots, A_n$ fits G and has rank at most k. We conclude that $k \geq \mathrm{rk}_2(A) \geq \mathrm{minrk}_2(G)$.

It remains to prove the claim. Fix an i and suppose to the contrary that e_i is *not* in the subspace W spanned by the vectors in $S \cup \{e_j : j \in N(i)\}$. Recall that the *dual* of W, denoted by $W^\perp$ denotes the set of vectors orthogonal to every vector in W, i.e., $W^\perp = \{v : v \cdot w = 0 \text{ for all } w \in W\}$. It is well-known that $W^{\perp\perp} = W$. Therefore, the assumption $e_i \notin W$ implies that there is a vector $x \in W^\perp$ such that $x \cdot e_i \overset{(*)}{\neq} 0$. On the other hand, since $x \in W^\perp$, we have that x is orthogonal to every vector in $S \cup \cup\{e_j : j \in N(i)\}$. It follows that (i) the encoding for x equals 0^k, and (ii) the side information x_j available to receiver R_i equals 0 for all $j \in N(i)$. This violates the correctness of the encoding because the input 0^n also satisfies (i) and (ii), yet Equation (*) shows that it differs from x in coordinate i. □

4. General codes

In this section, we prove lower bounds for the class of general randomized INDEX codes. The main technical statement is a direct-sum result for the *information cost* of a randomized INDEX code. See [8] for the basic information theory notions and facts used in this section.

4.1. Direct sum for information cost

Definition 11 (Information Cost)**.** Let $\mathcal{C}$ be a randomized index code for G. Let R denote the public random string of $\mathcal{C}$, and let $E(x, R)$ denote the encoding of x in $\mathcal{C}$.[3] Let X be uniformly distributed in $\{0,1\}^n$. The *information cost* of $\mathcal{C}$, denoted by $\mathrm{icost}(\mathcal{C})$, equals $I(X; E(X) \mid R) = H(X \mid R) - H(X \mid E(X), R)$.

It can be seen that the information cost of deterministic INDEX codes is just the entropy of the codewords, and thus is closely related to the length of the code.

Theorem 12. *Let $G_1, G_2, \ldots, G_k$ be vertex-induced subgraphs of a directed graph G such that:*

1. *The vertex sets of $G_1, G_2, \ldots, G_k$ are pairwise disjoint.*
2. *For any $i < j$ and vertices $v_i \in V(G_i)$ and $v_j \in V(G_j)$, there is no directed edge in G from v_i to v_j.*

Let $\mathcal{C}$ be a δ-error randomized INDEX *code for G. Then, there exist δ-error randomized* INDEX *codes $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ for $G_1, G_2, \ldots, G_k$ such that* $\mathrm{icost}(\mathcal{C}) \geq \sum_i \mathrm{icost}(\mathcal{C}_i)$.

Proof. Let $E(x, R)$ be encoding function of the INDEX code $\mathcal{C}$. Let X be uniformly distributed on $\{0,1\}^n$ and let E denote the random variable $E(X, R)$. By definition, $\mathrm{icost}(\mathcal{C}) = I(X\,;E \mid R)$. Define $U_0 = V \backslash \bigcup_{i=1}^{k} V_i$, and let $U_j = U_0 \cup (\bigcup_{i=1}^{j} V_i)$ for $j = 1 \ldots k$. By the chain rule for conditional entropy,

$$I(X\,;E \mid R)$$
$$= I(X[U_0]\,;E \mid R) + \sum_{j=1}^{k} I(X[V_j]\,;E \mid X[U_{j-1}], R)$$
$$\geq \sum_{j=1}^{k} I(X[V_j]\,;E \mid X[U_{j-1}], R)$$

For each j we will show that the expression $I(X[V_j]\,;E(X,R) \mid X[U_{j-1}], R)$ within the above sum is the information cost of an INDEX code $\mathcal{C}_j$ for G_j. The proof of this is based on a reduction lemma proven in [5].

Define an INDEX code $\mathcal{C}_j$ for G_j using the code $\mathcal{C}$ as follows. Let $x^j \in \{0,1\}^{|V_j|}$ denote the source input. Loosely speaking, x^j will be mapped to the vertices in G_j, and the inputs corresponding to the vertices in the other graphs will be generated using a combination of private and public random strings. Formally, let Y have the same distribution as $X[U_{j-1}]$, and let Z have the same distribution as $X[V \backslash U_j]$. The public random string for $\mathcal{C}_j$ consists of (R, Y) while Z will be part of the private randomness of the encoder. The encoding of x^j in $\mathcal{C}_j$ is defined by mapping x^j to V_j, Y to U_{j-1}, and Z to $V \backslash U_j$, and then use E to encode the combined input.

Let $i \in V_j$ be any coordinate and consider what is needed to recover the bit corresponding to coordinate i. By the property of G_j, it can be seen that the neighbors of i in G are either among the neighbors of i in V_j or belong to U_{j-1}. Now, the values for the former are part of the side information for coordinate i while the values for the latter can be found in the public random string Y. This means that the receiver in the INDEX coding problem for G_j can apply the decoding function D_i to recover the i-th coordinate. The routine calculations similar to [5] can be used to show that the error of this code is at most δ, and that the information cost equals $I(X[V_j]\,;E(X,R) \mid X[U_{j-1}], R)$, completing the proof of the theorem. □

4.2. Lower bound for randomized codes

Theorem 6 can now be shown as a simple application of the above Theorem 12.

Theorem 6 (restated) *The length of any δ-error randomized* INDEX *code for G is at least* $\mathrm{MAIS}(G) \cdot (1 - H_2(\delta))$*, where* $\mathrm{MAIS}(G)$ *is the size of the maximum acyclic induced subgraph of G and $H_2(\cdot)$ is the binary entropy function.*

Proof sketch. Let G' be a maximal acyclic induced subgraph of G. Let $u_1, u_2, \ldots, u_k$ denote the vertices of G' such that there is no edge from u_i to u_j whenever $i < j$. Apply Theorem 12 where G_j is a graph with a single vertex u_j. We have $\mathrm{icost}(\mathcal{C}) \geq \sum_j \mathrm{icost}(\mathcal{C}_j)$. Now, since $\mathcal{C}_j$ is a INDEX code for a single vertex graph, therefore, it encodes just a single bit that can be decoded with probability of error at most δ. By the classical Fano's inequality in information theory, it must have at least $1 - H_2(\delta)$ bits of information. □

5. On the tightness of the minrank bound

In this section, we provide supporting evidence for our conjecture that $\mathrm{minrk}_2(G)$ is a lower bound on the minimum length of *non-linear* INDEX codes for arbitrary graphs G.

[3] E also depends on the sender's private randomness which is being suppressed for ease of presentation.

Let $\mathcal{C}$ be an INDEX code for G. Let $D_1, \ldots, D_n$ be the n decoding functions of $\mathcal{C}$. Fix a codeword $c \in \mathcal{C}$, and for each index $i \in [n]$, we denote by D_i^c the function induced by fixing c as input to D_i: $D_i^c(x[N(i)]) = D_i(c, x[N(i)])$. Although D_i^c is applied only to the side information bits $x[N(i)]$, it will be convenient for us to view it as acting on the whole input x with the restriction that it depends only on the set of coordinates $N(i)$. Thus, from now on, $D_i^c : \{0,1\}^n \to \{0,1\}$. Note that $D_i^c(x) = x_i$ for every x whose encoding $E(x)$ equals c.

Proposition 13. *If $|\{x \mid D_i^c(x) = x_i \ \forall i\}| \leq M$ for every codeword $c \in \mathcal{C}$, then $\operatorname{len}(\mathcal{C}) \geq \lceil n - \log M \rceil$.*

5.1. Semi-linearly-decodable codes

An INDEX code $\mathcal{C}$ is said to be *k-linearly-decodable*, if for every codeword $c \in \mathcal{C}$, at least k of the decoding functions $D_1^c, \ldots, D_n^c$ are linear. Note that the smaller k is, the less restricted is the class of k-linearly-decodable codes. When $k = n$, these codes are simply called *linearly-decodable*, while 0-linearly-decodable are unrestricted codes. Our upper bound (Theorem 5) is a linearly-decodable INDEX code (and thus also k-linearly-decodable, for any k).

Our goal is to obtain lower bounds on the length of k-linearly-decodable codes for a value of k as small as possible.

Theorem 14. *Let c be a codeword in a k-linearly decodable code $\mathcal{C}$ with side information graph G, where $k \geq n-2$. Then, $|\{x \mid D_i^c(x) = x_i \ \forall i\}| \leq 2^{n-\operatorname{minrk}_2(G)}$.*

Proposition 13 immediately implies that the length of $\mathcal{C}$ is at least $\operatorname{minrk}_2(G)$, proving Theorem 8.

The rest of this section is devoted to the proof of Theorem 14. Fix a graph G. We say that a function f *fits* an index i if $f(x) = g(x) + x_i$ for some function g that depends only on $N(i)$. To simplify the notation, let e_i denote i-th standard basis vector and write $f = g + e_i$ so that $f(x) = g(x) + e_i \cdot x = g(x) + x_i$.

Fix a k-linearly-decodable code $\mathcal{C}$ for G and a codeword $c \in \mathcal{C}$. Let $D_1^c, \ldots, D_n^c$ be the n decoding functions associated with c. Note that each function $D_i^c + e_i$ fits index i, for all i. If the decoding function D_i^c is linear for some fixed i, then $D_i^c(x) = d \cdot x$ for some vector d. It can be seen that the value of d in every coordinate outside of $N(i)$ equals 0. Thus, we can also say that the vector $d + e_i$ fits[4] index i.

To motivate the proof of Theorem 14, consider the following simple argument for linearly-decodable codes i.e. $k = n$. Let A be the $n \times n$ Boolean matrix, whose rows are $d_1 + e_1, \ldots, d_n + e_n$. Since $d_i + e_i$ fits index i, it follows that A fits G, so $\operatorname{rk}_2(A) \geq \operatorname{minrk}_2(G)$. Next, observe that $D^c(x) = x_i$ if and only if $(d_i + e_i) \cdot x = 0$. Therefore, $\{x \mid D^c(x) = x_i \ \forall i\}$ is a linear subspace denoting the *kernel* of the matrix A. By standard linear algebra, its dimension is at most $n - \operatorname{rk}_2(A) \leq n - \operatorname{minrk}_2(G)$, and therefore the size of W is at most $2^{n-\operatorname{minrk}_2(G)}$.

To deal with the case $k < n$, we would like to generalize the above argument. For the rest of this section, we define the following notions: (i) a set S of k indices such that the decoding functions $D_i^c(x) = d_i \cdot x$ are linear for each $i \in S$ (ii) the subspace $W_S = \bigcap_{i \in S} \{x \mid (d_i + e_i) \cdot x = 0\}$ (iii) the function $f_j = D_j^c + e_j$ for each $j \in [n] \setminus S$.

The key idea of our proof is to view the function f_j in the ±1 world and consider its Fourier representation. Since $D_j^c(x)$ depends only on $N(j)$, the *characters* that have non-zero weight in the Fourier representation will be shown to be associated with vectors that fit index j.

We now come to an important notion that will be used in the proof. Let T be a subset of the indices in $[n] \setminus S$. Let $\mathcal{H} = \langle h_j : j \in T \rangle$ be a collection of $|T|$ vectors, not necessarily distinct, such that the vector h_j associated with $j \in T$ fits index j. Extending the definition, we say that $\mathcal{H}$ *fits* T. Define $q_{\mathcal{H}}(T) = \dim(W_S \cap \{x \mid h_j \cdot x = 0 \ \forall j \in T\}$ and let $q(T)$ denote the maximum value of $q_{\mathcal{H}}(T)$ over all collections $\mathcal{H}$ that fit T.

Proposition 15. *1. $q(\emptyset) = \dim(W_S)$.*

2. *For every $j \in T$, $q(T \setminus \{j\}) \in \{q(T), q(T)+1\}$.*

3. *More generally, $q(T) \leq q(T') \leq q(T) + |T| - |T'|$ for any $T' \subseteq T$.*

4. *$q(T) \leq \dim(W_S) \leq q(T) + |T|$.*

5. *If $q(T) = \dim(W_S) - |T|$, then $q(T') = \dim(W_S) - |T'|$ for every $T' \subseteq T$.*

6. *Suppose $q(\{j\}) = \dim(W_S)$ for every $j \in T$. Then $q(T) = dim(W_S)$ as well.*

7. *Let $T = [n] \setminus S$, then $q(T) \leq n - \operatorname{minrk}_2(G)$*

Proof. Part 1 follows simply by definition. Part 2 follows from the standard linear algebra fact adding a single constraint to any subspace can only decrease its dimension, but by at most 1; an inductive argument yields Part 3. Setting $T' = \emptyset$ in Part 3 and then using Part 1 yields Part 4.

For Part 5, note that the Part 4 implies that $\dim(W_S) - |T'| \leq q(T')$. By Part 3, $q(T') \leq q(T) +$

[4] This is consistent with our earlier notation, namely, a matrix A fits G if and only if the i-th row of A fits index i, for all i.

$|T| - |T'| = \dim(W_S) - |T'|$, using the premise of Part 5. Therefore, $q(T') = \dim(W) - |T'$ as well.

For Part 6, the premise says that there exist vectors h_j for all $j \in T$ such that $h_j \cdot x = 0$ for all $x \in W_S$. Define the collection $\mathcal{H} = \langle h_j : j \in T\rangle$. It can be seen that $q_{\mathcal{H}}(T) = \dim(W)$ which is the maximum value that $q(T)$ can attain by Part 4.

Finally, for Part 7, let $\mathcal{H} = \langle h_j : j \in T\rangle$ be the collection of vectors such that $q_{\mathcal{H}}(T) = q(T)$. Recall that $q_{\mathcal{H}}(T)$ is the dimension of the subspace $V = W \cap \{x \mid h_j \cdot x = 0\ \forall j \in T\}$. The vectors in $\mathcal{H}$ fit T, so let A be the matrix whose rows consist of the vectors in $\mathcal{H}$ together with the decoding vectors associated with the indices in S. It follows that A fits G. Since its kernel equals V, we conclude:

$$q_{\mathcal{H}}(T) = \dim(V) = n - \mathrm{rk}_2(A) \leq n - \mathrm{minrk}_2(G) \quad \square$$

The following lemma is the main technical result that will be used to prove Theorem 14.

Lemma 16. *Let S, W_S and the functions f_j for $j \notin S$ be as defined above. For any $T \subseteq [n] \setminus S$, $|T| \leq 2$,*

$$|W_S \cap \{x \mid f_j(x) = 0\ \forall j \in T\}| \leq 2^{q(T)}.$$

Applying the above lemma with $T = [n]\backslash S$, and then using $q(T) \leq n - \mathrm{minrk}_2(G)$ (Proposition 15, Part 7) immediately yields Theorem 14. Unfortunately, we currently do not know how to prove the lemma (or some suitable weaker version of it) for $|T| > 2$.

We first prove a stronger version of Lemma 16 for the special case when $q(T)$ has the smallest possible value $\dim(W_S) - |T|$ (Proposition 15, Part 4), but the size of T can be arbitrary. In this case, the bound given by Lemma 16 is always achieved with equality.

Lemma 17. *Using the notation of Lemma 16, if $q(T) = \dim(W_S) - |T|$, then*

$$|W_S \cap \{x \mid f_j(x) = 0\ \forall j \in T\}| = 2^{\dim(W_S)-|T|}.$$

Proof Sketch. The case when T is empty follows by the definition of dimension. For a non-empty T, we write

$$\begin{aligned}&|W_S \cap \{x \mid f_j(x) = 0\ \forall j \in T\}| \\ &= \sum_{x\in W_S} \prod_{j\in T} (1 - f_j(x))\end{aligned}$$

We will show that the latter expression equals $2^{\dim(W_S)-|T|}$ using Fourier analysis.

For each j, consider the function $\tilde{f}_j(x) = 1 - 2f_j(x)$ which is just a mapping $0 \mapsto 1$ and $1 \mapsto -1$ of the value $f_j(x)$. The Fourier transform of $\tilde{f}_j(x)$ is a linear combination of the characters $(-1)^{h\cdot x}$. The crucial property that can be shown is the following: since f_j is a function that fits index j, then the characters with non-zero coefficients in Fourier transform correspond to those vectors h that fit index j. Thus, $\tilde{f}_j(x) = \sum_{h \text{ fits } j} c_h(-1)^{h\cdot x}$, for some choice of c_h's. Using simple algebra, the expression

$$\prod_{j\in T}(1 - f_j(x)) = \prod_{j\in T}\left(\frac{1 + \tilde{f}_j(x)}{2}\right)$$

within the above sum can be rewritten as sum of 2 expressions

1. $\frac{|W_S|}{2^{|T|}} = 2^{\dim(W_S)-|T|}$ and
2. a weighted sum of terms of the form $(-1)^{(\sum_{j\in T'} h_j)\cdot x}$ over all choices of $\emptyset \neq T' \subseteq T$ and collections $\mathcal{H} = \langle h_j : j \in T'\rangle$ of vectors that fit T'.

For any fixed $\mathcal{H}$ that fits T', it suffices to show that the corresponding term $(-1)^{(\sum_{j\in T'} h_j)\cdot x}$ summed over all $x \in W_S$ equals 0.

Since $q(T) = \dim(W_S) - |T|$, Proposition 15, Part 5 implies that $q(T') = \dim(W_S) - |T'|$. It follows that the vectors $\{h_j \mid j \in T'\}$ are jointly independent of the subspace W_S, therefore their sum $\sum_{j\in T'} h_j$ does not belong to W_S. This means that $(\sum_{j\in T'} h_j) \cdot x$ is *balanced* on W_S: for half the vectors in W_S it will evaluate to 0 and for the other half it will evaluate to 1. We conclude

$$\sum_{x\in W_S} (-1)^{\sum_{j\in T'}(\sum_j h_j)\cdot x} = 0,$$

finishing the proof of the lemma. $\square$

We can now prove Lemma 16:

Proof of Lemma 16. We prove the lemma by induction on the size of T. The case $|T| = 0$, meaning $T = \emptyset$, follows simply from the fact that $q(\emptyset) = \dim(W_S)$ (Proposition 15, Part 1) and then applying Lemma 17. Assume that the statement of the lemma holds for all T such that $|T| \leq t$. We will prove it for $|T| = t + 1$, conditioned on $t + 1 \leq 2$.

For $i \in T$, let $T_{-i} = T \setminus \{i\}$. By Proposition 15, Part 2, for every $j \in T$, $q(T_{-j}) \in \{q(T), q(T) + 1\}$. We split our analysis into two cases.

Case 1: For some $i \in T$ $q(T_{-i}) = q(T)$. In this case

$$\begin{aligned}&|W_S \cap \{x \mid f_j(x) = 0\ \forall j \in T\}| \\ &\leq |W_S \cap \{x \mid f_j(x) = 0\ \forall j \in T_{-i}\}| \\ &\leq 2^{n-q(T_{-i})} = 2^{n-q(T)}\end{aligned}$$

where the second inequality follows from the induction hypothesis and the last equality follows from our assumption in Case 1.

Case 2: For all $i \in T$, $q(T_{-i}) = q(T) + 1$. This is the case we know how to handle only for $|T| = 1, 2$. Suppose, first, that $|T| = 1$. Then, by the assumption of this case, $q(\emptyset) = q(T) + 1$. Since $q(\emptyset) = \dim(W_S)$ (Proposition 15, Part 1) we obtain $q(T) = \dim(W_S) - 1$. Hence, the statement follows in this case from Lemma 17.

Consider now the case $|T| = 2$ and let $T = \{i, j\}$. By the assumption of this case, $q(\{i\}) = q(\{j\}) = q(\{i, j\}) + 1$. By Proposition 15, Part 2, either both $q(\{i\})$ and $q(\{j\})$ equal $q(\emptyset) = \dim(W_S)$ or both are 1 less than $\dim(W_S)$. The first case is impossible because by Proposition 15, Part 5, $q(\{i, j\}) = \dim(W_S)$ as well violating the assumption of this case. Therefore, $q(\{i\}) = q(\{j\}) = \dim(W_S) - 1$ implying that $q(\{i, j\}) = \dim(W_S) - 2$. Hence, the statement follows in this case once again from Lemma 17. □

6. Lower bounds for restricted graphs

In this section we show that for certain natural classes of graphs, the minrank bound is tight w.r.t. *arbitrary* INDEX codes.

Theorem 9 (restated) *Let G be any graph, which is either a DAG, a perfect graph, an odd hole, or an odd anti-hole. Then, the length of any* INDEX *code for G is at least* $\mathrm{minrk}_2(G)$.

The proofs for DAGs and perfect graphs are simple and deferred to the full version of the paper. Below, we prove the theorem for odd holes; the case of anti-holes uses similar notions and ideas and is once again deferred to the full version. In order to prove the bound for odd holes, we need to study some combinatorial properties of the *confusion graph* associated with INDEX coding.

Definition 18 (Confusion graph)**.** The *confusion graph* $C(G)$ associated with INDEX coding for a directed graph G (abbreviated "confusion graph for G") is an *undirected* graph on $\{0, 1\}^n$ such that x and x' are adjacent if for some i, we have $x[N(i)] = x'[N(i)]$ but $x_i \neq x'_i$.

If x and x' are adjacent in $C(G)$, then no INDEX code $\mathcal{C}$ for G can map x and x' to the same codeword, implying $\log \chi(C(G))$ is a lower bound on $\mathrm{len}(\mathcal{C})$.

Notation. Let $\mathbf{0}$ and $\mathbf{1}$ denote, respectively, the all-zero and the all-one vectors. Let $\mathbf{1}_S$ denotes the characteristic vector of a set $S \subseteq [n]$.

Lemma 19. *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 adjacent in $C(G)$.*

2. *If S is an independent set in G, then the set $X_S = \{\mathbf{1}_T \mid T \subseteq S\}$ 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 $\mathbf{1}_S$ and $\mathbf{1}_T$ are adjacent in $C(G)$.*

The proof of the lemma is deferred to the full version.

Let G be an odd hole on $2n + 1$ nodes ($n \geq 2$). Let $\mathcal{C}$ be any INDEX code for G. We will prove that the number of codewords in $\mathcal{C}$ is at least 2^n, implying that $\mathrm{len}(\mathcal{C}) \geq n + 1 = \mathrm{minrk}_2(G)$ (as noted in Section 2).

Consider the following coloring of G: $S_1 = \{1, 3, \ldots, 2n - 1\}$, $S_2 = \{2, 4, \ldots, 2n\}$ and $S_3 = \{2n + 1\}$. For each $i \in \{1, 2, 3\}$, since S_i is an independent set, then by Part 2 of Lemma 19, $\mathcal{C}$ must use $2^{|S_i|}$ different codewords to encode inputs in X_{S_i}. Since $|S_1| = |S_2| = n$, this already implies $|\mathcal{C}| \geq 2^n$. Assume, to the contradiction, that $|\mathcal{C}| = 2^n$.

Since S_1, S_2, S_3 are pairwise disjoint, then the sets $X_{S_1}, X_{S_2}, X_{S_3}$ have only $\mathbf{0}$ as a common input and are otherwise pairwise disjoint. Since $|\mathcal{C}| = 2^n$, and no codeword can encode two different inputs in X_{S_i} ($i = 1, 2, 3$), then there must be at least one codeword encoding a nonzero input from X_{S_1}, a nonzero input from X_{S_2}, and a nonzero input from X_{S_3}. We call these inputs x_1, x_2, x_3.

We view x_1, x_2, x_3 as characteristic vectors of sets $T_1, T_2, T_3 \subseteq [n]$. Since $x_1, x_2, x_3 \neq \mathbf{0}$, then $T_1, T_2, T_3 \neq \emptyset$. Furthermore, they are all independent and pairwise disjoint. Since the only nonzero vector in X_{S_3} is e_{2n+1}, $T_3 = \{2n + 1\}$.

Since x_1, x_2, x_3 are encoded by the same codeword, no two of them can be connected by an edge in the confusion graph. Consider any $i \in T_1$. By Part 3 of Lemma 19, i must have a neighbor $j \in T_2$. Similarly, both i and j must have neighbors in T_3. Since $T_3 = \{2n + 1\}$, both are neighbors of $2n + 1$. We conclude that $(i, j, 2n + 1)$ forms a triangle in G. However, all odd holes are triangle-free. This is a contradiction, and thus $|\mathcal{C}| > 2^n$.

The above theorem provides a tight lower bound on the *length* of INDEX codes for odd holes, but not on their *size*. Our upper bound (Theorem 5) gives a code whose size is 2^{n+1}, while the above proof only shows a lower bound of $|\mathcal{C}| > 2^n$. Optimal code size lower bounds are important for deriving lower bounds on the average encoding length and on the information cost. In the full version of this paper, we give tight lower bounds (i.e., 2^{n+1}) on the size of INDEX codes for odd holes; the proof for $n \geq 7$ involves a more involved combinatorial argument while proof for the pentagon is by brute force computer simulations.

7. Conclusions

In this paper, we explored upper and lower bounds on the length of INDEX codes for $\{0,1\}^n$ with side information graph G. We identified a measure on graphs, the *minrank*, which we showed to characterize the length of INDEX codes for natural classes of graphs (DAGs, perfect graphs, odd holes, and odd anti-holes). We also proved that minrank characterizes the minimum length of natural types of INDEX codes (linear, linearly-decodable, and semi-linearly-decodable) for *arbitrary* graphs. For general codes and general graphs, we were able to obtain a weaker bound in terms of the maximum acyclic induced subgraph. Finally, we proved a direct sum theorem for the information cost of INDEX codes with side information.

The general question, i.e., whether minrank is a lower bound on the length of *any* INDEX code for *any* graph, remains open. Perhaps one could relax the conjecture and consider fields other than $GF(2)$.

The minrank by itself is an interesting subject of study. We know that for undirected graphs, it is bounded from below by the Shannon capacity and from above by the chromatic number of the graph complement. It would be interesting to explore further properties of minrank with respect to other graph measures such as the Lovász Theta function.

References

[1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Trans. Inform. Theory*, 46:1204–1216, 2000.

[2] N. Alon and A. Orlitsky. Source coding and graph entropies. *IEEE Transactions on Information Theory*, 42(5):1329–1339, 1996.

[3] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *J. ACM*, 49(4):496–511, 2002.

[4] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. The sketching complexity of pattern matching. In *Proceedings of the 8th International Workshop on Randomization and Computation (RANDOM)*, pages 261–272, 2004.

[5] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Computer and System Sciences*, 68(4):702–732, 2004.

[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 Transactions on Information Theory*, 52(6):2825–2830, 2006. Earlier version appeared in INFOCOM '98.

[7] A. E. Brouwer and C. A. van Eijl. On the p-ranks of the adjacency matrices of strongly regular graphs. *Journal of Algebraic Combinatorics*, 1:329–346, 1992.

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[9] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 24(4):736–750, 1995.

[10] W. H. Haemers. An upper bound for the Shannon capacity of a graph. *Algebraic methods in Graph Theory*, 25:267–272, 1978.

[11] W. H. Haemers. On some problems of Lovász concerning the shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(2):231–232, 1979.

[12] P. Koulgi, E. Tuncel, S. L. Regunathan, and K. Rose. On zero-error source coding with decoder side information. *IEEE Transactions on Information Theory*, 49(1):99–111, 2003.

[13] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999.

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

[15] R. Peeters. On the p-ranks of the adjacency matrices of distance-regular graphs. *Journal of Algebraic Combinatorics*, 15(2):127–149, 2002.

[16] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[17] H. S. Witsenhausen. The zero-error side information problem and chromatic numbers. *IEEE Transactions on Information Theory*, 22(5):592–593, 1976.

[18] A. Wyner. A theorem on the entropy of certain binary sequences and applications II. *IEEE Transactions on Information Theory*, IT-19:772–777, 1973.

[19] A. Wyner and J. Ziv. A theorem on the entropy of certain binary sequences and applications I. *IEEE Transactions on Information Theory*, IT-19:769–771, 1973.

[20] R. W. Yeung and Z. Zhang. Distributed source coding for satellite communications. *IEEE Trans. Inform. Theory*, 45:1111–1120, 1999.