Networks, Matroids, and Non-Shannon Information Inequalities

Randall Dougherty, Chris Freiling, and Kenneth Zeger, *Fellow, IEEE*

***Abstract*—We define a class of networks, called matroidal networks, which includes as special cases all scalar-linearly solvable networks, and in particular solvable multicast networks. We then present a method for constructing matroidal networks from known matroids. We specifically construct networks that play an important role in proving results in the literature, such as the insufficiency of linear network coding and the unachievability of network coding capacity. We also construct a new network, from the Vámos matroid, which we call the Vámos network, and use it to prove that Shannon-type information inequalities are in general not sufficient for computing network coding capacities. To accomplish this, we obtain a capacity upper bound for the Vámos network using a non-Shannon-type information inequality discovered in 1998 by Zhang and Yeung, and then show that it is smaller than any such bound derived from Shannon-type information inequalities. This is the first application of a non-Shannon-type inequality to network coding. We also compute the exact routing capacity and linear coding capacity of the Vámos network. Finally, using a variation of the Vámos network, we prove that Shannon-type information inequalities are insufficient even for computing network coding capacities of multiple-unicast networks.**

***Index Terms*—Flow, information theory, matroids, multiple unicast, network coding.**

I. Introduction

In this paper, a *network* is a finite, directed, acyclic multigraph with node set ν and edge set ϵ, together with a finite set μ called the *message set*, a *source mapping*

$$S : \nu \to 2^{\mu}$$

and a *receiver mapping*

$$R : \nu \to 2^{\mu}.$$

For every node x, if $S(x)$ is nonempty, then x is called a *source*, and if $R(x)$ is nonempty, then x is called a *receiver*. The elements of $S(x)$ are called the *messages generated by* x and the elements of $R(x)$ are called the *messages demanded by* x. For convenience in definitions of capacity, we will assume that for each message m, every receiver demanding m is reachable from at least one source generating m.

An *alphabet* is a finite set $\mathcal{A}$ with at least two elements. For each network node x, let $\mathrm{In}(x)$ denote the union of the set of messages generated by x with the set of in-edges of x, and let $\mathrm{Out}(x)$ denote the union of the set of messages demanded by x with the set of out-edges of x.

Let k and n be positive integers, called the *source dimension* and the *edge capacity*, respectively. For every node x, fix an ordering of $\mathrm{In}(x)$ such that all messages in the resulting list occur before the edges in the list; the resulting ordered list is called the *input list* of x. For every edge $e = (x, y)$, an *edge function* is a map

$$f_e : (\mathcal{A}^k)^{\alpha} \times (\mathcal{A}^n)^{\beta} \to \mathcal{A}^n$$

where α and β are the number of messages and edges, respectively, in the input list of x (note that α and β are functions of x, whereas k and n are constants). For every $x \in \nu$ and $m \in R(x)$, a *decoding function* is a map

$$f_{x,m} : (\mathcal{A}^k)^{\alpha} \times (\mathcal{A}^n)^{\beta} \to \mathcal{A}^k$$

where α and β are the number of messages and edges, respectively, in the input list of x.

Given an alphabet $\mathcal{A}$, a (k, n) *code*[1] for a network is an assignment of edge functions and decoding functions to the network's edges and receivers, respectively. A *message assignment* is a map $a : \mu \to \mathcal{A}^k$. For any (k, n) code and for any message assignment, we recursively define the function

$$c : \epsilon \to \mathcal{A}^n$$

as follows. For every edge $e = (x, y)$, let

$$c(e) = f_e(a(x_1), \ldots, a(x_{\alpha}), c(x_{\alpha+1}), \ldots, c(x_{\alpha+\beta}))$$

where $x_1, \ldots, x_{\alpha}$ are the messages generated by x and $x_{\alpha+1}, \ldots, x_{\alpha+\beta}$ are the in-edges of x. We say that each edge e *carries* the symbol vector $c(e)$.

For a given network, (k, n) code, receiver x, and message m demanded by x, if for every message assignment $a : \mu \to \mathcal{A}^k$ we have

$$f_{x,m}(a(x_1), \ldots, a(x_{\alpha}), c(x_{\alpha+1}), \ldots, c(x_{\alpha+\beta})) = a(m)$$

then we say that x's *demand* m *is satisfied*. In other words, the receiver x can recover an arbitrary instance of the message m

[1]Sometimes called a *fractional code* [17] or simply a *code*.

Manuscript received January 24, 2006; revised December 27, 2006. This work was supported by the Institute for Defense Analyses, Air Force Office of Scientific Research, the National Science Foundation, and the UCSD Center for Wireless Communications. The material in this paper was presented in part at the Second Workshop on Network Coding, Theory, and Applications, Boston, MA, April 2006.

R. Dougherty is with the Center for Communications Research, San Diego, CA 92121-1969 USA (e-mail: rdough@ccrwest.org).

C. Freiling is with the Department of Mathematics, California State University, San Bernardino, San Bernardino, CA 92407-2397 USA (e-mail: cfreilin@csusb.edu).

K. Zeger is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: zeger@ucsd.edu).

Communicated by G. Kramer, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2007.896862

0018-9448/$25.00 © 2007 IEEE

generated by its source. A (k, n) code is said to be a (k, n) *solution* if every demand of every receiver is satisfied.

Informally, a network coding solution allows each receiver to deduce its demanded messages from its in-edges and source messages by having information propagate from the sources through the network. Each edge is allowed to be used at most once (i.e., at most n symbols can travel across each edge). Special codes of interest include *linear codes*, where the edge functions and decoding functions are linear, and *routing codes*, where the edge functions and decoding functions simply copy input components to output components. Special networks of interest include *multicast* networks, where there is only one source node and every receiver demands all of the source messages, and *multiple-unicast* networks, where each network message is generated by exactly one source node and is demanded by exactly one receiver node. The network coding terminology used here generally follows that of [3].

If a network has a (k, n) solution over some alphabet, then we say the ratio k/n is an *achievable coding rate* for the network. A network is said to be *solvable* if it has a (k, n) solution for the case $k = n = 1$. (Note that any (k, k) coding solution on alphabet $\mathcal{A}$ yields a $(1, 1)$ coding solution on alphabet $\mathcal{A}^k$, so we do not need to distinguish between scalar and vector solvability.) A network is said to be *scalar-linearly solvable* if it has a linear (k, n) solution for the case $k = n = 1$, or *vector-linearly solvable* if it has a linear (k, n) solution for the case $k = n$ (here we do need to distinguish).

An important goal in network coding is to find an achievable coding rate which is as large as possible for a network. The *coding capacity of a network with respect to (or over) an alphabet $\mathcal{A}$ and a class $\mathcal{C}$ of network codes* (a related definition appears in [32, p. 339]) is

$$\sup\{k/n : \exists\, (k, n) \text{ coding solution in } \mathcal{C} \text{ over } \mathcal{A}\}.$$

If $\mathcal{C}$ consists of all network codes, then we simply refer to the above quantity as the *coding capacity* of the network with respect to $\mathcal{A}$. If the class $\mathcal{C}$ of network codes consists of all routing codes or all linear codes, then the coding capacity is referred to as the *routing capacity* or *linear coding capacity*, respectively. (In all cases, if the alphabet $\mathcal{A}$ is not mentioned, the capacity is taken to be the supremum of the capacities over all alphabets $\mathcal{A}$.) The coding capacity of a given network is said to be *achievable* if there is some (k, n) solution for the network for which k/n equals the capacity.

Ahlswede, Cai, Li, and Yeung [2] exhibited a network whose linear coding capacity is larger than its routing capacity. Li, Yeung, and Cai [18] showed in the special case of a multicast network, the coding capacity and the linear coding capacity are equal. It was shown in [3] that for all networks, the coding capacity is independent of the alphabet size. Clearly, the routing capacity is also independent of the alphabet size. However, it was shown in [5] that the linear coding capacity of a network can depend on the alphabet size and the largest linear coding capacity of a network over any finite-field alphabet can be smaller than the network's coding capacity. It was also shown in [3] that the routing capacity is always rational, achievable, and computable by an algorithm.

Although the routing capacity of an arbitrary network is always computable, there is no known computationally efficient algorithm for such a task. Unfortunately, it is not even presently known whether or not there exist algorithms that can compute the coding capacity or the linear coding capacity of an arbitrary network. In fact, computing the exact coding capacity or linear coding capacity of even relatively simple networks can be a nontrivial task. At present, very few exact coding capacities have been rigorously derived in the literature. It is also known that the coding capacity might not be achievable [6].

As an alternative to determining exact coding capacities, it can be useful to determine bounds on the coding capacity and linear coding capacity of a network. One approach to obtaining capacity bounds (and possibly exact capacities) is to use information-theoretic entropy arguments. The basic idea is to assume a network's source messages are independent and identically distributed (i.i.d.) uniform random variables on some finite alphabet and then to use standard information-theory identities and inequalities to derive bounds on the largest possible ratio of the source dimension k to the edge capacity n.

Standard information inequalities are generally "Shannon-type" inequalities, which can be derived as special cases of the nonnegativity of conditional mutual information. These were the only known types of information inequalities until Zhang and Yeung in 1998 published a non-Shannon-type information inequality [35]. Some other results on non-Shannon-type information inequalities have been given by Lněnička [19], Makarychev, Makarychev, Romashchenko, and Vereshchagin [20], Matúš [23], Zhang [33], and Zhang and Yeung [34]. Previously, non-Shannon-type inequalities have been applied by Matúš to probability theory [22], by Chen and Yeung to group theory [4], and by Zhang to an information-theoretic optimization problem [33].

However, it has been an open question (e.g., see [14]) whether standard Shannon-type information-theoretic identities and inequalities are sufficient for computing the exact coding capacity of an arbitrary network, or whether they are sufficient for obtaining the best possible capacity bounds from entropy arguments. We answer these questions in the negative.

Specifically, we construct a network (from the well-known Vámos matroid) which we call the Vámos network, and demonstrate that no collection of Shannon-type information inequalities can produce an upper bound on the coding capacity which is as small as an upper bound obtainable using the Zhang–Yeung non-Shannon-type information inequality. To prove this result, we first show that Shannon-type information inequalities can only produce a coding capacity upper bound as low as 1 (Theorem VI.1), and then show that a non-Shannon-type information inequality argument can produce a coding capacity upper bound of $10/11$ (Theorem VI.2). Additionally, for the Vámos network, we compute the exact routing capacity (Theorem VI.4) and the exact linear coding capacity over every finite field (Theorem VI.8).

We note that Adler, Harvey, Kleinberg, Jain, and Rasala Lehman [1], [14] have recently given an interesting algorithmic procedure for determining upper bounds on the coding capacity of multiple-unicast networks. They mention the possibility that their bounds could be improved by the use of non-Shannon

inequalities (in fact, they conjecture that their bound is not sharp for this reason). We will demonstrate in what follows that this can indeed be accomplished. In Section VIII, we will give a specific multiple-unicast network and prove that there is a bound on its capacity using a non-Shannon-type inequality that is strictly better than any bound that can be obtained from Shannon-type inequalities alone. It can be shown [13] that the algorithm in [1], [14] is based on purely Shannon-type information inequalities and network entropy conditions, although this fact was not explicitly stated there. Thus, the capacity bounding algorithm in [1], [14] is not in general optimal.

The Vámos network is one of many networks that are closely related to matroids. The field of matroid theory has had many interesting results discovered over the last several decades. We explore the connection between matroids and networks and present a method of constructing networks from matroids. In addition to the Vámos network, we demonstrate that some specific known networks can be constructed from matroids. These include the Butterfly network from [2], and parts of networks used to establish the insufficiency of linear network coding in [5] and the unachievability of network coding capacity in [6].

Our use of the Vámos matroid was motivated by the important connection between non-Shannon-type information inequalities and the Vámos matroid, as presented by Hammer, Romashchenko, Shen, and Vereshchagin [12], and based partly on the work of Matúš [21], and Matúš and Studený [24]. Related ideas for building networks from matroids were explored by El Rouayheb, Georghiades, and Sprintson [8].

The paper is organized as follows. Sections II, III, and IV give overviews and lemmas relating to information-theoretic inequalities, networks, and matroids, respectively. Section V describes how to construct networks from matroids and gives various examples and demonstrates that Shannon-type inequalities cannot give a capacity upper bound smaller than 1 for a matroidal network. Section VI discusses the coding capacity, routing capacity, and linear coding capacity of the Vámos network. In particular, a non-Shannon-type information inequality is used to obtain a tighter upper bound on the capacity of the Vámos network than is achievable using only Shannon-type information inequalities. Section VII discusses converting arbitrary matroidal networks into multiple-unicast matroidal networks. Section VIII gives a multiple-unicast variation of the Vámos network and uses it to show that Shannon-type information inequalities are insufficient to compute the coding capacity of arbitrary multiple-unicast networks. Section IX mentions some open questions.

II. Information Inequalities

Let A, B, and C be collections of discrete random variables over alphabet $\mathcal{A}$, and let p be the probability mass function of A. Denote the *entropy* of A by

$$H(A) = -\sum_u p(u) \log_{|\mathcal{A}|} p(u),$$

the *conditional entropy* of A given B by

$$H(A|B) = H(A,B) - H(B), \tag{1}$$

the *mutual information* between A and B by

$$I(A;B) = H(A) - H(A|B), \tag{2}$$

and the *conditional mutual information* between A and B given C by[2]

$$I(A;B|C) = H(A|C) - H(A|B,C). \tag{3}$$

We will make use of the following basic information-theoretic facts [32]:

$$0 = H(\emptyset) \tag{4}$$
$$0 \le H(A) = H(A|\emptyset) \tag{5}$$
$$0 \le H(A|B) \tag{6}$$
$$0 \le I(A;B) \tag{7}$$
$$H(A,B|C) \le H(A|C) + H(B|C) \tag{8}$$
$$H(A|B,C) \le H(A|B) \le H(A,C|B) \tag{9}$$
$$I(A;B) = H(A) + H(B) - H(A,B) \tag{10}$$
$$I(A;B|C) = H(A,C) + H(B,C) - H(C) - H(A,B,C) \tag{11}$$
$$I(A;B,C) = I(B;A|C) + I(A;C). \tag{12}$$

It is obvious from (10) and (11) that $I(A;B)$ and $I(A;B|C)$ are symmetric in A and B.

Definition II.1: Let q be a positive integer, and let $S_1, \ldots, S_k$ be subsets of $\{1, \ldots, q\}$. Let $\alpha_i \in \boldsymbol{R}$ for $1 \le i \le k$. A linear inequality of the form

$$\alpha_1 H(\{A_i : i \in S_1\}) + \cdots + \alpha_k H(\{A_i : i \in S_k\}) \ge 0$$

is called an *information inequality* if it holds for all jointly distributed random variables $A_1, \ldots A_q$.

As an example, taking $q = 2$, $S_1 = \{1\}$, $S_2 = \{2\}$, $S_3 = \emptyset$, $S_4 = \{1,2\}$, $\alpha_1 = \alpha_2 = 1$, $\alpha_4 = -1$, and using (8) shows that $H(A_1) + H(A_2) - H(A_1, A_2) \ge 0$ is an information inequality; this can be more succinctly expressed using (10) as $I(A_1; A_2) \ge 0$.

Since all conditional entropies and all conditional mutual informations can be written as linear combinations of joint entropies, any valid linear inequality involving conditional entropies and conditional mutual informations will also be called an information inequality. The textbook [32] refers to information inequalities as "the laws of information theory."

The information inequalities in (5)–(9) were originally given in 1948 by Shannon [28] and can all be obtained as special cases (e.g., see [32]) of the inequality

$$I(A;B|C) \ge 0 \tag{13}$$

or equivalently (by (11)) of the inequality

$$H(A,C) + H(B,C) \ge H(C) + H(A,B,C). \tag{14}$$

[2]We will often use abbreviations such as "A, B" for $A \cup B$, "A, x" for $A \cup \{x\}$, and "$A - x$" for $A - \{x\}$.

A *Shannon-type information inequality* is any information inequality that is (or can be rearranged[3] to be) a finite sum of the form

$$\sum_i \alpha_i I(A_i; B_i|C_i) \geq 0 \tag{15}$$

where each α_i is a nonnegative real number. Virtually every known result in information theory that makes use of an information inequality only makes use of Shannon-type information inequalities.[4]

Any information inequality that cannot be expressed in the form (15) will be called a *non-Shannon-type information inequality*. It is known [32, p. 308] that all unconstrained information inequalities containing three or fewer random variables are Shannon-type inequalities. The first known non-Shannon-type information inequality was published in 1998 by Zhang and Yeung and is stated in the following theorem.

Theorem II.2: [35] The following is a non-Shannon-type information inequality:

$$2I(C;D) \leq I(A;B)+I(A;C,D)+3I(C;D|A)+I(C;D|B).$$

Let $\mathcal{S}$ be a finite set and let

$$f : 2^{\mathcal{S}} \to \boldsymbol{R}$$

be a function. Conditions (P1)–(P3) below are called the *polymatroidal axioms* for f (and (P4) is an alternate version, as explained by Lemma II.3).

(P1) $f(\emptyset) = 0$.

(P2) If $A \subseteq B \subseteq \mathcal{S}$, then $f(A) \leq f(B)$.

(P3) If $A, B \subseteq \mathcal{S}$, then $f(A \cup B)+f(A \cap B) \leq f(A)+f(B)$.

(P4) If $A, B, C \subseteq \mathcal{S}$,
then $f(A \cup C)+f(B \cup C) \geq f(C)+f(A \cup B \cup C)$.

Lemma II.3 below appears to be part of the information-theory folklore, and was proven in part by Fujishige [10] (also see Yeung [32, p. 297] and Welsh [30, p. 342]). For completeness, we provide a proof of the lemma here.

Lemma II.3: Conditions (P1)–(P3) hold if and only if conditions (P1) and (P4) hold.

Proof: Suppose (P1)–(P3) hold. Then

$$\begin{aligned} &f(A \cup C) + f(B \cup C) \\ &\quad \geq f((A \cup C) \cup (B \cup C)) \\ &\qquad + f((A \cup C) \cap (B \cup C)) \quad \text{[from (P3)]} \\ &\quad = f(A \cup B \cup C) + f(C \cup (A \cap B)) \\ &\quad \geq f(A \cup B \cup C) + f(C) \qquad \text{[from (P2)]} \end{aligned}$$

which gives (P4).

[3]We allow replacement of 0 by $H(\emptyset)$. This seemingly trivial technicality is needed, for example, in order to be able to assert that $I(A;B) \geq 0$ is of the form $I(A;B|\emptyset) \geq 0$. Yeung [32] calls the inequalities (5)–(7) and (13) the "basic Shannon inequalities."

[4]The constraints imposed on random variables by Shannon-type information inequalities define a region referred to in [32] as the LP bound.

Now suppose (P1) and (P4) hold. Then

$$\begin{aligned} &f(A \cap B) + f(A \cup B) \\ &\quad = f(A \cap B) + f(A \cup B \cup (A \cap B)) \\ &\quad \leq f(A \cup (A \cap B)) + f(B \cup (A \cap B)) \quad \text{[from (P4)]} \\ &\quad = f(A) + f(B) \end{aligned}$$

which gives (P3). Finally, suppose $A \subseteq B$. Then

$$\begin{aligned} &f(B) - f(A) \\ &\quad = f((B - A) \cup A) - f(A) \quad [\text{from } A \subseteq B] \\ &\quad = f((B - A) \cup A) + f(B \cup A) \\ &\qquad - f(A) - f(B \cup A) \\ &\quad = f((B - A) \cup A) + f(B \cup A) \\ &\qquad - f(A) - f((B - A) \cup B \cup A) \\ &\quad \geq 0 \quad \text{[from (P4)]} \end{aligned}$$

which gives (P2). □

Lemma II.4: Let $\mathcal{S}$ be a finite collection of jointly related discrete random variables. Then the polymatroidal axioms hold when f is replaced by the entropy function H.

Proof: It follows from (4), (14), and Lemma II.3. □

The polymatroid axioms are closely related to matroids via the matroid rank function, and so Lemma II.4 expresses a connection between matroids and information theory.

III. Network Fundamentals

If a network has nodes n_i and n_j (on diagrams these will usually be marked just i and j), then an edge between them will be written as $e_{i,j}$.

For any node $x \in \nu$ and any $S \subseteq \mathrm{Out}(x)$, we call the ordered pair $(\mathrm{In}(x), S)$ a *dependency* of the network. This terminology reflects the fact that the out-edges and demands of each node are deterministic functions of the in-edges and messages generated at the node. Using these, one can deduce further dependencies. For more on this, see [1], [14], [15].

In order to compute capacity bounds for networks, we will compute various joint entropies, where we take the network messages to be independent uniform random variables. In that case, given a network code, which determines the vectors carried by the edges from the messages, we will write $H(x)$ to denote the joint entropy of any collection x of edges and messages.

Lemma III.1: If a network has a (k, n) coding solution over alphabet $\mathcal{A}$, and the message components are independent random variables uniformly distributed over $\mathcal{A}$ (and entropies are computed using logarithms to base $|\mathcal{A}|$), then the following conditions hold.

(N1) (source rates) $H(x) = k|x|$ for any $x \subseteq \mu$.

(N2) (edge capacities) $H(x) \leq n$ for any $x \in \epsilon$.

(N3) (node input/output functional dependencies)
$H(\mathrm{In}(x)) = H(\mathrm{In}(x) \cup \mathrm{Out}(x))$ for any $x \in \nu$.

Proof: Conditions (N1) and (N2) are trivially true. Condition (N3) follows from the fact that the vector of alphabet symbols carried on each out-edge of a node or demanded by the

node must be a deterministic function of the node's messages and in-edges. □

We call conditions (N1)–(N3) the *network entropy conditions*.

Definition III.2: We define a (k, n) *polymatroid assignment* to a network $\mathcal{N}$ to be a map

$$\sigma : 2^{\mu \cup \epsilon} \rightarrow \boldsymbol{R}$$

such that conditions (N1)–(N3) hold when H is replaced by σ and the polymatroidal axioms for σ hold. For all $x, y \subseteq \mu \cup \epsilon$ we write

$$\sigma(y|x) = \sigma(x, y) - \sigma(x).$$

Thus, (4)–(9) hold when H is replaced by σ. The *polymatroid upper bound on the capacity* of network $\mathcal{N}$ is the quantity

$$\sup\{k/n : \exists (k, n) \text{ polymatroid assignment to } \mathcal{N}\}.$$

Remark III.3: If a network has a (k, n) coding solution over alphabet $\mathcal{A}$, then the network has a (k, n) polymatroid assignment.

To see this, let the message components be independent random variables uniformly distributed over $\mathcal{A}$, and take σ to be the entropy function H, in which case the assertion follows immediately from (14) and Lemma III.1.

Note that by Remark III.3, the terminology "polymatroid upper bound on the capacity" of a network is justified, since the coding capacity is the supremum of all k/n such that there exists a (k, n) coding solution over alphabet $\mathcal{A}$. Also, note that a network may have many polymatroid assignments that are not entropy functions, so it is feasible that the polymatroid upper bound might be larger than a bound obtained using entropy arguments. The purpose of the polymatroid assignments is to make precise the meaning of "bounds that are derivable from Shannon-type information inequalities." Indeed, if an upper bound is derived from Shannon-type information inequalities and uses no information about entropy other than what is contained in these inequalities and the network entropy conditions, then it should also be an upper bound for every polymatroid assignment. Thus, we may say (somewhat loosely) that the polymatroid upper bound on capacity is *the best upper bound on the network coding capacity obtainable using only Shannon-type information inequalities.*

Example III.4: To illustrate the calculation of a coding capacity bound for a network, consider a (k, n) coding solution to the Butterfly network[5] shown in Fig. 1. Assume that the network messages x and y are independent, k-dimensional, random vectors with uniformly distributed components. Since the presumed solution must allow node n_5 to deduce message y from its inputs x and z, it must be the case (via Remark III.3) that

$$H(y|x, z) = 0. \tag{16}$$

[5]The network's common name, due to its appearance.

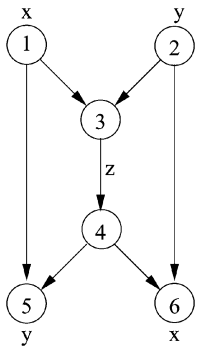


Fig. 1. The Butterfly network has source nodes n_1 and n_2 generating k-dimensional messages x and y, respectively. Receiver nodes n_5 and n_6 demand messages y and x, respectively. The n-dimensional vector carried on edge $e_{3,4}$ is denoted by z.

We have

$$\begin{aligned}
2k &= H(x) + H(y) && \text{[from (N1)]} \\
&= H(x, y) && \text{[from indep. of } x \text{ and } y\text{]} \\
&\leq H(x, y, z) && \text{[from (9)]} \\
&= H(x, z) + H(y|x, z) && \text{[from (1)]} \\
&= H(x, z) && \text{[from (16)]} \\
&\leq H(x) + H(z) && \text{[from (8)]} \\
&\leq k + n && \text{[from (N1) and (N2)]}
\end{aligned}$$

which implies $k/n \leq 1$, and therefore the coding capacity of the network is at most 1. A well-known solution for this network (over any alphabet carrying Abelian group operation $+$) is achieved with $k = n = 1$ and $z = x + y$, which implies 1 is an achievable rate and therefore the coding capacity is at least 1. The upper bound on the coding capacity was computed using only Shannon-type information inequalities and network entropy conditions, and in this example was exactly equal to the coding capacity (and also the linear coding capacity).

Although Shannon-type information inequalities were sufficient to compute the best possible upper bound on coding capacity for the network in Example III.4, an important point of this paper is to demonstrate that such inequalities are not sufficient in general. In what follows, we will exploit Theorem II.2 to obtain an upper bound on the coding capacity of the Vámos network and we will then show that this bound is strictly tighter than any such bound obtainable using only Shannon-type information inequalities.

In Lemma III.6, we will provide a useful extension of the network entropy condition (N3). Since we wish to apply this condition when the entropy H is replaced by a polymatroid assignment, the lemma will be written for arbitrary functions. Lemma III.6 appears in equivalent form as [14, Lemma 7]. First we need a definition.

Definition III.5: Let $\mathcal{N}$ be a network with message set μ and edge set ϵ. A set $C \subseteq \mu \cup \epsilon$ is a *cutset* for a node x in $\mathcal{N}$ if for every source w that generates a message not in C, every path[6] from w to x contains an edge in C.

For example, $C = \{e_{3,4}, x\}$ is a cutset for node n_5 in Fig. 1.

[6]This includes paths of length 0; so x cannot be a source for a message not in C.

Lemma III.6: Let $\mathcal{N}$ be a network with message set μ, node set ν, and edge set ϵ. If a function $f : 2^{\mu\cup\epsilon} \to \bar{\boldsymbol{R}}$ satisfies

$$f(\mathrm{In}(x)) = f(\mathrm{In}(x) \cup \mathrm{Out}(x)), \qquad \forall\, x \in \nu \tag{17}$$

and the polymatroid axioms and C is a cutset for a node y, then

$$f(\mathrm{Out}(y) \cup C) = f(C).$$

Proof: First, note that $\forall A, B, C \subseteq \mu \cup \epsilon$

$$\text{if } f(A \cup C) = f(C), \quad \text{then } f(A \cup B \cup C) = f(B \cup C). \tag{18}$$

This follows from

$$\begin{aligned} f(A \cup B \cup C) &\geq f(B \cup C) && \text{[from (P2)]} \\ &\geq f(A \cup B \cup C) \\ &\quad + f(C) - f(A \cup C) && \text{[from (P4)]} \\ &= f(A \cup B \cup C). \end{aligned}$$

Also, note from (P4) that $\forall\, A, B, C \subseteq \mu \cup \epsilon$

$$\begin{aligned} &\text{if } f(A \cup C) = f(B \cup C) = f(C), \\ &\quad \text{then } f(A \cup B \cup C) = f(C). \end{aligned} \tag{19}$$

Order the nodes in ν as $x_1, x_2, \ldots, x_t$ such that $i < j$ whenever (x_i, x_j) is an edge. Let $C \subseteq \mu \cup \epsilon$. We will prove by induction on j that if C is a cutset for x_j, then $f(C \cup \mathrm{Out}(x_j)) = f(C)$. Suppose $f(C \cup \mathrm{Out}(x_i)) = f(C)$ for all $i < j$ such that C is a cutset for x_i, and denote the in-edges of x_j by

$$(x_{i_1}, x_j), \ldots, (x_{i_u}, x_j).$$

Suppose C is a cutset for x_j, and let

$$I = \{s : 1 \leq s \leq u,\ (x_{i_s}, x_j) \notin C\}.$$

Then, for each $s \in I$, the set C is also a cutset for x_{i_s}, so

$$\begin{aligned} &f(C) = f(C \cup \mathrm{Out}(x_s)) \quad (\forall\, s \in I) \\ &\qquad \text{[from the induction hypothesis]} \\ &\therefore f(C) = f\left(C \cup \bigcup_{s \in I} \mathrm{Out}(x_{i_s})\right) && \text{[from(19)]} \\ &\qquad \geq f(C \cup \mathrm{In}(x_j)) && \text{[from (P2)]} \\ &\qquad \geq f(C) && \text{[from (P2)]} \end{aligned}$$

so

$$f(C) = f(C \cup \mathrm{In}(x_j)). \tag{20}$$

Therefore

$$\begin{aligned} &f(C \cup \mathrm{Out}(x_j)) \\ &\quad = f(C \cup \mathrm{In}(x_j) \cup \mathrm{Out}(x_j)) && \text{[from (20), (18)]} \\ &\quad = f(C \cup \mathrm{In}(x_j)) && \text{[from (17), (18)]} \\ &\quad = f(C) && \text{[from (20)].} \end{aligned}$$

□

IV. Matroid Fundamentals

We review here various definitions and results in matroid theory, as they will prove useful in the remainder of the paper. For a detailed introduction to matroid theory, the reader is referred to [26] or [30]. A *matroid* $\mathcal{M}$ is an ordered pair $(\mathcal{S}, \mathcal{I})$, where $\mathcal{S}$ is a finite set and $\mathcal{I}$ is a set of subsets of $\mathcal{S}$ satisfying the following three conditions.

(I1) $\emptyset \in \mathcal{I}$.

(I2) If $I \in \mathcal{I}$ and $J \subseteq I$, then $J \in \mathcal{I}$.

(I3) If $I, J \in \mathcal{I}$ and $|J| < |I|$, then $\exists e \in I - J$ such that $J \cup \{e\} \in \mathcal{I}$.

The set $\mathcal{S}$ is called the *ground set* and the matroid $\mathcal{M} = (\mathcal{S}, \mathcal{I})$ is called a matroid *on* $\mathcal{S}$. The members of $\mathcal{I}$ are called *independent sets* and any subset of $\mathcal{S}$ not in $\mathcal{I}$ is called a *dependent set*. A maximal independent set of a matroid is called a *base* of the matroid and a minimal dependent set is called a *circuit*. It can be easily shown that all bases are of the same cardinality.

There are many equivalent definitions of a matroid. One such alternate definition, which is particularly useful for us, uses the notion of a rank function (similar equivalent definitions using circuits or bases also exist).

For any matroid $\mathcal{M} = (\mathcal{S}, \mathcal{I})$ and any $X \subseteq \mathcal{S}$, let

$$\mathcal{I}|X = \{I \subseteq X : I \in \mathcal{I}\}$$

and let

$$\mathcal{M}|X = (X, \mathcal{I}|X).$$

Then $\mathcal{M}|X$ is a matroid (called the *restriction of* $\mathcal{M}$ *to* X) and the *rank* of X, denoted $r(X)$, is the size of a base of $\mathcal{M}|X$.

Lemma IV.1: [26, pp. 22–23] If r is the rank function of a matroid with ground set $\mathcal{S}$, then the following three conditions hold.

(R1) If $X \subseteq \mathcal{S}$, then $0 \leq r(X) \leq |X|$.

(R2) If $X \subseteq Y \subseteq \mathcal{S}$, then $r(X) \leq r(Y)$.

(R3) If $X, Y \subseteq \mathcal{S}$, then $r(X \cup Y) + r(X \cap Y) \leq r(X) + r(Y)$.

We will refer to (I1)–(I3) as the *independence axioms* of a matroid and to (R1)–(R3) as the *rank axioms* of a matroid. The following lemma shows that the rank axioms suffice to define a matroid.

Lemma IV.2: [26, p. 23] Let $\mathcal{S}$ be a set and let $r : 2^{\mathcal{S}} \to \{0, 1, 2, \ldots\}$ be a mapping satisfying (R1)–(R3). Let

$$\mathcal{I} = \{X \subseteq \mathcal{S} : r(X) = |X|\}.$$

Then $(\mathcal{S}, \mathcal{I})$ is a matroid having rank function r.

Some useful facts about matroids are summarized in the following lemmas.

Lemma IV.3: [26, p. 25] Let $\mathcal{M} = (\mathcal{S}, \mathcal{I})$ be a matroid with rank function r and suppose that $X \subseteq \mathcal{S}$. Then

(a) $X \in \mathcal{I}$ if and only if $|X| = r(X)$;

(b) X is a base if and only if $|X| = r(X) = r(M)$;

(c) X is a circuit if and only if $X \neq \emptyset$; and, for all $u \in X$, $r(X - \{u\}) = |X| - 1 = r(X)$.

The following lemma follows immediately from Lemmas IV.1 and IV.3(a).

Lemma IV.4: The rank function of any matroid satisfies the polymatroid axioms (P1)–(P3).

Lemma IV.5: Let r be the rank function of a matroid with ground set $\mathcal{S}$, and let X, Y, Z be subsets of $\mathcal{S}$. If $r(X) = r(X \cup Y)$, then $r(X \cup Z) = r(X \cup Y \cup Z)$.

Proof:

$$\begin{aligned} 0 &\geq r(X \cup Y \cup Z) + r(X) - r(X \cup Z) - r(X \cup Y) \\ &\qquad \text{[from (P4), Lemmas IV.4 and II.3]} \\ &= r(X \cup Y \cup Z) - r(X \cup Z) \qquad [\text{from } r(X) = r(X \cup Y)] \\ &\geq 0 \qquad \text{[from (R2)].} \end{aligned}$$

□

One important example of a matroid is obtained from graph theory. If $\mathcal{S}$ is the set of edges of a finite undirected graph, and $\mathcal{I}$ is the collection of all subforests (i.e., cycle-free subgraphs) of the graph, then $(\mathcal{S}, \mathcal{I})$ is a matroid. The spanning forests and cycles of the graph are, respectively, the bases and circuits in the matroid. The rank in the matroid of any subgraph determined by a subset of $\mathcal{S}$ is the number of edges in a spanning forest of the subgraph.

Another example of a matroid is obtained from linear algebra. Suppose A is an $m \times n$ matrix over a field F. If $\mathcal{S} = \{1, \ldots, n\}$ and $\mathcal{I}$ is the set of all $X \subseteq \mathcal{S}$ such that the multiset of columns of A indexed by the elements of X is linearly independent in the vector space $\mathrm{V}(n, F)$, then $(\mathcal{S}, \mathcal{I})$ is a matroid, called the *vector matroid* of A.

For example, if

$$A = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

over the field $\boldsymbol{R}$ of real numbers and the columns of A are indexed $1, 2, 3, 4$ as labeled above the matrix, then $(\mathcal{S}, \mathcal{I})$ is a vector matroid with $\mathcal{S} = \{1, 2, 3, 4\}$ and

$$\mathcal{I} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}.$$

A very useful collection of example matroids is the family of *uniform matroids* $U_{m,n}$, defined as follows. The ground set of $U_{m,n}$ is the set $\{1, \ldots, n\}$, and a subset of the ground set is independent if and only if it has size at most m.

Two matroids $(\mathcal{S}, \mathcal{I})$ and $(\mathcal{S}', \mathcal{I}')$ are said to be *isomorphic* if there exists a bijection $f : \mathcal{S} \to \mathcal{S}'$ such that $I \in \mathcal{I}$ if and only if $f(I) \in \mathcal{I}'$. If a matroid $\mathcal{M}$ is isomorphic to the vector matroid of a matrix over a field F, then $\mathcal{M}$ is said to be *representable over* F or *F-representable*. A matroid is *representable* if it is representable over some field.

A *geometric depiction* of any particular rank-$(m+1)$ matroid is a diagram in $\boldsymbol{R}^m$ consisting of nodes and undirected edges, where the nodes are in one-to-one correspondence with the matroid's ground set elements, and a collection of j of the matroid's ground set elements is dependent if and only if it corresponds to points in the diagram that are depicted as lying on a common $(j-2)$-dimensional plane.[7] Geometric depictions will be given to describe matroids in Figs. 5, 6, 8, and 10.

[7]A "plane" is sometimes drawn, by necessity, as a circle or other curved item.

A. Matroid Amalgams

Here we review various matroid terminology and results in the literature that will be used in Section VII. This will allow us to convert matroidal networks into multiple-unicast matroidal networks by means of a minor alteration. For a given matroid with ground set $\mathcal{S}$, the *closure* of an arbitrary set $X \subseteq \mathcal{S}$ is the set

$$\mathrm{cl}(X) = \{x \in \mathcal{S} : r(X \cup \{x\}) = r(X)\}.$$

A set $X \subseteq \mathcal{S}$ is said to be a *flat* if $X = \mathrm{cl}(X)$. If X and Y are flats in a given matroid, then (X, Y) is a *modular pair of flats* if

$$r(X) + r(Y) = r(X \cup Y) + r(X \cap Y).$$

A flat X is called a *modular flat* if (X, Y) is a modular pair of flats for all flats Y. A matroid is *modular* if all its flats are modular.

Suppose matroids $\mathcal{M}_1$ and $\mathcal{M}_2$ have ground sets $\mathcal{S}_1$ and $\mathcal{S}_2$, and let

$$\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$$

and

$$T = \mathcal{S}_1 \cap \mathcal{S}_2.$$

Furthermore, suppose that $\mathcal{M}_1|T = \mathcal{M}_2|T$. If $\mathcal{M}$ is a matroid on $\mathcal{S}$ such that $\mathcal{M}|\mathcal{S}_1 = \mathcal{M}_1$ and $\mathcal{M}|\mathcal{S}_2 = \mathcal{M}_2$, then $\mathcal{M}$ is called an *amalgam* of $\mathcal{M}_1$ and $\mathcal{M}_2$. Let r_1, r_2, and r be the rank functions of $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}$, respectively, and for all $X \subseteq \mathcal{S}$, let

$$\eta : 2^{\mathcal{S}} \to \boldsymbol{R}$$

and

$$\zeta : 2^{\mathcal{S}} \to \boldsymbol{R}$$

be defined by

$$\eta(X) = r_1(X \cap \mathcal{S}_1) + r_2(X \cap \mathcal{S}_2) - r(X \cap T) \tag{21}$$

$$\zeta(X) = \min\{\eta(Y) : X \subseteq Y\}. \tag{22}$$

If a function $f : 2^{\mathcal{S}} \to \boldsymbol{R}$ satisfies (P3), then f is called *submodular*.

Lemma IV.6: [26, p. 413] If ζ, as given in (22), is submodular on $2^{\mathcal{S}}$, then it is the rank function of a matroid on $\mathcal{S}$ which is an amalgam of $\mathcal{M}_1$ and $\mathcal{M}_2$.

When ζ is submodular, the (unique) matroid on $\mathcal{S}$ that has ζ as its rank function is called the *proper amalgam* of $\mathcal{M}_1$ and $\mathcal{M}_2$.

Lemma IV.7: [26, p. 416] Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be matroids and let T denote the intersection of their ground sets. If the restriction matroids $\mathcal{M}_1|T$ and $\mathcal{M}_2|T$ are identical and $\mathcal{M}_1|T$ is a modular matroid, then the proper amalgam of $\mathcal{M}_1$ and $\mathcal{M}_2$ exists.

V. Networks From Matroids

In this section, we give a method for building networks from matroids. The method involves a number of choices and hence does not produce a unique network.

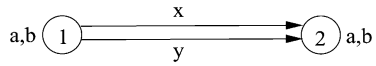


Fig. 2. A network which is matroidal with respect to more than one matroid. Messages a and b are generated by source n_1 and are demanded by receiver n_2.

Definition V.1: Let $\mathcal{N}$ be a network with message set μ, node set ν, and edge set ϵ. Let $\mathcal{M} = (\mathcal{S}, \mathcal{I})$ be a matroid with rank function r. The network $\mathcal{N}$ is a *matroidal network* associated with $\mathcal{M}$ if there exists a function $f : \mu \cup \epsilon \to \mathcal{S}$ such that the following conditions are satisfied:

(M1) f is one-to-one on μ;

(M2) $f(\mu) \in \mathcal{I}$;

(M3) $r(f(\mathrm{In}(x))) = r(f(\mathrm{In}(x) \cup \mathrm{Out}(x)))$, for every $x \in \nu$.

It follows from (M1), (M2), (I2), and Lemma IV.3(a) that

$$r(f(S)) = |S|, \qquad \text{for all } S \subseteq \mu. \tag{23}$$

We call the function f the *network-matroid mapping*. Condition (M1) assigns unique matroid ground set elements to the network messages, and condition (M2) assures that the network messages correspond to an independent set. Condition (M3) reflects the fact that the out-edges of each network node are completely determined by the in-edges and source messages of the node.

The following is a more flexible, but equivalent form of (M3): For every $x \in \nu$ and for any $S \subseteq \mathrm{Out}(x)$, we have

$$r(f(\mathrm{In}(x))) = r(f(\mathrm{In}(x) \cup S)).$$

To see this, note that

$$\begin{aligned} r(f(\mathrm{In}(x))) &= r(f(\mathrm{In}(x) \cup \mathrm{Out}(x))) && \text{[from (M3)]} \\ &\geq r(f(\mathrm{In}(x) \cup S)) && \text{[from (R2)]} \\ &\geq r(f(\mathrm{In}(x))) && \text{[from (R2)].} \end{aligned}$$

Example V.2: A matroid witnessing a network being matroidal need not be unique. Consider the network shown in Fig. 2. If we take $f(a) = f(x) = 1$ and $f(b) = f(y) = 2$, then f is a network-matroid mapping over the uniform matroid $U_{2,2}$ with ground set $\{1, 2\}$, and if we take $f(a) = f(x) = 1$, $f(b) = 2$, and $f(y) = 3$, then f is a network-matroid mapping over the uniform matroid $U_{2,3}$ with ground set $\{1, 2, 3\}$.

The following fact about matroidal networks will be used in Theorem VI.1.

Lemma V.3: For any matroidal network, the polymatroid upper bound on the capacity is at least 1.

Proof: Let $\mathcal{M}$ be a matroid with ground set $\mathcal{S}$ and let $\mathcal{N}$ be a matroidal network associated with $\mathcal{M}$, and having message set μ and edge set ϵ. We will give a $(1, 1)$ polymatroid assignment to $\mathcal{N}$. Let

$$f : \mu \cup \epsilon \to \mathcal{S}$$

be a network-matroid mapping for $\mathcal{N}$ and $\mathcal{M}$ and let r be the rank function of $\mathcal{M}$. Define the composition function

$$g = r \circ f.$$

The function g satisfies conditions (R1)–(R3) in Lemma IV.1, and, by Lemma IV.4, these conditions imply (P1)–(P3) if we replace f by g. Hence, by Lemma II.3, the function g satisfies condition (P4) and hence (14) when H is replaced by g. Also, $g(A) = |A|$ for any $A \subseteq \mu$, by (M1), (M2), and Lemma IV.3 (a), so network condition (N1) is satisfied with $k = 1$, when H is replaced by g. Similarly, $g(x) \leq 1$ for each $x \in \epsilon$, by (R1), so network condition (N2) is satisfied with $n = 1$, when H is replaced by g. Furthermore, for any node x and any $y \in \mathrm{Out}(x)$, we have

$$g(\mathrm{In}(x) \cup \{y\}) = g(\mathrm{In}(x))$$

by the equivalent form of (M3) preceeding this lemma, which implies network condition (N3), when H is replaced by g. Thus, the network conditions in (N1)–(N3) are satisfied with $k = n = 1$. So, by Definition III.2, the polymatroid upper bound on the capacity is at least 1. □

So, to show that Shannon inequalities are insufficient for computing coding capacity, it suffices to find a matroidal network that has capacity less than 1. This is accomplished using the Vámos network.

Next, we easily demonstrate that a large class of interesting networks are matroidal.

Theorem V.4: If a network is scalar-linearly solvable over some finite field, then the network is matroidal. In fact, the network is associated with a representable matroid.

Proof: Fix a scalar-linear solution to the network over finite field F, and let $a_1, \ldots, a_m$ be the network messages. Let $x_1, x_2, \ldots$ be the message and edge variables. For each i, the variable x_i can be written as a linear combination

$$c_1^{(i)} a_1 + \cdots + c_m^{(i)} a_m$$

of the messages, where $c_j^{(i)} \in F$, for all j. Form a matrix with a column

$$C^{(i)} = \begin{pmatrix} c_1^{(i)} \\ c_2^{(i)} \\ \vdots \\ c_m^{(i)} \end{pmatrix}$$

for each x_i. Let $\mathcal{M}$ be the vector matroid for this matrix and let r be the rank function of $\mathcal{M}$. Let $f(x_i) = i$, for all i. The function f is clearly one-to-one, giving (M1). If x_i is message a_j, then $C^{(i)}$ has all components zero except the jth component. The m columns associated with such messages are clearly independent, giving (M2). To prove (M3), suppose $x_i \in \mathrm{Out}(y)$; then x_i is a linear combination of the elements of $\mathrm{In}(y)$, so $C^{(i)}$ is the same linear combination of the $C^{(j)}$ for $x_j \in \mathrm{Out}(y)$. Therefore

$$r(f(\{x_i\} \cup \mathrm{Out}(y))) = r(f(\mathrm{Out}(y))).$$ □

Theorem V.4 suggests a technique for obtaining a network that has a good chance of not being scalar-linearly solvable. That is, choose a network that is matroidal over a nonrepresentable matroid. The Vámos matroid defined in Section V-F is the smallest example of a nonrepresentable matroid [26, p.512], providing inspiration to define and study a " Vámos network."

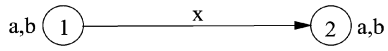


Fig. 3. An unsolvable nonmatroidal network. Messages a and b are generated by source n_1 and are demanded by receiver n_2. The symbol carried on edge $e_{1,2}$ is x.

The following corollary follows immediately from Theorem V.4 and the fact that all solvable multicast networks are scalar-linearly solvable over some finite field [18].

Corollary V.5: All solvable multicast networks are matroidal.

The following simple lemma immediately gives the capacity upper bound in Example III.4. It is a slight variation of a lemma that appears elsewhere (e.g., [32, p. 328], [5]).

Lemma V.6: Suppose a network, with message set μ and edge set ϵ, has a message z which is demanded by a node y and is generated only by source node x. Let B be a network edge. If every path in the network from x to y passes through B, then the polymatroid upper bound for the coding capacity of the network is at most 1.

Proof: Let σ be a (k, n) polymatroid assignment to the network. Note that $\{B, \mu - z\}$ is a cutset for y. Then

$$\begin{aligned}
k &= \sigma(\mu) - \sigma(\mu - z) && \text{[from (N1)]}\\
&= \sigma(z|\mu - z) && \text{[from (1)]}\\
&\le \sigma(z|\mu - z) + \sigma(B|z, \mu - z) && \text{[from (6)]}\\
&= \sigma(z, B|\mu - z) && \text{[from (1)]}\\
&= \sigma(B|\mu - z) + \sigma(z|B, \mu - z) && \text{[from (1)]}\\
&= \sigma(B|\mu - z) && \text{[from Lemma III.6]}\\
&\le \sigma(B) && \text{[from (9)]}\\
&\le n && \text{[from (N2)].}
\end{aligned}$$

Thus, any (k, n) polymatroid assignment satisfies $k/n \le 1$, and therefore the polymatroid upper bound on the capacity of the network is at most 1. □

We note that a special case of Lemma V.6 is when the network has a unique directed path from x to y.

A. The M-Network

A trivial example of an unsolvable network that is not matroidal is shown in Fig. 3. Note that if the network were matroidal with network-matroid mapping f and matroid rank function r, then

$$\begin{aligned}
2 &= r(f(a), f(b)) && \text{[from (M1), (M2)]}\\
&\le r(f(a), f(b), f(x)) && \text{[from (R2)]}\\
&= r(f(x)) && \text{[from (M3)]}\\
&\le 1 && \text{[from (R1)]}
\end{aligned}$$

which gives a contradiction.

Here, we demonstrate that not all solvable networks are matroidal. We call the network shown in Fig. 4 the *M-network* (due to its shape). The M-network was discussed in [25] as an example of a network with no scalar linear solution, but with a simple vector linear solution.

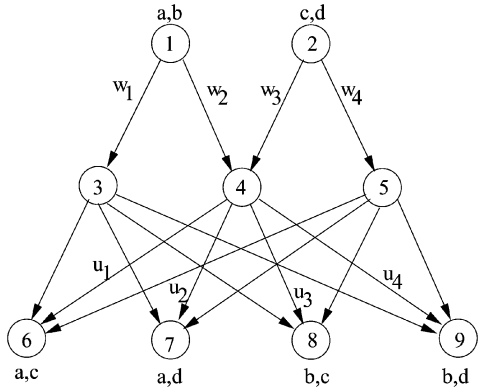


Fig. 4. The M-network. Messages a and b are generated by source n_1 and messages c and d are generated by source n_2. The four messages a, b, c, d are demanded in various pairs at the receivers n_6, n_7, n_8, and n_9. The edges $e_{1,3}$, $e_{1,4}$, $e_{2,4}$, $e_{2,5}$, $e_{4,6}$, $e_{4,7}$, $e_{4,8}$, and $e_{4,9}$, are denoted by w_1, w_2, w_3, w_4, u_1, u_2, u_3, and u_4, respectively.

Lemma V.7: The following is a Shannon-type information inequality:

$$I(C; D) \le I(A; B) + H(C|A) + H(D|B).$$

Proof: Let the notation $y \ge_S x$ mean that $y - x \ge 0$ is a Shannon-type information inequality (similarly for $\le_S$). It is easy to see that $\ge_S$ is transitive, since the sum of two Shannon-type inequalities is a Shannon-type inequality. Now, using

$$\begin{aligned}
I(A; B) + H(D|B) &= I(A; B) + I(A; D|B)\\
&\quad + H(D|A, B) && \text{[from (3)]}\\
&\ge_S I(A; B) + I(A; D|B) && \text{[from (6)]}\\
&= I(A; B, D) && \text{[from (12)]}\\
&= I(A; D) + I(A; B|D) && \text{[from (12)]}\\
&\ge_S I(A; D) && \text{[from (15)]} \quad (24)
\end{aligned}$$

we obtain

$$\begin{aligned}
I(C; D) &\le_S I(A; D) + H(C|A)\\
&\quad \text{[from (24) with } A \to D,\ B \to A,\ D \to C]\\
&\le_S I(A; B) + H(C|A) + H(D|B) && \text{[from (24)].}
\end{aligned}$$

□

Theorem V.8: The M-network is solvable, but is not matroidal.

Proof: A two-dimensional vector routing solution for the M-network was given in [25], so it remains to show the network is not matroidal.

Suppose, to the contrary, that the M-network is matroidal. Let r be the rank function of the associated matroid, let f be the network-matroid mapping, let $g = r \circ f$, and let

$$\begin{aligned}
A &= (a, b)\\
B &= (c, d)\\
C &= (w_1, a)\\
D &= (w_4, c).
\end{aligned}$$

Then we have

$$g(A,C) = g(A) \quad \text{[from (M3) and (R2)]} \tag{25}$$
$$g(B,D) = g(B) \quad \text{[from (M3) and (R2)]} \tag{26}$$
$$g(A,B) = g(A) + g(B) \quad \text{[from (23)]}. \tag{27}$$

By Lemma IV.4, the rank function of any matroid obeys the polymatroidal axioms (P1)–(P3), so we can apply Lemma V.7 (letting the map g play the role of the entropy H) to obtain

$$\begin{aligned} 0 &\le g(C) + g(D) - g(C,D) \quad \text{[from (R1) and (R3)]} \\ &\le g(A) + g(B) - g(A,B) \\ &\quad + g(A,C) - g(A) + g(B,D) - g(B) \\ &\quad \text{[from Lemma V.7]} \\ &= 0 \quad \text{[from (25)–(27)]}. \end{aligned} \tag{28}$$

Thus, we have

$$\begin{aligned} g(w_1,a) + g(w_4,c) &= g(C) + g(D) \\ &= g(C,D) && \text{[from (28)]} \\ &= g(a,c,w_1,w_4) \\ &\le g(w_1,u_1,w_4,a,c) && \text{[from (R2)]} \\ &= g(w_1,u_1,w_4) && \text{[from (M3) at } n_6\text{]} \\ &\le 3 && \text{[from (R1)]}. \end{aligned}$$

Similar arguments yield

$$g(w_1,a) + g(w_4,d) \le 3 \quad \text{[from (M3) at } n_7\text{]}$$
$$g(w_1,b) + g(w_4,c) \le 3 \quad \text{[from (M3) at } n_8\text{]} \tag{29}$$
$$g(w_1,b) + g(w_4,d) \le 3 \quad \text{[from (M3) at } n_9\text{]}. \tag{30}$$

Since $\{w_1,w_2,w_3,w_4\}$ is a cutset for node n_6 and for node n_9, we have

$$g(w_1,w_2,w_3,w_4,a,c) = g(w_1,w_2,w_3,w_4) \quad \text{[from Lemma III.6]} \tag{31}$$
$$g(w_1,w_2,w_3,w_4,b,d) = g(w_1,w_2,w_3,w_4) \quad \text{[from Lemma III.6]}. \tag{32}$$

Therefore

$$\begin{aligned} 4 &= g(a,b,c,d) && \text{[from(23)]} \\ &\le g(w_1,w_2,w_3,w_4,a,b,c,d) && \text{[from (R2)]} \\ &= g(w_1,w_2,w_3,w_4,b,d) && \text{[from (31) and Lemma IV.5]} \\ &= g(w_1,w_2,w_3,w_4) && \text{[from (32)]} \\ &\le g(w_1) + g(w_2) + g(w_3) + g(w_4) && \text{[from (R3)]} \end{aligned}$$

and

$$0 \le g(w_i) \le 1, \quad \text{for all } i \quad \text{[from (R1)]}$$

so

$$g(w_1) = g(w_2) = g(w_3) = g(w_4) = 1. \tag{33}$$

We also have

$$\begin{aligned} &g(w_1,a) + g(w_1,b) \\ &\quad \ge g(w_1,a,b) + g(w_1) && \text{[from (R3)]} \\ &\quad \ge g(a,b) + g(w_1) && \text{[from (R2)]} \\ &\quad = 3 && \text{[from (23), (33)]}. \end{aligned} \tag{34}$$

By a similar argument

$$g(w_4,c) + g(w_4,d) \ge 3. \tag{35}$$

Adding (29) and (30), subtracting (35), and then dividing by 2 gives

$$g(w_1,b) \le 1.5. \tag{36}$$

Similar arguments show that

$$g(w_1,a), g(w_4,c), g(w_4,d) \le 1.5. \tag{37}$$

However, in order to satisfy (34)–(37), it must be the case that

$$g(w_1,a) = g(w_1,b) = g(w_4,c) = g(w_4,d) = 1.5.$$

But since the rank function r is integer-valued, we have obtained a contradiction. Thus, the M-network is not matroidal. □

For Lemma V.9 and Theorem V.10, F is a finite field. The following lemma is given in [31, Theorem 7.3].

Lemma V.9: Let $L : F^m \to F^n$ be a linear map, and let x be a uniformly distributed random variable on F^m. Then $L(x)$ is uniformly distributed on the range of L, and $H(L(x)) = \mathsf{dim}(\mathsf{range}(L)) \cdot \log|F|$.

Proof: For each y in the range of L, $f^{-1}(y)$ is a coset of the kernel of L. All such cosets have the same cardinality, and each element of F^m has the same probability for x, so each element of the range of f has the same probability for $L(x)$. So $L(x)$ is uniformly distributed on the range of L. This range is a subspace of F^n, say of dimension d, so it has cardinality $|F|^d$; hence, we have $H(L(x)) = d \cdot \log|F|$. □

The two-dimensional vector-linear solution to the M-network given in [25] is a simple routing solution and easily extends to a vector-linear solution over any even vector dimension. We next show that no other vector dimensions are possible for vector-linear solutions to the M-network.

Theorem V.10: The M-network does not have any vector-linear solutions of odd vector dimension.

Proof: Suppose we have a vector-linear solution of dimension k over the field F. We assume all logarithms for entropies are taken to be base $|F|$ so that $\log|F| = 1$. Any edge variable, or any finite collection of messages and edge variables, is a linear function of the messages. Hence, if the messages are independent uniform random variables over F^k, then Lemma V.9 implies that $H(x_1,x_2,\ldots)$ is an integer for any messages and/or edge variables $x_1,x_2,\ldots$. But if we let

$$g(x_1,x_2,\ldots) = H(x_1,x_2,\ldots)/k$$

then the proof of Theorem V.8 can be repeated to give $g(w_1,a) = 1.5$ and hence, $H(w_1,a) = 1.5k$. Therefore, $1.5k$ must be an integer, so k must be even. □

In particular, the M-network does not have a scalar-linear solution.

B. Method for Constructing Networks From Matroids

We will next describe a method that can be useful for constructing a matroidal network associated with a matroid. Such constructions allow us to transfer various interesting properties

of matroids to networks. As matroid theory is a field rich in important results, the goal in constructing matroidal networks is to obtain some analogues for networks.

Let $\mathcal{M} = (\mathcal{S},\mathcal{I})$ be a matroid with rank function r. Let $\mathcal{N}$ denote the network to be constructed, μ its message set, ν its node set, and ϵ its edge set.

The construction will simultaneously construct the network $\mathcal{N}$, the function

$$f : \mu \cup \epsilon \to \mathcal{S},$$

and an auxiliary function

$$g : \mathcal{S} \to \nu,$$

where for each $x \in \mathcal{S}$, either

(i) $g(x)$ is a source node with message m and $f(m) = x$; or
(ii) $g(x)$ is a node with in-degree 1 and whose in-edge e satisfies $f(e) = x$.

The construction is carried out in four stages; each stage can be completed in many ways.

Step 1: Create network source nodes $n_1, n_2, \ldots, n_{r(\mathcal{S})}$ and corresponding messages $m_1, m_2, \ldots, m_{r(\mathcal{S})}$. Choose any base $B = \{b_1, \ldots, b_{r(\mathcal{S})}\}$ for $\mathcal{M}$ and let $f(m_i) = b_i$ and $g(b_i) = n_i$.

Step 2: (to be repeated until it is no longer possible). Find a circuit $\{x_0, \ldots, x_j\}$ in $\mathcal{M}$, such that $g(x_1), \ldots, g(x_j)$ have been already defined, but $g(x_0)$ has not yet been defined. Then we will add the following:

(i) a new node y and edges $e_1, \ldots, e_j$, such that e_i connects $g(x_i)$ to y, and we define $f(e_i) = x_i$.
(ii) another new node n_0 with a single in-edge e_0 connecting y to n_0, and we let $f(e_0) = x_0$ and $g(x_0) = n_0$.

Step 3: (to be repeated as many times as desired). If $\{x_0, \ldots, x_j\}$ is a circuit in $\mathcal{M}$ and $g(x_0)$ is a source node with message m_0, then add to the network a new receiver node y which demands the message m_0 and which has in-edges $e_1, \ldots, e_j$ where e_i connects $g(x_i)$ to y and where $f(e_i) = x_i$.

Step 4: (to be repeated as many times as desired). Choose a base $B = \{x_1, \ldots, x_{r(\mathcal{S})}\}$ of $\mathcal{M}$ and create a receiver node y that demands all of the network messages, and such that y has in-edges $e_1, \ldots, e_{r(\mathcal{S})}$ where e_i connects $g(x_i)$ to y. Let $f(e_i) = x_i$.

Note that after each of the preceding steps, the network constructed so far is matroidal with respect to $\mathcal{M}$.

It is clear that after Step 2, the function g has been completely determined. This is because for each $x \in \mathcal{S}$, one can always create a circuit containing x and some subset of the starting base B.

It is possible that some circuits cannot be used in Step 3 since they have no element which is mapped by g to a source message. Hence, after this stage of the construction there may be dependencies in $\mathcal{M}$ which are not reflected in the properties of the network $\mathcal{N}$. The final stage (Step 4), however, can at least assure us that all of the independencies in $\mathcal{M}$ are reflected in the properties of $\mathcal{N}$.

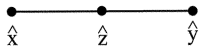


Fig. 5. Geometric depiction of the rank-2 uniform matroid $U_{2,3}$, which can be used to construct the Butterfly network. The matroid has ground set $\{\hat{x}, \hat{y}, \hat{z}\}$ and a set is independent if and only if it does not have three collinear points in the figure (i.e., iff it has size at most 2).

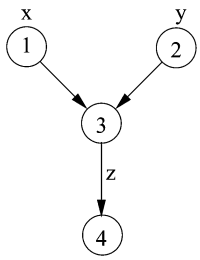
C. The Butterfly Network

The Butterfly network in Fig. 1 is matroidal associated with the rank-2 uniform matroid $U_{2,3}$ geometrically depicted in Fig. 5. The network-matroid mapping (from the network sources and edges to the matroid) constructed along with the network has been partially given[8] in Fig. 1. This network is known to have a linear solution over any ring alphabet (by taking $z = x + y$). One can easily check that the conditions (M1)–(M3) hold.

To illustrate the construction of a network from a matroid, we next show the steps from Section V-B involved in the construction of the Butterfly network.

Step 1: We choose a matroid base $B = \{\hat{x}, \hat{y}\}$ and network messages x and y, and we assign $f(x) = \hat{x}$ and $f(y) = \hat{y}$, and $g(\hat{x}) = n_1$ and $g(\hat{y}) = n_2$.

Step 2: The only circuit in the matroid is $\{\hat{x}, \hat{y}, \hat{z}\}$, and $g(\hat{x})$ and $g(\hat{y})$ have already been defined, but $g(\hat{z})$ has not yet been defined. We add a new node n_3 and edges $e_{1,3}$ and $e_{2,3}$, and we define $f(e_{1,3}) = \hat{x}$ and $f(e_{2,3}) = \hat{y}$. We add another new node n_4 with a single in-edge $e_{3,4}$ and we let $f(e_{3,4}) = \hat{z}$ and $g(\hat{z}) = n_4$.



Step 3: The only circuit in the matroid is $\{\hat{x}, \hat{y}, \hat{z}\}$ and $g(x) = n_1$ is a source node with message m_1. We add a new receiver node n_5 which demands the message m_1 and has in-edges $e_{1,5}$ and $e_{4,5}$. We repeat this step once more with the same circuit $\{\hat{x}, \hat{y}, \hat{z}\}$, but this time using the source node $g(\hat{y}) = n_2$ with message m_2. We add a new receiver node n_6

[8]Any edge coming from a node with only one input will not be labeled in diagrams, and it can be assumed that any such label equals the label of the unique input to the node.

TABLE I
DEPENDENCIES IN THE UNIFORM MATROID $U_{2,3}$ THAT ARE REFLECTED IN THE BUTTERFLY NETWORK. THE SECOND COLUMN INDICATES SETS OF VARIABLES IN THE BUTTERFLY NETWORK CORRESPONDING TO DEPENDENT SETS IN THE $U_{2,3}$ MATROID. THE THIRD COLUMN INDICATES AT WHICH NODES IN THE BUTTERFLY NETWORK THE CORRESPONDING DEPENDENCY IS ENFORCED

Step	Variables	Nodes	Type
1	$\{x, y\}$	n_1, n_2	message
2	$\{x, y, z\}$	n_3, n_4	circuit
3	$\{x, y, z\}$	n_5	circuit
3	$\{x, y, z\}$	n_6	circuit
4	none	not used	

which demands the message m_2 and has in-edges $e_{2,6}$ and $e_{4,6}$. The result is the Butterfly network.

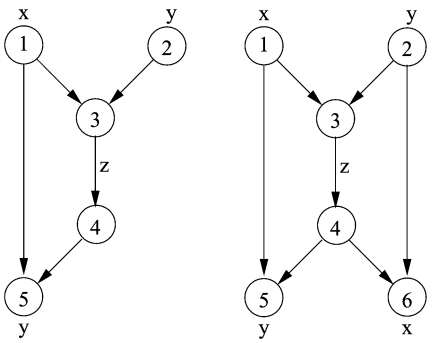


Table I lists the dependencies in the uniform matroid $U_{2,3}$ which are directly reflected in the Butterfly network.

D. The Fano Network

Fig. 6 is a geometric depiction of the well-known Fano matroid [26]. The network shown in Fig. 7, which we call the *Fano network*, is a matroidal network associated with the Fano matroid and is constructed using the technique described in Section V-B. The network-matroid mapping is partially shown in Fig. 7, where the mapping on the unlabeled edges is given by the usual convention. The network-matroid mapping is the identity function on the network source messages a, b, and c. It is easy to see that there exists a dependency between any three network variables if and only if the corresponding three matroid elements are dependent. Table II lists the dependencies in the Fano matroid which are directly reflected in the Fano network.

The Fano matroid is known to be F-representable over a finite field F if and only if F has characteristic two [26]. Correspondingly, the Fano network was shown in [6], to be solvable if and only if the alphabet size is an integer power of two. It, in fact, has a linear solution over any finite field of characteristic two (by taking $w = a+b$, $x = a+c$, $y = b+c$, and $z = a+b+c$). The Fano network was used as a building block to construct a network whose coding capacity cannot be achieved by the network. The Fano network was also used as a building block in [5] to construct a solvable network that is not linearly solvable (in a very general sense).

E. The Non-Fano Network

Fig. 8 is a geometric depiction of the well-known non-Fano matroid [26]. The network shown in Fig. 9, which we call the *non-Fano network*, is a matroidal network associated with the non-Fano matroid and is constructed using the technique described in Section V-B. The network-matroid mapping is partially shown in Fig. 9 where the mapping on the unlabeled edges is given by the usual convention.

Table III lists the dependencies in the non-Fano matroid which are directly reflected in the non-Fano network. The non-Fano matroid is known [26] to be F-representable over a finite field F if and only if F has odd characteristic. Correspondingly, the non-Fano network was shown in [6], to be

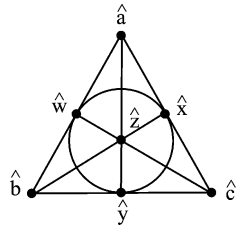


Fig. 6. Geometric depiction of the Fano matroid. The matroid has ground set $\{\hat{a}, \hat{b}, \hat{c}, \hat{w}, \hat{x}, \hat{y}, \hat{z}\}$ and has rank 3. Any three elements of the ground set are dependent if and only if they are collinear in the diagram (where we pretend that points on the drawn circle are also "collinear").

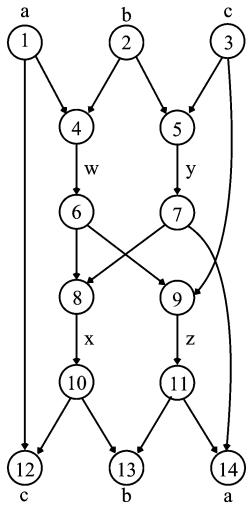


Fig. 7. The Fano network. Messages a, b, and c are emitted by sources n_1, n_2, and n_3, respectively, and are demanded by receivers n_{14}, n_{13}, and n_{12}, respectively. The edges $e_{4,6}$, $e_{5,7}$, $e_{8,10}$, and $e_{9,11}$ are labeled according to the network-matroid mapping by their corresponding ground set elements in the Fano matroid shown in Fig. 6.

TABLE II
DEPENDENCIES IN THE FANO MATROID THAT ARE REFLECTED IN THE FANO NETWORK. THE SECOND COLUMN INDICATES SETS OF VARIABLES IN THE FANO NETWORK CORRESPONDING TO DEPENDENT SETS IN THE FANO MATROID. THE THIRD COLUMN INDICATES AT WHICH NODES IN THE FANO NETWORK THE CORRESPONDING DEPENDENCY IS ENFORCED

Step	Variables	Nodes	Type
1	$\{a, b, c\}$	n_1, n_2, n_3	message
2	$\{a, b, w\}$	n_4, n_6	circuit
2	$\{b, c, y\}$	n_5, n_7	circuit
2	$\{w, x, y\}$	n_8, n_{10}	circuit
2	$\{c, w, z\}$	n_9, n_{11}	circuit
3	$\{a, c, x\}$	n_{12}	circuit
3	$\{b, x, z\}$	n_{13}	circuit
3	$\{a, y, z\}$	n_{14}	circuit
4	none	not used	

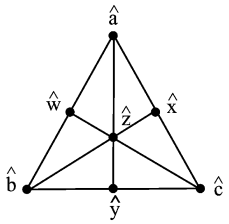


Fig. 8. Geometric depiction of the non-Fano matroid. The matroid has ground set $\{\hat{a}, \hat{b}, \hat{c}, \hat{w}, \hat{x}, \hat{y}, \hat{z}\}$ and has rank 3. Any three elements of the ground set are dependent if and only if they are collinear in the diagram.

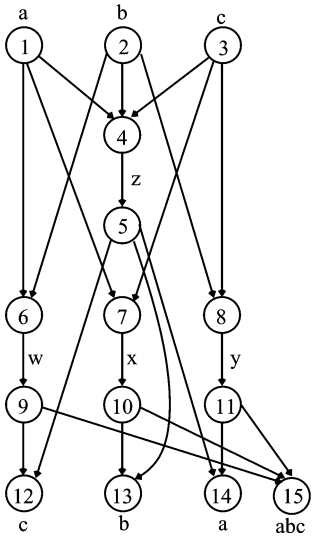


Fig. 9. The non-Fano network. Messages a, b, and c are emitted by sources n_1, n_2, and n_3, respectively, and are demanded by receivers n_{14}, n_{13}, and $\{n_{12}, n_{15}\}$, respectively. The edges $e_{4,5}$, $e_{6,9}$, $e_{7,10}$, and $e_{8,11}$ are labeled according to the network-matroid mapping by their corresponding ground set elements in the non-Fano matroid shown in Fig. 8.

TABLE III
DEPENDENCIES IN THE NON-FANO MATROID THAT ARE REFLECTED IN THE NON-FANO NETWORK. THE SECOND COLUMN INDICATES SETS OF VARIABLES IN THE NON-FANO NETWORK CORRESPONDING TO DEPENDENT SETS IN THE NON-FANO MATROID. THE THIRD COLUMN INDICATES AT WHICH NODES IN THE NON-FANO NETWORK THE CORRESPONDING DEPENDENCY IS ENFORCED

Step	Variables	Nodes	Type
1	$\{a, b, c\}$	n_1, n_2, n_3	message
2	$\{a, b, c, z\}$	n_4, n_5	circuit
2	$\{a, b, w\}$	n_6, n_9	circuit
2	$\{a, c, x\}$	n_7, n_{10}	circuit
2	$\{b, c, y\}$	n_8, n_{11}	circuit
3	$\{c, w, x\}$	n_{12}	circuit
3	$\{b, x, z\}$	n_{13}	circuit
3	$\{a, y, z\}$	n_{14}	circuit
4	$\{w, x, y\}$	n_{15}	independent set

solvable if and only if the alphabet size is odd.[9] It, in fact, has a linear solution over any alphabet of odd cardinality (by taking $w = a + b$, $x = a + c$, $y = b + c$, and $z = a + b + c$). The non-Fano network was used as a building block to construct a network whose coding capacity cannot be achieved by the network. The non-Fano network was also used as a building block in [5] to construct a solvable network that is not linearly solvable (in a very general sense).

[9]Actually, a slight variation of the non-Fano network was used in [6]; the variation consisted of removing the demands a and b from node n_{15}. However, the statements here about the solvability of the non-Fano network are true, since it can be shown that the non-Fano network is CSLS-equivalent (see Definition VII.6) to the variant network.

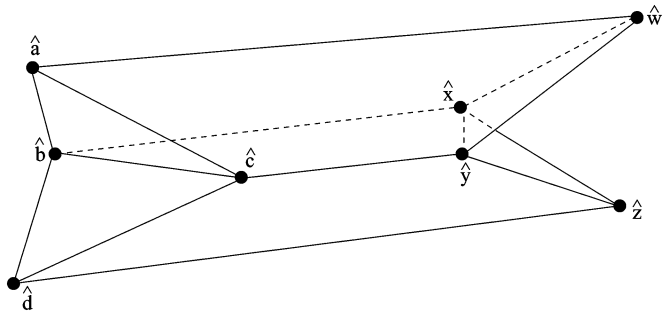


Fig. 10. A three-dimensional geometric depiction of the Vámos matroid.

F. The Vámos Network

The *Vámos matroid* is an eight-element rank-4 matroid $(\mathcal{S}, \mathcal{I})$ with

$$\mathcal{S} = \{\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{w}, \hat{x}, \hat{y}, \hat{z}\}$$

and whose dependent sets are the four-element sets which are coplanar in the three-dimensional drawing in Fig. 10 (i.e., precisely $\{\hat{b}, \hat{c}, \hat{x}, \hat{y}\}$, $\{\hat{a}, \hat{c}, \hat{w}, \hat{y}\}$, $\{\hat{a}, \hat{b}, \hat{w}, \hat{x}\}$, $\{\hat{c}, \hat{d}, \hat{y}, \hat{z}\}$, and $\{\hat{b}, \hat{d}, \hat{x}, \hat{z}\}$) and all subsets of $\mathcal{S}$ of cardinality at least 5. Note that $\{\hat{a}, \hat{d}, \hat{w}, \hat{z}\}$ is not considered a coplanar set in Fig. 10.

One of the interesting properties of the Vámos matroid is the following.

Theorem V.11: [26, p. 170] The Vámos matroid is not representable.

We call the network shown in Fig. 11 the *Vámos network*; it is a matroidal network associated with the Vámos matroid[10] and constructed using the technique described in Section B. The network has 17 nodes and 4 message variables. Nodes $n_9, \ldots, n_{13}$ are receiver nodes, each demanding one source message, except for n_{11}, which demands two source messages. The network has 4 hidden source nodes, each generating exactly one of the messages a, b, c, d. As depicted in Fig. 11, source messages are carried on hidden edges from their hidden source to various other network nodes (e.g., message c is carried by hidden edges from its hidden source to nodes n_1, n_5, n_7, n_{10}, and n_{12}).

The network-matroid mapping $f : \mu \cup \epsilon \to \mathcal{S}$ defined along with the network from the matroid in Fig. 11 is determined by: $f(u) = \hat{u}$ for all $u \in \{a, b, c, d, w, x, y, z\}$. Table IV lists the dependencies in the Vámos matroid which are directly reflected in the Vámos network.

> *Note 1:* As depicted in Fig. 11, several of the message variables a, b, c, d appear above some of the nodes. This is simply a convenience that makes the depiction easier to draw. When this happens, it is understood that there is an unshown edge from the appropriate source node to the node in question. So, for example, node n_1 actually has four in-edges (not shown), one from each source node (also not shown).

[10]It should be emphasized that there are many other networks that are associated with the Vámos matroid.

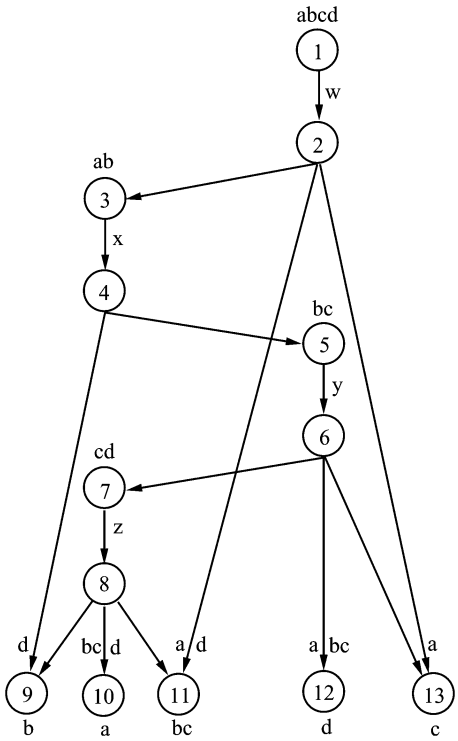


Fig. 11. The Vámos network. A message variable a, b, c, or d labeled above a node indicates an in-edge (not shown) from the source node (not shown) generating the message. Demand variables are labeled below the receivers n_9–n_{13} demanding them. The edges $e_{1,2}$, $e_{3,4}$, $e_{5,6}$, and $e_{7,8}$ are denoted by w, x, y, and z, respectively.

TABLE IV
DEPENDENCIES IN THE VÁMOS MATROID THAT ARE REFLECTED IN THE VÁMOS NETWORK. THE SECOND COLUMN INDICATES SETS OF VARIABLES IN THE VÁMOS NETWORK CORRESPONDING TO DEPENDENT SETS IN THE VÁMOS MATROID. THE THIRD COLUMN INDICATES AT WHICH NODES IN THE VÁMOS NETWORK THE CORRESPONDING DEPENDENCIES ARE ENFORCED

Step	Variables	Nodes	Type
1	$\{a,b,c,d\}$	hidden	message
2	$\{a,b,c,d,w\}$	n_1, n_2	circuit
2	$\{a,b,x,w\}$	n_3, n_4	circuit
2	$\{b,c,x,y\}$	n_5, n_6	circuit
2	$\{c,d,y,z\}$	n_7, n_8	circuit
3	$\{b,d,x,z\}$	n_9	circuit
3	$\{a,b,c,d,z\}$	n_{10}	circuit
3	$\{a,b,c,d,y\}$	n_{12}	circuit
3	$\{a,c,w,y\}$	n_{13}	circuit
4	$\{a,d,w,z\}$	n_{11}	independent set

Note 2: To obtain the Vámos network, Step 1 was used on the base $\{\hat{a},\hat{b},\hat{c},\hat{d}\}$, then Step 2 was used on the circuits $\{\hat{a},\hat{b},\hat{c},\hat{d},\hat{w}\}$, $\{\hat{a},\hat{b},\hat{w},\hat{x}\}$, $\{\hat{b},\hat{c},\hat{x},\hat{y}\}$, and $\{\hat{c},\hat{d},\hat{y},\hat{z}\}$, creating nodes $n_1,\ldots,n_8$. Step 3 was then used on the circuits $\{\hat{b},\hat{d},\hat{x},\hat{z}\}$, $\{\hat{a},\hat{b},\hat{c},\hat{d},\hat{z}\}$, $\{\hat{a},\hat{b},\hat{c},\hat{d},\hat{y}\}$, and $\{\hat{a},\hat{c},\hat{w},\hat{y}\}$, creating nodes n_9, n_{10}, n_{12}, n_{13}. Finally. Step 4 was used to create node n_{11} using the base $\{\hat{a},\hat{d},\hat{w},\hat{z}\}$. Note that Step 4 says that node n_{11} should demand all 4 messages a, b, c, d but since a and d are inputs to n_{11}, we did not bother demanding them there. This does not affect the matroidality of the network. Notice that Steps 3 and 4 are not used as much as they could have been. For example, $\{\hat{a},\hat{b},\hat{c},\hat{y},\hat{z}\}$ is a circuit in the Vámos matroid that is never used. Therefore, the Vámos network might not reflect all the properties of the Vámos matroid that it could have. The reason these stages were not completed was to make the network simpler, while still reflecting enough of the Vámos matroid to suit our purposes.

VI. BOUNDS ON CAPACITIES OF THE VÁMOS NETWORK

In general, the routing capacity of an arbitrary network can in principle be determined using a linear programming approach [3],[11] although the computational complexity can be prohibitive for even relatively small networks. It thus appears to be generally nontrivial to efficiently determine the routing capacity. In addition, there are presently no known techniques for computing the coding capacity or the linear coding capacity of an arbitrary network.[12] In fact, the linear coding capacity of a network depends, in general, on the finite-field alphabet used [5], whereas the routing capacity and coding capacity do not depend on the alphabet size [3]. However, somewhat surprisingly, the exact routing capacity and linear coding capacity of the Vámos network can be computed, and the linear coding capacity of the Vámos network turns out to be independent of the finite field alphabet.

In what follows, we first determine the polymatroid upper bound on the coding capacity of the Vámos network. Then we show that the upper bound on the coding capacity of the Vámos network can be improved if we allow the use of non-Shannon-type information inequalities. Specifically, we exploit the Zhang–Yeung non-Shannon-type information inequality given in Theorem II.2 and obtain a smaller upper bound on the coding capacity of the Vámos network than is obtainable using Shannon-type information inequalities. To the best of our knowledge, this is the first published application of a non-Shannon-type information inequality to network coding. Finally, we compute the exact routing capacity and the exact linear coding capacity of the Vámos network.

A. Bounds on Coding Capacity of the Vámos Network

Theorem VI.1: The polymatroid upper bound on the coding capacity of the Vámos network is 1.

Proof: By Lemma V.3, the polymatroid upper bound is greater than or equal to 1. Since there is a unique path in the Vámos network from the source node n_1 to the node n_{12} which demands message d, the bound cannot be greater than 1 (by Lemma V.6). □

The following theorem demonstrates that non-Shannon-type information inequalities can give tighter upper bounds on a network's capacity than can only Shannon-type information inequalities. In particular, Shannon-type information inequalities do not by themselves guarantee that the Vámos network is unsolvable, whereas adding one non-Shannon inequality indeed confirms the unsolvability of the network (since the coding capacity is strictly smaller than 1).

Theorem VI.2: The coding capacity of the Vámos network is at most $10/11$.

[11]This is analogous to the algorithm for achieving multicommodity flow capacity given in [9].

[12]An exception is for multicast networks, where it is known that the coding capacity equals the linear coding capacity and is computable [18].

Proof: Consider a (k, n) solution to the network. Assume that the network messages a, b, c, d are independent, k-dimensional, random vectors with uniformly distributed components and assume each edge in the network has capacity n. Let w, x, y, z denote the random variables carried by edges $e_{1,2}$, $e_{3,4}$, $e_{5,6}$, $e_{7,8}$, respectively. We have

$$H(y|b,c,x)=0 \quad \text{[from (N3) at } n_5\text{]} \tag{38}$$
$$H(b,c|a,d,w,z)=0 \quad \text{[from (N3) at } n_{11}\text{]} \tag{39}$$
$$H(a|b,c,d,z)=0 \quad \text{[from (N3) at } n_{10}\text{]} \tag{40}$$
$$H(b|d,x,z)=0 \quad \text{[from (N3) at } n_9\text{]} \tag{41}$$
$$H(z|c,d,y)=0 \quad \text{[from (N3) at } n_7\text{]} \tag{42}$$
$$H(d|a,b,c,y)=0 \quad \text{[from (N3) at } n_{12}\text{]} \tag{43}$$
$$H(x|a,b,w)=0 \quad \text{[from (N3) at } n_3\text{]} \tag{44}$$
$$H(c|a,w,y)=0 \quad \text{[from (N3) at } n_{13}\text{]} \tag{45}$$
$$H(w,x,y,z|a,b,c,d)=0 \quad \text{[from (1), Lemma III.6].} \tag{46}$$

Note that (46) can alternatively be obtained by seeing that w, x, y, z are deterministic functions of the network messages a, b, c, d. Then we obtain

$$\begin{aligned}
&2I(b,x;c,y)\\
&\quad=2H(c,y)+2H(b,x)\\
&\qquad-2H(b,c,x,y) && \text{[from (10)]}\\
&\quad=2H(c,y)+2H(b,x)\\
&\qquad-2H(b,c,x) && \text{[from (1), (38)]}\\
&\quad\geq 2H(c,y)-2H(c) && \text{[from(8)]}\\
&\quad=2H(c,y)-2k && \text{[from (N1)].}
\end{aligned} \tag{47}$$

$$\begin{aligned}
&I(d,z;a,w)\\
&\quad=H(d,z)+H(a,w)\\
&\qquad-H(a,d,w,z) && \text{[from (10)]}\\
&\quad=H(d,z)+H(a,w)\\
&\qquad-H(a,b,c,d,w,z) && \text{[from (1), (39)]}\\
&\quad=H(d,z)+H(a,w)\\
&\qquad-H(a,b,c,d) && \text{[from(46)]}\\
&\quad=H(d,z)+H(a,w)-4k && \text{[from (N1)]}
\end{aligned} \tag{48}$$

$$\begin{aligned}
&I(d,z;b,c,x,y)\\
&\quad=H(d,z)+H(b,c,x,y)\\
&\qquad-H(b,c,d,x,y,z) && \text{[from[(10)]}\\
&\quad=H(d,z)+H(b,c,x,y)\\
&\qquad-H(a,b,c,d,x,y,z) && \text{[from (1), (40), Lemma IV.5]}\\
&\quad=H(d,z)+H(b,c,x,y)\\
&\qquad-H(a,b,c,d) && \text{[from Lemma III.6]}\\
&\quad=H(d,z)+H(b,c,x)\\
&\qquad-H(a,b,c,d) && \text{[from (1), (38)]}\\
&\quad=H(d,z)+H(b,c,x)-4k && \text{[from (N1)]}\\
&\quad\leq H(d,z)+H(b)+H(c)+H(x)-4k && \text{[from (8)]}\\
&\quad\leq H(d,z)+(2k+n)-4k && \text{[from (N1), (N2)]}\\
&\quad=H(d,z)+n-2k
\end{aligned} \tag{49}$$

$$\begin{aligned}
&3I(b,x;c,y|d,z)\\
&\quad=3H(b,d,x,z)+3H(c,d,y,z)\\
&\qquad-3H(d,z)-3H(b,c,d,x,y,z) && \text{[from (11)]}\\
&\quad=3H(b,d,x,z)+3H(c,d,y,z)\\
&\qquad-3H(d,z)-3H(a,b,c,d,x,y,z)\\
&\qquad && \text{[from (1), (40), Lemma IV.5]}\\
&\quad=3H(b,d,x,z)+3H(c,d,y,z)\\
&\qquad-3H(d,z)-3H(a,b,c,d) && \text{[from Lemma III.6]}\\
&\quad=3H(d,x,z)+3H(c,d,y,z)\\
&\qquad-3H(d,z)-3H(a,b,c,d) && \text{[from (1), (41)]}\\
&\quad=3H(d,x,z)+3H(c,d,y)\\
&\qquad-3H(d,z)-3H(a,b,c,d) && \text{[from (1), (42)]}\\
&\quad\leq 2(H(d)+H(x)+H(z))\\
&\qquad+H(d,z)+H(x)\\
&\qquad+(H(c)+H(d)+H(y))\\
&\qquad+2H(c,y)+2H(d)\\
&\qquad-3H(d,z)-3H(a,b,c,d) && \text{[from (8)]}\\
&\quad\leq 2(k+2n)+H(d,z)+n\\
&\qquad+(2k+n)+2H(c,y)+2k\\
&\qquad-3H(d,z)-3(4k) && \text{[from (N1), (N2)]}\\
&\quad=2H(c,y)-2H(d,z)+6n-6k
\end{aligned} \tag{50}$$

$$\begin{aligned}
&I(b,x;c,y|a,w)\\
&\quad=H(a,b,x,w)+H(a,c,w,y)-H(a,w)\\
&\qquad-H(a,b,c,w,x,y) && \text{[from (11)]}\\
&\quad=H(a,b,x,w)+H(a,c,w,y)-H(a,w)\\
&\qquad-H(a,b,c,d,w,x,y) && \text{[from (1), (43), Lemma IV.5]}\\
&\quad=H(a,b,w,x)+H(a,c,w,y)-H(a,w)\\
&\qquad-H(a,b,c,d) && \text{[from Lemma III.6]}\\
&\quad=H(a,b,w)+H(a,c,w,y)-H(a,w)\\
&\qquad-H(a,b,c,d) && \text{[from (1), (44)]}\\
&\quad=H(a,b,w)+H(a,w,y)-H(a,w)\\
&\qquad-H(a,b,c,d) && \text{[from (1), (45)]}\\
&\quad\leq (H(a)+H(b)+H(w))\\
&\qquad+(H(a)+H(w)+H(y))\\
&\qquad-H(a,w)-H(a,b,c,d) && \text{[from (8)]}\\
&\quad\leq (2k+n)+(k+2n)\\
&\qquad-H(a,w)-4k && \text{[from (N1), (N2)]}\\
&\quad=3n-k-H(a,w).
\end{aligned} \tag{51}$$

Letting $A=(d,z)$, $B=(a,w)$, $C=(b,x)$, and $D=(c,y)$ in Theorem II.2 gives

$$\begin{aligned}
2I(b,x;c,y)\leq{}& I(d,z;a,w)+I(d,z;b,c,x,y)\\
&+3I(b,x;c,y|d,z)+I(b,x;c,y|a,w)
\end{aligned} \tag{52}$$

and then substituting (47)–(51) into (52) yields

$$\begin{aligned}
2H(c,y)-2k\leq{}& H(d,z)+H(a,w)-4k+H(d,z)+n-2k\\
&+2H(c,y)-2H(d,z)+6n-6k\\
&+3n-k-H(a,w)
\end{aligned} \tag{53}$$

or equivalently,

$$k/n \leq 10/11.$$

Therefore, the coding capacity of the Vámos network can be at most $10/11$. □

Corollary VI.3: Shannon-type information inequalities and the network entropy conditions (i.e., the polymatroid upper bound on the capacity of a network) are in general insufficient for determining the coding capacity of a network.

In the terminology used in [32] (see also [29]), Theorem VI.1 says that the LP bound gives an upper limit of 1 on the capacity of the Vámos network, while Theorem VI.2 says that the entropy bound $\mathcal{R}_{\text{out}}$ gives a strictly better upper limit on this capacity.

B. Routing Capacity of the Vámos Network

Theorem VI.4: The routing capacity of the Vámos network is $2/5$.

Proof: Consider any (k, n) routing solution to the Vámos network. The demands at nodes n_{10} and n_{12} require edges $e_{3,4}$ and $e_{5,6}$ to each carry all k components of messages a and d. The demand at node n_9 requires that each of the k components of message b be carried on at least one of the edges $e_{3,4}$ or $e_{5,6}$. Thus, at least one of the edges $e_{3,4}$ or $e_{5,6}$ must carry at least $k/2$ components of message b. Such an edge has capacity n and must carry a total of at least $2k + (k/2)$ message components, implying that $5k/2 \leq n$, or equivalently

$$k/n \leq 2/5.$$

Thus, the routing capacity is at most $2/5$.

We next give a routing code that achieves a rate of $2/5$. The code has message dimensions equal to $k = 2$ and edge capacities equal to $n = 5$. Let each of the messages' two components be denoted using subscripts 1 and 2. To describe the code, we list below the scalar components carried by decision-critical edges in the networkas follows:

$$\begin{aligned} e_{1,2} &: b_1, c, d \\ e_{3,4} &: b_1, a, d \\ e_{5,6} &: b_2, a, d \\ e_{7,8} &: b_2, a. \end{aligned}$$

It is straightforward to verify that the routing code implied by these conditions meets the networks' demands. □

C. Linear Coding Capacity of the Vámos Network

In this subsection, we will use a version of the Ingleton inequality for ranks of vector spaces to compute an upper bound for the linear capacity of the Vámos network. Then we will show that the upper bound can be acheived. For the reader's convenience, we will provide a proof of the Ingleton inequality (54) here. This proof is due to Hammer, Romashchenko, Shen, and Vereshchagin [12].

Definition VI.5: A random variable Z is said to be a *common information* for random variables X and Y if the following three conditions hold:

$$\begin{aligned} H(Z|X) &= 0 \\ H(Z|Y) &= 0 \\ H(Z) &= I(X;Y). \end{aligned}$$

The above definition can be found, for example, in [12, p. 461] (see also [11] and [33, p. 51]).

Lemma VI.6: Let A, B, C, D be random variables with a common information. If A and B have a common information, then

$$I(A,B) \leq I(A;B|C) + I(A;B|D) + I(C;D). \tag{54}$$

Proof: Let E be a common information for A and B; then

$$\begin{aligned} &I(A;B|C) + I(A;B|D) + I(C;D) \\ &\quad = I(A,E;B|C) + I(A,E;B|D) + I(C;D) \\ &\qquad\qquad [\text{from } H(E|A) = 0] \\ &\quad \geq I(E;B|C) + I(E;B|D) + I(C;D) \\ &\quad = H(E|C) + H(E|D) + I(C;D) \quad [\text{from } H(E|B) = 0] \\ &\quad \geq H(E) \quad [\text{from Lemma V.7, replacing} \\ &\qquad\qquad A, B, C, D \text{ by } C, D, E, E] \\ &\quad = I(A;B). \end{aligned}$$ □

Lemma VI.7: Let F be a finite field, let A and B be F-valued matrices with m columns, and let W be a random variable uniformly distributed over F^m. If $X = AW$ and $Y = BW$, then X and Y have a common information.

Proof: Let C be a matrix whose row space is the intersection of the row spaces of A and B, and let $Z = CW$. Since the row space of C is a subspace of the row space of A, the matrix

$$\begin{bmatrix} A \\ C \end{bmatrix}$$

has rank equal to the rank of A, and therefore by Lemma V.9

$$H(AW, CW) = H(AW)$$

and so

$$H(Z|X) = H(CW|AW) = 0.$$

Similarly, $H(Z|Y) = 0$. Now, let A', B', and C' denote the row spaces of A, B, and C, respectively. Note that the pointwise sum $A' + B'$ is the span of $A' \cup B'$. We have

$$\begin{aligned} H(Z) &= H(CW) \\ &= \dim(C') \cdot \log|F| \qquad [\text{from Lemma V.9}] \\ &= \dim(A' \cap B') \cdot \log|F| \\ &= \Big(\dim(A') + \dim(B') \\ &\qquad - \dim(A' + B')\Big) \cdot \log|F| \end{aligned}$$

$$
\begin{aligned}
&= \left(\mathsf{rank}\,(A) + \mathsf{rank}\,(B) \right. \\
&\qquad \left. - \mathsf{rank}\left(\begin{bmatrix} A \\ B \end{bmatrix} \right) \right) \cdot \log |F| \\
&= H(X) + H(Y) - H(X,Y) \\
&\qquad\qquad \text{[from Lemma V.9]} \\
&= I(X;Y).
\end{aligned}
$$

Thus, Z is a common information for X and Y. □

Theorem VI.8: The linear coding capacity of the Vámos network over every finite field is $5/6$.

Proof: First, we demonstrate that $5/6$ is an upper bound on the linear coding capacity of the Vámos network. Consider an arbitrary (k,n) linear solution over a finite field F for the Vámos network.

Since we are assuming a linear solution to the network, the symbol vector carried on any edge in the network is a linear combination of the network source messages. Since x and y are linear functions of the four network messages, each of which is uniformly distributed over F, we may apply Lemmas VI.7 and VI.6, by taking $m = 4k$, $A = (b,x)$, $B = (c,y)$, $C = (d,z)$, and $D = (a,w)$ to give

$$
\begin{aligned}
I(b,x;c,y) \le I(b,x;c,y|d,z) + I(b,x;c,y|a,w) \\
+ I(d,z;a,w). \quad (55)
\end{aligned}
$$

But we have:

$$
\begin{aligned}
I(b,x;c,y) &\ge H(c,y) - k && \text{[from (47)]} \\
I(b,x;c,y|d,z) &= H(d,x,z) + H(c,d,y) - H(d,z) \\
&\quad - H(a,b,c,d) && \text{[from (50)]} \\
&\le H(d) + H(x) + H(z) + H(c,y) \\
&\quad + H(d) - H(d,z) - H(a,b,c,d) \\
&\qquad\qquad \text{[from (8)]} \\
&\le H(c,y) - H(d,z) + 2n - 2k \\
&\qquad\qquad \text{[from (N1), (N2)]} \\
I(b,x;c,y|a,w) &\le 3n - k - H(a,w) && \text{[from (51)]} \\
I(d,z;a,w) &= H(d,z) + H(a,w) - 4k && \text{[from (48)].}
\end{aligned}
$$

Plugging these into (55)) gives $5n \ge 6k$, so

$$k/n \le 5/6.$$

Thus, the linear coding capacity of the Vámos network is at most $5/6$ over any finite field F.

Next, we demonstrate a $(5,6)$ linear network solution for the Vámos network, which thus establishes $5/6$ as a lower bound for the linear coding capacity. The solution is valid over any alphabet which is an Abelian group with operation $+$ (i.e., in particular, over any finite field alphabet). To describe the code, we list below the six scalar components carried by various edges in the network.

$e_{1,2}$:
$b_1 + d_1 + c_5$
$b_2 + d_2 + c_4$
$b_3 + d_3$
$b_4 + d_4$
$b_5 + d_5$
c_3

$e_{3,4}$:
$a_1 + d_1 + c_5$
$a_2 + d_2 + c_4$
$a_3 + d_3$
$a_4 + d_4$
$a_5 + d_5$
b_1

$e_{5,6}$:
$a_1 + b_1 + c_1 + d_1 + c_5$
$a_2 + b_2 + c_2 + d_2 + c_4$
$a_3 + b_3 + c_3 + d_3 + b_1$
$a_4 + b_4 + c_4 + d_4$
$a_5 + b_5 + c_5 + d_5$
b_2

$e_{7,8}$:
$a_1 + b_1 + c_1$
$a_2 + c_2$
$a_3 + b_3 + b_1$
$a_4 + b_4$
$a_5 + b_5$.

It is straightforward to verify that the network demands can be (linearly) met using this code. Thus, the linear coding capacity of the Vámos network is at least $5/6$ for any finite field alphabet. □

We note that an alternative method to obtain the upper bound of $5/6$ on the linear coding capacity of the Vámos network is to write each edge function and each decoding function as arbitrary linear combinations (with matrix coefficients) of their inputs, and then to use linear algebra to obtain an inequality that bounds the ratio k/n. This approach, however, appears to require substantially more calculations that the proof given above.

VII. Creating Multiple-Unicast Matroidal Networks

In [7], a technique was given for converting arbitrary networks into multiple-unicast networks. The conversion procedure preserves the solvability and linear solvability properties of the original network. In this section, we show that the conversion process also preserves the property of a network being matroidal. Then, in Section VIII, we use this conversion technique to create a multiple-unicast variation of the Vámos network which witnesses the insufficiency of using Shannon-type information inequalities for computing the coding capacity of a multiple-unicast network.

We want to convert a given network into an equivalent network in which each message has only one source and one receiver. Eliminating multiple sources for a given message is easy. Simply add a new node to be the sole source for this message, together with an edge from this node to each of the old sources of the message.

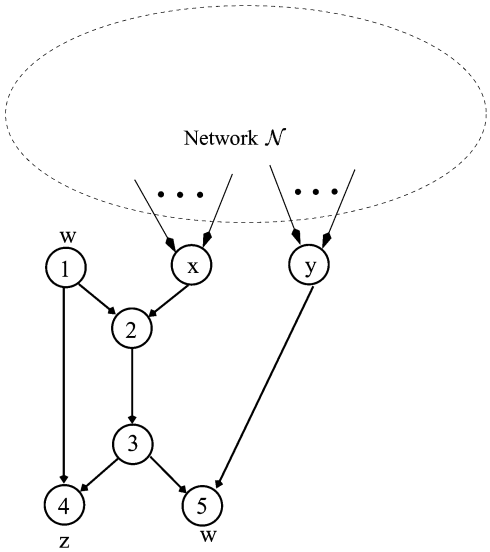


Fig. 12. The (n_x, n_y, z)-gadgetized version of $\mathcal{N}$ is the network $\mathcal{N}'$, built from network $\mathcal{N}$ by adding a gadget consisting of five new nodes with some incident edges, a new message w, and the message z being demanded at node n_4 instead of at nodes n_x and n_y.

Remark VII.1: When this elimination process is carried out on a matroidal network M, the resulting network is still matroidal. Indeed, if f is a map witnessing that M is matroidal, and m is the message for which multiple sources are being eliminated, we may simply extend f so that for each new edge e, $f(e) = f(m)$.

We eliminate multiple receivers by the following construction.

Definition VII.2: Let $\mathcal{N}$ be an arbitrary network with edges ϵ, nodes ν, and messages μ, such that nodes n_x and n_y each demand message z. Without loss of generality, let w be a new message not already in μ, and let $n_1, \ldots, n_5$ be new nodes not already in ν. The (n_x, n_y, z)-*gadgetized version* of $\mathcal{N}$ is the network $\mathcal{N}'$ with nodes

$$\nu \cup \{n_1, n_2, n_3, n_4, n_5\},$$

edges

$$\epsilon \cup \{e_{1,4}, e_{1,2}, e_{2,3}, e_{3,4}, e_{3,5}, e_{x,2}, e_{y,5}\},$$

messages

$$\mu \cup \{w\},$$

and with message z being demanded at n_4 instead of at n_x and n_y (as shown in Fig. 12).

The proof of the following lemma relies in large part on the material from Section VI-A.

Lemma VII.3: If a matroidal network has nodes n_x and n_y that each demand the same message z, then the (n_x, n_y, z)-gadgetized version of the network is also matroidal.

Proof: Suppose network $\mathcal{N}_1$ has edge set ϵ_1, node set ν_1, message set μ_1, with nodes $n_x, n_y \in \nu_1$ each demanding message $z \in \mu_1$, and suppose $\mathcal{N}_1$ is associated with the matroid $\mathcal{M}_1 = (\mathcal{S}_1, \mathcal{I}_1)$ with the network-matroid mapping $f_1 : \mu_1 \cup \epsilon_1 \to \mathcal{S}_1$. Let $\mathcal{N}$ be the (n_x, n_y, z)-gadgetized version of network $\mathcal{N}_1$. The network $\mathcal{N}$ has node set

$$\nu = \nu_1 \cup \{n_1, n_2, n_3, n_4, n_5\},$$

edge set

$$\epsilon = \epsilon_1 \cup \{e_{1,4}, e_{1,2}, e_{2,3}, e_{3,4}, e_{3,5}, e_{x,2}, e_{y,5}\},$$

and message set

$$\mu = \mu_1 \cup \{w\}$$

where $w \notin \mu_1$ and $w \notin \mathcal{S}_1$. Let $\mathcal{M}_2 = (\mathcal{S}_2, \mathcal{I}_2)$ be the matroid with ground set $\mathcal{S}_2 = \{u, w, f_1(z)\}$, where $u \notin \mathcal{S}_1$, and where the independent sets of $\mathcal{M}_2$ are the proper subsets of $\mathcal{S}_2$. Then, $\mathcal{M}_2$ is isomorphic to the uniform matroid $U_{2,3}$.

Let

$$T = \mathcal{S}_1 \cap \mathcal{S}_2 = \{f_1(z)\}.$$

Then the restriction matroids $\mathcal{M}_1|T$ and $\mathcal{M}_2|T$ are identical and are both isomorphic to the free matroid $U_{1,1}$, which is trivially modular. Thus, Lemma IV.7 implies that there exists a proper amalgam $\mathcal{M} = (\mathcal{S}, \mathcal{I})$ of $\mathcal{M}_1$ and $\mathcal{M}_2$. The matroid $\mathcal{M}$ has a ground set $\mathcal{S} = \mathcal{S}_1 \cup \{w, u\}$. (It can be shown that a set $X \subseteq \mathcal{S}$ is independent in the proper amalgam if and only if $X \cap \mathcal{S}_1$ is independent in $\mathcal{M}_1$, $X \cap \mathcal{S}_2$ is independent in $\mathcal{M}_2$, and either $(X \cup \{f_1(z)\}) \cap \mathcal{S}_1$ is independent in $\mathcal{M}_1$ or $(X \cup \{f_1(z)\}) \cap \mathcal{S}_2$ is independent in $\mathcal{M}_2$.)

Define a mapping $f : \mu \cup \epsilon \to \mathcal{S}$ by

$$f(v) = \begin{cases} f_1(v), & \text{if } v \in \mu_1 \cup \epsilon_1 \\ w, & \text{if } v \in \{w, e_{1,2}, e_{1,4}\} \\ u, & \text{if} \in \{e_{2,3}, e_{3,4}, e_{3,5}\} \\ f_1(z), & \text{if } v \in \{e_{x,2}, e_{y,5}\}. \end{cases}$$

We know that conditions (M1)–(M3) are satisfied by f_1 for $\mathcal{N}_1$; we will now show that these conditions are also satisfied by f for $\mathcal{N}$ (thus implying $\mathcal{N}$ is matroidal).

- Since $f(w) = w \notin \mu_1$ and f_1 is one-to-one on μ_1, the mapping f must be one-to-one on μ, and therefore condition (M1) is satisfied by f.
- Let r, r_1, and r_2 be the rank functions of $\mathcal{M}, \mathcal{M}_1$, and $\mathcal{M}_2$, respectively. For any $Y \subseteq \mathcal{S}$, if $f(\mu) \subseteq Y$, then

$$\begin{aligned} \eta(Y) &= r_1(Y \cap \mathcal{S}_1) + r_2(Y \cap \mathcal{S}_2) - r(Y \cap T) && \text{[from (21)]} \\ &\geq r_1(f(\mu) \cap \mathcal{S}_1) + r_2(f(\mu) \cap \mathcal{S}_2) - r(Y \cap T) \\ & && \text{[from } f(\mu) \subseteq Y \text{ and (R2)]} \\ &= r_1(f_1(\mu_1)) + r_2(\{w, f_1(z)\}) - r(\{f_1(z)\}) \\ &= |f(\mu_1)| + 2 - 1 && \text{[from (23)]} \\ &= |\mu_1| + 1 && \text{[from (M1)]} \\ &= |\mu|. && (56) \end{aligned}$$

Thus

$$\begin{aligned} |f(\mu)| &= |\mu| && \text{[from (M1)]} \\ &\leq r(f(\mu)) && \text{[from (22), (56)]} \\ &\leq |f(\mu)| && \text{[from (R1)]} \end{aligned}$$

so, by Lemma IV.3(a), $f(\mu) \in \mathcal{I}$. This shows that f satisfies condition (M2).

- Note that at node n_x, we have

$$\begin{aligned} r(f(\mathrm{In}(n_x) \cup e_{x,2})) &= r(f(\mathrm{In}(n_x)) \cup f_1(z)) \\ &= r_1(f(\mathrm{In}(n_x)) \cup f_1(z)) \\ &= r_1(f(\mathrm{In}(n_x))) \qquad \text{[from (M3)]} \end{aligned}$$

(using that fact that message z was demanded at n_x in $\mathcal{N}_1$) and the same reasoning holds at node n_y. Likewise, at node n_2, we have

$$\begin{aligned} r(f(\mathrm{In}(n_2) \cup e_{2,3})) &= r(f(\{e_{1,2}, e_{x,2}, e_{2,3}\})) \\ &= r(\{w, f_1(z), u\}) \\ &= r_2(\{w, f_1(z), u\}) \\ &= 2 \\ &= r_2(\{w, f_1(z)\}) \\ &= r_2(f(\mathrm{In}(n_2))) \\ &= r(f(\mathrm{In}(n_2))) \end{aligned}$$

and the same reasoning holds at nodes n_4 and n_5. Thus, condition (M3) holds for f, and so f is a network-matroid mapping for $\mathcal{N}$ and $\mathcal{M}$. □

Definition VII.4: A *multiple-unicast version* of a network $\mathcal{N}$ is a network constructed from $\mathcal{N}$ by eliminating multiple sources as described earlier and then repeatedly applying the construction in Lemma VII.3 until every message is demanded by exactly one node.

The following theorem follows immediately, by induction, from Remark VII.1 and Lemma VII.3.

Theorem VII.5: Every multiple-unicast version of a matroidal network is matroidal.

The following definition was given in [7]. ("CSLS" stands for "coding solvability and linear solvability.")

Definition VII.6: Two networks $\mathcal{N}$ and $\mathcal{N}'$ are *CSLS-equivalent* if the following two conditions hold.

1) For any alphabet $\mathcal{A}$, $\mathcal{N}$ is solvable over $\mathcal{A}$ if and only if $\mathcal{N}'$ is solvable over $\mathcal{A}$.
2) For any finite field F and any positive integer k, $\mathcal{N}$ is vector solvable over F in dimension k if and only if $\mathcal{N}'$ is vector solvable over F in dimension k.

Lemma VII.7: [7] Every multiple-unicast version of a network is CSLS-equivalent to that network.

The following corollary follows immediately from Theorem VII.5 and Lemma VII.7.

Corollary VII.8: Every matroidal network is CSLS-equivalent to a multiple-unicast matroidal network.

VIII. A Multiple-Unicast Version of the Vámos Network

The algorithm given by Adler, Harvey, Kleinberg, Jain, and Rasala Lehman [1], [14] for computing coding capacity bounds of networks applies as stated only to multiple-unicast networks.

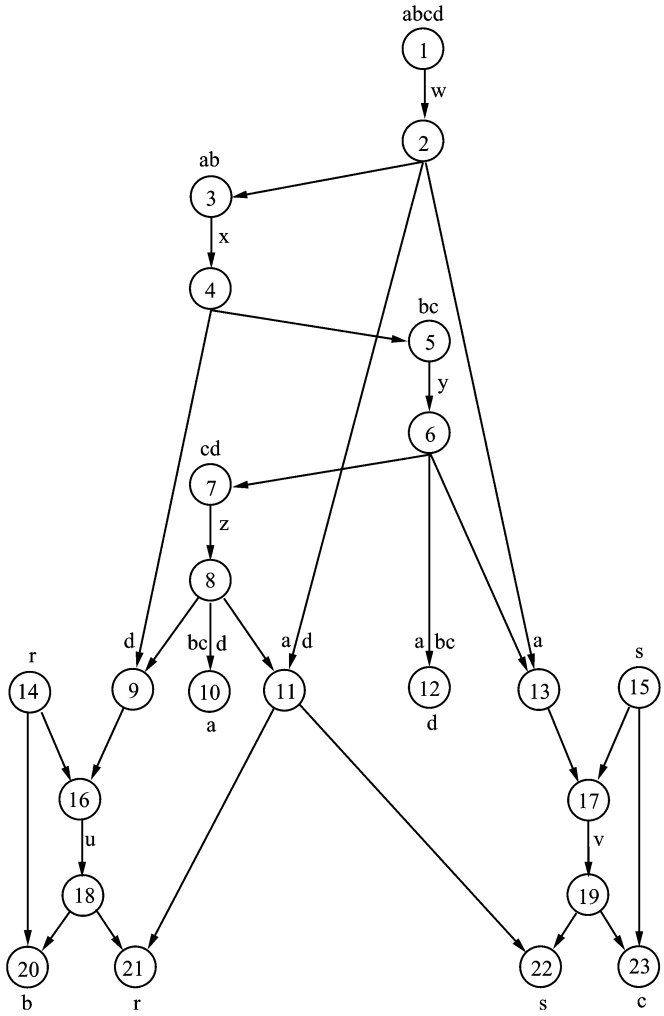


Fig. 13. The Multiple-Unicast Vámos network. This network was constructed by adding to the bottom of the Vámos network, two gadgets consisting of the 10 nodes $n_{14}, \ldots, n_{23}$ and their adjacent edges. Two new source messages, r and s, have been added, the demand b at node n_9 has been deleted, the demand c at node n_{13} has been deleted, the demands b and c at node n_{11} have been deleted, and new demands at nodes $n_{20}, \ldots, n_{23}$ have been added. The edges $e_{16,18}$ and $e_{17,19}$ are denoted by u and v, respectively.

Since the Vámos network is not multiple-unicast, one might wonder whether Shannon-type information inequalities suffice for computing the best possible coding capacity bounds for such networks.

We give (via Theorem VII.5) a multiple-unicast version of the Vámos network and show that the best possible upper bound on its coding capacity produced by Shannon-type information inequalities is strictly larger than the coding capacity. A consequence of this result is the fact that the algorithm given in [1], [14] is not optimal, in the sense that it cannot always yield the best possible coding capacity bound. The construction of a multiple-unicast network from the Vámos network is based upon a method described in [7].

We refer to the network in Fig. 13 as the Multiple-Unicast Vámos network. Clearly, using Lemma V.6 one can see that the coding capacity of the Multiple-Unicast Vámos network is at most 1 since, for example, there is a unique path from node n_{14} (generating message r) to node n_{21} (demanding message r).

Theorem VI.2 showed that the coding capacity of the Vámos network is at most $10/11$. Theorem VIII.2 gives an analogous upper bound (slightly larger than $10/11$ but still less than 1) for the Multiple-Unicast Vámos network.

Lemma VIII.1: The Multiple-Unicast Vámos network is matroidal.

Proof: It follows immediately from Lemma VII.3 and the fact that the Vámos network is matroidal (see Section V-F). □

Theorem VIII.2: The coding capacity of the Multiple-Unicast Vámos network is at most $12/13$.

Proof: The proof is the same as that of Theorem VI.2, except for a few changes which we will describe. To obtain (50), we used the fact that $H(b|d,x,z)=0$ (at node n_9). To see that $H(b|d,x,z)=0$ still holds in the Multiple-Unicast Vámos network, note that

$$H(b,d,x,z|a,b,c,d)=0 \quad \text{[from Lemma III.6]} \tag{57}$$

and hence

$$\begin{aligned}
0 &\le H(b|d,x,z) && \text{[from (6)]}\\
&= H(b|d,r,x,z)+I(b;r|d,x,z) && \text{[from (3)]}\\
&= I(b;r|d,x,z) && \text{[from Lemma III.6]}\\
&\le I(b;r|d,x,z)+I(r;d,x,z) && \text{[from (7)]}\\
&= I(r;b,d,x,z) && \text{[from (12)]}\\
&\le I(r;a,b,c,d)+H(r|r)\\
&\quad +H(b,d,x,z|a,b,c,d) && \text{[from Lemma V.9]}\\
&= I(r;a,b,c,d) && \text{[from def. of } H(r|r) \text{ and (57)]}\\
&= 0 && \text{[from indep. of } a,b,c,d,r\text{]}. \qquad (58)
\end{aligned}$$

To obtain (51) in the proof of Theorem VI.2, we used the fact that $H(c|a,w,y)=0$ (at node n_{13}). A similar argument as above shows that $H(c|a,w,y)=0$ still holds in the Multiple-Unicast Vámos network.

To obtain (48) in the proof of Theorem VI.2, we used the fact that $H(b,c|a,d,w,z)=0$ (at node n_{11}). However, this condition no longer holds in the Multiple-Unicast Vámos network. Instead, we can obtain the following:

$$\begin{aligned}
0 &\le H(r,s|a,d,u,v,w,z) && \text{[from (6)]}\\
&\le H(r|a,d,u,w,z)\\
&\quad +H(s|a,d,v,w,z) && \text{[from (8), (9)]}\\
&= 0 && \text{[from Lemma III.6 at } n_{21},\ n_{22}\text{]} \qquad (59)\\
0 &\le H(b,c|r,s,u,v) && \text{[from (6)]}\\
&\le H(b|r,u)+H(c|s,v) && \text{[from (8), (9)]}\\
&= 0 && \text{[from Lemma III.6 at } n_{20},\ n_{23}\text{]}. \qquad (60)
\end{aligned}$$

Then we have

$$\begin{aligned}
0 &\le H(b,c,r,s|a,d,u,v,w,z) && \text{[from (6)]}\\
&= H(r,s|a,d,u,v,w,z)\\
&\quad +H(b,c|a,d,r,s,u,v,w,z) && \text{[from (1)]}\\
&\le H(r,s|a,d,u,v,w,z)\\
&\quad +H(b,c|r,s,u,v) && \text{[from (9)]}\\
&= 0 && \text{[from (59) and (60)]}
\end{aligned}$$

which implies

$$H(b,c,r,s|a,d,u,v,w,z)=0. \tag{61}$$

In the proof of Theorem VI.2, we used the fact that

$$H(a,d,w,z)=H(a,b,c,d)=4k$$

to obtain (48). Alternatively, however, in the Multiple-Unicast Vámos network, we can use the inequality

$$\begin{aligned}
H(a,d,w,z) &\ge H(a,d,u,v,w,z)-H(u,v) && \text{[from (8)]}\\
&= H(a,b,c,d,r,s,u,v,w,z)-H(u,v)\\
&\quad -H(b,c,r,s|a,d,u,v,w,z) && \text{[from (1)]}\\
&= H(a,b,c,d,r,s,u,v,w,z)\\
&\quad -H(u,v) && \text{[from (61)]}\\
&\ge H(a,b,c,d,r,s)-H(u,v) && \text{[from (9)]}\\
&\ge H(a,b,c,d,r,s)-2n && \text{[from (N2)]}\\
&= 6k-2n && \text{[from (N1)]} \qquad (62)
\end{aligned}$$

to obtain

$$\begin{aligned}
&I(d,z;a,w)\\
&\quad = H(d,z)+H(a,w)-H(a,d,w,z) && \text{[from (10)]}\\
&\quad \le H(d,z)+H(a,w)+2n-6k && \text{[from (62)]}
\end{aligned}$$

as a replacement for (48). This then results in an extra $2n-2k$ on the right-hand side of (53), which in turn implies

$$k/n \le 12/13.$$

Thus, the coding capacity of the Multiple-Unicast Vámos network can be at most $12/13$. □

IX. Open Questions

The exact coding capacity of the Vámos network remains an open question. In particular, is the coding capacity of the Vámos network strictly greater than its linear coding capacity?

Can the network-matroid construction be modified so that all matroid dependencies are reflected in the network?

Acknowledgment

The authors thank Raymond Yeung and Ying-On Yan for their efforts in making available the Information Theoretic Inequality Prover (ITIP). The ITIP software, together with significant human processing, led to the proof of Theorem VI.2. We thank Nick Harvey and April Rasala Lehman for confirming that their algorithm in [14] can be derived from network inequalities and purely Shannon-type information inequalities. We thank Salim El Rouayheb, Costas Georghiades, and Alexander Sprintson for providing a copy of some of their work on network coding. Finally, we thank the two anonymous reviewers and the Associate Editor Gerhard Kramer for providing very thorough suggestions for improving the manuscript.

References

[1] M. Adler, N. J. A. Harvey, K. Jain, R. D. Kleinberg, and A. L. Rasala, "On the capacity of information networks," in *Proc. ACM-SIAM Symp. Discrete Algorithms (SODA 06)*, Miami, FL, Jan. 2006.

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

[3] J. Cannons, R. Dougherty, C. Freiling, and K. Zeger, "Network routing capacity," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 777–788, Mar. 2006.

[4] T. H. Chan and R. W. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1992–1995, Jul. 2002.

[5] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.

[6] R. Dougherty, C. Freiling, and K. Zeger, "Unachievability of network coding capacity," *IEEE Trans. Inf. Theory* , vol. 52, no. 6, pp. 2365–2372, Jun. 2006.

[7] R. Dougherty and K. Zeger, "Nonreversibility and equivalent constructions of multiple-unicast networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5067–5077, Nov. 2006.

[8] S. E. Rouayheb, C. N. Georghiades, and A. Sprintson, Dec. 2005, personal communication.

[9] L. R. Ford Jr. and D. R. Fulkerson, *Flows in Networks*. Princeton, NJ: Princeton Univ. Press, 1962.

[10] S. Fujishige, "Polymatroidal dependence structure of a set of random variables," *Inf. Contr.*, vol. 39, pp. 55–72, 1978.

[11] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Contr. Inf. Theory*, vol. 2, no. 2, pp. 149–162, 1973.

[12] D. Hammer, A. E. Romashchenko, A. Shen, and N. K. Vereshchagin, "Inequalities for Shannon entropy and Kolmogorov complexity," *J. Comp. Syst. Sci.*, vol. 60, pp. 442–464, 2000.

[13] N. J. A. Harvey and A. R. Lehman, Dec. 2005, personal communication.

[14] N. J. A. Harvey, R. D. Kleinberg, and A. R. Lehman, "On the capacity of information networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2345–2364, Jun. 2006.

[15] G. Kramer and S. A. Savari, "Edge-cut bounds on network coding rates," *J. Network Syst. Manag.*, vol. 14, no. 1, pp. 49–67, Mar. 2006.

[16] A. W. Ingleton, *Representation of Matroids in Combinatorial Mathematics and Its Applications*, D. J. A. Welsh, Ed. London, U.K.: Academic, 1971, pp. 149–167.

[17] A. L. Rasala and E. Lehman, "Network information flow: Does the model need tuning?," in *Proc. Symp. Discrete Algorithms (SODA)*, Vancouver, BC, Canada, Jan. 2005, pp. 499–504.

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

[19] R. Lněnička, "On the tightness of the Zhang-Yeung inequality for Gaussian vectors," *Commun. Inf. Syst.*, vol. 3, no. 1, pp. 41–46, Jun. 2003.

[20] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Commun. Inf. Syst.*, vol. 2, no. 2, pp. 147–166, Dec. 2002.

[21] F. Matúš, "Extreme convex set functions with many nonnegative differences," *Discr. Math.*, vol. 135, pp. 177–191, 1994.

[22] F. Matúš, "Conditional independences among four random variables III: Final conclusion," *Comb., Probab., Comput.*, vol. 8, pp. 269–276, 1999.

[23] F. Matúš, "Inequalities for Shannon entropies and adhesivity of polymatroids," in *Proc. Canadian Workshop on Information Theory*, Montreal, QC, Canada, 2005, pp. 28–31.

[24] F. Matúš and M. Studený, "Conditional independence among four random variables I," *Comb., Probab., Comput.*, vol. 4, pp. 269–278, 1995.

[25] M. Médard, M. Effros, T. Ho, and D. Karger, "On coding for non-multicast networks," in *Proc. 41st Annu. Allerton Conf. Communication Control and Computing*, Monticello, IL, Oct. 2003.

[26] J. G. Oxley, *Matroid Theory*. New York: Oxford Univ. Press, 1992.

[27] I. Satake, *Linear Algebra*. New York: Marcel Dekker, 1975.

[28] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Oct. 1948.

[29] L. Song, R. W. Yeung, and N. Cai, "Zero-error network coding for acyclic networks," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3129–3139, Dec. 2003.

[30] D. J. A. Welsh, *Matroid Theory*. London, U.K.: Academic, 1976.

[31] R. Yeung, R. S.-Y. Li, N. Cai, and Z. Zhang, *Foundations and Trends in Communication and Information Theory: Network Coding Theory*. Boston, MA: Now Publishers, 2006.

[32] R. W. Yeung, *A First Course in Information Theory*. Norwell, MA: Kluwer, 2002.

[33] Z. Zhang, "On a new non-Shannon type information inequality," *Commun. Inf. Syst.*, vol. 3, no. 1, pp. 47–60, Jun. 2003.

[34] Z. Zhang and R. W. Yeung, "A non-Shannon-type conditional inequality of information quantities," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1982–1985, Nov. 1997.

[35] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Inf. Theory*, vol. 44, no. 4, pp. 1440–1452, Jul. 1998.