Max-Flow Min-Cut Theorem for Rényi Entropy in Communication Networks

Maximilien Gadouleau and Søren Riis
School of Electronic Engineering and Computer Science
Queen Mary, University of London
Emails: (soren.riis,mgadouleau)@eecs.qmul.ac.uk

***Abstract*—A symbolic approach to communication networks, where the topology of the underlying network is contained in a set of formal terms, was recently introduced. Many communication problems can be recast as dispersion problems in this setup. The so-called min-cut of a term set represents its number of degrees of freedom. For any assignment of function symbols, its dispersion measures the amount of information sent to the destinations. It was proved that the maximum dispersion asymptotically reaches the min-cut of the term set. In this paper, we refine this result in two ways. First, we prove a max-flow min-cut theorem for the Rényi entropy with order less than one, given that the inputs are equiprobably distributed; conversely, there is no max-flow min-cut theorem for Rényi entropy with order greater than one. Second, although linear coding functions have the practical appeal of low complexity, we prove that they are insufficient in general to reach the min-cut. More specifically, there exist term sets which have an arbitrarily large dispersion for non-linear coding functions, yet limited dispersion when linear coding functions are considered. Conversely, we show that if there is a solution based on low degree polynomials, then there exists a linear solution.**

I. Introduction

A multi-user communication problem is given by a network with prescribed sets of sources and of destinations, where the destinations request messages sent by the sources. The general problem of determining whether all the demands of the destinations can be satisfied at the same time (i.e. whether the problem is solvable) is crucial, and as such has been widely studied [1], [2] via techniques from network coding [3], [4].

In [5], the communication problem was represented via a set of terms, which are concepts from *logic*. A term is built on variables representing the messages sent by the sources, and on function symbols representing coding functions at the intermediate nodes. A term thus formally represents all the possible operations undergone by messages from the sources to the destinations. This novel representation is actually more general than network coding, and hence offers not only a generalization of results in network coding, but also a reformulation in terms of flows.

A natural question is to determine the maximum amount of information that can be transmitted to the destinations. This problem can also be generalized for term sets. The *min-cut* of a term set an be viewed as the number of degrees of freedom of the set and hence represents the information bottlenecks on the network. One measure of information introduced in [5] is the dispersion, which is (the logarithm of) the number of possible outputs, once the coding functions have been assigned. It was proved in [5] that the maximum dispersion is asymptotically given by the min-cut of the term set.

In this paper, we thus introduce different measures of performance for the non-cooperative case where the inputs are uniformly distributed based on the Rényi entropy [6]. These measures characterize the amount of information that can be inferred about the input messages when those are uniformly distributed. We show that the dispersion is a special case of the Rényi entropy. We prove a variant of the max-flow min-cut theorem for the Rényi entropy with $0 \leq \alpha < 1$, thus strengthening the result for the dispersion. Conversely, the Rény entropy for $\alpha > 1$ does not necessarily reach the min-cut. Therefore, the Rényi entropy is sensitive to information bottlenecks that cannot be taken into account via the min-cut approach. However, the case of the Shannon entropy, where $\alpha = 1$, remains open.

In order to simplify the combinations operated at each intermediate node and the decoding at each destination, linear network coding only considers linear coding functions [7], [8]. However, it is known that linear network coding is not optimal in general [9], [10]. In this paper, we generalize the inefficiency of linear network coding to the case of communication networks based on terms sets. In particular, we design a family of term sets with arbitrarily large min-cut which can be achieved by using non-linear coding functions, whereas the maximum dispersion achieved by linear functions is only equal to 2. Conversely, we prove that if maximum dispersion equal to the min-cut value can be achieved using coding and decoding functions based on polynomials of fixed degree, then it can be achieved using linear functions only.

The rest of the paper is organized as follows. Section II reviews the concepts and the main results given in [5] on communication networks based on term sets. Section III then proves the max-flow min-cut theorem for the Rényi entropy for these networks. Section IV then investigates the dispersion of linear coding functions. Due to space restrictions, some claims are given without proofs, which can all be found in [5].

II. Communication networks based on term sets

A. Term sets

Let $X = \{x_1, x_2, \ldots, x_k\}$ be a set of *variables* and consider a set of *function symbols* $\{f_1, f_2, \ldots, f_l\}$ with respective arities (numbers of arguments) $d_1, d_2, \ldots, d_l$. A *term* is defined

to be an object obtained from applying function symbols to variables recursively.

We say that u is a subterm of t if the term u appears in the definition of t. Furthermore, u is a *direct subterm* of t if $t = f_j(v_1, \ldots, u, \ldots, v_{d_j})$, and f_j is referred to as the *principal function* of t.

We shall consider finite term sets, typically referred to as $\Gamma = \{t_1, t_2, \ldots, t_r\}$. We denote the set of variables that occur in terms in Γ as Γ_{var} and the collection of subterms of one or more terms in Γ as Γ_{sub}. Since each variable is a subterm and since a term is a subterm of itself we have the inclusions $\Gamma_{\text{var}} \subseteq \Gamma_{\text{sub}}$ and $\Gamma \subseteq \Gamma_{\text{sub}}$.

We now define a term-cut, which can be viewed as replacing some subterms in the definition of a term by variables.

Definition 1 (Term-cut): A set of subterms $s_1, s_2, \ldots, s_\rho \in \Gamma_{\text{sub}}$ provides a *term-cut* of size ρ for Γ if all the terms can be expressed syntactically by applying function symbols to $s_1, s_2, \ldots, s_\rho$.

A minimal term-cut for Γ is a term-cut with minimum size, referred to as the *min-cut* of Γ. The min-cut can hence be viewed as the number of degrees of freedom of the term set. Clearly the min-cut is no more than the number of variables k and the number of terms r.

These concepts can be graphically explained as follows.

Definition 2 (The graph G_Γ): For a given term set Γ, the directed graph $G_\Gamma = (V, E, S, T)$ is defined to have vertex set $V = \Gamma_{\text{sub}}$, edge set $E = \{(u, v) : u \,\text{directsubtermof}\, v\}$, source set $S = \Gamma_{\text{var}}$, and target set $T = \Gamma$.

A set of subterms is a term-cut for Γ if and only if it is a vertex cut that separates $S = \Gamma_{\text{var}}$ from $T = \Gamma$ in the directed graph G_Γ. Notice that $S \cap T$ is non-empty if Γ contains one or more terms that are variables.

B. Dispersion

So far, we have treated function symbols as abstract entities. We now study the case when these function symbols are assigned explicit values.

Definition 3 (Model): Let A be a finite set with $|A| \geq 2$, referred to as the alphabet. A *model* for Γ over A is an assignment of the function symbols $\psi = \{\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_l\}$, where $\bar{f}_i : A^{d_i} \to A$ for all $1 \leq i \leq l$.

Once all the function symbols f_i are assigned coding functions $\bar{f}_i$, then by composition each term $t_j \in \Gamma$ is assigned a function $\bar{t}_j : A^k \to A$. In order to simplify notations, we shall write functions by the way they map a tuple $\mathbf{a} = (a_1, a_2, \ldots, a_k) \in A^k$, and we typically write tuples in bold face. We shall abuse notations and also denote the *induced mapping* of the model as $\psi : A^k \to A^r$, defined as

$$\psi(\mathbf{a}) = \big(\bar{t}_1(\mathbf{a}), \bar{t}_2(\mathbf{a}), \ldots, \bar{t}_r(\mathbf{a})\big).$$

Note that the definition of the induced mapping depends on the ordering of terms in Γ. However, our performance measures for models and induced mappings will not depend on a particular ordering.

We are especially interested in how ψ disperses its outputs, and how much information about the inputs can be obtained from the outputs. For any $\mathbf{b} \in A^r$, we denote the pre-image of $\mathbf{b}$ as $\text{pre}(\mathbf{b}) = \{\mathbf{a} \in A^k : \psi(\mathbf{a}) = \mathbf{b}\}$. The image and the *dispersion* of ψ are respectively defined as

$$\begin{aligned} \text{image}(\psi) &:= \{\mathbf{b} \in A^r : |\text{pre}(\mathbf{b})| \geq 1\}, \\ \gamma(\psi) &:= \log_{|A|} |\text{image}(\psi)|. \end{aligned}$$

The dispersion is the analogue of the value of a flow for information transfer on networks based on term sets.

We finally define the dispersion of Γ over A as the maximal Γ-dispersion over all models for Γ over A, and we denote this value by $\gamma(\Gamma, |A|)$ as this quantity clearly depends on A via its cardinality only.

C. Max-flow min-cut theorem for dispersion of term sets

Theorem 1 (Max-flow min-cut theorem for dispersion): Let Γ be a term set with min-cut of ρ, then for any alphabet A, $\gamma(\Gamma, |A|) \leq \rho$. Conversely, $\lim_{|A| \to \infty} \gamma(\Gamma, |A|) = \rho$.

This theorem is proved using dynamic routing, which we review below. According to the directed graph version of Menger's theorem [11], there exists a family P of ρ vertex-disjoint directed paths from $S = \Gamma_{\text{var}}$ to $T = \Gamma$. First of all, let us assume that the term set is *diversified*, i.e. a distinct function symbol g_v is associated to each subterm $v \in \Gamma_{\text{sub}}$. In that case, *routing* can be used.

Definition 4 (Routing): If u_j is the direct subterm of v on the same path in P, then we let $\bar{g}_v(a_1, a_2, \ldots, a_d) = a_j$. Otherwise, i.e. if v does not belong to any path in P, then $\bar{g}_v(a_1, a_2, \ldots, a_d) = 1$.

Note that our definition of routing depends on the set of paths P, and hence is not unique. However, the dispersion of routing does not depend on the choice of P. Using routing, all points of the form $(a_1, a_2, \ldots, a_\rho, 1, \ldots, 1) \in A^k$ are mapped to $(a_1, a_2, \ldots, a_\rho, 1, \ldots, 1) \in A^r$, thus yielding a Γ-dispersion of ρ.

The construction of coding functions for routing assumed that each subterm v was assigned a distinct coding function. However, in general distinct subterms might be assigned the same function symbol (e.g., $f(x, y)$ and $f(y, x)$). In general, one can use *dynamic routing*, defined below.

For $|A| > |\Gamma_{\text{sub}}|$, there exist two sets B and R with $1 \leq |R| \leq |\Gamma_{\text{sub}}|$ such that $|A| = |(\Gamma_{\text{sub}} \times B) \cup R|$ and the union is disjoint. We shall abuse notation slightly and assume $A = (\Gamma_{\text{sub}} \times B) \cup R$. By construction, a tuple $\mathbf{a} = (a_1, a_2, \ldots, a_{d_j}) \in A^{d_j}$ either has an element in R or has each $a_i = (u_i, b_i) \in \Gamma_{\text{sub}} \times B$.

Definition 5 (Dynamic routing): For each subterm $v \in \Gamma_{\text{sub}}$, select a coding function $\bar{g}_v$ over B using routing, as in Definition 4. We define the functions $\bar{f}_j(a_1, a_2, \ldots, a_{d_j})$ over A as follows. If each a_i is of the form $a_i = (u_i, b_i) \in \Gamma_{\text{sub}} \times B$, let s denote the term $s = f_j(u_1, u_2, \ldots, u_{d_j})$; then if $s \in \Gamma_{\text{sub}}$

$$\bar{f}_j(a_1, a_2, \ldots, a_{d_j}) = (s, \bar{g}_s(b_1, b_2, \ldots, b_{d_j})) \in \Gamma_{\text{sub}} \times B.$$

Otherwise, let $\bar{f}_j(a_1, a_2, \ldots, a_{d_j}) = r$ for some $r \in R$.

Remark that the headers u_i of the inputs then indicate to the coding function $\bar{f}_j$ which subterm v it is located on, and

hence which function $\bar{g}_v$ to use. We say an input message a_j for the variable x_j is *correctly formatted* if it is of the form (x_j, b_j) where $b_j \in B$, and we denote the set of all corrected inputs as $I := \{\mathbf{a} \in (\Gamma_{\text{var}} \times B)^k : a_j = (x_j, b_j)\}$ and the set of correctly formatted outputs as $J := \{\mathbf{a} \in (\Gamma \times B)^r : a_j = (t_j, b_j)\}$. The idea behind dynamic routing is that if all inputs are correctly formatted (i.e. have the correct headers) then the coding functions $\bar{f}_j$ mimic the behavior of the routing functions $\bar{g}_v$. Thus, correctly formatted messages in I are mapped to correctly formatted outputs in J (as long as they are mapped by the functions $\bar{g}_v$), while other messages are mapped to an "error message" in R.

III. Max-flow min-cut theorem for Rényi entropy

Let Γ be a set of r terms built on k variables, and let ψ be a model for Γ over an alphabet A. Once A and ψ are fixed, the flow of data from the inputs $\mathbf{a} \in A^k$ to the outputs $\psi(\mathbf{a}) \in A^r$ can be viewed as the transmission of a random variable $\bar{\mathbf{a}}$ taking values in A^k through the channel operating the induced mapping ψ. Its capacity C_ψ is easily computed:

$$\begin{aligned} C_\psi &:= \sup_{\bar{\mathbf{a}}} I(\bar{\mathbf{a}}; \psi(\bar{\mathbf{a}})) \\ &= \sup_{\bar{\mathbf{a}}} H(\psi(\bar{\mathbf{a}})) = \gamma(\psi), \end{aligned} \tag{1}$$

where H denotes the Shannon entropy. The maximum is reached when $\bar{\mathbf{a}}$ has the following probability distribution: for each $\mathbf{b} \in \text{image}(\psi)$, select $a(\mathbf{b}) \in \text{pre}(\mathbf{b})$ and let $\mathbb{P}\{\bar{\mathbf{a}} = a(\mathbf{b})\} = |\text{image}(\psi)|^{-1}$. Eq. (1) shows that the capacity of the channel is given by the dispersion of the model considered. Thus, the max-flow min-cut theorem for the dispersion states that the channel capacity asymptotically converges to the min-cut ρ of the term set, i.e.

$$\sup_{A,\psi} C_\psi = \sup_{A,\psi} \gamma(\psi) = \rho.$$

We note that the capacity is achieved for a specific input random variable which is not uniformly distributed over all inputs. This represents the capacity achieved when the sources are *cooperative*, i.e. they agree on a coding scheme for the input. We are now interested in the case where the sources are non-cooperative, and we assume that the inputs are uniformly distributed over A^r.

This opens the question of the most accurate measure of performance for a term set. If the input $\bar{\mathbf{a}} \in A^k$ is uniformly distributed, then $\psi(\bar{\mathbf{a}})$ is a random variable with values in A^r, where for all $\mathbf{b} \in A^r$

$$p_{\mathbf{b}} = \mathbb{P}\{\psi(\bar{\mathbf{a}}) = \mathbf{b}\} = |A|^{-k} |\text{pre}(\mathbf{b})|.$$

The normalized Rényi Entropy over A of order α of the random variable $\psi(\bar{\mathbf{a}})$, which we will simply denote as $H_\alpha(\psi)$, is thus given for $0 < \alpha < 1$ or $1 < \alpha < \infty$ by [6]

$$\begin{aligned} H_\alpha(\psi) &:= \frac{1}{1-\alpha} \log_{|A|} \sum_{\mathbf{b} \in A^r} p_{\mathbf{b}}^\alpha \\ &= \frac{\alpha}{\alpha - 1} \left(k - \frac{1}{\alpha} \log_{|A|} \sum_{\mathbf{b} \in A^r} |\text{pre}(\mathbf{b})|^\alpha \right). \end{aligned}$$

Three other special cases need close attention.

First, when $\alpha = 0$, the Rényi Entropy is the logarithm of the cardinality of the number of outcomes, which is often referred to as the *Hartley entropy*. In our case, the Hartley entropy of ψ is equal to its dispersion, which is equal to the channel capacity:

$$H_0(\psi) := \log_{|A|} |\{\mathbf{b} \in A^r : p_{\mathbf{b}} > 0\}| = \gamma(\psi).$$

Second, we remark that $\log_{|A|} |\text{pre}(\mathbf{b})|$ is the uncertainty about the input when the message $\mathbf{b}$ is received. Hence the variable $k - \log_{|A|} |\text{pre}(\mathbf{b})|$ is the amount of information (counted in symbols in A) that can be inferred about $\mathbf{a}$ from $\mathbf{b} = \psi(\mathbf{a})$. The *Shannon entropy*, obtained when $\alpha = 1$, is therefore the expected amount of information inferred from the term set about the input messages:

$$H_1(\psi) := -\sum_{\mathbf{b} \in A^r} p_{\mathbf{b}} \log_{|A|} p_{\mathbf{b}} = \mathbb{E}\left\{ k - \log_{|A|} |\text{pre}(\mathbf{b})| \right\}.$$

Third, when $\alpha = \infty$, the *min-entropy* quantifies the amount of information that can be inferred from any output, by considering the point $\mathbf{b} \in A^r$ with the most pre-images:

$$H_\infty(\psi) := -\log_{|A|} \max_{\mathbf{b} \in A^r} p_{\mathbf{b}} = \min_{\mathbf{b} \in A^r} \left\{ k - \log_{|A|} |\text{pre}(\mathbf{b})| \right\}.$$

Note that there exist models ψ_1 and ψ_2 such that $H_0(\psi_1) > H_0(\psi_2)$ and yet $H_\alpha(\psi_1) < H_\alpha(\psi_2)$ for some $\alpha > 0$. Therefore, having the highest dispersion does not guarantee to perform well for the other measures.

The max-flow min-cut theorem for the dispersion indicates that the Rényi entropy for $\alpha = 0$ tends to the min-cut of the term set. Similarly to the dispersion, we denote the maximum Rényi entropy over all models for Γ over A as $H_\alpha(\Gamma, |A|)$. We shall prove the following result.

Theorem 2 (Max-flow min-cut theorem for Rényi entropy): Let Γ be a term set built on k variables and with min-cut of ρ, then $\lim_{|A| \to \infty} H_\alpha(\Gamma, |A|) = \rho$ for all $0 \leq \alpha < 1$. Conversely, for any $\alpha > 1$, there exists a term set Γ with min-cut ρ for which $\lim_{|A| \to \infty} H_\alpha(\Gamma, |A|) < \rho$.

Let us first prove the max-flow min-cut theorem for the Rényi entropy with $\alpha < 1$.

Proposition 1 (Max-flow min-cut theorem for $\alpha < 1$): Let Γ be a term set built on k variables and with min-cut of ρ and for all $0 < \alpha < 1$ let $\beta = \rho + \frac{\alpha}{1-\alpha} k$ and $n = (2|\Gamma_{\text{sub}}|)^{\frac{\epsilon}{\beta}}$. Then for any alphabet A with $|A| \geq n$, $H_\alpha(\Gamma, |A|) \geq \rho - \epsilon$, which is achieved by dynamic routing.

Proof: Suppose that dynamic routing is used over an alphabet A (recall the notations from Section II-C). It is clear that any output of the form $((t_1, b_1), (t_2, b_2), \ldots, (t_\rho, b_\rho), (t_{\rho+1}, 1), \ldots, (t_r, 1)) \in (\Gamma \times B)^r$ has exactly $|B|^{k-\rho}$ pre-images, namely those of the form $((x_1, b_1), (x_2, b_2), \ldots, (x_k, b_k)) \in (\Gamma_{\text{var}} \times B)^k$, where $b_{\rho+1}, b_{\rho+2}, \ldots, b_k \in B$. Let us denote this set of $|B|^\rho$

outputs as C and compute the Rényi entropy of routing.

$$\begin{aligned} H_\alpha(\psi) &\geq \frac{1}{1-\alpha}\log_{|A|}\sum_{\mathbf{b}\in C}|B|^{\alpha(k-\rho)} - k\frac{\alpha}{1-\alpha} \\ &= \rho - \beta\log_{|A|}\left\{|\Gamma_{\rm sub}| + \frac{|R|}{|B|}\right\} \\ &\geq \rho - \beta\log_{|A|}(2|\Gamma_{\rm sub}|) \geq \rho - \epsilon. \end{aligned}$$

■

The Rényi entropy for large α is sensitive to some types of bottlenecks which cannot be handled with the graphic approach of max-flow min-cut. Indeed, we design below a family of term sets for which the Rényi entropy does not tend to the min-cut for $\alpha > 1$ and arbitrarily close to 1. Therefore, there is no max-flow min-cut theorem for the Rényi entropy with $\alpha > 1$.

For all $k \geq 2$, we define the set Γ_k of k^2 terms, built on k^2 variables x_j^a, $0 \leq a, j \leq k-1$ and on one function symbol f of arity k, to be

$$\Gamma_k = \{t_{i,j} = f(x_i^0, x_j^1, x_j^2, \ldots, x_j^{k-1}) : 0 \leq i, j \leq k-1\}.$$

Proposition 2 (No max-flow min-cut theorem for $\alpha > 1$): The term set Γ_k has min-cut of k^2. However, for $\alpha > \frac{k}{k-1}$ and all A,

$$H_\alpha(\Gamma, |A|) \leq \frac{(2k-1)\alpha - k}{\alpha - 1} < k^2.$$

Proof: First, we prove that Γ_k has min-cut of k^2 by constructing k^2 vertex-disjoint paths from $\Gamma_{k,\rm var}$ to Γ_k. We have $x_j^0 \in t_{j,j}$ for all j and if $a \geq 1$, $x_j^a \in t_{b,j}$ for any $0 \leq b \leq k-1$. Therefore, $(x_j^0, t_{j,j})$ for all j and $(x_j^a, t_{a+j \mod k,j})$ for all $a \geq 1$ and j form a set of k^2 vertex-disjoint paths.

Second, we give an upper bound on the Rényi entropy of any model ψ for Γ_k over an alphabet A. Consider the set $B = \{\mathbf{a} \in A^{k^2} : a_0^0 = a_1^0 = \ldots = a_{k-1}^0 = a^0\}$, then if $\mathbf{a} \in B$, $\bar{t}_{i,j}(\mathbf{a}) = \bar{t}_{i',j}(\mathbf{a})$ for all i, i', and j and only k terms are non necessarily equal. Therefore, $|B| = |A|^{k^2-k+1}$ while the size of the image of B is at most $|A|^k$. We have

$$\begin{aligned} H_\alpha(\psi) &\leq k^2\frac{\alpha}{\alpha-1} - \frac{1}{\alpha-1}\log_{|A|}\sum_{\mathbf{b}\in\psi(B)}|\mathrm{pre}(\mathbf{b})|^\alpha \\ &\leq k^2\frac{\alpha}{\alpha-1} - \frac{1}{\alpha-1}\log_{|A|}(|A|^k|A|^{\alpha(k-1)^2}) \quad (2) \\ &= \frac{(2k-1)\alpha - k}{\alpha-1}, \end{aligned}$$

where (2) follows the fact that since $\alpha > 1$, the summation is minimized when all terms are equal. ■

The case of the Shannon entropy ($\alpha = 1$) remains open. This is an important question, as a max-flow min-cut theorem for the Shannon entropy would mean that the amount of information obtained in the non-cooperative case is asymptotically equal to that in the cooperative case. We would like to illustrate the difficulty of treating the Shannon entropy case. It can be shown that for any fixed n and for any $\epsilon > 0$, there exists $0 < \alpha < 1$ such that $H_1(X) \geq H_\alpha(X) - \epsilon$ for any probability distribution X on n points. However, we show below that ϵ cannot be chosen independently of n.

Example 1: For $0 < \alpha < 1$ and for $n \geq 2$ consider the probability distribution $X = X_{\alpha,n}$ given by:

$$p_1 = p_2 = \ldots = p_{n-1} = \frac{1-\alpha}{n-1}, \quad p_n = \alpha.$$

We obtain

$$\begin{aligned} H_1(X) &= -(1-\alpha)\log_n\left\{\frac{1-\alpha}{n-1}\right\} - \alpha\log_n\alpha, \\ H_\alpha(X) &= \frac{1}{(1-\alpha)}\log_n\left\{(n-1)\left(\frac{1-\alpha}{n-1}\right)^\alpha + \alpha^\alpha\right\}. \end{aligned}$$

It is it not hard to show that $\lim_{n\to\infty} H_1(X) = 1-\alpha$ while $\lim_{n\to\infty} H_\alpha(X) = 1$, and hence

$$\lim_{\alpha\to 1}\lim_{n\to\infty} H_1(X) = 0 \neq 1 = \lim_{\alpha\to 1}\lim_{n\to\infty} H_\alpha(X).$$

We finish this section by comparing our results to the max-flow min-cut for the one-to-one dispersion given in [5]. The *one-to-one dispersion* of a model $\gamma_{\rm one}(\psi)$ is the logarithm of the number of images with exactly one pre-image, i.e. it corresponds to the cases where all the information about the input can be inferred from the given output. In [5], it is shown that the one-to-one dispersion of a term set also tends to the min-cut. It can be easily shown that the one-to-one dispersion is no more than the dispersion; however, it is a measure independent from any other Rényi entropy: for any $0 < \alpha \leq \infty$, there exist models ψ_1 and ψ_2 for which $\gamma_{\rm one}(\psi_1) < H_\alpha(\psi_1)$ while $\gamma_{\rm one}(\psi_2) > H_\alpha(\psi_2)$. Therefore, the max-flow min-cut theorem for the one-to-one dispersion cannot be viewed as a special case of Theorem 2.

IV. Dispersion of linear coding functions

A. Linear versus non-linear coding functions

We now consider the important class of linear coding functions. First, scalar linear functions are defined when A is organized as a field $\mathbb{F}_q$ for some prime power q, and the coding functions $\bar{f} : \mathbb{F}_q^d \to \mathbb{F}_q$ are linear i.e. can be written as $\bar{f}(a_1, a_2, \ldots, a_d) = \sum_{i=1}^d a_i b_i$ for $b_1, b_2, \ldots, b_d \in \mathbb{F}_q$. A more general class are the matrix-linear coding functions that are defined when A is organized as a finite vector space V and the coding functions $\bar{f} : V^d \to V$ are linear i.e can be written as $\bar{f}(a_1, a_2, \ldots, a_d) = \sum_{i=1}^d F_i a_i$ where $F_1, F_2, \ldots, F_d$ are linear maps from V to V.

Clearly, if the coding functions are linear, then so is the induced mapping of the corresponding model, and hence the dispersion of a linear model is always an integer. In particular, if a term set Γ with min-cut ρ is not solvable, then the dispersion of any linear model for Γ is at most $\rho - 1$. However, by the max-flow min-cut theorem for the dispersion there exist non-linear models with dispersion arbitrarily close to ρ, and hence the highest dispersion may not always be achieved by linear models. These considerations are strengthened in Example 2 below, which exhibits a solvable term set for which linear models are not optimal for all alphabet sizes.

Example 2: Consider the term set

$$\Gamma := \{f(x,y), f\big(c(x),y\big), f\big(x,c(y)\big), c(z)\}$$

with min-cut of 3. Then Γ has a linear solution if A is organized as a field with at least 4 elements. However, Γ has a non-linear solution over $A = \mathbb{F}_2$ but does not have any linear solution over $\mathbb{F}_2$.

We now significantly strengthen this result by designing a solvable term set for which linear functions perform extremely poorly in terms of dispersion. Let $X = \{x_1, x_2, \ldots, x_k\}$ be a set of variables and consider $k+1$ functions h_i of these variables: $h_i(x_1, x_2, \ldots, x_k) = h_i(X)$. We then define the set Γ built on the k variables in X and the function symbols $h_1, h_2, \ldots, h_{k+1}$ together with f of arity $k+1$ and $g_1, g_2, \ldots, g_{k+1}$ of arities all equal to 1 by $\Gamma = \{t_1, t_2, \ldots, t_{k+1}\}$, where

$$t_i = f\Big(g_i\big(h_1(X)\big), h_2(X), h_3(X), \ldots, h_{k+1}(X)\Big).$$

Proposition 3: There exists $n \in \mathbb{N}$ such that for any A with $|A| \geq n$, Γ defined above is solvable over A, i.e. $\gamma(\Gamma, |A|) = k$. However, any linear model for Γ has dispersion at most 2.

The construction above can be easily generalized to obtain the following result.

Corollary 1: For any integers $k \geq l \geq 2$, there exists a solvable set Γ with dispersion k over all alphabets of sufficient large size where l is the maximal Γ-dispersion that can be achieved by (matrix) linear coding functions.

We finally remark that because linear maps disperse information uniformly, the Rényi entropy of a linear map does not depend on the coefficient α and is hence equal to its dispersion: for any linear model ψ for Γ over A, we have $H_\infty(\psi) = H_\alpha(\psi) = \gamma(\psi)$ for all $0 \leq \alpha \leq \infty$.

B. Perfect dispersion

In Section IV-A, we showed that there could be a huge difference between the dispersion achievable by linear coding functions and non-linear coding functions. In this section we show that if the min-cut is achievable by the use of coding functions of low degree (i.e. constant degree independently of the size of the underlying field), then the min-cut is actually achievable by the use of linear coding functions.

More precisely, let $\Gamma = \{t_1, t_2, \ldots, t_r\}$ be a term set built on the variables $\{x_1, x_2, \ldots, x_k\}$ and with min-cut of ρ. Let ψ be a model for Γ over an alphabet A with perfect Γ-dispersion of ρ. Then *decoding functions* for ψ are functions $\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_\rho : A^r \to A$ such that there exist $i_1, i_2, \ldots, i_\rho$ for which

$$(a_{i_1}, a_{i_2}, \ldots, a_{i_\rho}) = \Big(\bar{d}_1\big(\psi(\mathbf{a})\big), \bar{d}_2\big(\psi(\mathbf{a})\big), \ldots, \bar{d}_\rho\big(\psi(\mathbf{a})\big)\Big).$$

Theorem 3 (Low-degree solutions imply linear solutions): Let Γ be a term set built on the variables $x_1, x_2, \ldots, x_k$ and coding functions $f_1, f_2, \ldots, f_l$ with min-cut of ρ. Assume that there exist coding functions defined by fixed polynomials $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_l \in \mathbb{Z}[a_1, a_2, \ldots, a_k]$ with dispersion ρ for arbitrarily large fields F of characteristic q. Then there exist scalar linear coding functions over all sufficiently large fields F of characteristic q that achieve perfect dispersion of ρ.

Proof: Consider a solution ψ based on polynomials of fixed degree. Since the coding and decoding functions are all polynomials of fixed degrees, their compositions are also polynomials of fixed degrees. Suppose q^k is greater than all polynomial degrees and let L be the linear part operator: for any multivariate polynomial $\bar{p}$ over $\mathbb{F}_{q^k}$, $L(\bar{p})$ is the linear part of $\bar{p}$. Then it is easy to check that $L(\bar{p}_1 \circ \bar{p}_2) = L(\bar{p}_1) \circ L(\bar{p}_2)$ if the polynomial $\bar{p}_1 \circ \bar{p}_2$ has degree less than q^k. Let us now consider the model $L(\psi)$, defined as taking the linear part of each coding function in ψ. The induced mapping of $L(\psi)$ is then equivalent to taking the linear part of the induced mapping of ψ. Since $\bar{d}_j\big(\psi(\mathbf{a})\big) = a_{i_j}$ for all j, we have $L(\bar{d}_j) \circ L(\psi) = L(\bar{d}_j \circ \psi) = a_j$, which forms a linear solution with linear decoding functions. ∎

We remark that the results in Section IV-A imply that conversely, for each characteristic q there exist term sets with solutions which require that at least some of the involved coding functions (including decoding functions) should be given by polynomials of degree at least $|F|$. Example 3 gives a term set with no linear solution, and where a solution is given by polynomials whose degrees depend on the size of the alphabet.

Example 3: Let

$$\Gamma = \left\{f\big(f(x_1,x_2), f(x_2,x_1)\big), g\big(g(x_1,x_2), g(x_2,x_1)\big)\right\}.$$

This term set has no (scalar) linear solution over fields of characteristic 2, yet it has non-linear solutions over fields of size divisible by 4.

REFERENCES

[1] R. Dougherty, C. Freiling, and K. Zeger, "Linearity and solvability in multicast networks," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2243–2256, October 2004.

[2] ——, "Networks, matroids, and non-Shannon information inequalities," *IEEE Transactions on Information Theory*, vol. 53, no. 6, pp. 1949–1969, June 2007.

[3] R. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Info. Theory*, vol. 45, no. 3, pp. 1111–1120, May 1999.

[4] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Info. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[5] S. Riis and M. Gadouleau, "Max-flow min-cut theorems for multi-user networks based on term sets from logic," *submitted*, avaible at http://arxiv.org/abs/1012.5224.

[6] A. Rényi, "On measures of information and entropy," in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1961, p. 547561.

[7] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Info. Theory*, vol. 49, no. 2, pp. 371–381, February 2003.

[8] R. Kötter and M. Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.

[9] S. Riis, "Linear versus non-linear boolean functions in network flow," in *Proc. CISS*, Princeton, NJ, March 2004.

[10] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2745–2759, August 2005.

[11] K. Menger, "Zur allgemeinen kurventhorie," *Fundamenta Mathematicae*, vol. 10, pp. 95–115, 1927.