Index Coding via Linear Programming

Anna Blasiak* Robert Kleinberg† Eyal Lubetzky‡

Abstract

Index Coding has received considerable attention recently motivated in part by applications such as fast video-on-demand and efficient communication in wireless networks and in part by its connection to Network Coding. Optimal encoding schemes and efficient heuristics were studied in various settings, while also leading to new results for Network Coding such as improved gaps between linear and non-linear capacity as well as hardness of approximation. The basic setting of Index Coding encodes the side-information relation, the problem input, as an undirected graph and the fundamental parameter is the broadcast rate β, the average communication cost per bit for sufficiently long messages (i.e. the non-linear vector capacity). Recent nontrivial bounds on β were derived from the study of other Index Coding capacities (e.g. the scalar capacity β_1) by Bar-Yossef *et al* (2006), Lubetzky and Stav (2007) and Alon *et al* (2008). However, these indirect bounds shed little light on the behavior of β: there was no known polynomial-time algorithm for approximating β in a general network to within a nontrivial (i.e. $o(n)$) factor, and the exact value of β remained unknown for *any graph* where Index Coding is nontrivial.

Our main contribution is a direct information-theoretic analysis of the broadcast rate β using linear programs, in contrast to previous approaches that compared β with graph-theoretic parameters. This allows us to resolve the aforementioned two open questions. We provide a polynomial-time algorithm with a nontrivial approximation ratio for computing β in a general network along with a polynomial-time decision procedure for recognizing instances with $\beta = 2$. In addition, we pinpoint β precisely for various classes of graphs (e.g. for various Cayley graphs of cyclic groups) thereby simultaneously improving the previously known upper and lower bounds for these graphs. Via this approach we construct graphs where the difference between β and its trivial lower bound is linear in the number of vertices and ones where β is uniformly bounded while its upper bound derived from the naive encoding scheme is polynomially worse.

1 Introduction

In the Index Coding problem a server holds a set of messages that it wishes to broadcast over a noiseless channel to a set of receivers. Each receiver is interested in one of the messages and has side-information comprising some subset of the other messages. Given the side-information map as an input, the objective is to devise an optimal encoding scheme for the messages (e.g., one minimizing the broadcast length) that allows all the receivers to retrieve their required information.

This notion of source coding that optimizes the encoding scheme given the side-information map of the clients was introduced by Birk and Kol [6] and further developed by Bar-Yossef *et al.*

*Department of Computer Science, Cornell University, Ithaca NY 14853. E-mail: ablasiak@cs.cornell.edu.
Supported by an NDSEG Graduate Fellowship, an AT&T Labs Graduate Fellowship, and an NSF Graduate Fellowship.

†Department of Computer Science, Cornell University, Ithaca NY 14853. E-mail: rdk@cs.cornell.edu. Supported in part by NSF grant CCF-0729102, AFOSR grant FA9550-09-1-0100, a Microsoft Research New Faculty Fellowship, a Google Research Grant, and an Alfred P. Sloan Foundation Fellowship.

‡Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA. Email: eyal@microsoft.com.

in [5]. Motivating applications include satellite transmission of large files (e.g. video on demand), where a slow uplink may be used to inform the server of the side-information map, namely the identities of the files currently stored at each client due to past transmissions. The goal of the server is then to issue a shortest possible broadcast that allows every client to decode its target file while minimizing the overall latency. See [5, 6, 9] and the references therein for further applications of the model and an account of various heuristic/rigorous Index Coding protocols.

The basic setting of the problem (see [2]) is formalized as follows: the server holds n messages $x_1, \ldots, x_n \in \Sigma$ where $|\Sigma| > 1$, and there are m receivers $R_1, \ldots, R_m$. Receiver R_j is interested in one message, denoted by $x_{f(j)}$, and knows some subset $N(j)$ of the other messages. A solution of the problem must specify a finite alphabet Σ_P to be used by the server, and an encoding scheme $\mathcal{E} : \Sigma^n \to \Sigma_P$ such that, for any possible values of $x_1, \ldots, x_n$, every receiver R_j is able to decode the message $x_{f(j)}$ from the value of $\mathcal{E}(x_1, \ldots, x_n)$ together with that receiver's side-information. The minimum encoding length $\ell = \lceil \log_2 |\Sigma_P| \rceil$ for messages that are t bits long (i.e. $|\Sigma| = 2^t$) is denoted by $\beta_t(G)$, where G refers to the data specifying the communication requirements, i.e. the functions $f(j)$ and $N(j)$. As noted in [19], due to the overhead associated with relaying the side-information map to the server the main focus is on the case $t \gg 1$ and namely on the following *broadcast rate*.

$$\beta(G) \triangleq \lim_{t\to\infty} \frac{\beta_t(G)}{t} = \inf_t \frac{\beta_t(G)}{t} \tag{1.1}$$

(The limit exists by sub-additivity.) This is interpreted as the average asymptotic number of broadcast bits needed per bit of input, that is, the asymptotic broadcast rate for long messages. In Network Coding terms, β is the *vector* capacity whereas β_1 is a *scalar* capacity.

An important special case of the problem arises when there is exactly one receiver for each message, i.e. $m = n$ and $f(j) = j$ for all j. In this case, the side-information map $N(j)$ can equivalently be described in terms of the binary relation consisting of pairs (i, j) such that $x_j \in N(i)$. These pairs can be thought of as the edges of a directed graph on the vertex set $[n]$ or, in case the relation is symmetric, as the edges of an undirected graph. This special case of the problem (which we will hereafter identify by stating that G *is a graph*) corresponds to the original Index Coding problem introduced by Birk and Kol [6], and has been extensively studied due to its rich connections with graph theory and Ramsey theory. These connections stem from simple relations between broadcast rates and other graph-theoretic parameters. Letting $\alpha(G), \overline{\chi}(G)$ denote the independence and clique-cover numbers of G, respectively, one has

$$\alpha(G) \le \beta(G) \le \beta_1(G) \le \overline{\chi}(G)\,. \tag{1.2}$$

The first inequality above is due to an independent set being identified with a set of receivers with no mutual information, whereas the last one due to [5, 6] is obtained by broadcasting the bitwise XOR of the vertices per clique in the optimal clique-cover of G.

1.1 History of the problem

The framework of graph Index Coding and its scalar capacity β_1 were introduced in [6], where Reed-Solomon based protocols hinging on a greedy clique-cover (related to the bound $\beta_1 \le \overline{\chi}$) were proposed and empirically analyzed. In a breakthrough paper [5], Bar-Yossef *et al.* proposed a new class of linear index codes based on a matrix rank minimization problem. The solution to this problem, denoted by $\mathrm{minrk}_2(G)$, was shown to achieve the optimal linear scalar capacity over $GF(2)$ and in particular to be superior to the clique-cover method, i.e. $\beta_1 \le \mathrm{minrk}_2 \le \overline{\chi}$. The parameter

minrk_2 was extended to general fields in [19], where arguments from Ramsey Theory showed that for any $\varepsilon > 0$ there is a family of graphs on n vertices where $\beta_1 \leq n^\varepsilon$ while $\text{minrk}_2 \geq n^{1-\varepsilon}$ for any fixed $\varepsilon > 0$. The first proof of a separation $\beta < \beta_1$ for graphs was presented by Alon *et al.* in [2]; the proof introduces a new capacity parameter β^* such that $\beta \leq \beta^* \leq \beta_1$ and shows that the second inequality can be strict using a graph-theoretic characterization of β^*. In addition, the paper studied hypergraph Index Coding (i.e. the general broadcasting with side information problem, as defined above), for which several hard instances were constructed — ones where $\beta = 2$ while β^* is unbounded and others where $\beta^* < 3$ while β_1 is unbounded. The first proof of a separation $\alpha < \beta$ for graphs is presented in a companion paper [7]; the proof makes use of a new technique for bounding β from below using a linear program whose constraints express information inequalities. The paper then uses lexicographic products to amplify this separation, yielding a sequence of graphs in which the ratio β/α tends to infinity. The same technique of combining linear programs with lexicographic products also leads to an unbounded multiplicative separation between non-linear and vector-linear Index Coding in hypergraphs.

As is clear from the foregoing discussion, the prior work on Index Coding has been highly successful in bounding the broadcast rate above and below by various parameters (all of which are, unfortunately, NP-hard to compute) and in coming up with examples that exhibit separations between these parameters. However it has been less successful at providing general techniques that allow the determination (or even the approximation) of the broadcast rate β for large classes of problem instances. The following two facts starkly illustrate this limitation. First, the exact value of $\beta(G)$ remained unknown for *every* graph G except those for which trivial lower and upper bounds $\alpha(G), \overline{\chi}(G)$ coincide. Second, it was not known whether the broadcast rate β could be approximated by a polynomial-time algorithm whose approximation ratio improves the trivial factor n (achieved by simply broadcasting all n messages) by more than a constant factor.[1]

In this paper, we extend and apply the linear programming technique recently introduced in [7] to obtain a number of new results on Index Coding, including resolving both of the open questions stated in the preceding paragraph. The following two sections discuss our contributions, first to the general problem of broadcasting with side information, and then to the case when G is a graph.

1.2 New techniques for bounding and approximating the broadcast rate

The technical tool at the heart of our paper is a pair of linear programs whose values bound β above and below. The linear program that supplies the lower bound was introduced in [7] and discussed above; the one that supplies the upper bound is strikingly similar, and in fact the two linear programs fit into a hierarchy defined by progressively strengthening the constraint set (although the relevance of the middle levels of this hierarchy to Index Coding, if any, is unclear).

Theorem 1. *Let G be a broadcasting with side information problem, having n messages and m receivers. There is an explicit sequence of n information-theoretic linear programs, each one a relaxation of its successors, whose respective solutions $b_1 \leq b_2 \leq \ldots \leq b_n$ are such that:*

(i) The broadcast rate β satisfies $b_2 \leq \beta \leq b_n$, and both of the inequalities can be strict.

(ii) When G is a graph, the extreme LP solutions b_1 and b_n coincide with the independence number $\alpha(G)$ and the fractional clique-cover number $\overline{\chi}_f(G)$ respectively.

As a first application of this tool, we obtain the following pair of algorithmic results.

[1] When G is a graph, it is not hard to derive a polynomial-time $o(n)$-approximation from (1.2).

Capacities compared	Best previous bounds in graphs	New separation results	Appears in Section
$\beta - \alpha$	$\Theta\left(n^{0.56}\right)$	$\Theta(n)$	2.4
β vs. $\overline{\chi}_f$	$\beta \leq n^{o(1)}$ $\overline{\chi}_f \geq n^{1-o(1)}$	$\beta = 3$ $\overline{\chi}_f = \Omega(n^{1/4})$	4.1
$\beta_1 - \beta$	≈ 0.32	$\Theta(n)$	2.4
$\beta^* - \beta$	—	$\Theta(n)$	2.4

Table 1: New separation results for Index Coding capacities in n-vertex graphs

Theorem 2. *Let G be a broadcasting with side information problem, having n messages and m receivers. Then there is a polynomial time algorithm which computes a parameter $\tau = \tau(G)$ such that $1 \leq \frac{\tau(G)}{\beta(G)} \leq O\big(n\frac{\log\log n}{\log n}\big)$. There is also a polynomial time algorithm to decide whether $\beta(G) = 2$.*

In fact, the $O\big(n\frac{\log\log n}{\log n}\big)$ approximation holds in greater generality for the *weighted* case, where different messages may have different rates (in the motivating applications this can correspond e.g. to a server that holds files of varying size). The generalization is explained in Section 3.2.

1.3 Consequences for graphs

In Section 5 we demonstrate the use of Theorem 1 to derive the exact value of $\beta(G)$ for various families of graphs by analyzing the LP solution b_2. As mentioned above, the exact value of $\beta(G)$ was previously unknown for any graph except when the trivial lower and upper bounds — $\alpha(G)$ and $\overline{\chi}(G)$ — coincide, as happens for instance when G is a perfect graph. Using the stronger lower and upper bounds b_2 and b_n, we obtain the exact value of $\beta(G)$ for all cycles and cycle-complements: $\beta(C_n) = n/2$ and $\beta(\overline{C_n}) = n/\lfloor \frac{n}{2} \rfloor$. In particular this settles the Index Coding problem for the 5-cycle investigated in [2, 5, 7], closing the gap between $b_2(C_5) = 2.5$ and $\beta^*(C_5) = 5 - \log_2 5 \approx 2.68$. These results also provide simple constructions of networks with gaps between vector and scalar Network Coding capacities.

We also use Theorem 1 to prove separation between broadcast rates and other graph parameters. Our results, summarized in Table 1, improve upon several of the best previously known separations. Prior to this work there were no known graphs G where $\beta_1(G) - \beta(G) \geq 1$. (For the more general setting of broadcasting with side information, multiplicative gaps that were logarithmic in the number of messages were established in [2].) In fact, merely showing that the 5-cycle satisfies $2 \leq \beta < \beta_1 = 3$ required the involved analysis of an auxiliary capacity β^*, discussed earlier in Section 1.1. With the help of our linear programming bounds (Theorem 1) we supply in Section 2.4 a family of graphs on n vertices where $\beta_1 - \beta$ is linear in n, namely $\beta = n/2$ whereas $\beta_1 = (1 - \frac{1}{5}\log_2 5 - o(1))n \approx 0.54n$.

We turn now to the relation between $\beta(G)$ and $\overline{\chi}_f(G)$, the upper bound provided by our LP hierarchy. As mentioned earlier, Lubetzky and Stav [19] supplied, for every $\varepsilon > 0$, a family of graphs on n vertices satisfying $\beta(G) \leq \beta_1(G) < n^{\varepsilon}$ while $\overline{\chi}_f(G) > n^{1-\varepsilon}$, thus implying that $\overline{\chi}_f(G)$ is not bounded above by any polynomial function of $\beta(G)$. We strengthen this result by showing that $\overline{\chi}_f(G)$ is not bounded above by *any* function of $\beta(G)$. To do so, we use a class of projective Hadamard graphs due to Erdős and Rényi to prove the following theorem in Section 4.1.

Theorem 3. *There exists an explicit family of graphs G on n vertices such that $\beta(G) = 3$ whereas the Index Coding encoding schemes based on clique-covers cost at least $\overline{\chi}_f(G) = \Theta(n^{1/4})$ bits.*

Recall the natural heuristic approach to Index Coding: greedily cover the side-information graph G by $r \geq \overline{\chi}(G)$ cliques and send the XORs of messages per clique for an average communication cost of r. A similar protocol based on Reed-Solomon Erasure codes was proposed by [6] and was empirically shown to be effective on large random graphs. Theorem 3 thus presents a hard instance for this protocol, namely graphs where $\beta = O(1)$ whereas $\overline{\chi}(G)$ is polynomially large.

2 Linear programs bounding the broadcast rate

In this section we present linear programs that bound the broadcast rate β below and above, using an information-theoretic analysis. We demonstrate this technique by determining $\beta(C_5)$ precisely; later, in Section 5, we determine β precisely for various infinite families of graphs.

2.1 The LP hierarchy

Numerous results in Network Coding theory bound the Network Coding rate (e.g., [1, 10, 15, 16, 23]) by combining entropy inequalities of two types. The first is purely information-theoretic and holds for any set of random variables; the second is derived from the graph structure. An important example of the second type of inequality, that we refer to as "decoding", enforces the following: if a set of edges A cuts off a set of edges B from all the sources, then any information on edges in B is determined by information on edges in A. We translate this idea to the setting of Index Coding in order to develop stronger lower bounds for the broadcast rate.

Definition 2.1. Given a broadcasting with side information problem and subsets of messages A, B, we say that A *decodes* B (denoted $A \rightsquigarrow B$) if $A \subseteq B$ and for every message $x \in B \setminus A$ there is a receiver R_j who is interested in x and knows only messages in A (i.e. $x_{f(j)} = x$ and $N(j) \subseteq A$).

Remark 2.2. For graphs, $A \rightsquigarrow B$ if $A \subseteq B$ and for every $v \in B \setminus A$ all the neighbors of v are in A.

If we consider the Index Coding problem on G and a valid solution $\mathcal{E}$, then the relation $A \rightsquigarrow B$ implies $H(A, \mathcal{E}(x_1, \ldots, x_n)) \geq H(B, \mathcal{E}(x_1, \ldots, x_n))$, since for each message in $B \setminus A$ there is a receiver who must be able to determine the message from only the messages in A and the public channel $\mathcal{E}(x_1, \ldots, x_n)$. (Here and in what follows we denote by $H(X, Y)$ the joint entropy of the random variables X, Y.) Combining these decoding inequalities with purely information-theoretic inequalities, one can prove lower bounds on the entropy of the public channel, a process formalized by a linear program (that we denote by $\mathcal{B}_2$) whose solution b_2 constitutes a lower bound on β. (See [7, 25] for more on information-theoretic LPs.) Interestingly, $\mathcal{B}_2$ fits into a hierarchy of n increasing linear programs such that the last LP in the hierarchy gives an *upper* bound on β.

Definition 2.3. For a broadcasting with side information problem on a set V of n messages, the *β-bounding LP hierarchy* is the sequence of LPs, denoted by $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \ldots, \mathcal{B}_n$ with solutions $b_1, b_2, \ldots, b_n$, given by:

k-th level of the LP hierarchy for the broadcast rate		
minimize $X(\emptyset)$		
subject to:		
$X(V) \geq n$		(*initialize*)
$X(\emptyset) \geq 0$		(*non-negativity*)
$X(S) + \lvert T \setminus S\rvert \geq X(T)$	$\forall S \subseteq T \subseteq V$	(*slope*)
$X(T) \geq X(S)$	$\forall S \subseteq T \subseteq V$	(*monotonicity*)
$X(A) \geq X(B)$	$\forall A, B \subseteq V \,:\, A \rightsquigarrow B$	(*decode*)
$\sum_{T \subseteq R} (-1)^{\lvert R \setminus T\rvert} X(T \cup Z) \leq 0$	$\forall R \subseteq V \,:\, 2 \leq \lvert R\rvert \leq k$ $\forall Z \subseteq V \,:\, Z \cap R = \emptyset$	($\lvert R\rvert$-*th order submodularity*)

Remark 2.4. The above defined 2^{nd} *order submodularity* inequalities are equivalent to the classical submodularity inequalities whereby $X(S) + X(T) \geq X(S \cap T) + X(S \cup T)$ for all S, T.

Theorem 1 traps β in the solution sequence of the above-defined hierarchy and characterizes its extreme values for graphs. The proofs of these results appear in Section 2.2, and in what follows we first outline the arguments therein and the intuition behind them.

As mentioned above, the parameter b_2 is the entropy-based lower bound via Shannon inequalities that is commonly used in the Network Coding literature. To see that indeed $\beta \geq b_2$ we interpret a solution to the broadcasting problem as a feasible primal solution to $\mathcal{B}_2$ via the assignment $X(A) = H(A \cup \mathcal{E}(x_1, \ldots, x_n))$. The proof that $\alpha(G) = b_1(G)$ for graphs is similarly based on constructing a feasible primal solution to $\mathcal{B}_1$, this time via the assignment $X(A) = |A| + \max\{|I| :$ I is an independent set disjoint from $A\}$. (The existence of this primal solution justifies the inequality $b_1 \leq \alpha$; the reverse inequality is an easy consequence of the decoding, initialization, and slope constraints.)

To establish that $\beta(G) \leq b_n(G)$ when G is a graph we will show that $b_n(G) = \overline{\chi}_f(G)$, the fractional clique-cover number of G, while $\overline{\chi}_f(G)$ is an upper bound on β. For a general broadcasting network G we will follow the same approach via an analog of $\overline{\chi}_f$ for hypergraphs. It turns out that there are two natural generalizations of cliques and clique-covers in the context of broadcasting with side information.

Definition 2.5. A *weak hyperclique* of a broadcasting problem is a set of receivers $\mathcal{J}$ such that for every pair of distinct elements $R_i, R_j \in \mathcal{J}$, $f(i)$ belongs to $N(j)$. A *strong hyperclique* is a subset of messages $T \subseteq V$ such that for any receiver R_j that desires $x_{f(j)} \in T$ we have that $T \subseteq N(j) \cup \{f(j)\}$.

A *weak fractional hyperclique-cover* is a function that assigns a non-negative weight to each weak hyperclique, such that for every receiver R_j, the total weight assigned to weak hypercliques containing R_j is at least 1. A *strong fractional hyperclique-cover* is defined the same way, except that the weights are assigned to strong hypercliques and the coverage requirement is applied to messages rather than receivers. In both cases, the *size* of the hyperclique-cover is defined to be the sum of all weights.

Observe that if T is any set of messages and $\mathcal{J}$ is the set of all receivers desiring a message in T, then T is a strong hyperclique if and only if $\mathcal{J}$ is a weak hyperclique. However, it is not the case that every weak hyperclique can be obtained from a strong hyperclique T in this way.

Observe also that if $\mathcal{J}$ is a weak hyperclique and each of the messages $x_{f(j)}$ ($R_j \in \mathcal{J}$) is a single scalar value in some field, then broadcasting the sum of those values provides sufficient information for each $R_j \in \mathcal{J}$ to decode $x_{f(j)}$. This provides an indication (though not a proof) that β is bounded above by the weak fractional hyperclique cover number. The proof of Theorem 1(i) in fact identifies b_n as being equal to the *strong* fractional hyperclique-cover number, which is obviously greater than or equal to its weak counterpart. The role of the n^{th}-order submodularity constraints is that they force the function $F(S) \triangleq X(\overline{S}) - |\overline{S}|$ to be a *weighted coverage function*. Using this representation of F it is not hard to extract a fractional set cover of V, and the sets in this covering are shown to be strong hypercliques using the decoding constraints.

Finally, we will show that one can have $\beta > b_2$ using a construction based on the Vámos matroid following the approach used in [11] to separate the corresponding Network Coding parameters. As for showing that one can have $\beta < b_n$, we will in fact show that one can have $\beta < b_3 \leq b_n$.

We believe that the other parameters $b_3, \ldots, b_{n-1}$ have no relation to β, e.g. as noted above we show that there is a broadcasting instance for which $\beta < b_3$ and thus b_3 is not a lower bound on β.

2.2 Proof of Theorem 1

In this section we prove Theorem 1 via a series of claims. The main inequalities involving the broadcast rate β are shown in §2.2.1 whereas the constructions demonstrating that these inequalities can be strict appear in §2.2.2.

2.2.1 Bounding the broadcast rate via the LP hierarchy

We begin by familiarizing ourselves with the framework of the LP-hierarchy through proving the following straightforward claim regarding the LP-solution b_1 and the graph independence number.

Claim 2.6. *If G is a graph then the LP-solution b_1 satisfies $b_1(G) = \alpha(G)$.*

Proof. In order to show that $b_1(G) \geq \alpha(G)$, let I be an independent set of maximal size in G. Now, $V \setminus I \rightsquigarrow V$ implies that $X(V \setminus I) \geq X(V) \geq n$ is true for any feasible solution. Additionally, $X(V \setminus I) \leq X(\emptyset) + |V \setminus I|$. Combining these together, we get $X(\emptyset) \geq |V| - |V \setminus I| = |I| = \alpha(G)$. To prove $b_1(G) \leq \alpha(G)$ we present a feasible solution to the primal attaining the value $\alpha(G)$,

$$X(S) = |S| + \max\{|I| \,:\, I \text{ is an independent set disjoint from } S\}\,, \tag{2.1}$$

We verify that the solution is feasible by checking that it satisfies all the constraints of $\mathcal{B}_1$. The fact that $X(V) = n$ implies the initialization constraint is satisfied. To prove the slope constraint, for $S \subseteq T \subseteq V$ let I, J be maximum-cardinality independent sets disjoint from S, T respectively. Note that J itself is disjoint from S, implying $|J| \leq |I|$. Thus we have

$$X(T) = |T| + |J| = |S| + |T \setminus S| + |J| \leq |S| + |T \setminus S| + |I| = X(S) + |T \setminus S|.$$

Note also that $I \setminus T$ is an independent set disjoint from T, hence it satisfies $|I \setminus T| \leq |J|$. Thus

$$X(T) = |T| + |J| \geq |T| + |I \setminus T| = |T \cup I| \geq |S \cup I| = |S| + |I| = X(S),$$

which verifies monotonicity. Finally, to prove decoding let A, B be any vertex sets such that $A \rightsquigarrow B$. Consider $G \setminus A$, the induced subgraph of G on vertex set $V \setminus A$. Every vertex of $B \setminus A$ is isolated in $G \setminus A$, and consequently if I is a maximum-cardinality independent set disjoint from B, then $I \cup (B \setminus A)$ is an independent set in $G \setminus A$. Therefore,

$$X(A) \geq |A| + |I| + |B \setminus A| = |B| + |I| = X(B)\,. \qquad \blacksquare$$

We next turn to showing that b_2 is a lower bound on the broadcast rate.

Claim 2.7. *The LP-solution b_2 satisfies $b_2(G) \leq \beta(G)$.*

Proof. Let G be a broadcasting with side information problem with n messages V and m receivers. Consider the message $P = \mathcal{E}(x_1, \ldots, x_n)$ that we send on the public channel to achieve β. Denote by H the entropy function normalized so that $H(x_i) = 1$ for all i. This induces a function from the power set of $V \cup P$ to $\mathcal{R}$ where $H(S) = |S|$ for any subset of messages S and $H(P) = \beta$.

Now, let $X(S) = H(S, P)$ for $S \subseteq V$. We will show that X satisfies all the constraints of the LP $\mathcal{B}_2$, implying X it is a feasible solution $\mathcal{B}_2$.

First, $X(V) \geq n$ since $H(V, P) = H(V)$ and our normalization has $H(V) = n$. Non-negativity holds because $H(P) \geq 0$. The $X(\cdot)$ values satisfy monotonicity and submodularity because entropy does. Slope is implied by the fact that entropy is submodular (that is, $H(S, P) + H(T \setminus S) \geq H(T, P)$) together with our normalization. Finally, decoding is satisfied because the coding solution is valid: each receiver R_j can determine its sought information from $N(j)$ and the public channel.

This solution gives $X(\emptyset) = H(P) = \beta$ and since the LP is stated as a minimization problem it implies that β is an upper bound on its solution b_2. $\blacksquare$

Next we prove that $\beta \leq b_n$. For every instance G of the broadcasting with side information problem, define $\overline{\chi}_f(G)$ to be the minimum size of a strong fractional hyperclique-cover; this parameter specializes to the fractional clique-cover number when G is a graph. To prove $\beta \leq b_n$ we first show that $\beta \leq \overline{\chi}_f$, and then that $\overline{\chi}_f = b_n$.

Claim 2.8. *For any broadcasting problem with side information, G, we have $\beta(G) \leq \overline{\chi}_f(G)$.*

Proof. Let $\mathcal{C}$ be the set of strong hypercliques in $G = (V, E)$. If $\overline{\chi}_f \leq w$ then there is a finite collection of ordered pairs $\{(S, x_S) : S \in \mathcal{C}\}$ where the x_S's are positive rational numbers satisfying

$$\sum_{S \in \mathcal{C}} x_S = w\,, \qquad \text{and} \qquad \sum_{S \in \mathcal{C}:\, x \in S} x_S \geq 1 \text{ for all } x \in V\,.$$

Let q be a positive integer such that each of the numbers x_S $(S \in \mathcal{C})$ is an integer multiple of $1/q$. Set $p = qw$, noting that p is also a positive integer. Letting $y_S = qx_S$ for every $S \in \mathcal{C}$, we have:

$$\sum_{S \in \mathcal{C}} y_S = p\,, \qquad \text{and} \qquad \sum_{S \in \mathcal{C}:\, x \in S} y_S \geq q \text{ for all } x \in V\,. \tag{2.2}$$

Replacing each pair (S, y_S) with y_S copies of the pair $(S, 1)$ if necessary, we can assume that $y_S = 1$ for every S. Similarly, replacing each S by a proper subset if necessary, we can assume that the inequality (2.2) is tight for every x. (Note that this step depends on the fact that the collection of strong hypercliques, $\mathcal{C}$, is closed under taking subsets.) Altogether we have a sequence of sets $S_1, S_2, \ldots, S_p$, each of which is a strong hyperclique in G, such that every message occurs in exactly q of these sets.

From such a set system it is easy to construct an index code where every message has q bits (i.e. $\Sigma = \{0,1\}^q$) and the broadcast utilizes p bits (i.e. $\Sigma_P = \{0,1\}^p$). Indeed, for each message $x \in V$ let $j_1(x) < j_2(x) < \cdots < j_q(x)$ denote the indices such that $x \in S_j$ for $j \in \{j_1(x), j_2(x), \ldots, j_q(x)\}$. If the bits of message x are denoted by $b_1(x), b_2(x), \ldots, b_q(x)$ then for each $1 \le i \le p$ the i-th bit of the index code is computed by taking the sum (modulo 2) of all bits $b_k(z)$ such that $z \in S_i$ and $i = j_k(z)$. Receiver $R = (S, x)$ is able to decode the k^{th} bit of x by taking the $j_k(x)$-th bit of the index code and subtracting various bits belonging to other messages $x' \in S_{j_k(x)}$. All of these bits are known to R since $S_{j_k(x)}$ is a strong hyperclique containing x. This confirms that $\beta(G) \le p/q = w$, as desired. ■

It remains to characterize the extreme upper LP solution:

Claim 2.9. *The LP-solution b_n satisfies $b_n(G) = \overline{\chi}_f(G)$.*

Proof. The proof hinges on the fact that the entire set of constraints of $\mathcal{B}_n$ gives a useful structural characterization of any feasible solution X. Once we have this structure it will be simple to infer the required result.

Lemma 2.10. *A vector X satisfies the slope constraint and the i-th order submodularity constraints for $i \in \{2, \ldots, n\}$ if and only if there exists a vector of non-negative numbers $w(T)$, defined for every non-empty set of messages T, such that $X(S) = |S| + \sum_{T:T \not\subseteq S} w(T)$ for all $S \subseteq V$.*

The proof of this fact is similar to a characterization of a weighted coverage function. While much of the proof is likely folklore, we include it in Section 6 for completeness.

Given this fact we now prove that $b_n(G) \ge \overline{\chi}_f(G)$ by showing that any solution X having the form stated in Lemma 2.10 is a fractional coloring of $\overline{G}$. Thus, for the remainder of this subsection, X refers to a solution of $\mathcal{B}_n$ having value $b_n(G)$ and w refers to the associated vector of non-negative numbers whose existence is guaranteed by Lemma 2.10.

Fact 2.11. *For every message $x \in V$, $\sum_{T \ni x} w(T) = 1$.*

To see this, observe that monotonicity and decoding imply that $X(V \setminus \{x\}) = X(V)$. Lemma 2.10 implies that the right-hand-side is n while the left-hand-side is $n - 1 + \sum_{T \ni x} w(T)$.

Fact 2.12. *For every receiver R_j, if x denotes $x_{f(j)}$, then $\sum_{T:\, x \in T \subseteq N(j) \cup \{x\}} w(T) = 1$.*

Indeed, monotonicity and decoding imply that $X(N(j) \cup \{x\}) = X(N(j))$. Lemma 2.10 implies that the right side and left side differ by $1 - \sum_{T:\, x \in T \subseteq N(j) \cup \{x\}} w(T)$.

For a message x, let $N(x) = \bigcap_{j:x=x_{f(j)}} N(j)$ be the intersection of the side information for every receiver who wants to know x. By combining Facts 2.11 and 2.12 we find that if $w(T)$ is positive then T is contained in $N(x) \cup \{x\}$ for every x in T. Thus, we can infer the following:

Corollary 2.13. *If $w(T) > 0$ then the set of receivers desiring messages in T is a strong hyperclique.*

Now, to prove $b_n(G) \le \overline{\chi}_f(G)$ we show that if a vector w gives a feasible fractional coloring then $X(S) = |S| + \sum_{T:T \not\subseteq S} w(T)$ is feasible for the LP $\mathcal{B}_n$. By the argument made in the proof of Claim 2.8 we can assume without loss of generality that $\sum_{T \ni u} w(T) = 1 \ \forall u \in V$. X has value equal to the fractional coloring because $X(\emptyset) = \sum_T w(T)$. Further, Lemma 2.10 implies that X satisfies the i-th order submodularity constraints and slope. It trivially satisfies initialization and non-negativity.

To show that X satisfies monotonicity it is sufficient to prove that $X(S \cup \{u\}) \geq X(S)$ for all $S \subseteq V, u \in V \setminus S$. By definition, we have $X(S \cup \{u\}) - X(S) = 1 - \sum_{T:\, u \in T \not\subseteq S} w(T)$. Additionally, we know $\sum_{T:\, u \in T \not\subseteq S} w(T) \leq \sum_{T: u \in T} w(T) = 1$, where the last equality is because w is a fractional coloring. Finally, for the decoding constraints, it is sufficient to show that $X(A) \geq X(A \cup \{x\})$ for $A = N(j)$ where R_j is a receiver who desires x. By definition of X, $X(A) - X(A \cup \{x\}) = \sum_{T:\, x \in T \subseteq N(j) \cup \{x\}} w(T) - 1$. Also, $\sum_{T:\, x \in T \subseteq N(j) \cup \{x\}} w(T) = \sum_{T \ni x} w(T) = 1$ because T with $w(T) > 0$ is a strong hyperclique. ■

2.2.2 Strict lower and upper bounds for the broadcast rate

Claim 2.14. *There exists a broadcasting with side information instance G for which $\beta(G) < b_3(G)$.*

Proof. The construction is an extremely simple instance with only three messages $\{a, b, c\}$ and three receivers $(\{a\}, b), (\{b\}, c)$, and $(\{c\}, a)$. It is easy to see that there is a valid solution in which $\Sigma = \{0, 1\}$, $\Sigma_P = \{0, 1\}^2$, the encoding function is given by $a \oplus b, b \oplus c$. Thus $\beta \leq 2$. However, using the 3$^{\text{rd}}$-order submodularity constraint we have that

$$X(ab) + X(bc) + X(ac) + X(\emptyset) \geq X(abc) + X(a) + X(b) + X(c).$$

Combining that with decoding inequalities

$$X(a) \geq X(ab)\,, \quad X(b) \geq X(bc)\,, \quad X(c) \geq X(ac)\,,$$

together with the initialization inequality $X(abc) \geq 3$ now gives us that $b_3 = X(\emptyset) \geq 3$. ■

Claim 2.15. *There exists a broadcasting instance G for which $b_2(G) < \beta(G)$.*

Proof. We construct an instance G such that $\beta \geq \frac{45}{11}$ whereas $b_2 = 4$. The instance is constructed from the well-known Vámos matroid, used e.g. in [11] to show that Shannon inequalities do not suffice to prove tight bounds for network coding problems.

Definition 2.16. The Vámos matroid is an eight-element rank-four matroid whose ground set is $E = \{a, b, c, d, w, x, y, z\}$ and whose dependent sets are all the subsets of cardinality at least five as well as the four-element sets $\{b, c, x, y\}, \{a, c, w, y\}, \{a, b, w, x\}, \{c, d, y, z\}$, and $\{b, d, x, z\}$.

We use the following framework of [7] (where it was used to establish a polynomial gap between linear and non-linear index coding rates) to obtain a broadcasting instance from the Vámos matroid. Given a matroid M with a rank function r, create a message for each element of the ground-set of M, and create receivers for each circuit. That is, receivers are in one-to-one correspondence with pairs (T, e) such that $r(T, e) = r(T)$, and there is no subset $T' \subset T$ for which $r(T', e) = r(T')$. (This amounts to having $|C|$ receivers for each circuit C in M.) Using this construction, the matroid rank function can be used to give a primal solution to our LP $\mathcal{B}_2$ and thus an upper bound on b_2. A similar yet slightly more complicated transformation from matroids to broadcasting instances was given in [21]. Our construction is (essentially) a subset of the one appearing there.

We begin by showing that $b_2 = 4$. First, we prove that the function $r'(A) = r(A) + 4$ is a valid primal solution to $\mathcal{B}_2$. To show this we must simply verify all the constraints of our LP. Initialization and non-negativity are trivial: $r'(V) = r(V) + 4 = 8$ and $r'(\emptyset) = 4 \geq 0$. Slope is satisfied since adding a single element to a subset of a matroid can increase the rank by at most one. It is well known that rank functions (and therefore rank functions offset by a constant) satisfy monotonicity and submodularity. Decoding holds as we defined our receivers precisely so that they represent matroid dependencies. Since r' gives a feasible primal solution, we infer that $b_2 \le r'(\emptyset) = 4$.

To see that $b_2 \ge 4$, let A be the set of messages corresponding to a four-element independent set. Now, applying the decoding inequalities for each circuit that contains A and combining them using submodularity, we get that $X(A) \ge X(E)$. Summing and canceling that with the slope constraint $X(\emptyset) + 4 \ge X(A)$ and the initialization constraint $X(E) \ge 8$ gives $X(\emptyset) \ge 4$.

We will next show that $\beta \ge \frac{45}{11}$ by combining the inequalities used in $\mathcal{B}_2$ with a non-Shannon-type inequality due to Zhang and Yeung [26].

Theorem 2.17 ([26])**.** *The following is a non-Shannon-type information inequality:*

$$\begin{aligned} 3H(AC) &+ 3H(AD) + 3H(CD) + H(BC) + H(BD) \\ &\ge 2H(C) + 2H(D) + H(AB) + H(A) + H(BCD) + 4H(ACD)\,. \end{aligned} \tag{2.3}$$

Following the approach of [11], we apply the above inequality to sets $A = \{d, z\}$, $B = \{a, w\}$, $C = \{b, x\}$ and $D = \{c, y\}$. Observe that the rank function does not satisfy this inequality since the sets on the left-hand-side are each dependent sets of size four, giving a total rank of 33, yet on the right-hand-side we have six sets with rank 4 and five with rank 2, giving a total rank of 34.

Let T be a variable that ranges over all of the four-element dependent sets appearing on the left side of (2.3) and let $P = \mathcal{E}(x_1, \dots, x_n)$ be our optimal encoding. Summing and rearranging the following inequalities will produce the desired result:

$$\begin{aligned}
&11 \times [H(P) + 3 \ge H(T, P)] && \text{(slope , decode)}\\
&11H(T, P) \ge 6H(V, P) + 2H(bx, P) + 2H(cy, P) + H(dz, P) && \text{(non-Shannon , decode)}\\
&4 \times [H(P) + 2 \ge H(ad, P)] && \text{(slope)}\\
&H(P) + 2 \ge H(ab, P) && \text{(slope)}\\
&2 \times [H(ad, P) + H(bx, P) \ge H(V, P) + H(P)] && \text{(submod , decode)}\\
&2 \times [H(ad, P) + H(cy, P) \ge H(V, P) + H(P)] && \text{(submod , decode)}\\
&H(ab, P) + H(dz, P) \ge H(V, P) + H(P) && \text{(submod , decode)}\\
&11 \times [H(V, P) \ge 8] && \text{(initialize)}\,.
\end{aligned}$$

Altogether, $\beta \ge H(P) \ge \frac{45}{11}$ while $b_2 = 4$, completing the proof of the Claim 2.15. ■

2.3 The broadcast rate of the 5-cycle

As stated in Theorem 1, whenever the LP-solution b_2 equals $\overline{\chi}_f$ we obtain that β is precisely this value, hence one may compute the broadcast rate (previously unknown for any graph) via a chain of entropy-inequalities. We will demonstrate this in Section 5 by determining β for several families of graphs, in particular for cycles and their complements (Theorem 5.1). These seemingly simple cases were previously studied in [2, 5] yet their β values were unknown before this work.

To give a flavor of the proof of Theorem 5.1, we provide a proof-by-picture for the broadcast rate of the 5-cycle (Figure 1), illustrating the intuition behind choosing the set of inequalities one may combine for an analytic lower bound on β. The inequalities in Figure 1 establish that $\beta(C_5) \ge \frac{5}{2}$, thus matching the upper bound $\beta(C_5) \le \overline{\chi}_f(C_5) = \frac{5}{2}$.

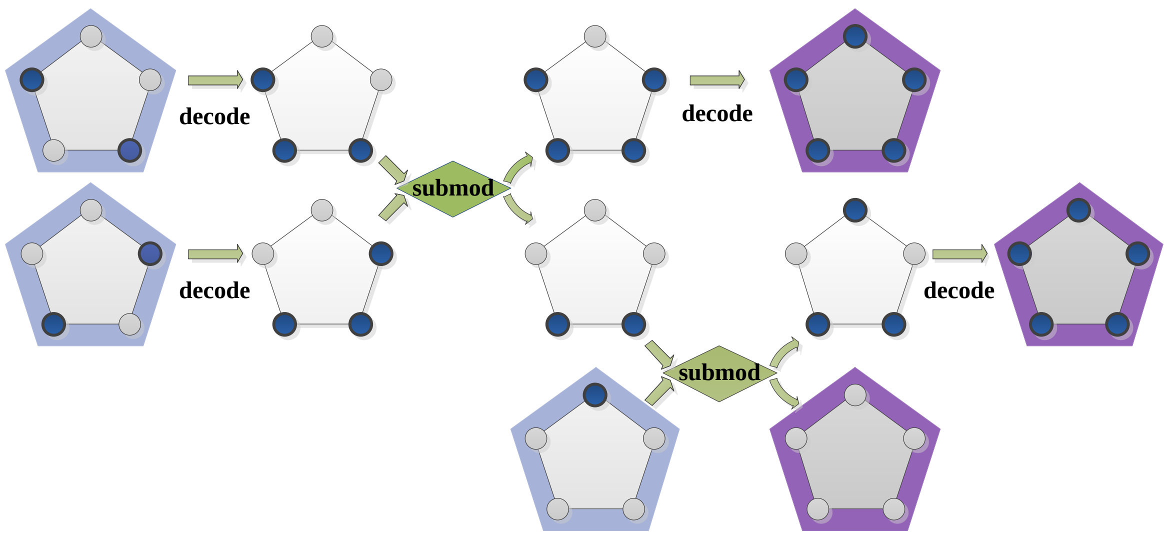

Figure 1: A proof-by-picture that $\beta(C_5) = \frac{5}{2}$. Variables marked by highlighted subsets of vertices, e.g. the first submodularity application applies the LP constraint $X(\{3,4,5\}) + X(\{2,3,4\}) \geq X(\{2,3,4,5\}) + X(\{3,4\})$. Final outcome is a proof that $\beta(C_5) \geq X(\emptyset)$ with $3X(\emptyset)+5 \geq X(\emptyset)+10$.

We note that odd cycles on $n \geq 5$ vertices as well as their complements constitute the first examples for graphs where the independence number α is strictly smaller than β. Corollary 2.18 will further amplify the gap between these parameters.

2.4 Corollaries for vector/scalar index codes

Prior to this work and its companion paper [7] there was no known family of graphs where $\alpha \neq \beta$, and one could conjecture that for long enough messages the broadcast rate in fact converges to the independence number, the largest set of receivers that are pairwise oblivious. We now have that the 5-cycle provides an example where $\alpha = 2$ while $\beta = \frac{5}{2}$, however here the difference $\beta - \alpha < 1$ could potentially be attributed to integer-rounding, e.g. it could be that $\alpha = \lfloor \beta \rfloor$.

Such was also the case for the best known difference between the vector capacity β and the scalar capacity β_1. The best lower bound on $\beta_1 - \beta$ in any graph was again attained by the 5-cycle where it was slightly less than $\frac{1}{3}$, and again in the constrained setting of graph Index Coding we could conjecture that $\beta_1 = \lceil \beta \rceil$.

The following corollary of the above mentioned results refutes these suggestions by amplifying both these gaps to be linear in n. The separation between α and β was further strengthened in the companion paper [7], where we obtained a gap of a polynomial factor between these parameters.

Corollary 2.18. *There exists a family of graphs G on n vertices for which $\beta(G) = n/2$ while $\alpha(G) = \frac{2}{5}n$ and $\beta_1(G) = (1 - \frac{1}{5}\log_2 5 + o(1))n \approx 0.54n$. Moreover, we have $\beta^*(G) = (1-o(1))\beta_1(G)$.*

To prove this result we will use the direct-sum capacity β^*. Recall that this capacity is defined to be $\beta^*(G) = \lim_{t\to\infty} \frac{1}{t}\beta_1(t\cdot G) = \inf_t \frac{1}{t}\beta_1(t \cdot G)$ where $t \cdot G$ denotes the disjoint union of t copies of G. This parameter satisfies $\beta \leq \beta^* \leq \beta_1$. Similarly we let $G + H$ denote the disjoint union of the graphs G, H. We need the following simple lemma.

Lemma 2.19. *The parameters β and β^* are additive with respect to disjoint unions, that is for any two graphs G, H we have $\beta(G+H) = \beta(G) + \beta(H)$ and $\beta^*(G+H) = \beta^*(G) + \beta^*(H)$.*

Proof of lemma. The fact that β^* is additive w.r.t. disjoint unions follows immediately from the results of [2]. Indeed, it was shown there that for any graph G on n vertices $\beta^*(G) = \log_2 \chi_f(\mathfrak{C}(G))$ where $\mathfrak{C} = \mathfrak{C}(G)$ is an appropriate undirected Cayley graph on the group $\mathbb{Z}_2^n$. Furthermore, it was shown that $\mathfrak{C}(G+H) = \mathfrak{C}(G) \vee \mathfrak{C}(H)$, where $\vee$ denotes the OR-graph-product. It is well-known (see, e.g., [13, 17]) that the fractional chromatic number is multiplicative w.r.t. this product, i.e. $\chi_f(G \vee H) = \chi_f(G)\chi_f(H)$ for any two graphs G, H. Combining these statements we deduce that

$$2^{\beta^*(G+H)} = \chi_f(\mathfrak{C}(G+H)) = \chi_f(\mathfrak{C}(G) \vee \mathfrak{C}(H)) = \chi_f(\mathfrak{C}(G))\chi_f(\mathfrak{C}(H)) = 2^{\beta^*(G)+\beta^*(H)}\,.$$

We shall now use this fact to show that β is additive. The inequality $\beta(G+H) \leq \beta(G) + \beta(H)$ follows from concatenating the codes for G and H and it remains to show a matching upper bound.

As observed by [19], the Index Coding problem for an n-vertex graph G with messages that are t bits long has an equivalent formulation as a problem on a graph with tn vertices and messages that are 1-bit long; denote this graph by G_t (formally this is the t-blow-up of G with independent sets, i.e. the graph on the vertex set $V(G) \times [t]$, where (u, i) and (v, j) are adjacent iff $uv \in E(G)$). Under this notation $\beta_t(G) = \beta_1(G_t)$. Notice that $(G+H)_t = G_t + H_t$ for any t and furthermore that $s \cdot G_t$ is a spanning subgraph of G_{st} for any s and t, in particular implying that $\beta_1(s \cdot G_t) \geq \beta_1(G_{st})$.

Fix $\varepsilon > 0$ and let t be a large enough integer such that $\beta(G+H) \geq \beta_t(G+H)/t - \varepsilon$. Further choose some large s such that $\beta^*(G_t) \geq \beta_1(s \cdot G_t)/s - \varepsilon$ and $\beta^*(H_t) \geq \beta_1(s \cdot H_t)/s - \varepsilon$. We now get

$$\beta(G+H) + \varepsilon \geq \beta_1(G_t + H_t)/t \geq \beta^*(G_t + H_t)/t = \beta^*(G_t)/t + \beta^*(H_t)/t\,,$$

where the last inequality used the additivity of β^*. Since

$$\beta^*(G_t)/t \geq \beta_1(s \cdot G_t)/st - \varepsilon \geq \beta_1(G_{st})/st - \varepsilon \geq \beta(G) - \varepsilon$$

and an analogous statement holds for $\beta^*(H_t)/t$, altogether we have $\beta(G+H) \geq \beta(G) + \beta(H) - 3\varepsilon$. Taking $\varepsilon \to 0$ completes the proof of the lemma. ∎

Proof of Corollary 2.18. Consider the family of graphs on $n = 5k$ vertices given by $G = k \cdot C_5$. It was shown in [2] that $\beta^*(C_5) = 5 - \log_2 5$, which by definition implies that $\beta^*(G) = (5 - \log_2 5)k$ and $\beta_1(G) = \beta^*(G) + o(k)$. At the same time, clearly $\alpha(G) = 2k$ and combining the fact that $\beta(C_5) = \frac{5}{2}$ with Lemma 2.19 gives $\beta(G) = 5k/2 = n/2$, as required. ∎

The above result showed that the difference between the broadcast rate β and the Index Coding scalar capacity β_1 can be linear in the number of messages. We now wish to use the gap between β and β_1 to infer a gap between the vector and scalar Network Coding capacities.

Corollary 2.20. *For any $k \geq 1$ there exists a Network Coding instance on $5k+2$ vertices where the ratio between the vector and scalar-linear capacities is precisely 1.2 while the ratio between the vector and scalar capacities converges to $1 - \frac{1}{2}\log_2 5 \approx 1.07$ as $k \to \infty$.*

Proof. It is well known (e.g. [22]) that an n-vertex graph Index Coding instance G can be translated into a capacitated network H on $2n+2$ vertices via a reduction that preserves linear encoding. It thus suffices to bound the ratio of the corresponding Index Coding capacities.

For $k \geq 1$ consider the graph G consisting of k disjoint 5-cycles. Corollary 2.18 established that $\beta(G) = 5k/2$ whereas $\beta_1(G) = (5 - \log_2 5 + o(1))k$ where the $o(1)$-term tends to 0 as $k \to \infty$. At the same time, it was shown in [5] that the scalar-linear Index Coding capacity over $GF(2)$ coincides with a parameter denoted by $\mathrm{minrk}_2(G)$, and as observed in [19] this extends to any finite field $\mathbb{F}$ as follows: For a graph $H = (V, E)$ we say that a matrix B indexed by V over $\mathbb{F}$ is a *representation* of H over $\mathbb{F}$ if it has nonzero diagonal entries ($B_{uu} \neq 0$ for all $u \in V$) whereas $B_{uv} = 0$ for any $u \neq v$ such that $uv \notin E$. The smallest possible rank of such a matrix over $\mathbb{F}$ is denoted by $\mathrm{minrk}_{\mathbb{F}}(H)$. For the 5-cycle we have $\mathrm{minrk}_{\mathbb{F}}(C_5) \leq \overline{\chi}(C_5) = 3$ by the linear clique-cover encoding and this is tight by as $\mathrm{minrk}_{\mathbb{F}}(C_5) \geq \lceil \beta(C_5) \rceil = 3$. Finally, $\mathrm{minrk}_{\mathbb{F}}$ is clearly additive w.r.t. disjoint unions of graphs by its definition and thus $\mathrm{minrk}_{\mathbb{F}}(G) = 3k$ as required. ■

3 Approximating the Broadcast Rate

This section is devoted to the proof of Theorem 2, on polynomial-time algorithms for approximating β and deciding whether $\beta = 2$. Working in the setting of a general broadcast network is somewhat delicate and we begin by sketching the arguments that will follow.

In the simpler case of undirected graphs, a $o(n)$-approximation to β is implied by results of [3, 8, 24] that together give a polynomial time procedure that finds either a small clique-cover or a large independent set (see Remark 3.1). To get an approximation for the general broadcasting problem we will apply a similar technique using analogues of independent sets and clique-covers that give lower and upper bounds respectively on the general broadcasting rate. The analogue of an independent set is an *expanding sequence* — a sequence of receivers where the i^{th} receiver's desired message is unknown to receivers $1, \ldots, i-1$. The clique-cover analogue is a weak fractional hyperclique-cover (see Definition 2.5). In the remainder of this section, whenever we refer to hypercliques or hyperclique-covers we always mean weak hypercliques and weak hyperclique-covers.

We will prove that there is a polynomial time algorithm that outputs an expanding sequence of size k or reports a fractional hyperclique-cover of size $O\left(kn^{1-1/k}\right)$; the approximation follows by setting k appropriately. We will argue that either we can partition the graph and apply induction or else the side-information map is dense enough to deduce existence of a small fractional hyperclique-cover. The proof of the latter step deviates significantly from the techniques used for graphs, and seems interesting in its own right. We will give a simple procedure to randomly sample hypercliques and use it to produce a valid weight function for the hyperclique-cover by defining the weight of a hyperclique to be proportional to the probability it is sampled by the procedure.

To prove the second part of Theorem 2 we will prove that a structure called an *almost alternating cycle* (AAC) constitutes a minimal obstruction to obtaining a broadcast rate of 2. The proof makes crucial use of Theorem 1, calculating the parameter b_2 for AAC's to prove that their broadcast rate is strictly greater than 2. Furthermore, the proof reduces finding an AAC to finding the transitive closure of a particular relation, which is polynomial time computable.

3.1 Approximating the broadcast rate in general networks

We now present a nontrivial approximation algorithm for β for a general network described by a hypergraph (that is, the most general framework where there are $m \geq n$ receivers).

Remark 3.1. In the setting of undirected graphs a slightly better approximation algorithm for β is a consequence of a result of Boppana and Halldorsson [8], following the work of Wigderson [24].

In [8] the authors showed an algorithm that finds either a "large" clique or a "large" independent set in a graph (where the size guarantee involves the Ramsey number estimate). A simple adaptation of this result (Proposition 2.1 in the Alon-Kahale [3] work on approximating α via the ϑ-function) gives a polynomial-time algorithm for finding an independent set of size $t_k(m) = \max\left\{s : \binom{k+s-2}{k-1} \le m\right\}$ in any graph satisfying $\overline{\chi}(G) \ge n/k + m$. In particular, taking $m = n/k$ with $k = \frac{1}{2}\log n$ we clearly have $t_k(m) \ge k$ for any sufficiently large n and obtain that either $\overline{\chi}(G) < 4n/\log n$ or we can find an independent set of size $\frac{1}{2}\log n$ in polynomial-time.

We use the following notation: the n message streams are identified with the elements of $[n] = V$. The data consisting of the pairs $\{(N(j), f(j))\}_{j=1}^m$ is our *directed hypergraph* instance. When referring to the hypergraph structure itself (rather than the corresponding index coding problem) we will refer to elements of V as *vertices* and we will refer to pairs $(N(j), f(j))$ as *directed hyperedges*. For notational convenience, we denote $S(j) = N(j) \cup \{f(j)\}$.

An *expanding sequence* of size k is a sequence of receivers $j_1, \ldots, j_k$ such that

$$f(j_\ell) \notin \bigcup_{i<\ell} S(i) \tag{3.1}$$

for $1 \le \ell \le k$. For a hypergraph G, let $\alpha(G)$ denote the maximum size of an expanding sequence.

Lemma 3.2. *Every hypergraph G satisfies the bound $\beta(G) \ge \alpha(G)$.*

Proof. The proof is by contradiction. Let $j_1, \ldots, j_k$ be an expanding sequence and suppose that there is an index code that achieves rate $r < k$. Let $J = \{j_1, \ldots, j_k\}$. For $b = \log_2 |\Sigma|$ we have

$$|\Sigma|^k = 2^{bk} > 2^{br} \ge |\Sigma_P|.$$

Let us fix an element $x_i^* \in \Sigma$ for every $i \notin \{f(j) : j \in J\}$, and define Ψ to be the set of all $\vec{x} \in \Sigma^n$ that satisfy $x_i = x_i^*$ for all $i \notin \{f(j) : j \in J\}$. The cardinality of Ψ is $|\Sigma|^k$, so the Pigeonhole Principle implies that the function $\mathcal{E}$, restricted to Ψ, is not one-to-one. Suppose that $\vec{x}$ and $\vec{y}$ are two distinct elements of Ψ such that $\mathcal{E}(\vec{x}) = \mathcal{E}(\vec{y})$. Let i be the smallest index such that $x_{f(j_i)} \ne y_{f(j_i)}$. Denoting j_i by j, we have $x_k = y_k$ for all $k \in N(j)$, because $N(j)$ does not contain $f(j_\ell)$ for any $\ell \ge i$, and the components with indices $j_i, j_{i+1}, \ldots, j_k$ are the only components in which $\vec{x}$ and $\vec{y}$ differ. Consequently receiver j is unable to distinguish between message vectors $\vec{x}, \vec{y}$ even after observing the broadcast message, which violates the condition that j must be able to decode message $f(j)$. ■

Lemma 3.3. *Let $\psi_f(G)$ denote the minimum weight of a fractional weak hyperclique-cover of G. Every hypergraph G satisfies the bound $\beta(G) \le \psi_f(G)$.*

Proof. The linear program defining $\psi_f(G)$ has integer coefficients, so G has a fractional hyperclique cover of weight $w = \psi_f(G)$ in which the weight $w(\mathcal{J})$ of every hyperclique $\mathcal{J}$ is a rational number. Assume we are given such a fractional hyperclique-cover, and choose an integer d such that $w(\mathcal{J})$ is an integer multiple of $1/d$ for every $\mathcal{J}$. Let $\mathcal{C}$ denote a multiset of hypercliques containing $d \cdot w(\mathcal{J})$ copies of $\mathcal{J}$ for every hyperclique $\mathcal{J}$. Note that the cardinality of $\mathcal{C}$ is $d \cdot w$.

For any hyperclique $\mathcal{J}$, let $f(\mathcal{J})$ denote the set $\bigcup_{j\in\mathcal{J}}\{f(j)\}$. For each $i \in [n]$, let $\mathcal{C}_i$ denote the sub-multiset of $\mathcal{C}$ consisting of all hypercliques $\mathcal{J} \in \mathcal{C}$ such that $i \in f(\mathcal{J})$. Fix a finite field $\mathbb{F}$ such that $|\mathbb{F}| > dw$. Define $\Sigma = \mathbb{F}^d$ and $\Sigma_P = \mathbb{F}^{d\cdot w}$. Let $\{\xi_P^{\mathcal{J}}\}_{\mathcal{J}\in\mathcal{C}}$ be a basis for the dual vector space

Σ_P^* and let $\{\xi_i^{\mathcal{J}}\}_{\mathcal{J}\in\mathcal{C}_i}$ be a set of dual vectors in Σ^* such that any d of these vectors constitute a basis for Σ^*. (The existence of such a set of dual vectors is guaranteed by our choice of $\mathbb{F}$ with $|\mathbb{F}| > dw \geq d$.)

The encoding function is defined to be the unique linear function satisfying

$$\xi_P^{\mathcal{J}}(\mathcal{E}(x_1,\ldots,x_n)) = \sum_{i\in f(\mathcal{J})} \xi_i^{\mathcal{J}}(x_i) \qquad \forall \mathcal{J}.$$

For each receiver j, if $i = f(j)$, the set of dual vectors $\xi_i^{\mathcal{J}}$ with $j \in \mathcal{J}$ compose a basis of Σ^*, hence to prove that j can decode message x_i it suffices to show that j can determine the value of $\xi_i^{\mathcal{J}}(x_i)$ whenever $j \in \mathcal{J}$. This holds because the public channel contains the value of $\sum_{\ell\in f(\mathcal{J})} \xi_\ell^{\mathcal{J}}(x_\ell)$, and receiver j knows that value of $\xi_\ell^{\mathcal{J}}(x_\ell)$ for every $\ell \neq i$ in $f(\mathcal{J})$ because $\ell \in N(j)$. ∎

We now turn our attention to bounding the ratio $\psi_f(G)/\alpha(G)$ for a hypergraph G. Our goal is to show that this ratio is bounded by a function in $o(n)$. To begin with, we need an analogue of the lemma that undirected graphs with small maximum degree have small fractional chromatic number.

Lemma 3.4. *If G is a hypergraph with n vertices, and d is a natural number such that for every receiver j, $|S(j)| + d \geq n$, then $\psi_f(G) \leq 4d + 2$.*

Proof. Let us define a procedure for sampling a random subset $T \subseteq [n]$ and a random hyperclique $\mathcal{J}$ as follows. Let π be a uniformly random permutation of $[n + d]$, let i be the least index such that $\pi(i + 1) > n$, and let T be the set $\{\pi(1), \pi(2), \ldots, \pi(i)\}$. (If $\pi(1) > n$ then $i = 0$ and T is the empty set.) Now let $\mathcal{J}$ be the set of all j such that $f(j) \in T \subseteq S(j)$. (Note that $\mathcal{J}$ is indeed a hyperclique.)

For any hyperclique $\mathcal{J}$ let $p(\mathcal{J})$ denote the probability that $\mathcal{J}$ is sampled by this procedure and let $w(\mathcal{J}) = (4d + 2) \cdot p(\mathcal{J})$. We claim that the weights $w(\cdot)$ define a fractional hyperclique-cover of G, or equivalently, that for every receiver j, $\mathbb{P}(f(j) \in T \subseteq S(j)) \geq \frac{1}{4d+2}$. Let $U(j)$ denote the set $\{f(j)\} \cup ([n] \setminus S(j)) \cup ([n + d] \setminus [n])$. The event $\mathcal{E} = \{f(j) \in T \subseteq S(j)\}$ occurs if and only if, in the ordering of $U(j)$ induced by π, the first element of $U(j)$ is $f(j)$ and the next element belongs to $[n + d] \setminus [n]$. Thus,

$$\mathbb{P}(\mathcal{E}) = \frac{1}{|U(j)|} \cdot \frac{d}{|U(j)| - 1}.$$

The bound $\mathbb{P}(\mathcal{E}) \geq \frac{1}{4d+2}$ now follows from the fact that $|U(j)| \leq 2d + 1$. ∎

Lemma 3.5. *If G is a hypergraph and $\alpha(G) \leq k$, then $\psi_f(G) \leq 6k \cdot n^{1-1/k}$. Moreover, there is a polynomial-time algorithm, whose input is a hypergraph G and a natural number k, that either outputs an expanding sequence of size $k + 1$ or reports (correctly) that $\psi_f(G) \leq 6k \cdot n^{1-1/k}$.*

Proof. The proof is by induction on k. In the base case $k = 1$, either G itself is a hyperclique or there is some pair of receivers j, j' such that $f(j)$ is not in $S(j')$. In that case, the sequence $j_1 = j', j_2 = j$ is an expanding sequence of size 2.

For the induction step, for each hyperedge j define the set $D(j) = \{f(j)\} \cup ([n] \setminus S(j))$ and let j_1 be a hyperedge such that $|D(j)|$ is maximum. If $|D(j_1)| \leq n^{1-1/k} + 1$, then the bound $|S(j)| + n^{1-1/k} \geq n$ is satisfied for every j and Lemma 3.4 implies that $\psi_f(G) < 4n^{1-1/k} + 2 \leq$

$6n^{1-1/k}$. Otherwise, partition the vertex set of G into $V_1 = [n] \setminus S(j_1)$ and $V_2 = S(j_1)$, and for $i = 1, 2$ define G_i to be the hypergraph with vertex set V_i and edge set E_i consisting of all pairs $(N(j) \cap V_i, f(j))$ such that $(N(j), f(j))$ is a hyperedge of G with $f(j) \in V_i$. (We will call such a structure the *induced sub-hypergraph of G on vertex set V_i*.) If G_1 contains an expanding sequence $j_2, j_3, \ldots, j_{k+1}$ of size k, then the sequence $j_1, j_2, \ldots, j_{k+1}$ is an expanding sequence of size $k+1$ in G. (Moreover, if an algorithm efficiently finds the sequence $j_2, j_3, \ldots, j_{k+1}$ then it is easy to efficiently construct the sequence $j_1, \ldots, j_{k+1}$.) Otherwise, by the induction hypothesis, G_1 has a fractional hyperclique-cover of weight at most $6(k-1)|V_1|^{1-1/(k-1)} \leq 6(k-1)|V_1|n^{-1/k}$. Continuing to process the induced sub-hypergraph on vertex set V_2 in the same way, we arrive at a partition of $[n]$ into disjoint vertex sets $W_1, W_2, \ldots, W_\ell$ of cardinalities $n_1, \ldots, n_\ell$, respectively, such that for $1 \leq i < \ell$, the induced sub-hypergraph on W_i has a fractional clique-cover of weight at most $6(k-1)n_i n^{-1/k}$, and for $i = \ell$ the induced sub-hypergraph on W_i satisfies the hypothesis of Lemma 3.4 with $d = n^{1-1/k}$ and consequently has a fractional hyperclique-cover of weight at most $6n^{1-1/k}$. The lemma follows by summing the weights of these hyperclique-covers. ■

Combining Lemmas 3.2, 3.3, 3.5, we obtain the approximation algorithm asserted by Theorem 2.

3.2 Extending the algorithm to networks with variable source rates

The aforementioned approximation algorithm for β naturally extends to the setting where each source in the broadcast network has its own individual rate. Namely, the n message streams are identified with the elements of $[n] = V$, where message stream i has a rate r_i, and the problem input consists of the vector $(r_1, \ldots, r_n)$ and the pairs $\{(N(j), f(j))\}_{j=1}^m$. Thus the input is a *weighted directed hypergraph* instance. An index code for a weighted hypergraph consists of the following:

- Alphabets Σ_P and Σ_i for $1 \leq i \leq n$,
- An encoding function $\mathcal{E} : \prod_{i=1}^n \Sigma_i \to \Sigma_P$,
- Decoding functions $\mathcal{D}_j : \Sigma_P \times \prod_{i \in N(j)} \Sigma_i \to \Sigma_{f(j)}$.

The encoding and decoding functions are required to satisfy

$$\mathcal{D}_j(\mathcal{E}(\sigma_1, \ldots, \sigma_n), \sigma_{N(j)}) = \sigma_{f(j)}$$

for all $j = 1, \ldots, m$ and all $(\sigma_1, \ldots, \sigma_n) \in \prod_{i=1}^n \Sigma_i$. Here the notation $\sigma_{N(j)}$ denotes the tuple obtained from a complete n-tuple $(\sigma_1, \ldots, \sigma_n)$ by retaining only the components indexed by elements of $N(j)$. An index code *achieves* rate $r \geq 0$ if there exists a constant $b > 0$ such that $|\Sigma_i| \geq 2^{b \cdot r_i}$ for $1 \leq i \leq n$ and $|\Sigma_P| \leq 2^{b \cdot r}$. If so, we say that rate r is *achievable*. If G is a weighted hypergraph, we define $\beta(G)$ to be the infimum of the set of achievable rates.

The first step in generalizing the proof given in the previous subsection to the case where the r_i's are non-uniform is to properly extend the notions of hypercliques and expanding sequences. A weak fractional hyperclique cover of a weighted hypergraph will now assign a weight $w(\mathcal{J})$ to every weak hyperclique $\mathcal{J}$ such that for every receiver j, $\sum_{\mathcal{J} \ni j} w(\mathcal{J}) \geq r_{f(j)}$ (cf. Definition 2.5 corresponding to $r_{f(j)} = 1$). As before, the weight of a fractional weak hyperclique-cover is given by $\sum_{\mathcal{J}} w(\mathcal{J})$ and for a weighted hypergraph G we let $\psi_f(G)$ denote the minimum weight of a fractional weak hyperclique-cover. An expanding sequence $j_1, \ldots, j_k$ is defined as before (see Eq. 3.1) except now we associate such a sequence with the weight $\sum_{\ell=1}^k r_{f(j_\ell)}$ and the quantity $\alpha(G)$ will denote the maximum weight of an expanding sequence (rather than the maximum cardinality).

With these extended defintions, the proofs in the previous subsection carry unmodified to the weighted hypergraph setting with the single exception of Lemma 3.5, where the assumption that the hypergraph is unweighted was essential to the proof. In what follows we will qualify an application of that lemma via a dyadic partition of the vertices of our weighted hypergraph according to their weights r_i.

Assume without loss of generality that $0 \le r_i \le 1$ for every vertex $i \in [n]$, and partition the vertex of set G into subsets $V_1, V_2, \ldots$ such that V_s contains all vertices i such that $2^{-s} < r_i \le 2^{1-s}$. Let G_s denote the induced hypergraph on vertex set V_s. For each of the nonempty hypergraphs G_s, run the algorithm in Lemma 3.5 for $k = 1, 2, \ldots$ until the smallest value of $k(s)$ for which an expanding sequence of size $k(s)+1$ is not found. If G_s° denotes the unweighted version of G_s, then we know that

$$\alpha(G_s) \ge 2^{-s}\alpha(G_s^\circ) \ge 2^{-s}k(s)$$
$$\psi_f(G_s) \le 2^{1-s}\psi_f(G_s^\circ) \le 2^{-s}\cdot 12k(s)n^{1-1/k(s)}.$$

In addition, for each $i \in V_s$ the set of hyperedges containing i constitutes a hyperclique, which implies the trivial bound

$$\psi_f(G_s) \le \sum_{i\in V_s} r_i \le 2^{1-s}|V_s|.$$

Combining these two upper bounds for $\psi_f(G_s)$, we obtain an upper bound for $\psi_f(G)$:

$$\psi_f(G) \le \sum_{s=1}^{\infty}\psi_f(G_s) \le \sum_{s=1}^{\infty} 2^{-s}\cdot\min\left\{12k(s)n^{1-1/k(s)},\, 2|V_s|\right\}. \tag{3.2}$$

We define $\tau(G)$ to be the right side of (3.2). We have described a polynomial-time algorithm to compute $\tau(G)$ and have justified the relation $\psi_f(G) \le \tau(G)$, so it remains to show that $\tau(G)/\alpha(G) \le cn\left(\frac{\log\log n}{\log n}\right)$ for some constant c.

The bound $\tau(G) \le n$ follows immediately from the definition of τ, so if $\alpha(G) \ge \frac{\log n}{\log\log n}$ there is nothing to prove. Assume henceforth that $\alpha(G) < \frac{\log n}{\log\log n}$, and define w to be the smallest integer such that $2^w\cdot\alpha(G) > \frac{\log n}{2\log\log n}$. We have

$$\begin{aligned}\tau(G) &\le \sum_{s=1}^{w} 2^{-s}\cdot 12k(s)n^{1-1/k(s)} \;+\; \sum_{s=w+1}^{\infty} 2^{1-s}\cdot|V_s| \\ &\le 12n\sum_{s=1}^{w}2^{-s}k(s)n^{-1/k(s)} \;+\; 2^{-w}\cdot n \\ &< 12n\alpha(G)\sum_{s=1}^{w}n^{-1/k(s)} \;+\; 2n\alpha(G)\left(\frac{\log\log n}{\log n}\right),\end{aligned} \tag{3.3}$$

with the last line derived using the relations $2^{-s}k(s) \le \alpha(G_s) \le \alpha(G)$ and $2^{-w} < \alpha(G)\left(\frac{2\log\log n}{\log n}\right)$. Applying once more the fact that $2^{-s}k(s) \le \alpha(G)$, we find that $n^{-1/k(s)} \le n^{-1/(2^s\cdot\alpha(G))}$. Substituting this bound into (3.3) and letting α denote $\alpha(G)$, we have

$$\frac{\tau(G)}{\alpha(G)} \le 2n\left(\frac{\log\log n}{\log n}\right) \;+\; 12n\left(n^{-1/2\alpha} + n^{-1/4\alpha} + \cdots + n^{-1/2^w\alpha}\right).$$

In the sum appearing on the right side, each term is the square of the one following it. It now easily follows that the final term in the sum is less than 1/2, so the entire sum is bounded above by twice its final term. Thus

$$\frac{\tau(G)}{\alpha(G)} \leq 2n \left(\frac{\log\log n}{\log n} \right) \; + \; 24n \cdot n^{-1/2^w \alpha}. \tag{3.4}$$

Our choice of w ensures that $2^w \alpha \leq \frac{\log n}{\log\log n}$ hence $n^{-2^{-w}a} \leq n^{-\log\log n/\log n} = (\log n)^{-1}$. By substituting this bound into (3.4) we obtain

$$\frac{\tau(G)}{\alpha(G)} \leq n \left(\frac{2\log\log n}{\log n} + \frac{24}{\log n} \right) ,$$

as desired.

3.3 Proof of Theorem 2, determining whether the broadcast rate equals 2

Let G be an undirected graph with independence number $\alpha = 2$. Clearly, if $\overline{G}$ is bipartite then $\overline{\chi}(G) = 2$ and so $\beta(G) = 2$ as well. Conversely, if $\overline{G}$ is not bipartite then it contains an odd cycle, the smallest of which is induced and has $k \geq 5$ vertices since the maximum clique in $\overline{G}$ is $\alpha(G) = 2$. In particular, Theorem 5.1 implies that $\beta(G) \geq \beta(\overline{C_k}) = \frac{k}{\lfloor k/2 \rfloor} > 2$. We thus conclude the following:

Corollary 3.6. *Let G be an undirected graph on n vertices whose complement $\overline{G}$ is nonempty. Then $\beta(G) = 2$ if and only if $\overline{G}$ is bipartite.*

A polynomial time algorithm for determining whether $\beta = 2$ in undirected graphs follows as an immediate consequence of Corollary 3.6. However, for broadcasting with side information in general — or even for the special case of directed graphs (the main setting of [5,6]) — it is unclear whether such an algorithm exists. In this section we provide such an algorithm, accompanied by a characterization theorem that generalizes the above characterization for undirected graphs. To state our characterization we need the following definitions. As in Section 3.1 we use $S(j)$ to denote the set $N(j) \cup \{f(j)\}$. Additionally, we introduce the notation $T(j)$ to denote the complement of $S(j)$ in the set of messages. When referring to the message desired by receiver R_j, we abbreviate $x_{f(j)}$ to $x(j)$. Henceforth, when referring to a hypergraph $G = (V, E)$, we assume that for each edge $j \in E$, the hypergraph structure specifies the vertex $f(j)$ and both of the sets $S(j), T(j)$.

Definition 3.7. If $G = (V, E)$ is a directed hypergraph and S is a set, a function $F : V \to S$ is said to be *G-compatible* if for every edge $j \in E$, there are two *distinct* elements $t, u \in S$ such that F maps every element of $T(j)$ to t, and it maps $f(j)$ to u.

Definition 3.8. If $G = (V, E)$ is a directed hypergraph, an *almost alternating (2n+1)-cycle* in G is a sequence of vertices $v_{-n}, v_{-n+1}, \ldots, v_n$, and a sequence of edges $j_0, \ldots, j_n$, such that for $i = 0, \ldots, n$, the vertex $f(j_i)$ is equal to v_{i-n} and the set $T(j_i)$ contains v_i, as well as v_{i+1} if $i < n$.

Theorem 3.9. *For a directed hypergraph G the following are equivalent:*

(i) $\beta(G) = 2$

(ii) There exists a set S and a G-compatible function $F : V \to S$.

(iii) G contains no almost alternating cycles.

Furthermore there is a polynomial-time algorithm to decide if these equivalent conditions hold.

Proof. **(i)⇒(iii):** The contrapositive statement says that if G contains an almost alternating cycle then $\beta(G) > 2$. Let $v_{-n}, \ldots, v_n$ be the vertices of an almost alternating $(2n+1)$-cycle with edges

$j_0, \ldots, j_n$. To prove $\beta(G) > 2$ we manipulate entropy inequalities involving the random variables $\{x_i : -n \le i \le n\}$ and y, where x_i denotes the message associated to vertex v_i normalized to have entropy 1, and y denotes the public channel. For $S \subseteq \{y, x_{-n}, \ldots, x_n\}$, let $H(S)$ denote the entropy of the joint distribution of the random variables in S, and let $\overline{H}(S)$ denote $H(\overline{S})$. Let $S_{i:j}$ denote the set $\{x_i, x_{i+1}, \ldots, x_j\}$.

For $0 \le i \le n-1$, we have

$$H(y) + (2n-2) \ge \overline{H}(\{x_{i-n}, x_i, x_{i+1}\}) = \overline{H}(\{x_i, x_{i+1}\}) = \overline{H}(S_{i:i+1})\,, \tag{3.5}$$

where the second equation holds because receiver j_i can decode message $x_{i-n} = x(j_i)$ given the value y and the values x_k for $k \in N(j_i)$. Using submodularity we have that for $0 < j < n$,

$$\overline{H}(S_{0:j}) + \overline{H}(S_{j:j+1}) \ge \overline{H}(S_{0:j+1}) + \overline{H}(x_j) = \overline{H}(S_{0:j+1}) + \overline{H}(\emptyset) = \overline{H}(S_{0:j+1}) + 2n + 1\,. \tag{3.6}$$

Summing up (3.6) for $j = 1, \ldots, n-1$ and canceling terms that appear on both sides, we obtain

$$\sum_{j=0}^{n-1} \overline{H}(S_{j:j+1}) \ge \overline{H}(S_{0:n}) + (n-1)(2n+1)\,. \tag{3.7}$$

Summing up (3.5) for $i = 0, \ldots, n-1$ and combining with (3.7) we obtain

$$nH(y) + n(2n-2) \ge \overline{H}(S_{0:n}) + (n-1)(2n+1)\,. \tag{3.8}$$

Now, observe that

$$\overline{H}(S_{0:n}) + n - 1 \ge \overline{H}(x_0, x_n) \ge \overline{H}(x_n) \ge \overline{H}(\emptyset) = 2n + 1\,. \tag{3.9}$$

Summing (3.8) and (3.9), we obtain

$$nH(y) + 2n^2 - n - 1 \ge 2n^2 + n$$

and rearranging we get $H(y) \ge 2 + n^{-1}$, from which it follows that $\beta(G) \ge 2 + n^{-1}$.

(iii)⇒(ii): Define a binary relation $\sharp$ on the vertex set V by specifying that $v \sharp w$ if there exists an edge j such that $\{v, w\} \subseteq T(j)$. Let $\sim$ denote the transitive closure of $\sharp$. Define F to be the quotient map from V to the set S of equivalence classes of $\sim$. We need to check that F is G-compatible. For every edge $j \in E$, the definition of relation $\sharp$ trivially implies that F maps all of $T(j)$ to a single element of S. The fact that it maps $f(j)$ to a *different* element of S is a consequence of the non-existence of almost alternating cycles. A relation $f(j) \sim v$ for some $v \in T(j)$ would imply the existence of a sequence $v_0, \ldots, v_n$ such that $v_0 = f(j), v_n = v$, and $v_i \sharp v_{i+1}$ for $i = 0, ..., n-1$. If we choose j_i for $0 \le i < n$ to be an edge such that $T(j_i)$ contains v_i, v_{i+1} (such an edge exists because $v_i \sharp v_{i+1}$) and we set $j_n = j$ and $v_{i-n} = f(j_i)$ for $i = 0, \ldots, n-1$, then the vertex sequence $v_{-n}, \ldots, v_n$ and edge sequence $j_0, \ldots, j_n$ constitute an almost alternating cycle in G.

Computing the relation $\sim$ and the function F, as well as testing that F is G-compatible, can easily be done in polynomial time, implying the final sentence of the theorem statement.

(ii)⇒(i): If $F : V \to S$ is G-compatible, we may compose F with a one-to-one mapping from S into a finite field $\mathbb{F}$, to obtain a function $\phi : V \to \mathbb{F}$ that is G-compatible. The public channel broadcasts two elements of $\mathbb{F}$, namely:

$$y = \sum_v x_v$$
$$z = \sum_v \phi(v) x_v$$

Receiver R_j now decodes message $x(j)$ as follows. Let c denote the unique element of $\mathbb{F}$ such that $\phi(v) = c$ for every v in $T(j)$. Using the pair (y, z) from the public channel, R_j can form the linear combination

$$cy - z = \sum_v [c - \phi(v)] x_v .$$

We know that every $v \in T(j)$ appears with coefficient zero in this sum. For every $v \in N(j)$, receiver R_j knows the value of x_v and can consequently subtract off the term $[c - \phi(v)]x_v$ from the sum. The only remaining term is $[c - \phi(x(j))]x(j)$. The coefficient $c - \phi(x(j))$ is nonzero, because ϕ is G-compatible. Therefore R_j can decode $x(j)$. ■

4 The gap between the broadcast rate and clique cover numbers

4.1 Separating the broadcast rate from the extreme LP solution b_n

In this section we prove Theorem 3 that shows a strong form of separation between β and its upper bound $b_n = \overline{\chi}_f$. Not only can we have a family of graphs where $\beta = O(1)$ while $\overline{\chi}_f$ is unbounded, but one can construct such a family where $\overline{\chi}_f$ grows polynomially fast with n.

Proof of Theorem 3. The following family of graphs (up to a small modification) was introduced by Erdős and Rényi in [12]. Due to its close connection to the (Sylvester-)Hadamard matrices when the chosen field has characteristic 2 we refer to it as the *projective-Hadamard* graph $H(\mathbb{F}_q)$:

1. Vertices are the non-self-orthogonal vectors in the 2-dimensional projective space over $\mathbb{F}_q$.
2. Two vertices are adjacent iff their corresponding vectors are non-orthogonal.

Let q be a prime-power. We claim that the projective-Hadamard graph $H(\mathbb{F}_q)$ on $n = n(q)$ vertices satisfies $\beta = 3$ while $\overline{\chi}_f = \Theta(n^{1/4})$. The latter is a well-known fact which appears for instance in [4, 20]. Showing that $\overline{\chi}_f \geq (1 - o(1))n^{1/4}$ is straightforward and we include an argument establishing this for completeness.

The fact that $\beta \geq 3$ follows from the fact that the standard basis vectors form an independent set of size 3. A matching upper bound will follow from the $\mathrm{minrk}_{\mathbb{F}}$ parameter defined in Section 2.4: Let $\mathbb{F}$ be some finite field and let $\ell = \mathrm{minrk}_{\mathbb{F}}(G)$ be the length of the optimal linear encoding over $\mathbb{F}$ for the Index Coding problem of a graph G with messages taking values in $\mathbb{F}$. Broadcasting $\ell \lceil \log_2 |\mathbb{F}| \rceil$ bits allows each receiver to recover his required message in $\mathbb{F}$ and so clearly $\beta \leq \ell$. It thus follows that $\lceil \beta(G) \rceil \leq \mathrm{minrk}_{\mathbb{F}}(G)$ for any graph G and finite field $\mathbb{F}$.

Here, dealing with the projective-Hadamard graph H, let B be the Gram matrix over $\mathbb{F}_q$ of the vectors corresponding to the vertices of H. By definition the diagonal entries are nonzero and whenever two vertices u, v are nonadjacent we have $B_{uv} = 0$. In particular B is a representation for H over $\mathbb{F}_q$ which clearly has rank 3 as the standard basis vectors span its entire row space. Altogether we deduce that $\beta(H) = 3$ whereas $\overline{\chi}_f = \Theta(n^{1/4})$, as required.

The fact that $\overline{\chi}_f \geq (1-o(1))n^{1/4}$ will follow from a straightforward calculation showing that the clique-number of H is at most $(1+o(1))q^{3/2} = (1+o(1))n^{3/4}$.

Consider the following multi-graph G which consists of the entire projective space:

1. Vertices are all vectors of the 2-dimensional projective space over $\mathbb{F}_q$.
2. Two (possibly equal) vertices are adjacent iff their corresponding vectors are orthogonal.

Clearly, G contains the complement of the Hadamard graph $H(\mathbb{F}_q)$ as an induced subgraph and it suffices to show that $\alpha(G) \leq (1+o(1))q^{3/2}$.

It is well-known (and easy) that G has $N = q^2+q+1$ vertices and that every vertex of G is adjacent to precisely $q+1$ others. Further observe that for any $u,v \in V(G)$ precisely one vertex of G belongs to $\{u,v\}^\perp$ (as u,v are linearly independent vectors). In other words, the codegree of any two vertices in G is 1. We conclude that G is a strongly-regular graph (see e.g. [14] for more details on this special class of graphs) with codegree parameters $\mu = \nu = 1$ (where μ is the codegree of adjacent pairs and ν is the codegree of non-adjacent ones). There are thus precisely 2 nontrivial eigenvalues of G given by $\frac{1}{2}((\mu-\nu) \pm \sqrt{(\mu-\nu)^2+4(q+1-\nu)}) = \pm\sqrt{q}$, and in particular the smallest eigenvalue is $\lambda_N = -\sqrt{q}$. Hoffman's eigenvalue bound (stating that $\alpha \leq \frac{-m\lambda_m}{\lambda_1-\lambda_m}$ for any regular m-vertex graph with largest and smallest eigenvalues λ_1, λ_m resp., see e.g. [14]) now shows

$$\alpha(G) \leq \frac{-N\lambda_N}{(q+1)-\lambda_N} = \frac{(q^2+q+1)\sqrt{q}}{q-\sqrt{q}+1} = q^{3/2}+q+\sqrt{q}\,,$$

as required. ∎

In addition to demonstrating a large gap between $\overline{\chi}_f$ and β on the projective-Hadamard graphs, we show that even in the extreme cases where G is a triangle-free graph on n vertices, in which case $\overline{\chi}_f(G) \geq n/2$, one can construct Index Coding schemes that significantly outperform $\overline{\chi}_f$. We prove this in Section 4.2 by providing a family of triangle-free graphs on n vertices where $\beta \leq \frac{3}{8}n$.

4.2 Broadcast rates for triangle-free graphs

In this section we study the behavior of the broadcast rate for triangle-free graphs, where the upper bound b_n on β is at least $n/2$. The first question in this respect is whether possibly $\beta = b_n$ in this regime, i.e. for such sparse graphs one cannot improve upon the fractional clique-cover approach for broadcasting. This is answered by the following result.

Theorem 4.1. *There exists an explicit family of triangle-free graphs on n vertices where $\overline{\chi}_f \geq n/2$ whereas the broadcast rate satisfies $\beta \leq \frac{3}{8}n$.*

The following lemma will be the main ingredient in the construction:

Lemma 4.2. *For arbitrarily large integers n there exists a family $\mathcal{F}$ of subsets of $[n]$ whose size is at least $8n/3$ and has the following two properties: (i) Every $A \in \mathcal{F}$ has an odd cardinality. (ii) There are no distinct $A,B,C \in \mathcal{F}$ that have pairwise odd cardinalities of intersections.*

Remark 4.3. For n even, a simple family $\mathcal{F}$ of size $2n$ with the above properties is obtained by taking all the singletons and all their complements. However, for our application here it is crucial to obtain a family $\mathcal{F}$ of size strictly larger than $2n$.

Remark 4.4. The above lemma may be viewed as a higher-dimensional analogue of the Odd-Town theorem: If we consider a graph on the odd subsets with edges between those with an odd cardinality of intersection, the original theorem looks for a maximum independent set while the lemma above looks for a maximum triangle-free graph.

Proof of lemma. It suffices to prove the lemma for $n = 6$ by super-additivity (we can partition a ground-set $[N]$ with $N = 6m$ into disjoint 6-tuples and from each take the original family $\mathcal{F}$).

Let $U_1 = \big\{\{x\} : x \in [5]\big\}$ be all singletons except the last, and $U_2 = \big\{A\cup\{6\} : A \subset [5]\,,\, |A| = 2\big\}$. Clearly all subsets given here are odd.

We first claim that there are no triangles on the graph induced on U_2. Indeed, since all subsets there contain the element 6, two vertices in U_2 are adjacent iff their corresponding 2-element subsets A, A' are disjoint, and there cannot be 3 disjoint 2-element subsets of $[5]$.

The vertices of U_1 form an independent set in the graph, hence the only remaining option for a triangle in the induced subgraph on $U_1 \cup U_2$ is of the form $\{x\}, (A \cup \{6\}), (A' \cup \{6\})$. However, to support edges from $\{x\}$ to the two sets in U_2 we must have that x belongs to both sets, and since $x \neq 6$ by definition we must have $x \in A \cap A'$. However, we must also have $A \cap A' = \emptyset$ for the two vertices in U_2 to be adjacent, contradiction.

To conclude the proof observe that adding the extra set $[5]$ does not introduce any triangles, since U_1 is an independent set while $[5]$ is not adjacent to any vertex in U_2 (its intersection with any set $(A \cup \{6\}) \in U_2$ contains precisely 2 elements). Altogether we have $|\mathcal{F}| = 5 + \binom{5}{2} + 1 = \frac{8}{3}n$. ■

Proof of Theorem 4.1. Let $\mathcal{F}$ be the family provided by the above lemma and consider the graph G whose N vertices are the elements of $\mathcal{F}$ with edges between A, B whose cardinality of intersection is odd. By definition the graph G is triangle-free and we have $\overline{\chi}_f(G) \geq N/2$.

Next, consider the binary matrix M indexed by the vertices of G where $M_{A,B} = |A \cap B| \pmod 2$. All the diagonal entries of M equal 1 by the fact that $\mathcal{F}$ is comprised of odd subsets only, and clearly M is a representation of G over $GF(2)$. At the same time, M can be written as FF^{T} where F is the $N \times n$ incidence-matrix of the ground-set $[n]$ and subsets of $\mathcal{F}$. In particular we have that $\mathrm{rank}(M) \leq \mathrm{rank}(F) \leq n$ over $GF(2)$. This implies that $\mathrm{minrk}_2(G) \leq n$ and the proof is now concluded by the fact that $\beta(G) \leq \mathrm{minrk}_2(G)$. ■

Remark 4.5. The construction of the family of subsets $\mathcal{F}$ in Lemma 4.2 relied on a triangle-free 15-vertex base graph H which is equivalent to the Peterson graph with 5 extra vertices added to it, each one adjacent to one of the independent sets of size 4 in the Peterson graph.

Having discussed the relation between β and b_n for sparse graphs we now turn our attention to the analogous question for the other extreme end, namely whether $\beta = b_1$ when $b_1 = \alpha$ attains its smallest possible value (other than in the complete graph) of 2.

4.3 Graphs with a broadcast rate of nearly 2

We now return to the setting of undirected graphs, where the class of $\{G : \beta(G) = 2\}$ is simply the complements of nonempty bipartite graphs, where in particular Index Coding is trivial. It turns out that extending this class to $\{G : \beta(G) < 2 + \varepsilon\}$ for any fixed small $\varepsilon > 0$ already turns this family of graphs to a much richer one, as the following simple corollary of Theorem 1 shows. Recall that the Kneser graph with parameters (n, k) is the graph whose vertices are all the k-element subsets of $[n]$ where two vertices are adjacent iff their two corresponding subsets are disjoint.

Corollary 4.6. *Fix $0 < \varepsilon < \frac{1}{2}$ and let G be the complement of the Kneser(n, k) graph on $N = \binom{n}{k}$ vertices for $n = (2 + \varepsilon)k$. Then $\beta(G) \leq 2 + \varepsilon$ whereas $\overline{\chi}(G) \geq (\varepsilon/2) \log N$.*

Proof. Using topological methods, Lovász [18] proved that the Kneser graph with parameters (n, k) has chromatic number $n - 2k + 2$, in our case giving that $\overline{\chi}(G) = \varepsilon k + 2 \leq (\varepsilon/2) \log N$ (with the last inequality due to the fact that $N \geq [\mathrm{e}(2+\varepsilon)]^k$ and so $k \geq \frac{1}{2} \log N$). At the same time, it is well known that G satisfies $\overline{\chi}_f = n/k$ (its maximum clique corresponds to a maximum set of intersecting k-subsets, which has size $\omega = \binom{n-1}{k-1}$ by the Erdős-Ko-Rado Theorem, and being vertex-transitive it satisfies $\overline{\chi}_f = N/\omega$). The bound $\beta \leq b_n = \overline{\chi}_f$ given in Theorem 1 thus completes the proof. ■

5 Establishing the exact broadcast rate for families of graphs

5.1 The broadcast rate of cycles and their complements

The following theorem establishes the value of β for cycles and their complements via the LP framework of Theorem 1.

Theorem 5.1. *For any integer $n \geq 4$ the n-cycle satisfies $\beta(C_n) = n/2$ whereas its complement satisfies $\beta(\overline{C_n}) = n/\lfloor n/2 \rfloor$. In both cases $\beta_1 = \lceil \beta \rceil$ while $\alpha = \lfloor \beta \rfloor$.*

Proof. As the case of n even is trivial with all the inequalities in (1.2) collapsing into an equality (which is the case for any perfect graph), assume henceforth that n is odd. We first show that $\beta(C_n) = n/2$. Putting $n = 2k + 1$ for $k \geq 2$, we aim to prove that $b_2 \geq k + 1/2$, which according to Theorem 1 will imply the required result since clearly $\overline{\chi}_f = k + 1/2$.

Denote the vertices V of the cycle by $0, 1, \ldots, 2k$. Further define:

$$
\begin{aligned}
E &= \{i \ : \ i \equiv 0 \bmod 2 \ , \ i \neq 2k\} && \text{(Evens)}\,, \\
O &= \{i \ : \ i \equiv 1 \bmod 2\} && \text{(Odds)}\,, \\
E^+ &= \{i \ : \ i \leq 2k - 2\} && \text{(Evens decoded)}\,, \\
O^+ &= \{i \ : \ 1 \leq i \leq 2k - 1\} && \text{(Odds decoded)}\,, \\
M &= \{i \ : \ 1 \leq i \leq 2k - 2\} && \text{(Middle)}\,.
\end{aligned}
$$

Next, consider the following constraints in the LP $\mathcal{B}_2$:

$$
\begin{aligned}
X(\emptyset) + k &\geq X(E) && \text{(slope)} \\
X(\emptyset) + k &\geq X(O) && \text{(slope)} \\
X(\emptyset) + 1 &\geq X(\{2k\}) && \text{(slope)} \\
X(E) &\geq X(E^+) && \text{(decode)} \\
X(O) &\geq X(O^+) && \text{(decode)} \\
X(E^+) + X(O^+) &\geq X(V) + X(M) && \text{(submod , decode)} \\
X(M) + X(\{2k\}) &\geq X(V) + X(\emptyset) && \text{(submod , decode)} \\
2X(V) &\geq 2(2k + 1) && \text{(initialize)}\,.
\end{aligned}
$$

Summing and canceling we get $2X(\emptyset) + 2k + 1 \geq 4k + 2$, implying $X(\emptyset) \geq k + 1/2$. The main idea of this proof, as with the ones to follow, is that we input some sets of vertices and then apply decoding to the sets as well as combine them together using submodularity to eventually output $X(V)$ and $X(\emptyset)$.

It remains to treat complements of odd cycles. Let $H = \overline{\mathsf{AAC}_n}$ be the complement of a directed odd almost-alternating cycle on n vertices (as defined in Section 3.3). Treating $\overline{C_n}$ as a directed graph (replacing each edge with a bi-directed pair of edges) it is clearly a spanning subgraph of H, hence $\beta(\overline{C}_n)$ is at least as large as $\beta(H)$. The proof in Section 3.3 establishes that $\beta(H) \geq \frac{n}{\lfloor n/2 \rfloor}$, translating to a lower bound on $\beta(\overline{C_n})$. The matching upper bound follows from the fact that due to Theorem 1 we have $\beta(\overline{C_n}) \leq \overline{\chi}(\overline{C_n}) = \frac{n}{\lfloor n/2 \rfloor}$. ■

5.2 The broadcast rate of cyclic Cayley Graphs

In this section we demonstrate how the same framework of the proof of Theorem 5.1 may be applied with a considerably more involved sequence of entropy-inequalities to establish the broadcast rate of two classes of Cayley graphs of the cyclic group $\mathbb{Z}_n$. Recall that a *cyclic Cayley graph* on n vertices with a set of generators $G \subseteq \{1, 2, \dots, \lfloor n/2 \rfloor\}$ is the graph on the vertex set $\{0, 1, 2, \dots, n-1\}$ where (i, j) is an edge iff $j - i \equiv g \pmod{n}$ for some $g \in G$.

Theorem 5.2. *For any $n \geq 4$, the 3-regular Cayley graph of $\mathbb{Z}_n$ has broadcast rate $\beta = n/2$.*

Theorem 5.3 (Circulant graphs)**.** *For any integers $n \geq 4$ and $k < \frac{n-1}{2}$, the Cayley graph of $\mathbb{Z}_n$ with generators $\{\pm 1, \dots, \pm k\}$ has broadcast rate $\beta = n/(k+1)$.*

To simplify the exposition of the proofs of these theorems we make use of the following definition.

Definition 5.4. A *slice* of size i in $\mathbb{Z}_n$ indexed by x is the subset of i contiguous vertices on the cycle given by $\{x + j \pmod{n} : 0 \leq j < i\}$.

Proof of Theorem 5.2. It is not hard to see that for a cyclic Cayley graph to be 3-regular it must have two generators, 1 and $n/2$, and n must be even. If n is not divisible by four, then it is easy to check that there is an independent set of size $n/2$ and $\overline{\chi}_f$ is also $n/2$. Thus, it immediately follows that $\beta = n/2$. For 3-regular cyclic Cayley graphs where n is divisible by four, α is strictly less than $n/2$. So to prove that $\beta = n/2$ we use the LP $\mathcal{B}_2$ to show $b_2 \geq n/2$, implying $\beta \geq n/2$.

Let $0, 1, 2, \dots, 4k-1$ be the vertex set of the graph. We assume that any solution X has cyclic symmetry. That is, $X(S) = X(\{s + i | s \in S\})$ for all $i \in [0, 4k-1]$. This assumption is without loss of generality because we can take any LP solution X and find a new one X' that is symmetric and has the same value by setting $X'(S) = \frac{1}{4k} \sum_{i=0}^{4k-1} X(\{s + i | s \in S\})$. All the constraints are feasible for X' because each is simply the average of $4k$ feasible constraints.

In our proof we will be using the following subsets of vertices:

$$\begin{aligned} [i] &= \{0, 1, 2, \dots, i-1\} \ \text{(a slice of size } i) \\ D &= \{0, 2, \dots, 2k-4, 2k-2, 2k+1, 2k+3, \dots, 4k-5, 4k-3\} \\ D^+ &= \{0, 1, 2, \dots, 2k-4, 2k-3, 2k-2, 2k+1, 2k+2, 2k+3, \dots, 4k-4, 4k-3\}. \end{aligned}$$

Observe from Figure 2 that $D \rightsquigarrow D^+$. Also note that D^+ is missing only four vertices, two on each side almost directly across from each other, and $|D| = 2k-1$.

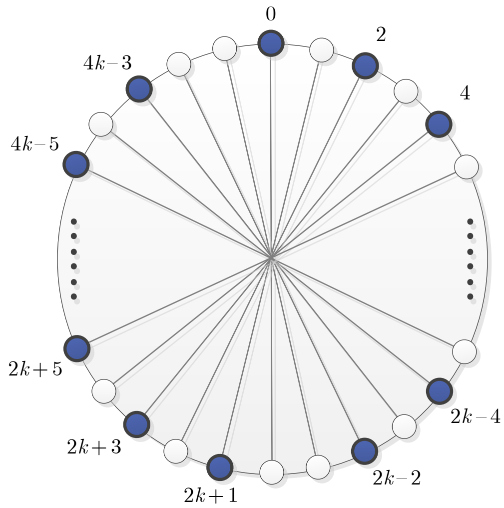

Figure 2: A 3-regular cyclic Cayley graph on $4k$ vertices. Highlighted vertices mark the set D used in the proof of Theorem 5.2.

Similar to our proof for the 5-cycle, we will prove $b_2 \geq n/2$ by listing a sequence of constraints in the LP $\mathcal{B}_2$ that sum and cancel to give us $X(\emptyset) \geq n/2$. However, this proof differs from the 5-cycle proof because we list inequalities implied not only by the constraints in our LP but also our assumption of cyclic symmetry. The fact that any two slices of size i have the same X value is used heavily in the sequence of inequalities that make up our proof.
First, we create $2k-1$ $X(D^+)$ terms on the right-hand-side:

$$(2k-2) + X(\emptyset) \geq X(D \setminus \{0\}) \qquad \text{(slope)}$$
$$X([1]) + X(D \setminus \{0\}) \geq X(D^+) + X(\emptyset) \qquad \text{(submod , decode)}$$
$$(2k-2)((2k-1) + X(\emptyset) \geq X(D^+)) \qquad \text{(slope , decode)}$$

Now, we apply submodularity to slices of size $i = 2 \ldots 2k$ and an $X(D^+)$ term — canceling all the $X(D^+)$ terms we created on the right-hand-side in the previous step. We pick our slices so that the union is a slice missing only two vertices, and the intersection is a slice of size $i-1$.

$$X(D^+) + X([2k]) \geq X([4k-2]) + X([2k-1])$$
$$X(D^+) + X([2k-1]) \geq X([4k-2]) + X([2k-2])$$
$$\vdots$$
$$X(D^+) + X([2]) \geq X([4k-2]) + X([1])$$

If we sum and cancel the inequalities listed so far we have:

$$2k(2k-2) + (2k-2)X(\emptyset) + X([2k]) \geq (2k-1)X([4k-2])$$

Now, we combine all $2k-1$ of the $X([4k-2])$ terms to get full cycles.

$$\begin{aligned} 2X([4k-2]) &\geq X(V) + X([4k-3]) \\ X([4k-3]) + X([4k-2]) &\geq X(V) + X([4k-4]) \\ X([4k-4]) + X([4k-2]) &\geq X(V) + X([4k-5]) \\ &\vdots \\ X([2k+1]) + X(H[4k-2]) &\geq H(V) + H([2k]) \end{aligned}$$

Now, we are left with:

$$2k(2k-2) + (2k-2)X(\emptyset) \geq (2k-2)X(V)$$

We can apply the constraint $X(V) \geq n$, yielding:

$$2k(2k-2) + (2k-2)X(\emptyset) \geq (2k-2)4k$$

thus $X(\emptyset) \geq 2k$ for any feasible solution, implying $b_2 \geq 2k = n/2$. ∎

Proof of Theorem 5.3. It is easy to check that $\overline{\chi}_f$ for these graphs is $n/(k+1)$, so it is sufficient to prove that $b_2 \geq n/(k+1)$. As we did in the proof of Theorem 5.2 we will assume that our solution X has cyclic symmetry. Suppose that $n \mod (k+1) \equiv j$. Now, consider dividing the cycle into sections of size $k+1$ and let S be the set of vertices consisting of the first k in each complete section ($|S| = k(n-j)/(k+1)$). Then by decoding $X(S) = X([-j])$ where $[-j]$ is a slice of size $n-j$. We will also use $[j]$ to denote a set of size j, as in the proof of Theorem 5.2. Observe that if $j = 0$ then this observation alone completes the proof.

Lemma 5.5. $(k+1)X[-j] + X[k] \geq (k+1)[-j-1] + X(\emptyset)$

Proof. The following inequalities are true by submodularity and the cyclic symmetry of X. In each inequality we apply submodularity to two slices, say of size s and t, $s \leq t$, overlapping such that their intersection is a slice of size $s-1$ and their union a slice of size $t+1$.

$$\begin{aligned} X([-j]) + X([-j]) &\geq X([-j+1]) + X([-j-1]) \\ X([-j]) + X([-j-1]) &\geq X([-j+1]) + X([-j-2]) \\ X([-j]) + X([-j-2]) &\geq X([-j+1]) + X([-j-3]) \\ &\vdots \\ X([-j]) + X([-j-(k-1)]) &\geq X([-j+1]) + X([-j-k]) \\ X([-j-k]) + X([k]) &\geq X(\emptyset) + X([-j+1]) \qquad \text{(submod , decode).} \end{aligned}$$

Adding up all of these inequalities gives us the desired inequality. ∎

Now, if we sum together the following string of inequalities we get the bound we want on $X(\emptyset)$.

Essentially, we iteratively apply our Lemma to get us to the trivial $j = 0$ case.

$$\begin{aligned}
k(n-j) + (k+1)X(\emptyset) &\geq (k+1)X([-j]) && \text{(slope , decode)}\\
jk + jX(\emptyset) &\geq jX([k]) && \text{(slope)}\\
(k+1)X([-j]) + X([k]) &\geq (k+1)X([-j-1]) + X(\emptyset) && \text{(by Lemma 5.5)}\\
(k+1)X([-j-1]) + X([k]) &\geq (k+1)X([-j-2]) + X(\emptyset) && \text{(by Lemma 5.5)}\\
&\vdots\\
(k+1)X([-1]) + X([k]) &\geq (k+1)X(V) + X(\emptyset) && \text{(by Lemma 5.5)}\\
(k+1)X(V) &\geq (k+1)n\,.
\end{aligned}$$

This completes the proof. ∎

5.3 The broadcast rate of specific small graphs

For any specific graph one can attempt to solve the second level of the LP-hierarchy directly to yield a possibly tight lower bound $\beta \geq b_2$. The following corollary lists a few examples obtained using an AMPL/CPLEX solver.

Fact 5.6. *The following graphs satisfy $b_2 = \beta = \overline{\chi}_f$:*

(1) Petersen graph (Kneser graph on $\binom{5}{2}$ vertices): $n = 10$, $\alpha = 4$ and $\beta = 5$.
(2) Grötzsch graph (smallest triangle-free graph with $\chi = 4$): $n = 11$, $\alpha = 5$ and $\beta = \frac{11}{2}$.
(3) Chvatal graph (smallest triangle-free 4-regular graph with $\chi = 4$): $n = 12$, $\alpha = 4$ and $\beta = 6$.

6 Coverage functions: a proof of Lemma 2.10

Lemma 2.10 (§ 2.2) will readily follow from establishing the following Lemmas 6.1 and 6.2, as it is easy to verify that the slope constraints and the i-th order submodularity constraints in our LP are equivalent to the inequalities in Eq. (6.1).

Lemma 6.1. *A vector X, indexed over all subsets of the groundset V, satisfies*

$$\forall R \neq \emptyset, \forall Z \cap R = \emptyset, \quad \sum_{T \subseteq R} (-1)^{|R \setminus T|} X(T \cup Z) \leq \begin{cases} 1 & \textit{if } |R| = 1 \\ 0 & \textit{otherwise} \end{cases} \tag{6.1}$$

if and only if it satisfies:

$$\forall R \subseteq V, R \neq \emptyset, \quad \sum_{T \subseteq R} (-1)^{|T|} (X(R \setminus T) - |R \setminus T|) \leq 0\,. \tag{6.2}$$

Lemma 6.2. *A vector X, indexed over all subsets of the ground-set V, satisfies (6.2) if and only if there exists a vector of non-negative numbers $w(T)$, defined for every non-empty vertex set T, such that $X(S) = |S| + \sum_{T : T \not\subseteq S} w(T)$ $\forall S \subseteq V$.*

Proof of Lemma 6.1. First, we claim that X satisfies (6.2) if and only if it satisfies:

$$\forall R \subseteq V, R \neq \emptyset, \quad \sum_{T \subseteq R} (-1)^{|R \setminus T|} X(T) \leq \begin{cases} 1 & \text{if } |R| = 1\,, \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

Starting with the inequalities (6.2), observe that we get an equivalent set of inequalities when we switch the roles of T and $R \backslash T$, as it is essentially summing over the complements of T instead of T. Additionally, for $|R| \geq 2$ we can remove the constant term because it is equal to the alternating sum $\pm \sum_{i=1}^{k} (-1)^k \binom{|R|}{k} k = 0$. If $|R| = 1$ then the constant term is one.

Now, we show the equivalence of (6.3) and (6.1). Clearly, if X satisfies (6.1) then it satisfies (6.3) because the inequalities in the latter are a subset of the inequalities in the former. Now, we show by induction on the size of Z that (6.3) implies (6.1). Our base case, $|Z| = 0$ holds trivially. We assume that (6.3) implies (6.1) for $|Z| < |Z^*|$ and show the following inequality holds:

$$\sum_{T \subseteq R^*} (-1)^{|R^* \backslash T|} X(T \cup Z^*) \leq \begin{cases} 1 & \text{if } |R| = 1 \\ 0 & \text{otherwise} \end{cases} \tag{$\star$}$$

By our inductive hypothesis, Eq. (6.3) implies the following two inequalities from (6.1):

$$R = R^* \cup \{z\}, \quad Z = Z^* \setminus \{z\}, \tag{I}$$

$$R = R^*, \qquad Z = Z^* \setminus \{z\} \tag{II}$$

for some $z \in Z^*$. It is easy to see that $(\star) - (\text{II}) = (\text{I})$, thus we can derive $(\star)$ from (I), (II). ■

Proof of Lemma 6.2. Suppose there exists a vector of non-negative numbers $w(T)$, defined for every non-empty vertex set T, such that $X(S) = |S| + \sum_{T: T \not\subseteq S} w(T)\ \forall S \subseteq V$ as in the statement of our Lemma. Then rearranging, we have:

$$X(\overline{S}) - |\overline{S}| = \sum_{T: T \not\subseteq \overline{S}} w(T) = \sum_{T: T \cap S \neq \emptyset} w(T) \ \ \forall \overline{S} \subseteq V$$

Now, define $F(S) = X(\overline{S}) - |\overline{S}|$.

Lemma 6.3. *The set function F satisfies*

$$\forall R \subseteq V, R \neq \emptyset, \ \sum_{T \subseteq R} (-1)^{|T|} F(\overline{R} \cup T) \leq 0. \tag{6.4}$$

if and only if there exists a vector of non-negative numbers $w(T)$, defined for every non-empty vertex set T, such that

$$F(S) = \sum_{T: T \cap S \neq \emptyset} w(T) \ \ \forall S \subseteq V. \tag{6.5}$$

Remark 6.4. A set function F is called a weighted set cover function if it can be written as in Eq. (6.5).

Plugging in $X(\overline{S}) - |\overline{S}|$ for $F(S)$ and noting that $\overline{\overline{R} \cup T} = R \backslash T$ it is easy to see that Lemma 6.3 implies our desired result. Thus, it remains to prove Lemma 6.3.

Proof of Lemma 6.3. In this proof we will be working with vectors and matrices whose rows and columns are indexed by subsets of V. Let $n = |V|, N = 2^n$. Expressing F and w as vectors with

$N-1$ components (ignoring the component corresponding to the empty set), this equation can be written in matrix form as

$$F = Mw,$$

where M is the $(N-1)$-by-$(N-1)$ matrix defined by

$$M_{TS} = \begin{cases} 1 & \text{if } T \cap S \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

We shall see below that M is invertible. It follows that F can be written as in Eq. (6.5) if and only if $M^{-1}F$ is a vector w with non-negative components.

To prove that M is invertible and to obtain a formula for the entries of the inverse matrix, let L be the N-by-N matrix defined by

$$L_{TS} = \begin{cases} 1 & \text{if } T \cap S \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

In other words, L is the matrix obtained from M by adding a top row and a left column consisting entirely of zeros. Let us define another matrix K by

$$K_{TS} = 1 - L_{TS} = \begin{cases} 1 & \text{if } T \cap S = \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

We can now begin to make progress on inverting these matrices, using the observation that both K and $K+L$ can be represented as Kronecker products of 2-by-2 matrices. Specifically,

$$K = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{\otimes n}, \quad K + L = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^{\otimes n}.$$

The inverse of $\left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ is $\left(\begin{smallmatrix} 0 & 1 \\ 1 & -1 \end{smallmatrix}\right)$. We may now make use of the fact that Kronecker products commute with matrix products, to deduce that

$$\begin{aligned} L \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes n} &= (K+L) \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes n} - K \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes n} \\ &= \left[\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \right]^{\otimes n} - \left[\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \right]^{\otimes n} \\ &= \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}^{\otimes n} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}^{\otimes n}. \end{aligned}$$

Examine the matrices occurring on the left and right sides of the equation above, and consider the submatrix obtained by deleting the left column and top row. On the right side, we obtain $-I$, where I denotes the $(N-1)$-by-$(N-1)$ identity matrix. On the left side we obtain $M \cdot A$, where A is the matrix obtained from $\left(\begin{smallmatrix} 0 & 1 \\ 1 & -1 \end{smallmatrix}\right)^{\otimes n}$ by deleting the left column and top row. This implies that M is invertible and its inverse is $-A$. Moreover, one can verify that the entries of $-A$ are given by

$$-A_{TS} = \begin{cases} 0 & \text{if } T \cup S \neq V \\ (-1)^{|T \cap S|} & \text{if } T \cup S = V. \end{cases}$$

Recall that a set function F can be expressed as it is in Eq. (6.5) if and only if $M^{-1}F$ has non-negative entries. Now that we have derived an expression for M^{-1} we find that this criterion is equivalent to stating that for all nonempty sets $R \subseteq V$,

$$\sum_{S\,:\,T\cup S=V} (-1)^{|T\cap S|} F(S) \leq 0.$$

This condition is equivalent to Eq. (6.2) because every set S such that $T \cup S = V$ can be uniquely written as the disjoint union of two sets $\overline{T}$ and $R = T \cap S$. This completes the proof of Lemma 6.3 and subsequently proves Lemmas 6.2 and 2.10. ■

7 Open problems

While our work sheds light on the relationship between the broadcast rate, β, the information-theoretic lower bound b_2, and other parameters of index coding problems — as well as on the computational complexity of computing or approximating β — it also leaves many appealing open questions on both of these topics. The following is a partial list of such questions.

- What is the largest possible (multiplicative) gap between β and the lower bound b_2? Section 2.2.2 gives an example where $\beta/b_2 \geq 45/44$, but we know of no examples with a greater multiplicative gap than this.
- Recalling that the linear program for b_2 contains exponentially many constraints, is there an efficient algorithm for computing b_2?
- Our results include a polynomial time algorithm for determining whether $\beta = 2$ for any broadcasting network. A major open problem is establishing the hardness of determining whether $\beta < C$ for a given graph G and real $C > 0$. While no such hardness result is known, presumably this problem is extremely difficult e.g. it is unclear whether it is even decidable.
- In an effort to approximate β, we give an efficient multiplicative $o(n)$-approximation algorithm for the general broadcasting problem. Can one obtain an approximation of β (even for case of undirected graphs) within a multiplicative constant of $n^{1-\varepsilon}$ for some fixed $\varepsilon > 0$?
- Using certain projective-Hadamard graphs introduced by Erdős and Rényi, we show that the broadcast rate can be uniformly bounded while its upper bound b_n is polynomially large. Is the scalar capacity β_1 of these graphs unbounded as the field characteristic q tends to ∞?

Acknowledgment. We thank Noga Alon for useful discussions.

References

[1] M. Adler, N. J. A. Harvey, K. Jain, R. Kleinberg, and A. R. Lehman, *On the capacity of information networks*, Proc. of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2006), pp. 241–250.

[2] N. Alon, A. Hassidim, E. Lubetzky, U. Stav, and A. Weinstein, *Broadcasting with side information*, Proc. of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008), pp. 823–832.

[3] N. Alon and N. Kahale, *Approximating the independence number via the θ-function*, Math. Programming **80** (1998), no. 3, Ser. A, 253–264.

[4] N. Alon and M. Krivelevich, *Constructive bounds for a Ramsey-type problem*, Graphs Combin. **13** (1997), no. 3, 217–225.

[5] 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.

[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** (2006), 2825–2830. An earlier version appeared in INFOCOM 1998.

[7] A. Blasiak, R. Kleinberg, and E. Lubetzky, *Lexicographic Products and the Power of Non-Linear Network Coding*, Proc. of the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2011). To appear.

[8] R. Boppana and M. M. Halldórsson, *Approximating maximum independent sets by excluding subgraphs*, BIT **32** (1992), no. 2, 180–196.

[9] M. A. R. Chaudhry and A. Sprintson, *Efficient algorithms for index coding*, IEEE Conference on Computer Communications Workshops (INFOCOM 2008), 2008, pp. 1–4.

[10] R. Dougherty, C. Freiling, and K. Zeger, *Insufficiency of linear coding in network information flow*, IEEE Trans. Inform. Theory **51** (2005), 2745–2759.

[11] R. Dougherty, C. Freiling, and K. Zeger, *Networks, matroids, and non-Shannon information inequalities*, IEEE Trans. Inform. Theory **53** (2007), no. 6, 1949–1969.

[12] P. Erdős and A. Rényi, *On a problem in the theory of graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **7** (1962), 623–641 (1963) (Hungarian, with Russian and English summaries).

[13] U. Feige, *Randomized graph products, chromatic numbers, and the Lovász ϑ-function*, Combinatorica **17** (1997), no. 1, 79–90. An earlier version appeared in Proc. of the 27th Annual ACM Symposium on Theory of computing (STOC 1995), pp. 635–640.

[14] C. Godsil and G. Royle, *Algebraic graph theory*, Graduate Texts in Mathematics, vol. 207, Springer-Verlag, New York, 2001.

[15] N. J. A. Harvey, R. Kleinberg, and A. R. Lehman, *On the capacity of information networks*, IEEE Trans. Inform. Theory **52** (2006), no. 6, 2345–2364.

[16] N. J. A. Harvey, R. Kleinberg, C. Nair, and Y. Wu, *A "Chicken & Egg" Network Coding Problem*, IEEE International Symposium on Information Theory (ISIT 2007), pp. 131–135.

[17] N. Linial and U. Vazirani, *Graph products and chromatic numbers*, Proc. of the 30th Annual IEEE Symposium on Foundations of Computer Science (FOCS 1989), pp. 124–128.

[18] L. Lovász, *Kneser's conjecture, chromatic number, and homotopy*, J. Combin. Theory Ser. A **25** (1978), 319–324.

[19] E. Lubetzky and U. Stav, *Non-linear index coding outperforming the linear optimum*, IEEE Trans. Inform. Theory **55** (2009), 3544–3551. An earlier version appeared in Proc. of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), pp. 161–167.

[20] D. Mubayi and J. Williford, *On the independence number of the Erdős-Rényi and projective norm graphs and a related hypergraph*, J. Graph Theory **56** (2007), no. 2, 113–127.

[21] S. El Rouayheb, A. Sprintson, and C. Georghiades, *A new construction method for networks from matroids*, IEEE International Symposium on Information Theory (ISIT 2009), pp. 2872–2876.

[22] S. El Rouayheb, A. Sprintson, and C. Georghiades, *On the relation between the Index Coding and the Network Coding problems*, IEEE International Symposium on Information Theory (ISIT 2008), pp. 1823–1827.

[23] L. Song, R. W. Yeung, and N. Cai, *Zero-error network coding for acyclic networks*, IEEE Trans. Inform. Theory **49** (2003), no. 12, 3129–3139.

[24] A. Wigderson, *Improving the performance guarantee for approximate graph coloring*, J. Assoc. Comput. Mach. **30** (1983), no. 4, 729–735.

[25] R. W. Yeung, *A first course in information theory*, Information Technology: Transmission, Processing and Storage, Kluwer Academic/Plenum Publishers, New York, 2002.

[26] Z. Zhang and R. W. Yeung, *On Characterization of Entropy Function via Information Inequalities*, IEEE Transactions on Information Theory **44** (1998), no. 4, 1440–1452.