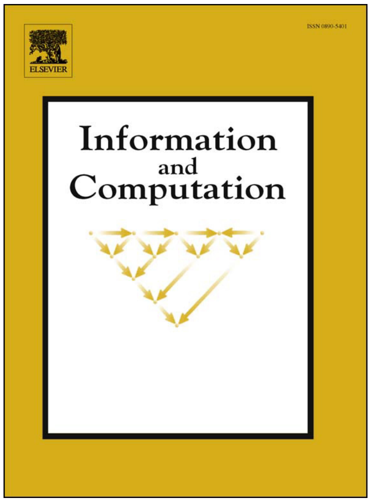
Max-flow min-cut theorems on dispersion and entropy measures for communication networks

Søren Riis, Maximilien Gadouleau



PII: S0890-5401(19)30022-7
DOI: https://doi.org/10.1016/j.ic.2019.03.004
Reference: YINCO 4421

To appear in: *Information and Computation*

Received date: 15 October 2014
Revised date: 19 September 2017

Please cite this article in press as: S. Riis, M. Gadouleau, Max-flow min-cut theorems on dispersion and entropy measures for communication networks, *Inf. Comput.* (2019), https://doi.org/10.1016/j.ic.2019.03.004

Max-Flow Min-Cut Theorems on Dispersion and Entropy Measures for Communication Networks

Søren Riis[1,*]

School of Electronic Engineering and Computer Science, Queen Mary University of London.

Maximilien Gadouleau

Department of Computer Science, Durham University.

Abstract

The paper presents four distinct new ideas and results for communication networks:

1) We show that relay-networks (i.e. communication networks where different nodes use the same coding functions) can be used to model dynamic networks, in a way, vaguely akin to Kripke's *possible worlds* from logic and philosophy. Link failures, point failures, changes in network topology during transmission, changes in receivers demands etc. can all be modelled, in a discrete fashion, by considering a multiverse where different possible worlds (hypothetical situations) are modelled as worlds existing in parallel. Nodes with the same labels might represent nodes in parallel worlds that behave, in the same way, as they are unaware of which world becomes the actual world.

2) We introduce *the term model*, which is a simple, graph-free symbolic approach to communication networks. We use the term model to create an algorithm (based on the max-flow min-cut algorithm) that calculates the information-theoretic limit for the capacity of a given communication network. We notice that different non-isomorphic communication networks might never-the-less be mathematically identical because they lead to the same term model. We illustrate the power of our formalism through a number of examples.

3) We state and prove variants of a theorem concerning the dispersion of information in single-receiver communications. The dispersion theorem resembles the max-flow min-cut theorem for commodity networks and states that the minimal cut value (the channel capacity) can be achieved asymptotically. The theorem is proved by combining Menger's theorem from graph theory with an argument that is similar to the proof of Shannon fundamental theorem for memoryless noisy channels. To prove the theorem we introduce a very weak kind

[*] Corresponding author

Email addresses: `s.riis@qmul.ac.uk` (Søren Riis), `m.r.gadouleau@durham.ac.uk` (Maximilien Gadouleau)

[1] This work was supported by EPSRC ref: EP/H016015/1.

of network coding (network coding lite), which we will refer to as routing with dynamic headers.

4) We show that the solvability of an abstract multi-user communication problem is equivalent to the solvability of a single-target communication in a suitable relay network.

In the paper, we develop a number of technical ramifications of these ideas and results. One technical result is a max-flow min-cut theorem for the Rényi entropy with order less than one, given that the sources are equiprobably distributed; conversely, we show that the max-flow min-cut theorem fails for the Rényi entropy with order greater than one. We leave the status of the theorem with regards to the ordinary Shannon Entropy measure (Rényi entropy of order one and the limit case between validity or failure of the theorem) as an open question. In non-dynamic static communication networks with a single receiver, a simple application of Menger's theorem shows that the optimal throughput can be achieved without proper use of network coding i.e. just by using ordinary packet-switching. This fails dramatically in relay networks with a single receiver. We show that even a powerful method like linear network coding fails miserably for relay networks. With that in mind, it is noticeable that our rather weak form of network coding (routing with dynamic headers) is asymptotically sufficient to reach capacity.

Keywords: Communication networks, flow, information theory, logic, entropy, dispersion

1. Introduction

There is an extensive literature on commodity networks and these play a central role in an array of applications: traffic routing, urban planning, scheduling of freight delivery, economical networks, communication through packet switching and much more. Though a typical commodity network has multiple sources and multiple sinks, the central theorem in the field - the max-flow min-cut theorem - only concerns the special case of commodity networks with a single source and a single sink.

Digital information in communication networks is radically different form ordinary commodities as it can be modified, copied or mixed during transmission. This idea goes beyond traditional routing and lays the foundation of Network Coding [1, 2]. In Network Coding intermediate nodes can combine the packets they receive and retransmit the combined versions towards their destinations. An instance of Network Coding is given by a network with prescribed sets of sources and of destinations, where the destinations request messages sent by the sources. In this paper we generalise network coding networks to relay networks (transfer networks) that are networks where different nodes might be required to use the same coding function. Our main result is a theorem that mathematically plays a similar role in the theory for relay networks as the max-flow min-cut theorem plays in the theory of commodity networks.

The general problem of determining whether all the demands of the receiver nodes in a communication network can be satisfied simultaneously has been widely studied [3, 4] since the discovery of network coding. The problem of solvability of multi-user communications exhibits many pathological examples, for instance there exist communication problems which are asymptotically solvable but not solvable for any finite alphabet [5]. Different methods have been proposed to determine whether a problem is solvable, including graph entropy [4] and guessing games [6].

1.1. Relay networks for communication networks

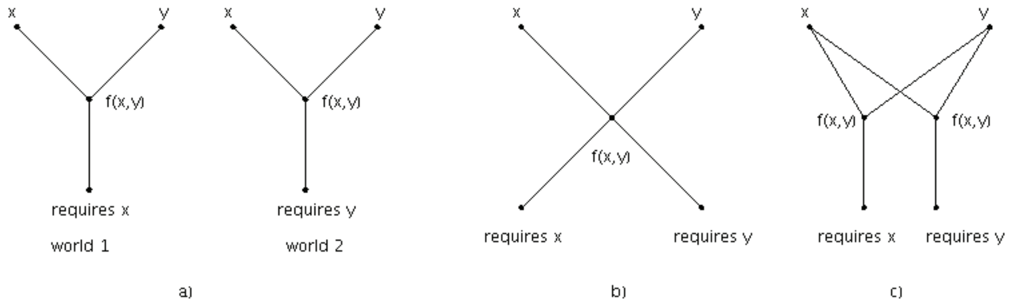


Figure 1: The receiver node requires message x or message y, but the internal node does not know which is the case. The dynamic communication problem is equivalent to the relay network in Figure 1b as well as the relay network in Figure 1c.

Let us consider a few basic examples that introduce dynamic communication networks and illustrate their link to relay networks. This will be properly formalised in Sections 7 and 8.

Example 1. *In Figure 1 we consider a dynamic network with a single receiver. This receiver might require the message x (world 1) or might require the message y (world 2). As usual in the literature on network coding we assume that the messages x and y are selected from some finite message space (finite alphabet A), that $f : A \times A \to A$ is a suitable coding function and that $f(x, y)$ is some message selected from A. And we assume that each edge has unit capacity i.e. allows the transmission of one message (e.g. a block of symbols)* [2]*. The communication problem in Figure 1 is unsolvable in the sense that it is impossible to satisfy the receivers demands (at a full rate where each receiver, receives one message in each step). The relay networks in Figure 1b and Figure 1c illustrate how the dynamic network can be represented as relay networks.*

[2] For a more detailed account see [1] or any other standard text on Network Coding

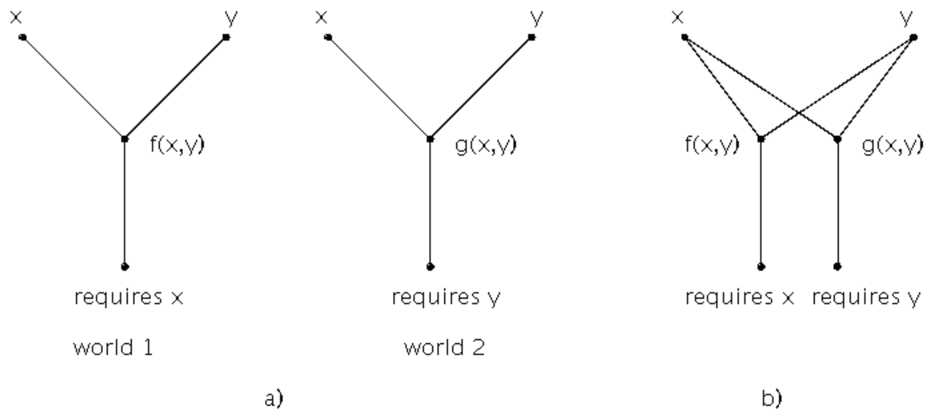


Figure 2: The receiver requires message x or message y and the internal node knows which is the case and can choose its coding function accordingly. The dynamic communication problem is equivalent to the relay network in Figure 2b.

Example 2. *In Figure 2, we consider the same network, however here we assume that the inner node knows the requirement of the receiver, so it can choose the routing accordingly by either applying the coding function $f : A \times A \to A$ given by $f(x, y) = x$ or by applying the coding function $g : A \times A \to A$ given by $g(x, y) = y$ depending on the situation. The situation can be represented as a static communication problem with two receivers as seen in Figure 2b.*

Example 3. *In Figure 3 the receiver has a preference between the messages so message z is preferred to message x that in turn is preferred to message y. The sender nodes are aware of this preference, but the inner node is oblivious to the situation and always broadcasts a message determined by a fixed coding function $f : A \times A \to A$.*

Example 4. *In Figure 4 there are two receivers. World 1 is the butterfly network that can be solved using network coding. In world 2 there is a link failure, and in conjunction with this link failure an important message z is being broadcast to both receivers. The inner node is not concerned with the situation and uses the same coding function $f : A \times A \to A$ in both cases. The communication problem is equivalent to the communication problem given by the relay network with 4 receivers given in Figure 4b.*

1.2. Term sets and relay networks

Mathematically, the problem in Example 4 can be described as the task of constructing a (coding) function $f : A \times A \to A$ such that $x \in A$ can be

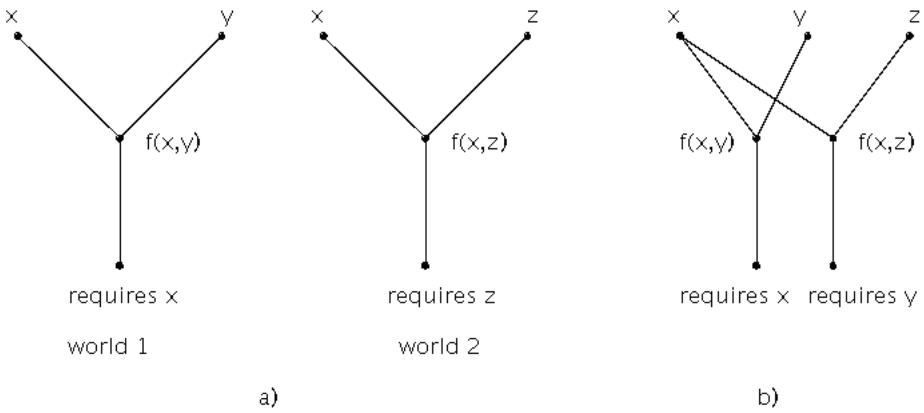


Figure 3: The receiver prefers message z to x and is the least interested in message y. The internal node does not know the preferences for the receiver.

reconstructed from the value $f(x, y) \in A$, and $z \in A$ can be reconstructed from the value $f(x, z) \in A$. Explicitly, we are looking for (decoding) functions $h_1, h_2 : A \to A$ such that the *term equations* $h_1(f(a_1, a_2)) = a_1$ and $h_2(f(a_1, a_2)) = a_2$ hold for all $a_1, a_2 \in A$. While this is impossible for finite alphabets, asymptotically for large alphabets A (e.g. by considering long blocks of bits) we can do very well by sending messages x, y and z such that a small *header* part of the messages indicates if the *body* of message is x, y or z [3]. The coding function f simply outputs the value of x or z according to this header information.

In general it turns out that a given communication problem given by a relay network can be expressed by a set of term equations. This graph-free symbolic approach cuts right to the mathematical part of the given communication problem. Non-isomorphic relay networks (e.g. the ones in Figure 1b and Figure 1c) might lead to the same set of term equations. One of the key points in working with term equations rather than the graphs is the flexibility this gives. We can for example translate the relay problem in Figure 3b that has two receiver nodes, to a problem that - like the dynamic version of it in Figure 3a - has only one receiver. In this problem the receiver has access to the terms $f(x, y)$ and $f(x, z)$ and requires the values of x and z. Or explicitly, the task:

(1) *Find functions f, h_1 and h_2 such that $h_1(f(x, y)) = x$ and $h_2(f(x, z)) = z$ hold for all $x, y, z \in A$.*

[3] In the actual example one bit suffices that is 0 for messages x and y and is 1 for message z

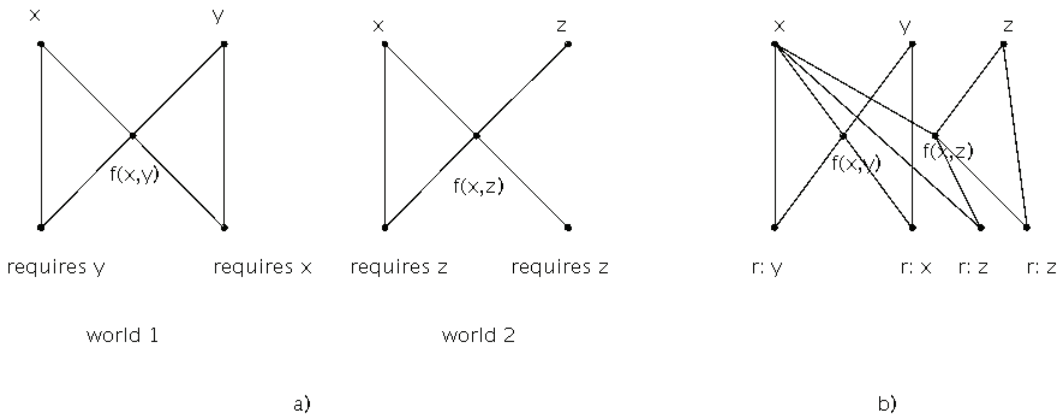


Figure 4: The receiver node requires message x or message y, but the internal node does not know which is the case. The relay network in b) is equivalent to the dynamic communication problem in a).

This task is logically equivalent to (if we only are concerned with solvability)

(2) *Find functions* f, h_1 *and* h_2 *such that* $h_1(f(x,y)) = x$ *and* $h_2(f(x,y)) = y$ *holds for all* $x, y \in A$.

Notice that the communication problem in Figure 1b gives rise to exactly the same task as (2). In general we are interested not just in solvability or unsolvability, but also in how much information that can be transmitted to the receivers. With regard to that type of question, the term equations in (1) and (2) are distinct.

Similarly, we can translate the problem in Figure 4b - a network communication problem with four receiver nodes - to a problem with only a single receiver with access to the terms $y_1, f(x_1, y_1), x_2, f(x_2, y_2)$ and $f(x_3, z)$ that requires the messages x_1, y_2 and z. More specifically the task is to

(3) *Find functions* h_1, h_2, h_3 *such that* $h_1(y_1, f(x_1, y_1)) = x_1, h_2(x_2, f(x_2, y_2)) = y_2$ *and* $h_3(f(x_3, z), x_3) = z$ *for all* $x_1, y_1, x_2, y_2, x_3, z \in A$.

While task (3) can not be achieved, it can asymptotically be achieved with regards to some measures as we can divide the message space into two parts: One part for messages x and y and one part for messages z. If messages are transmitted as blocks of bits, we can use one of the bits as a *flag* that controls the routing at the inner node. The remaining bits can then be used freely to send the body of the messages x, y and z.

In the paper we show that this type of routing - where a negligible part of the message serves as header information - can asymptotically achieve the capacity for single receiver relay networks. More specifically we show that the above example works for general relay networks. That is we will consider multi-user communication problems where each user is assigned a collection of variables (representing the required messages) and as well as a set of *terms* (as known from logic) expressing the relationship between the sent messages, the coding functions and the received messages. A term is built on variables representing the messages sent by the sources, and on function symbols representing coding functions at the intermediate nodes. A term thus formally represents all the possible operations undergone by messages from the sources to the destinations. This novel representation has several advantages. First, the topology of the network is contained in the term set; we can hence work without the help of the adjacency matrix of the network. Second, this graph-free framework makes computations easier to handle. Third, it is versatile and allows to convert a dynamic multi-user network process into a single static system. Fourth, it is actually more general than network coding, and hence offers not only a generalisation of results in network coding, but also a reformulation in terms of flows.

1.3. Max-flow min-cut for term sets

Next, and this is a crucial for our main results, we define the *min-cut* of a term set, which can be viewed as the channel capacity to the receiver and hence represents the information bottlenecks on the network. Conversely, to each choice of coding functions we associate a flow value, referred to as the *dispersion*, which quantifies the amount of information sent to the receivers. More precisely, the dispersion is the logarithm of the number of possible outputs of the term set, while the *one-to-one dispersion* is the logarithm of the number of outputs with exactly one pre-image, i.e. for which the input can be completely determined.

In the paper we show a max-flow min-cut theorem for the dispersion of term sets: the maximum dispersion and one-to-one dispersion of a term set are asymptotically equal to the value of its min-cut.

The term sets we consider may have distributed coding functions, which happens when different subterms use the same coding function. For instance, distributed coding functions occur in the term set associated to a multi-user communication problem, where a distributed function represents the same intermediate node in terms received by different users. Our proof of the max-flow min-cut theorem is based on a novel protocol, referred to as *dynamic routing*, which uses dynamic headers to eliminate distributed functions. Clearly, this comes at a cost in bandwidth equal to the size of the header; however, this is a constant given by the term set and becomes negligible when the alphabet size increases. Dynamic routing is interesting in its own sake, for unlike typical network coding approaches, such as random linear network coding [7], the manipulation of data is operated on headers only, and not on the whole packets.

If all sources are cooperative and can choose the optimal input distribution, then the maximum amount of information that can be inferred about the input from the received output is given by the min-cut of the term set. We thus introduce different measures of performance based on the Rényi entropy [8] for the non-cooperative case where the inputs are uniformly distributed. We show that the dispersion is a special case of the Rényi entropy, while the one-to-one dispersion is an independent performance measure.

The second main contribution is the max-flow min-cut theorem for the Rényi entropy with order $0 \leq \alpha < 1$, thus strengthening the result for the dispersion. Conversely, the Rényi entropy for $\alpha > 1$ does not necessarily reach the min-cut. Therefore, the Rényi entropy is sensitive to information bottlenecks that cannot be taken into account via the min-cut approach. However, the case of the Shannon entropy, where $\alpha = 1$, remains open.

In order to simplify the combinations operated at each intermediate node and the decoding at each destination, linear network coding only considers linear coding functions [9, 10]. It is known that linear network coding is not optimal in general [11, 12]; we generalize the inefficiency of linear network coding in our framework. In particular, we design a family of term sets with arbitrarily large min-cut where the maximum dispersion achieved by linear functions is only equal to 2. This can be intuitively explained by the fact that the dispersion of linear coding functions is equal to their min-entropy, which is the Rényi entropy of infinite order. Conversely, we prove that if maximum dispersion equal to the min-cut can be achieved using coding and decoding functions based on polynomials of fixed degree, then it can be achieved using linear functions only.

The third main contribution is the multi-user max-flow min-cut theorem. This shows that the maximum dispersion received by each receiver (user) can be asymptotically attained simultaneously. In other words, if a dispersion can be achieved locally, i.e. while disregarding the other users, it can be achieved globally, i.e. when the other users have to be accommodated as well. This result is then applied to multi-user communication problems such as satellite communication (the well-known butterfly network) and data storage.

Finally, our framework based on term sets is extended to simulate dynamic networks whose topologies may change over time. We view a dynamic network as possible "worlds", i.e. states in which the network is, and we allow the users to have requirements on the dispersion that change over time. A dynamic network can thus be modeled as one main term set, viewed as the union of all term sets for all users, possible worlds, and time-slots. Our last main contribution is the multi-user theorem for dynamic networks, which proves that if the demand (over all worlds and all time-slots) of each user can be satisfied locally, then they can all be satisfied globally.

1.4. Outline

The rest of the paper is organized in two main parts as follows. Firstly, Sections 2 to 6 study term sets and relay networks. Section 2 reviews some key concepts of logic and term sets and defines the analogues of flows and cuts in the new communication networks. Section 3 then proves the max-flow min-cut theorems for the dispersion and the one-to-one dispersion for these networks. The theorem for the Rényi entropy is given in Section 4. Section 5 then investigates the dispersion of linear coding functions. In order to illustrate the concepts and results of this paper, a case study of a simple term set is carried out in Section 6. Secondly, Sections 7 and 8 illustrate how relay networks, term sets, and the max-flow min-cut theorems can be applied to communication networks. In Section 7, we associate a term set to a multi-user communication problem, and prove the multi-user max-flow min-cut theorem. Our model is finally generalized to dynamic networks in Section 8. Section 9 then concludes and summarizes the paper.

2. Abstract Communication Channels based on logic

This section introduces a new type of abstract communication channels (1-1 user communication networks) based on term sets in logic and determines its main characteristics. We first review the basic concepts of logic and determine the analogue of a min-cut. We then view flows as transmission of data over a given alphabet, hence determining the analogue of max-flow.

2.1. Term sets

Let $X = \{x_1, x_2, \ldots, x_k\}$ be a set of *variables* and consider a set of *function symbols* $\{f_1, f_2, \ldots, f_l\}$ with respective arities (numbers of arguments) $d_1, d_2, \ldots, d_l$. A *term* is defined to be an object obtained from applying function symbols to variables recursively. For instance, if $k = 2$, $l = 3$, and the arities are given by $d_1 = 1$, $d_2 = d_3 = 2$, then the following are terms: $t_1 = f_2\big(f_1(x_1), x_2\big)$, $t_2 = f_1\Big(f_3\big(f_2(x_2, x_1), f_3(x_1, x_2)\big)\Big)$, $t_3 = f_1(x_1)$. For a broader introduction to first-order logic see for example [13]. Only Definition 3.2 on p.18 which introduces the notions of variables, function symbols, and terms is relevant for us.

We say that u is a subterm of t if the term u appears in the definition of t. For instance, t_3 is a subterm of t_1 as $t_1 = f_2(t_3, x_2)$, but it is not a subterm of t_2. Furthermore, u is a *direct subterm* of t if $t = f_j(v_1, \ldots, u, \ldots, v_{d_j})$, and f_j is referred to as the *principal function* of t.

We shall consider finite term sets, typically referred to as $\Gamma = \{t_1, t_2, \ldots, t_r\}$ built on variables $x_1, x_2, \ldots, x_k$ and function symbols $f_1, f_2, \ldots, f_l$ of respective arities $d_1, d_2, \ldots, d_l$. We denote the set of variables that occur in terms in Γ as Γ_{var} and the collection of subterms of one or more terms in Γ as Γ_{sub}; thus $\Gamma_{\mathrm{var}} \subseteq \Gamma_{\mathrm{sub}}$ and $\Gamma \subseteq \Gamma_{\mathrm{sub}}$.

Definition 1. *A* (network coding) communication channel *is given by a collection Γ of terms. The* requirement *of the channel is a collection of variables. If no such set is specified it is assumed that the channel requires all variables.*

A communication network *is a collection of receivers that each is assigned a channel with some requirement.*

A communication network is said to be a many-to-many (or multi-cast) network *if each receiver requires the same set of variables.*

Example 5. *The* butterfly network *has two receivers. One receiver is assigned a channel given by* $\{x, f(x,y)\}$ *and requires* y*. The other receiver is assigned a channel given by* $\{y, f(x,y)\}$ *and requires* x*.*

We rename variables and take the union of terms sets for the butterfly network we get a communication channel $\Gamma := \{x, f(x,y), w, f(z,w)\}$ *which requires* y, z *(or equivalently requires* x, y, z, w*).*

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

Definition 2 (Term-cut). *A set of subterms* $s_1, s_2, \ldots, s_\rho \in \Gamma_{\text{sub}}$ *provides a* term-cut *of size* ρ *for* Γ *if all the terms can be expressed syntactically by applying function symbols to* $s_1, s_2, \ldots, s_\rho$*.*

For a collection of terms Γ *and for a collection of variables* $U \subseteq \Gamma_{\text{var}}$ *we let* Γ^U *denote the set of terms that occur by substituting each variable* $x \in \Gamma_{\text{Var}} \setminus U$ *with the constant symbol* 0*.*

A set of subterms $s_1, s_2, \ldots, s_\rho \in \Gamma_{\text{sub}}$ *together with the symbol* 0 *provides a* term-cut with respect to the variables in U *of size* ρ *for* Γ *if all the terms can be expressed syntactically by applying function symbols to* $s_1, s_2, \ldots, s_\rho$ *and* 0*.*

A minimal term-cut for Γ is a term-cut with minimum size, referred to as the *min-cut* of Γ. The min-cut can hence be viewed as the number of degrees of freedom of the term set. Alternatively, the min-cut can be viewed as a measure of the channel capacity of the channel defined by Γ. Clearly, the min-cut is no more than the number of variables k since $\{x_1, x_2, \ldots, x_k\}$ is a term-cut for Γ; similarly, the min-cut is no more than the number of terms r.

Example 6. *Consider the communication channel given by the term set*

$$\Gamma_1 = \big\{h\big(f(x,y), g(z,w), f(y,x)\big), m\big(g(z,w), f(y,x)\big), g\big(f(x,y), g(z,w)\big), f\big(g(z,w), f(y,x)\big)\big\},$$

then the subterms $s_1 = f(x,y)$*,* $s_2 = g(z,w)$*, and* $s_3 = f(y,x)$ *form a term-cut for* Γ_1 *since we–in a purely syntactical way–can express the terms in* Γ_1 *by applying function symbols to* s_1*,* s_2*, and* s_3 *as*

$$\Gamma_1 = \{h(s_1, s_2, s_3), m(s_2, s_3), g(s_1, s_2), f(s_2, s_3)\}.$$

This shows that Γ_1 *has capacity* 3 *(a simple combinatorial argument shows there is no min-cut of value less than* 3*).*

The min-cut of Γ_1 *with regards to the variables* w, z *is* 1 *as*

$$\Gamma_1^{\{w,z\}} = \big\{h\big(f(0,0), g(z,w), f(0,0)\big), m\big(g(z,w), f(0,0)\big), g\big(f(0,0), g(z,w)\big), f\big(g(z,w), f(0,0)\big)\big\},$$

has a term cut with $s_1 = g(z,w)$ *as we can express each term in* $\Gamma_1^{\{w,z\}}$ *by applying function symbols to* s_1 *and* 0 *since*

$$\Gamma_1^{\{w,z\}} = \{h(f(0,0), s_1, f(0,0)), m(s_1, f(0,0)), g(f(0,0), s_1), f(s_1, f(0,0))\}.$$

Thus the value of the minimal term cut of Γ_1 is 1 with respect the variables w, z.

The concepts explained so far can be graphically explained as follows.

Definition 3 (The graph G_Γ). *For a given term set Γ, the directed graph $G_\Gamma = (V, E, S, T)$ is defined to have vertex set $V = \Gamma_{\text{sub}}$, edge set $E = \{(u, v) : u \text{ is a direct subterm of } v\}$, source set $S = \Gamma_{\text{var}}$, and target set $T = \Gamma$.*

In the graph G_Γ, each term is connected to all the variables it is built on; however, the graph is clearly acyclic. Notice that $S \cap T$ is non-empty if Γ contains one or more terms that are variables.

Example 7. *Consider the term set Γ_1 in Example 6. The graph G_{Γ_1} consists of a vertex for each subterm in*

$$\begin{aligned}\Gamma_{1,\text{sub}} \quad = \quad & \big\{x, y, z, w, f(x, y), f(y, x), g(z, w), h\big(f(x, y), g(z, w), f(y, x)\big), \\ & m\big(g(z, w), f(y, x)\big), g\big(f(x, y), g(z, w)\big), f\big(g(z, w), f(y, x)\big)\big\}\,.\end{aligned}$$

Furthermore, each variable in $\Gamma_{1,\text{var}} = \{x, y, z, w\}$ represents a source node and each term in Γ_1 represents a sink (or target) node. The graph G_{Γ_1} is then given as in Figure 5.

Assume that G is a directed graph with source set S and target set T. We say a set U of vertices is a *vertex cut*—commonly referred to as a separating set—if the removal of U leaves no *directed* path from S to T. If $S \cap T \neq \emptyset$ each single point in $S \cap T$ is considered to be a path from S to T. Proposition 1 below shows that term-cuts for Γ are equivalent to vertex cuts in G_Γ.

Proposition 1. *Assume Γ is a finite term set. A subset $C \subseteq \Gamma_{\text{sub}}$ is a term-cut for Γ if and only if C is a vertex cut that separates $S = \Gamma_{\text{var}}$ from $T = \Gamma$ in the directed graph G_Γ. Therefore, the min-cut of Γ is identical to the size of the minimal cut that separates S from T in the directed graph G_Γ.*

The proof relies on the following two technical lemmas.

Lemma 1. *A proper subterm is a subterm that is not equal to the term. We have the following:*

- *a) The subterm relation is transitive, i.e. if t_1 is a subterm of t_2 and t_2 is a subterm of t_3, then t_1 is a subterm of t_3.*
- *b) If t_1 is a proper subterm of t_2 and t_2 is a subterm of t_3, then t_1 is a proper subterm of t_3.*
- *c) The parsing of terms is unambiguous. More specifically if two terms t_1 and t_2 are identical, they have the same principal function symbol (say f). If we write $t_1 = f(u_1, u_2, \ldots, u_d)$ and $t_2 = f(u'_1, u'_2, \ldots, u'_d)$ then $u_j = u'_j$ for $j = 1, 2, \ldots, d$.*

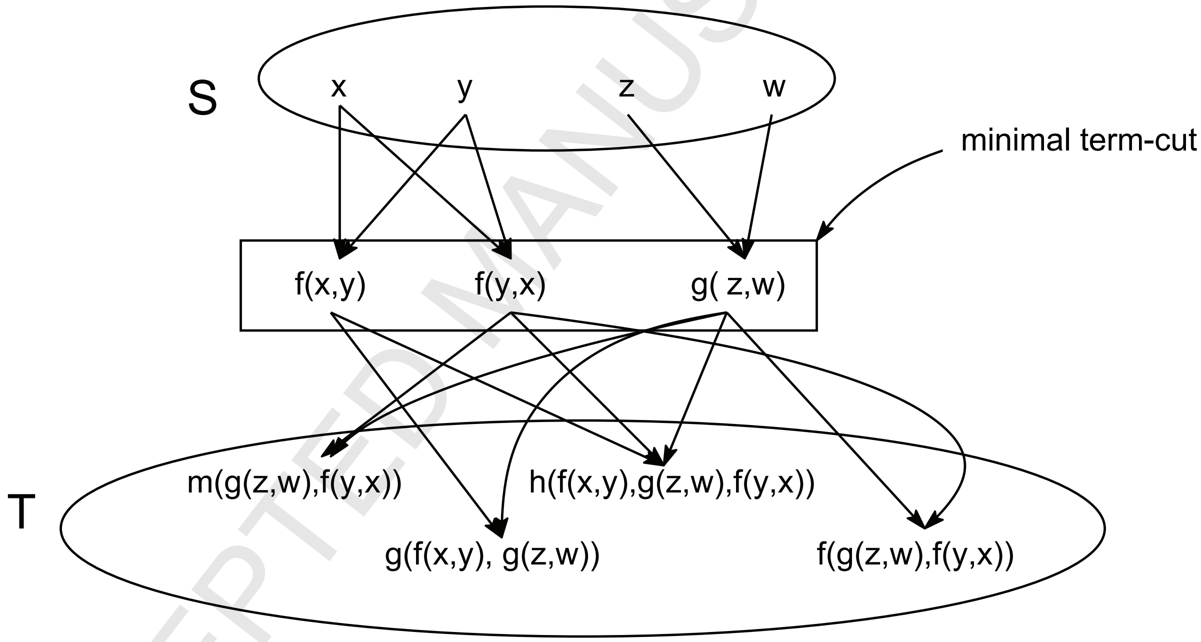


Figure 5: The graph G_{Γ_1}.

The proof of Lemma 1 is easy and hence omitted.

Lemma 2. *Assume t is a term and let $s_1, s_2, \ldots, s_\rho$ be a term-cut for t such that s_i is not a subterm of s_j for all $i \neq j$. Let u be a subterm of t, then there are two exclusive possibilities:*

- *i) $s_1, s_2, \ldots, s_\rho$ is a term-cut for u.*
- *ii) u is a proper subterm of some s_j, $j = 1, 2, \ldots, \rho$ (u might be a subterm of more than one s_j).*

Proof. We first show that the possibilities i) and ii) are exclusive. Suppose on the contrary that a subterm u satisfies both i) and ii). Then some s_i is a proper subterm of u, and by Lemma 1 it is a subterm of some s_j, which contradicts our assumption.

We now show that at least one of i) or ii) is satisfied. Assume u is chosen as a subterm of t that fails to satisfy both i) and ii) and such that it is not the proper subterm of some other subterm of t which also fails i) and ii). Since the term t satisfies i), u must be a proper subterm of t. Thus u occurs as a direct subterm in some subterm v of t, i.e. v can be written as $v = g(\ldots, u, \ldots)$ where g is the principal function symbol in v. Since the subterm relation is transitive (Lemma 1 part a), v does not satisfy ii), and as $v = g(\ldots, u, \ldots)$ with u not satisfying i), it follows from Lemma 1 part c, that v also fails to satisfy i). But this contradicts the assumption that u was not a subterm of a subterm that failed to satisfy both i) and ii).

We now prove Proposition 1.

Proof. Assume first that $C = \{s_1, s_2, \ldots, s_\rho\} \subseteq \Gamma_{\text{sub}}$ is a term-cut for Γ. If some term s_i is a subterm of s_j, remove s_j from C. What remains after having repeated this procedure is a set $C' \subseteq C$ of minimal subterms in C. Clearly C' is a term-cut for Γ. To keep the notation simple let us assume $C' = \{s_1, s_2, \ldots, s_{\rho'}\}$. It suffices to show that $C' \subseteq \Gamma_{\text{sub}}$ is a cut that separates $S = \Gamma_{\text{var}}$ from $T = \Gamma$ in the directed graph G_Γ.

Assume there is a path P from a variable $x \in \Gamma_{\text{var}}$ to a term $t \in \Gamma$ which does not intersect C'. Since C' satisfies the conditions for Lemma 2, each term in P satisfies either i) or ii). Since x satisfies ii) while t satisfies i), consider the first subterm u in P that satisfies i). The proper subterm of u in P satisfies ii), and hence is a proper subterm of some s_j. Thus u must be identical to that s_j, since otherwise we could not express u as a function of the subterms in the cut C'. This is not possible since P was assumed not to intersect $C' = \{s_1, s_2, \ldots, s_{\rho'}\}$.

To prove the converse, assume that $C = \{s_1, s_2, \ldots, s_\rho\} \subseteq \Gamma_{\text{sub}}$ is a cut that separates $S = \Gamma_{\text{var}}$ from $T = \Gamma$ in the directed graph G_Γ. Each subterm t that does not belong to the cut C and is on the same side of the cut as T has each of its arguments either in the cut or on the same side of the cut as T. A simple argument by induction shows that subterms that have all their arguments either in the cut or on the same side of the cut as T can be written on the form $g(s_1, s_2, \ldots, s_\rho)$. This shows that C defines a term-cut for Γ.

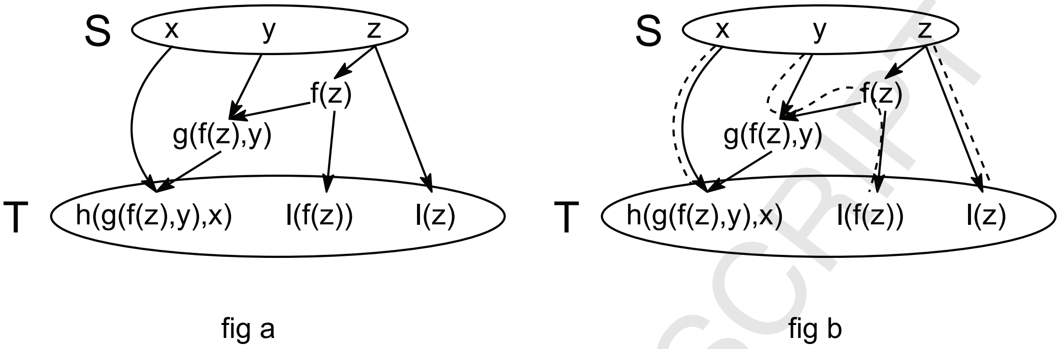


Figure 6: Directed graph G_Γ.

The fact that the size of a minimal term-cut is identical to the size of a minimal cut follows trivially from the first part of the proposition.

According to the directed graph version of Menger's theorem [14], there exists a family P of vertex-disjoint directed paths from $S = \Gamma_{\text{var}}$ to $T = \Gamma$ and a vertex cut C which consists of exactly one vertex from each path in P. Moreover, given the term set Γ, it is computationally feasible to find the exact value of the min-cut. By use of Dinic's algorithm [15] for finding max flows in networks with unit capacities–which terminates in $O(V\sqrt{E})$ time–a term-cut of minimal size for any term set Γ can be returned in time $O(|\Gamma_{\text{sub}}|^2)$.

We would like to emphasize that any term-cut $C \subseteq \Gamma_{\text{sub}}$ for Γ is always a vertex cut that separates S from T *when G_Γ is considered as a directed graph.* This is due to the antisymmetric nature of the subterm relation. Example 8 illustrates this distinction.

Example 8. *Consider the term set*

$$\Gamma = \Big\{h\Big(g\big(f(z), y\big), x\Big), l\big(f(z)\big), l(z)\Big\}.$$

The graph G_Γ has vertex set $V = \Big\{x, y, z, f(z), l\big(f(z)\big), g\big(f(z), y\big), h\Big(g\big(f(z), y\big), x\Big)\Big\}$, source set $S = \{x, y, z\}$ and target set $T = \Big\{h\Big(g\big(f(z), y\big), x\Big), l\big(f(z)\big), l(z)\Big\}$ and is displayed in Figure 6 a.

Viewed as a directed graph, G_Γ has a vertex cut (i.e. a term-cut for Γ) $C = \Big\{h\Big(g\big(f(z), y\big), x\Big), z\Big\}$ of size 2 and contains only 2 vertex-disjoint directed paths from S to T. Viewed as an undirected graph, G_Γ has a minimal vertex cut of size 3 and contains 3 vertex-disjoint undirected paths from S to T (see Figure 6 b).

2.2. Coding functions and their ability to disperse information in a single communication channel

So far, we have treated function symbols as abstract entities; we now assign them explicit values.

Definition 4 (Interpretation). *Let A be a finite set with $|A| \geq 2$, referred to as the alphabet. An* interpretation *for Γ over A is an assignment of the function symbols $\psi = \{\bar{f}_1, \bar{f}_2, \ldots, \bar{f}_l\}$, where $\bar{f}_i : A^{d_i} \to A$ for all $1 \leq i \leq l$.*

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

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

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

Example 9. *Consider Γ_1 introduced in Example 6 and let $A = \mathbb{F}_2$. The interpretation $\psi = \{\bar{f}, \bar{g}, \bar{h}, \bar{m}\}$ given by $\bar{f}(a_1, a_2) = a_1$, $\bar{g}(a_1, a_2) = a_1 + a_2$, $\bar{h}(a_1, a_2, a_3) = a_2 a_3 + 1$, $\bar{m}(a_1, a_2) = a_1 a_2$ induces the mapping*

$$\psi(a_1, a_2, a_3, a_4) = \big((a_3 + a_4)a_2 + 1, (a_3 + a_4)a_2, a_1 + a_3 + a_4, a_3 + a_4\big).$$

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

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

We now define the analogue of the value of a flow for information transfer on networks based on logic, which we refer to as the dispersion.

Definition 5. *The* Γ-dispersion *and* one-to-one Γ-dispersion *of an interpretation ψ for Γ over A are respectively defined as*

$$\begin{aligned} \gamma(\psi) &:= \log_{|A|} |\mathrm{image}(\psi)|, \\ \gamma_{\mathrm{one}}(\psi) &:= \log_{|A|} |\mathrm{one}(\psi)|. \end{aligned}$$

We define the worst-case (average-case) Γ-dispersion *(*one-to-one Γ-dispersion*) of an interpretation ψ for Γ over A with respect to the variables in $U \subseteq \Gamma_{\mathrm{var}}$ as the minimal (average) Γ-dispersion (one-to-one Γ-dispersion) for any setting of the variables in $\Gamma_{\mathrm{var}} \setminus U$.*

We remark that since $\mathrm{one}(\psi) \subseteq \mathrm{image}(\psi)$, we have $\gamma_{\mathrm{one}}(\psi) \leq \gamma(\psi)$ for all interpretations ψ. For instance, the interpretations in Example 9 has Γ_1-dispersion of $\log_2 6$, while $\psi(\mathbf{a} + (0,0,1,1)) = \psi(\mathbf{a})$ for all $\mathbf{a} \in A^4$ implies it has one-to-one Γ_1-dispersion $-\infty$.

We finally define the (one-to-one) dispersion of Γ over A as the maximal Γ-dispersion (one-to-one Γ-dispersion, respectively) over all interpretations for Γ over A, and we denote this value by $\gamma(\Gamma, |A|)$ (by $\gamma_{\mathrm{one}}(\Gamma, |A|)$, respectively) as this quantity clearly depends on A via its cardinality only. We say an interpretation ψ has perfect dispersion if it equals the Γ-dispersion for some finite alphabet.

Observation 1. *For perfect dispersion there is no difference between worst case dispersion and average dispersion. Both are identical to the value of the minimal term cut with regards to the required variables.*

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

The main purpose of this section is to prove the following max-flow min-cut theorem for the dispersion and the one-to-one dispersion of term sets.

Theorem 1 (Max-flow min-cut theorem for dispersion). *Let Γ be a term set with min-cut of ρ, then for any alphabet A,*

$$\gamma_{\mathrm{one}}(\Gamma, |A|) \leq \gamma(\Gamma, |A|) \leq \rho.$$

Conversely,

$$\lim_{|A| \to \infty} \gamma_{\mathrm{one}}(\Gamma, |A|) = \lim_{|A| \to \infty} \gamma(\Gamma, |A|) = \rho.$$

The first part of the max-flow min-cut theorem is easily proved.

Lemma 3. *Let Γ be a term set built on k variables and with min-cut of $\rho \leq k$. Then for all A, $\gamma_{\mathrm{one}}(\Gamma, |A|) \leq \gamma(\Gamma, |A|) \leq \rho$. Furthermore, if $\rho < k$, then $\gamma_{\mathrm{one}}(\Gamma, |A|) \leq \log_{|A|}(|A|^\rho - 1) < \rho$.*

Proof. Let C be a minimal term-cut for Γ. C can be viewed as a term set, hence let ψ_C be an interpretation for C over A. The size of the image of its induced mapping is at most $|A|^\rho$. Furthermore, let ψ_Γ be an interpretation for Γ over A. Since all terms of Γ can be expressed as functions of elements of C, the size of the image of ψ_Γ is at most that of ψ_C, hence $|\mathrm{image}(\psi_\Gamma)| \leq |A|^\rho$ and $\gamma(\Gamma, |A|) \leq \rho$.

Furthermore, if $\rho < k$, the average number of pre-images per element of $\mathrm{image}(\psi_\Gamma)$ is at least $|A|^{k-\rho} > 1$. Therefore, there exists an element with more than one pre-image, and $|\mathrm{one}(\psi_\Gamma)| \leq |A|^\rho - 1$.

3.1. Diversified term sets

We first prove the max-flow min-cut theorem for the dispersion in the specific case where each subterm has a distinct function symbol. More specifically, we define the diversified term set by assigning a new function symbol to each subterm that is not a variable.

Definition 6 (Diversified term set). *For any term set* Γ*, the* diversified term set Γ^{div} *is built on the same variables as* Γ *and its function symbols are obtained by replacing the principal function* g *of any* $u \in \Gamma_{\text{sub}} \backslash \Gamma_{\text{var}}$ *by a new function symbol* g_u *of the same arity as* g*.*

Example 10. *Recall the term set* Γ_1 *from Example 6:*

$$\Gamma_1 = \big\{h\big(f(x,y), g(z,w), f(y,x)\big), m\big(g(z,w), f(y,x)\big), g\big(f(x,y), g(z,w)\big), f\big(g(z,w), f(y,x)\big)\big\}\,,$$

then

$$\begin{aligned}\Gamma_1^{\text{div}} \quad = \quad & \big\{h_{h(f(x,y),g(z,w),f(y,x))}\big(f_{f(x,y)}(x,y), g_{g(z,w)}(z,w), f_{f(y,x)}(y,x)\big),\\ & m_{m(g(z,w),f(y,x))}\big(g_{g(z,w)}(z,w), f_{f(y,x)}(y,x)\big),\\ & g_{g(f(x,y),g(z,w))}\big(f_{f(x,y)}(x,y), g_{g(z,w)}(z,w)\big),\\ & f_{f(g(z,w),f(y,x))}\big(g_{g(z,w)}(z,w), f_{f(y,x)}(y,x)\big)\big\}\,.\end{aligned}$$

We can simplify the indices and rewrite the diversified term set as

$$\big\{h\big(f_1(x,y), g_1(z,w), f_2(y,x)\big), m\big(g_1(z,w), f_2(y,x)\big), g_2\big(f_1(x,y), g_1(z,w)\big), f_3\big(g_1(z,w), f_2(y,x)\big)\big\}\,.$$

We remark that before diversification, the same function symbol may be assigned to different subterms (e.g., $f(x,y)$ and $f(y,x)$ have the same principal function symbol in Γ_1). However, after diversification, there cannot be such overlap, as each subterm is assigned a distinct principal function. By definition, it is easily seen that the graph $G_{\Gamma^{\text{div}}}$ is isomorphic to G_Γ. In particular, Γ and Γ^{div} have the same min-cut.

For diversified term sets, maximal dispersion can be achieved via *routing*, which is defined in a similar way to the case of ordinary networks. Let $\Gamma = \{t_1, t_2, \dots, t_r\}$ be built on the variables $\{x_1, x_2, \dots, x_k\}$ and have min-cut of ρ. Let P be a set of ρ vertex-disjoint paths from Γ_{var} to Γ in G_Γ which, without loss, start in $x_1, x_2, \dots, x_\rho$ and end in $t_1, t_2, \dots, t_\rho$, respectively.

Definition 7 (Routing). *A distinct function symbol* g_v *is associated to each subterm* $v \in \Gamma_{\text{sub}}$*. If* u_j *is the direct subterm of* v *on the same path, then we let* $\bar{g}_v(a_1, a_2, \dots, a_d) = a_j$*. Otherwise, i.e. if* v *does not belong to any path in* P*, then* $\bar{g}_v(a_1, a_2, \dots, a_d) = 1$*.*

Note that our definition of routing depends on the set of paths P, and hence is not unique. However, the dispersion and one-to-one dispersion of routing do not depend on the choice of P. It is straightforward to verify that using routing, all points of the form $(a_1, a_2, \dots, a_\rho, 1, \dots, 1) \in A^k$ are mapped to

$(a_1, a_2, \dots, a_\rho, 1, \dots, 1) \in A^r$, thus yielding a Γ-dispersion of ρ. Furthermore, when $\rho = k$, the induced mapping (restricted to the first ρ coordinates) becomes the identity on A^ρ and hence $\gamma_{\text{one}}(\Gamma, |A|) = \rho$. However, routing has one-to-one dispersion $-\infty$ when $\rho < k$. In order to thwart this drawback, we define *one-to-one routing* below.

Definition 8 (One-to-one routing). *Let v be a subterm of the form $v = g_v(u_1, u_2, \dots, u_d)$, and denote the set of arguments equal to variables $x_{\rho+1}, x_{\rho+2}, \dots, x_k$ as $u_{i_1}, u_{i_2}, \dots, u_{i_m}$. We define the coding function $\bar{g}_v : A^d \to A$ as follows. If a path in P goes through v, denote the direct subterm of v on the same path as u_j; then, if $a_{i_1} = a_{i_2} = \dots = a_{i_m} = 1$, we let $\bar{g}_v(a_1, a_2, \dots, a_d) = a_j$. Otherwise, let $\bar{g}_v(a_1, a_2, \dots, a_d) = 1$.*

With one-to-one routing, it is straightforward to check that the $(|A| - 1)^\rho$ points of the form
$(a_1, a_2, \dots, a_\rho, 1, \dots, 1) \in A^k$ with $a_1 \neq 1, a_2 \neq 1, \dots, a_\rho \neq 1$ are mapped in a one-to-one fashion to $(a_1, a_2, \dots, a_\rho, 1, \dots, 1) \in A^r$, thus yielding a one-to-one Γ-dispersion of at least $\rho \log_{|A|}(|A| - 1)$.

We obtain the following max-flow min-cut result for diversified term sets. Recall that for any term set Γ, Γ^{div} is obtained by diversifying variables and function symbols.

Proposition 2. *Assume Γ is a term set built on k variables and with min-cut ρ. Let A be an alphabet of size $|A| \geq 2$, then $\gamma(\Gamma^{\text{div}}, |A|) = \rho$, and it is achieved by routing. Furthermore, if $\rho = k$, $\gamma_{\text{one}}(\Gamma^{\text{div}}, |A|) = \rho$ is achieved by routing, while if $\rho < k$, one-to-one routing yields*

$$\rho \log_{|A|}(|A| - 1) \leq \gamma_{\text{one}}(\Gamma^{\text{div}}, |A|) \leq \log_{|A|}(|A|^\rho - 1).$$

3.2. Dynamic routing

The construction of coding functions in Section 3.1 used the fact that each subterm v was assigned a distinct function symbol. However, in general distinct subterms might be assigned the same function symbol (e.g., $f(x, y)$ and $f(y, x)$). The proof of the general case relies on *dynamic routing*, defined below.

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

Definition 9 (Dynamic routing). *Consider the term set Γ^{div} first, which contains one function symbol g_v for each subterm $v \in \Gamma_{\text{sub}}$. Select coding functions $\bar{g}_v$ over B using routing, as in Definition 7. We then define the functions $\bar{f}_j(a_1, a_2, \dots, a_{d_j})$ over A as follows. If each a_i is of the form $a_i = (u_i, b_i) \in \Gamma_{\text{sub}} \times B$, let s denote the term $s = f_j(u_1, u_2, \dots, u_{d_j})$; then if $s \in \Gamma_{\text{sub}}$*

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

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

We can similarly define *dynamic one-to-one routing*. Remark that the headers u_i of the inputs then indicate to the coding function $\bar{f}_j$ which subterm v it is located on, and hence which function $\bar{g}_v$ to use. We say an input message a_j for the variable x_j is *correctly formatted* if it is of the form (x_j, b_j) where $b_j \in B$, and we denote the set of all correctly formatted inputs as $I := \{\mathbf{a} \in (\Gamma_{\text{var}} \times B)^k : a_j = (x_j, b_j)\}$. Moreover, the set of correctly formatted outputs is denoted as $J := \{\mathbf{a} \in (\Gamma \times B)^r : a_j = (t_j, b_j)\}$, and for all $\mathbf{a} \in J$, we denote the data part of $\mathbf{a}$ as $b(\mathbf{a}) = (b_1, b_2, \ldots, b_r) \in B^r$. The idea behind dynamic routing is that if all inputs are correctly formatted (i.e. have the correct headers) then the coding functions $\bar{f}_j$ mimic the behavior of the routing functions $\bar{g}_v$. Thus, correctly formatted messages in I are mapped–in a one-to-one fashion if one-to-one routing is used–to correctly formatted outputs in J (as long as they are mapped by the functions $\bar{g}_v$), while other messages will be mapped to an "error message" in R. We obtain the following lemma.

Lemma 4. *Let ψ be a dynamic routing interpretation for Γ over A based on the routing interpretation ϕ for Γ^{div} over B, then $\{\mathbf{a} \in J : b(\mathbf{a}) \in \text{image}(\phi)\} \subseteq \text{image}(\psi)$. Similarly, if ψ_{one} is a dynamic one-to-one routing interpretation for Γ over A based on the one-to-one routing interpretation ϕ_{one} for Γ^{div} over B, then $\{\mathbf{a} \in J : b(\mathbf{a}) \in \text{one}(\phi_{\text{one}})\} \subseteq \text{one}(\psi_{\text{one}})$.*

Lemma 4, together with Proposition 2, gives a lower bound on the dispersion and one-to-one dispersion. By choosing an appropriate alphabet size, we can prove the following quantitative version of the max-flow min-cut theorem for the dispersion.

Theorem 2. *Let Γ be a term set built on k variables and with min-cut of ρ. For $\epsilon < \rho$, let $n_1 := |\Gamma_{\text{sub}}|^{\rho/\epsilon}(1 - |\Gamma_{\text{sub}}|^{1-\rho/\epsilon})^{-\rho/\epsilon}$. Then for all $|A| \geq n_1$, $\gamma(\Gamma, |A|) \geq \rho - \epsilon$ and if $\rho = k$, $\gamma_{\text{one}}(\Gamma, |A|) \geq \rho - \epsilon$. These are achieved by dynamic routing.*

Moreover, for $\epsilon < \frac{\rho}{1+\log_{|\Gamma_{\text{sub}}|} 2}$, let $n_2 := |\Gamma_{\text{sub}}|^{\rho/\epsilon}(1 - 2|\Gamma_{\text{sub}}|^{1-\rho/\epsilon})^{-\rho/\epsilon}$. then for all $|A| \geq n_2$, if $\rho < k$, $\gamma_{\text{one}}(\Gamma, |A|) \geq \rho - \epsilon$ is achieved by dynamic one-to-one routing.

Proof. We only prove the case involving n_2, the other being proved similarly. Suppose A is an alphabet with $|A| \geq n_2$ and let ψ be the mapping induced by dynamic one-to-one routing for Γ over A. By Lemma 4 and Proposition 2, $|\text{one}(\psi)| \geq (|B| - 1)^\rho$. We have

$$\frac{|B| - 1}{|A|^{1-\epsilon/\rho}} \geq \frac{|A|^{\epsilon/\rho}}{|\Gamma_{\text{sub}}|}\left(1 - \frac{2|\Gamma_{\text{sub}}|}{|A|}\right) \geq \frac{|A|^{\epsilon/\rho}}{|\Gamma_{\text{sub}}|}\left(1 - \frac{2|\Gamma_{\text{sub}}|}{|\Gamma_{\text{sub}}|^{\rho/\epsilon}}\right) \geq 1,$$

where the successive inequalities follow from $|B| \geq \frac{|A|}{|\Gamma_{\text{sub}}|} - 1$, $n_2 \geq |\Gamma_{\text{sub}}|^{\rho/\epsilon}$, and the definition of n_2, respectively. Thus $\gamma_{\text{one}}(\Gamma, |A|) \geq \log_{|A|}(|B| - 1)^\rho \geq \rho - \epsilon$.

4. Max-flow min-cut for the Rényi entropy

4.1. Rényi entropy of an interpretation

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

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

where H denotes the Shannon entropy. The maximum is reached when $\bar{\mathbf{a}}$ has the following probability distribution: for each $\mathbf{b} \in \mathrm{image}(\psi)$, select $a(\mathbf{b}) \in \mathrm{pre}(\mathbf{b})$ (where $\mathrm{pre}(\mathbf{b})$ is the pre-image of $\mathbf{b}$ under ψ) and let $\mathbb{P}\{\bar{\mathbf{a}} = a(\mathbf{b})\} = |\mathrm{image}(\psi)|^{-1}$. Eq. (1) shows that the capacity of the channel is given by the dispersion of the interpretation considered. This justifies our study of the dispersion in Section 3. Thus, the max-flow min-cut theorem for the dispersion states that the channel capacity asymptotically converges to the min-cut ρ of the term set, i.e.

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

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

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

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

The normalized Rényi entropy over A of order $0 \le \alpha \le \infty$ of the random variable $\psi(\bar{\mathbf{a}})$, which we will simply denote as $H_\alpha(\psi)$, is thus given by [8]

$$\begin{aligned} H_\alpha(\psi) &:= \frac{1}{1-\alpha} \log_{|A|} \sum_{\mathbf{b} \in A^r} p_{\mathbf{b}}^\alpha \\ &= \frac{\alpha}{\alpha - 1} \left(k - \frac{1}{\alpha} \log_{|A|} \sum_{\mathbf{b} \in A^r} |\mathrm{pre}(\mathbf{b})|^\alpha \right) \quad 0 < \alpha < 1 \,\mathrm{or}\, 1 < \alpha < \infty. \end{aligned}$$

Three further special cases need close attention.

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

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

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

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

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

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

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

4.2. Rényi entropy and one-to-one dispersion

Although the dispersion is a special Rényi entropy, Proposition 3 below shows that the one-to-one dispersion cannot be viewed as a Rényi entropy. Even more strikingly, the second statement shows that the one-to-one dispersion can actually conflict with the other entropy measures. This can be intuitively explained as follows. The Rényi entropies measure to which degree the inputs have been mixed by the induced mapping ψ, for instance a high min-entropy guarantees that not too many inputs have been mapped to the same output. On the other hand, in order to guarantee a high one-to-one dispersion, the mixing has to be controlled so that one output is reserved for each input in the one-to-one pre-image. This control may significantly reduce the entropy if the size of the one-to-one image is very close to the total number of images, which occurs necessarily if the former is very near $|A|^\rho$.

Proposition 3. *First, for any term set Γ, there exists an interpretation ψ for Γ over any alphabet A such that $\gamma_{\mathrm{one}}(\psi) < H_\infty(\psi)$. Second, let $\Gamma_1 = \{f(x, y)\}$ be a term set with min-cut 1. Then for any $\alpha > 0$, there exists an interpretation ψ for Γ_1 such that $\gamma_{\mathrm{one}}(\psi) > H_\alpha(\psi)$. Furthermore, if an interpretation ψ for Γ_1 has maximal one-to-one Γ_1-dispersion, then $H_1(\psi)$ tends to zero for large $|A|$. Conversely, if $H_\alpha(\psi) = 1$ for some $\alpha > 0$ then $\gamma_{\mathrm{one}}(\psi) = -\infty$.*

Proof. First, if all the coding functions are constant, $\gamma_{\rm one}(\psi) = -\infty$. Second, one-to-one routing has maximal one-to-one Γ_1-dispersion equal to $\log_{|A|}(|A|-1)$ which tends to 1, while its Rényi entropy is given by

$$\frac{1}{1-\alpha}\log_{|A|}\{|A|-1+(|A|^2-|A|+1)^\alpha\} - \frac{2\alpha}{1-\alpha},$$

which tends to $1-\frac{\alpha}{1-\alpha}$ for $\alpha < \frac{1}{2}$. Conversely, if $H_\alpha(\psi) = 1$ then all outputs have exactly $|A|$ pre-images.

4.3. Max-flow min-cut theorem for the Rényi entropy

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

Theorem 3 (Max-flow min-cut theorem for the Rényi entropy). *Let Γ be a term set with min-cut of ρ, then*

$$\lim_{|A|\to\infty} H_\alpha(\Gamma,|A|) = \rho \quad \textit{for all} 0 \le \alpha < 1.$$

Conversely, for any $\alpha > 1$, there exists a term set Γ with min-cut ρ for which $\lim_{|A|\to\infty} H_\alpha(\Gamma,|A|) < \rho$.

The max-flow min-cut theorem for the Rényi entropy with $\alpha < 1$ is actually based on dynamic routing.

Proposition 4 (Max-flow min-cut theorem for $\alpha < 1$). *Let Γ be a term set built on k variables and with min-cut of ρ and for all $0 < \alpha < 1$ define $\beta = \rho + \frac{\alpha}{1-\alpha}k$ and $n_3 = (2|\Gamma_{\rm sub}|)^{\frac{\beta}{\epsilon}}$. Then for any alphabet A with $|A| \ge n_3$, $H_\alpha(\Gamma,|A|) \ge \rho - \epsilon$, which is achieved by dynamic routing.*

Proof. Suppose that dynamic routing is used over an alphabet A (recall the notations from Section 3.2). It is clear that any output of the form

$$((t_1,b_1),(t_2,b_2),\ldots,(t_\rho,b_\rho),(t_{\rho+1},1),\ldots,(t_r,1)) \in (\Gamma \times B)^r$$

has exactly $|B|^{k-\rho}$ pre-images, namely those of the form

$$((x_1,b_1),(x_2,b_2),\ldots,(x_k,b_k)) \in (\Gamma_{\rm var} \times B)^k,$$

where $b_{\rho+1}, b_{\rho+2}, \ldots, b_k \in B$. Let us denote this set of $|B|^\rho$ outputs as C and

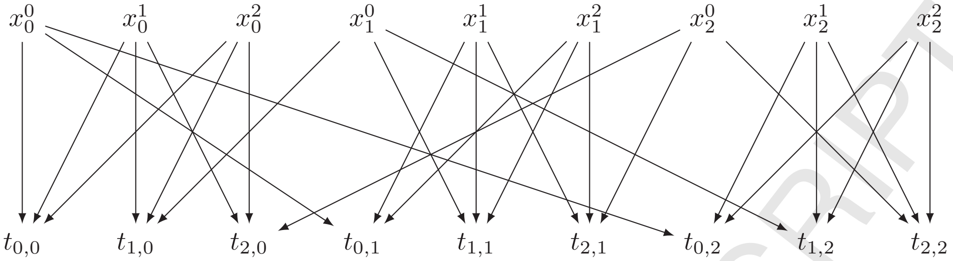


Figure 7: The term set Γ_3 used in Proposition 5

compute the Rényi entropy of dynamic routing.

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

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

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

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

The term set Γ_2 will be used in our case study in Section 6 (with the variables x, y, z, and w instead of x_0^0, x_0^1, x_1^1, and x_1^0 respectively), see Figure 8; the term set Γ_3 is illustrated in Figure 7.

Proposition 5 (No max-flow min-cut theorem for $\alpha > 1$). *The term set Γ_k has min-cut of k^2. However, for $\alpha > \frac{k}{k-1}$ and all A,*

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

Proof. First, we prove that Γ_k has min-cut of k^2 by constructing k^2 vertex-disjoint paths from $\Gamma_{k,\mathrm{var}}$ to Γ_k. We have $x_j^0 \in t_{j,j}$ for all j and if $a \geq 1$, $x_j^a \in t_{b,j}$

for any $0 \leq b \leq k-1$. Therefore, $(x_j^0, t_{j,j})$ for all j and $(x_j^a, t_{a+j \mod k, j})$ for all $a \geq 1$ and j form a set of k^2 vertex-disjoint paths.

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

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

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

The case of the Shannon entropy ($\alpha = 1$) remains open. This is an important question, as a max-flow min-cut theorem for the Shannon entropy would mean that the amount of information obtained in the non-cooperative case is asymptotically equal to that in the cooperative case. We note, however, that this is a "collective" measure of information, and does not indicate how much a given user knows about a given source.

We would like to highlight the difficulty of treating the Shannon entropy case. It can be shown that for any fixed n and for any $\epsilon > 0$, there exists $0 < \alpha < 1$ such that $H_1(X) \geq H_\alpha(X) - \epsilon$ for any probability distribution X on n points. However, we show below that ϵ cannot be chosen independently of n.

Example 11. *For $0 < \alpha < 1$ and for $n \geq 2$ consider the probability distribution $X = X_{\alpha,n}$ given by:*

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

We obtain

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

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

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

In contrast, we remark that the diversified term sets satisfy a much more general max-flow min-cut theorem. Indeed, when using routing, the number of pre-images of any output is a constant given by $|A|^{k-\rho}$. We obtain the following result.

Proposition 6. *For any term set* Γ *with min-cut of* ρ*, any* α*, and any alphabet* A*,*

$$H_\infty(\Gamma^{\mathrm{div}}, |A|) = H_\alpha(\Gamma^{\mathrm{div}}, |A|) = \rho.$$

Proposition 6 shows that the case of traditional network coding, where distributed coding functions are absent, is trivial. This motivates our study of term sets with distributed coding functions, which yield different types of bottlenecks that cannot be captured by the typical directed graph approach.

5. Dispersion of linear coding functions

5.1. Insufficiency of linear coding functions

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

Clearly, if the coding functions are linear, then so is the induced mapping of the corresponding interpretation. The structure of linear maps allows us to characterize their one-to-one dispersion in Proposition 7 below.

Proposition 7. *For any set* Γ *of terms built on* k *variables and with min-cut* ρ *and for any linear interpretation* ψ *for* Γ*, then* $\gamma_{\mathrm{one}}(\psi) = \rho$ *if and only if* $\gamma(\psi) = \rho = k$ *and* $\gamma_{\mathrm{one}}(\psi) = -\infty$ *otherwise.*

Proof. Let A be a vector space of dimension r over a finite field F, and let ψ be a linear interpretation for Γ over A. Since the induced mapping is linear, each point in its image has exactly $|\ker(\psi)| = |F|^d$ pre-images, where $d = \dim(\ker(\psi))$.

If $d = 0$, then each point in the image of ψ has exactly one pre-image, and hence $\gamma_{\mathrm{one}}(\psi) = k$. Since $\gamma_{\mathrm{one}}(\psi) \le \rho$ by Lemma 3, while $\rho \le k$, we obtain $\gamma_{\mathrm{one}}(\psi) = k = \rho$. Conversely, it is easily shown that if $\gamma(\psi) = \rho = k$, then $\gamma_{\mathrm{one}}(\psi) = \rho$. If $d \ge 1$, then each point in the image of ψ has more than one pre-image and $\gamma_{\mathrm{one}}(\psi) = -\infty$.

Because linear maps disperse information uniformly, the Rényi entropy of a linear map does not depend on the coefficient α and is hence equal to its dispersion: for any linear interpretation ψ for Γ over A, we have $H_\infty(\psi) =$

$H_\alpha(\psi) = \gamma(\psi)$ for all $0 \leq \alpha \leq \infty$. This shows a clear limitation of linear maps, as Theorem 3 shows that the min-entropy may not reach the min-cut. In particular, for the term set Γ_k introduced for Proposition 5, the min-cut is equal to k^2 while the min-entropy of any interpretation (and hence the dispersion of any linear interpretation) is upper bounded by $2k-1$.

As Theorem 1 indicates, although it is not always possible to reach the min-cut for any fixed finite alphabet, this can be achieved asymptotically. The class of term sets Γ can then naturally be divided into two disjoint classes whether there exist coding functions that achieve perfect dispersion equal to the min-cut. If Γ has perfect dispersion, we also say Γ is *solvable* and a *solution* is an interpretation with dispersion equal to the min-cut. Solvable term sets are easily found; conversely, we implicitly proved that the term sets Γ_k are not solvable for all k.

More generally, the dispersion of a scalar linear interpretation is always an integer, therefore if a term set Γ with min-cut ρ is not solvable, then the dispersion of any scalar linear interpretation for Γ is at most $\rho - 1$. However, by the max-flow min-cut theorem for the dispersion there exist non-linear interpretations with dispersion arbitrarily close to ρ, and hence the highest dispersion may not always be achieved by scalar linear interpretations.

We now significantly strengthen the considerations above by designing a solvable term set for which linear functions have dispersion bounded by a constant, while the min-cut can be arbitrarily large. Let $X = \{x_1, x_2, \dots, x_k\}$ be a set of variables and consider $k+1$ functions h_i of these variables: $h_i(x_1, x_2, \dots, x_k) = h_i(X)$. We then define the set Γ built on the k variables in X and the function symbols $h_1, h_2, \dots, h_{k+1}$ together with f of arity $k+1$ and $g_1, g_2, \dots, g_{k+1}$ of arities all equal to 1 by

$$\Gamma = \Big\{ f\Big(g_i\big(h_1(X)\big), h_2(X), h_3(X), \dots, h_{k+1}(X)\Big) : 1 \leq i \leq k+1 \Big\}.$$

Proposition 8. *There exists $n \in \mathbb{N}$ such that for any A with $|A| \geq n$, Γ defined above is solvable over A, i.e. $\gamma(\Gamma, |A|) = k$. However, any linear interpretation for Γ has dispersion at most 2.*

In order to prove Proposition 8, we first consider a term set related to Γ, where we convert the function symbols $h_1(X)$ into variables.

Lemma 5. *Consider the following set of $k+1$ terms built on $k+1$ variables $h_1, h_2, \dots, h_{k+1}$ and $k+2$ function symbols $f, g_1, g_2, \dots, g_{k+1}$:*

$$\Gamma' = \big\{ t_i = f\big(g_i(h_1), h_2, h_3, \dots, h_{k+1}\big) : 1 \leq i \leq k+1 \big\}.$$

Γ' has min-cut $k+1$, and hence dispersion arbitrarily close to $k+1$. However, if the coding functions are linear, then they have Γ'-dispersion at most 2.

Proof. We first prove that the min-cut of Γ' is $k+1$ by constructing $k+1$ vertex-disjoint paths from Γ'_{var} to Γ' in $G_{\Gamma'}$. There is a path $(h_1, g_1(h_1), t_1)$ and k paths (h_i, t_i) for $2 \leq i \leq k+1$; all these paths are clearly vertex-disjoint. By the max-flow min-cut theorem for the dispersion, Γ' has dispersion arbitrarily close to $k+1$.

Let ψ be a matrix linear interpretation for Γ', i.e. $\bar{g}_i(a_1) = G_i a_1$ for all $1 \leq i \leq k+1$ and $\bar{f}(\mathbf{a}) = \sum_{i=1}^{k+1} F_i a_i$. Consider the induced mapping ψ composed with the permutation π of A^{k+1} defined as $\pi(\mathbf{a}) = (a_1, a_2 - a_1, \ldots, a_{k+1} - a_1)$. Then we obtain

$$\pi \circ \psi(\mathbf{a}) = \left(F_1 G_1 a_1 + \sum_{i=2}^{k+1} F_i a_i, F_1(G_2 - G_1)a_1, \ldots, F_1(G_{k+1} - G_1)a_1 \right),$$

and $|\text{image}(\pi \circ \psi)| \leq A^2$. However, we have $|\text{image}(\pi \circ \psi)| = |\text{image}(\psi)|$, and hence the interpretation has dispersion at most 2.

We now prove Proposition 8.

Proof. By Theorem 1, there exists n such that for any A with $|A| \geq n$, there is an interpretation ψ for Γ' over A with one-to-one dispersion above k. Denote the pre-image of one(ψ) as I, then $|I| \geq |A|^k$. Furthermore, assign the coding functions $\bar{h}_i$ such that the inputs in A^k are mapped to $|A|^k$ elements in I. Therefore, the obtained interpretation for Γ has perfect dispersion of k. On the other hand, since the dispersion of an interpretation for Γ is no more than that of an interpretation for Γ', any linear interpretation for Γ has dispersion at most 2.

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

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

5.2. Low-degree solutions

In Section 5.1, we showed that there could be a huge difference between the dispersion achievable by linear coding functions and non-linear coding functions. In this section we show that if the min-cut is achievable by the use of coding functions of low degree (i.e. constant degree independently of the size of the underlying field), then the min-cut is actually achievable by the use of linear coding functions.

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

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

Theorem 4 (Low-degree solutions imply linear solutions). *Let Γ be a term set built on k variables and with min-cut of ρ. Assume that there exist coding and decoding functions defined by fixed polynomials $\bar{p}_1, \bar{p}_2, \ldots, \bar{p}_l \in \mathbb{Z}[a_1, a_2, \ldots, a_k]$ with dispersion ρ for arbitrarily large fields F of characteristic q. Then there exist scalar linear coding and decoding functions over all sufficiently large fields F of characteristic q that achieve perfect dispersion of ρ.*

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

We remark that the results in Section 5.1 imply that conversely, for each characteristic q there exist term sets with solutions which require that at least some of the involved coding functions (including decoding functions) should be given by polynomials of degree at least $|F|$. Example 12 gives a term set with no linear solution, and where a solution is given by polynomials whose degrees depend on the size of the alphabet.

Example 12. *Let*

$$\Gamma = \big\{ t_1 = f\big(f(x_1, x_2), f(x_2, x_1)\big), t_2 = g\big(g(x_1, x_2), g(x_2, x_1)\big) \big\} .$$

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

Proof. Let $\bar{f}$ and $\bar{g}$ be linear functions, i.e. $\bar{f}(a_1, a_2) = \alpha a_1 + \beta a_2$ and $\bar{g}(a_1, a_2) = \gamma a_1 + \delta a_2$, then $(\bar{t}_1, \bar{t}_2) = \big((\alpha^2 + \beta^2) a_1, (\gamma^2 + \delta^2) a_1\big)$ does not depend of a_2. Therefore, there are no linear solutions.

If $q = 2^k$ where k is even, first remark that the polynomial $u^{\sqrt{q}} + u$ has at most $\sqrt{q} < q$ roots in $\mathbb{F}_q$, hence there exists $\tau \in \mathbb{F}_q$ such that $\tau + \tau^{\sqrt{q}} \neq 0$. Then let $f(a_1, a_2) = a_1^{\sqrt{q}} + \tau a_2^{\sqrt{q}}$ and $g(a_1, a_2) = a_1$. We obtain

$$\begin{aligned} (\bar{t}_1, \bar{t}_2) &= \big((1 + \tau^{\sqrt{q}+1}) a_1 + (\tau + \tau^{\sqrt{q}}) a_2, a_1\big) \\ (a_1, a_2) &= \big(\bar{t}_2, (\tau + \tau^{\sqrt{q}})^{-1} (\bar{t}_1 - (1 + \tau^{\sqrt{q}+1}) \bar{t}_2)\big). \end{aligned}$$

Note that the degree of the solution depends on q, and hence Theorem 4 does not apply.

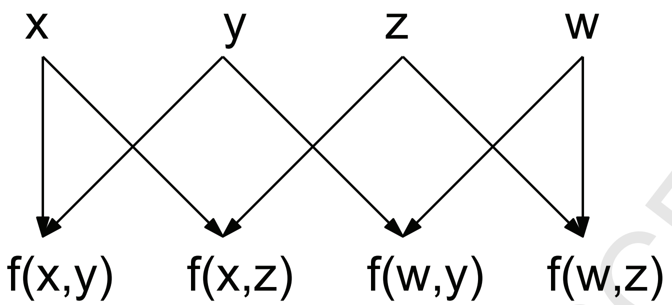


Figure 8: The graph G_Γ.

6. Case study involving a single coding function

The purpose of this section is double. First, we illustrate the different concepts introduced throughout the paper (distributed coding functions, insufficiency of linear coding functions, Rényi entropy, etc.). Second, although the results obtained so far are quite tight, we emphasize that for a specific term set, more can usually be said: tighter bounds can be derived, other types of functions can be considered such as linear functions over rings, etc.

Throughout this section, we consider the following term set:

$$\Gamma = \{f(x,y), f(x,z), f(w,y), f(w,z)\},$$

which actually is Γ_k introduced in Proposition 5 for $k = 2$. The graph G_Γ is given in Figure 8 below. The term set Γ can be viewed as an abstraction of a many-to-one cast with four sources and one user, where the intermediate node f corresponds to a relay which only picks up two signals at a time.

As seen previously, Γ has min-cut 4 ($\Gamma_{\rm var}$ is a term-cut of size 4; conversely $(x, f(x,y))$, $(z, f(x,z))$, $(y, f(w,y))$, and $(w, f(w,z))$ are 4 vertex-disjoint paths from $\Gamma_{\rm var}$ to Γ). According to the max-flow min-cut theorem for the dispersion, for any given $\epsilon > 0$ we can select a (coding) function $\bar{f} : A^2 \to A$ with Γ-dispersion at least $4-\epsilon$ for each sufficiently large alphabet A. By Proposition 2, the dispersion of the diversified term set is $\gamma_{\rm one}(\Gamma^{\rm div}, |A|) = \gamma(\Gamma^{\rm div}, |A|) = 4$ for all $|A| \geq 2$, and perfect one-to-one dispersion can be achieved by routing for the diversified case. In other words, all demands can be satisfied independently. However, Proposition 9 shows that the whole problem does not have any solution.

Proposition 9. *For any alphabet A, $\gamma(\Gamma, |A|) < 4$. In fact we have the tighter bound*

$$\gamma(\Gamma, |A|) \leq 4 - \log_{|A|}(1 - 2|A|^{-1} + 3|A|^{-2} - |A|^{-3})$$

Proof. We prove the tighter bound by refining the argument in the proof of Proposition 5. We partition the set of inputs A^4 into 4 parts C_1, C_2, C_3, and

C_4 defined as follows. The set C_1 consist of the inputs $\mathbf{a}$ where $a_1 \neq a_4$ and $a_2 \neq a_3$. This set contains $|A|^2(|A|-1)^2$ points hence the image of C_1 contains at most $|A|^2(|A|-1)^2$ points. The set C_2 consists of inputs where $a_1 = a_4$ and $a_2 \neq a_3$, and the set C_3 consists of inputs where $a_1 \neq a_4$ and $a_2 = a_3$. Finally, C_4 is the set of inputs where $a_1 = a_4$ and $a_2 = a_3$. The image of C_2 contains at most $|A|(|A|-1)$ points (same for C_3) while the image of C_4 contains at most $|A|$ points. Thus the range contains at most $|A|^2(|A|-1)^2+2|A|(|A|-1)+|A| = |A|^4-2|A|^3+3|A|^2-|A|$ outputs.

We proved in Section 5.1 that linear maps defined over fields are not optimal when the term set is not solvable. We extend this result below for linear maps defined over rings.

Proposition 10. *Assume A is organized as a ring and that $\bar{f} : A^2 \to A$ is a linear map in the usual sense of algebra, i.e. $\bar{f}(a_1, a_2) = r_1a_1 + r_2a_2$ for some $r_1, r_2 \in A$. Then $\bar{f}$ has Γ-dispersion at most 3, which is achieved for $r_1 = r_2 = 1$.*

Proof. We first prove the upper bound. Notice that the ring might not be commutative so we do not assume that $r_1r_2 = r_2r_1$. Then

$$\psi(a_1, a_2, a_3, a_4) = (r_1a_1 + r_2a_2, r_1a_1 + r_2a_3, r_1a_4 + r_2a_2, r_1a_4 + r_2a_3) \in A^4,$$

which is uniquely determined by the word $(r_1a_1+r_2a_2, r_2(a_2-a_3), r_1a_4+r_2a_3) \in A^3$. We conclude that the size of the image of ψ is no more than $|A|^3$. This proves the upper bound. For the lower bound let $\bar{f}(a_1, a_2) = a_1 + a_2$ and consider the set $C \subseteq A^4$ of inputs with last coordinate equal to 0; we have

$$\psi(a_1, a_2, a_3, 0) = (a_1 + a_2, a_1 + a_3, a_2, a_3) \in A^4$$

and hence the image of C has size $|A|^3$. Thus $\bar{f}$ has a Γ-dispersion equal to 3.

Similarly, if the alphabet is organized as a group G, we define the coding function $\bar{f}_G : G^2 \to G$ as $\bar{f}_G(\alpha, \beta) = \alpha\beta$ for all $\alpha, \beta \in G$. Then it can be easily shown that $\bar{f}_G$ has Γ-dispersion of 3 for any group G.

We would like to illustrate the difference between performance measures introduced in Section 4 by considering $A = \mathbb{F}_2$. In this case, there are $2^{2^2} = 16$ choices for the coding function $\bar{f}$. However, it can be easily shown that any function is equivalent in terms of the random variable $\psi(\mathbf{a})$ to one of the following four functions: $\bar{f}_0(a_1, a_2) = 0$, $\bar{f}_1(a_1, a_2) = a_1$, $\bar{f}_2(a_1, a_2) = a_1 + a_2$, and $\bar{f}_3(a_1, a_2) = a_1a_2$. We easily obtain that $\bar{f}_0$, $\bar{f}_1$, and $\bar{f}_2$ have dispersion 0, 2, and 3, respectively; since they are linear functions, they all have one-to-one dispersion equal to $-\infty$. On the other hand, the behavior of the non-linear function $\bar{f}_3$ is more complex, as its image consists of 10 elements: the 9 elements of its one-to-one image together with the all-zero vector, which has 7 pre-images. Therefore, $\bar{f}_3$ has dispersion $\gamma(\bar{f}_3) = \log_2 10 \equiv 3.32$ and one-to-one dispersion $\gamma_{\text{one}}(\bar{f}_3) = \log_2 9 \equiv 3.17$. On the other hand, $\bar{f}_3$ has Shannon entropy equal to

$H_1(\bar{f}_3) = 4 - \frac{7}{16}\log_2 7 \equiv 2.77$, which is lower than the Shannon entropy of $\bar{f}_2$. Furthermore, $\bar{f}_3$ has min-entropy of $H_\infty(\bar{f}_3) = 4 - \log_2 7 \equiv 1.19$, which is lower than those of $\bar{f}_1$ and $\bar{f}_2$.

We now consider the reverse illustration: we fix the coding function, but we change the alphabet on which it is defined. By use of computer calculations, we can show that the coding function $\bar{f}(a_1, a_2) = (a_1 - a_2)^2 + a_1 + a_2$ over the ring $\mathbb{Z}_3$ provides optimal Γ-dispersion of $\log_3 51 \equiv 3.58$, which attains the upper bound in Proposition 9. The same coding function also provides an optimal one-to-one Γ-dispersion of $\log_3 36 \equiv 3.26$. In general, we shall denote the interpretation for Γ based on the function $\bar{f}$ defined above over $\mathbb{Z}_{|A|}$ as $\psi_{|A|}$; we then have

$$\psi_{|A|}(x,y,z,w) = \left((x-y)^2 + x + y, (x-z)^2 + x + z, (w-y)^2 + w + y, (w-z)^2 + w + z\right). \tag{3}$$

Using an argument based on the Fourier transform [16], one can fully determine the number of elements of the image of that interpretation with a given number of pre-images when the alphabet size is a prime number: $|A| = p$. Thus the entropy $H_\alpha(\psi_p)$ can be determined for all α; its Γ-dispersion and one-to-one Γ-dispersion are respectively given by

$$\begin{aligned} \gamma(\psi_p) &= 4 - \log_p 2 + \log_p(1 + p^{-1} - p^{-2} + p^{-3}), \\ \gamma_{\text{one}}(\psi_p) &= 3 + \log_p 3 + 2\log_p(1 - p^{-1}). \end{aligned} \tag{4}$$

The behavior of the Rényi entropy for different values of α and the one-to-one dispersion of ψ_p is displayed in Figure 9. When p tends to infinity, then the Rényi entropy actually reaches the upper bound from Proposition 5. A more detailed argument is given in the next paragraph. On the other hand, by (4) the one-to-one dispersion of ψ_p tends to 3 on the primes p, which is below the min-cut. Thus, the interpretation ψ_p is asymptotically optimal in terms of Rényi entropy but not in terms of one-to-one dispersion. This illustrates the fundamental difference between the one-to-one dispersion and the different entropy measures.

Using an idea by Keevash and Sisask based on Fourier analysis [16], it is possible to show that the image of ψ_p for each prime number p is partitioned into 4 sets S_1, S_2, S_p, S_{3p-2} where:

- S_1 contains $3p(p-1)^2$ points with exactly one preimage
- S_2 contains $p(p-1)^2(p-3)/2$ points with exactly two preimages
- S_{p-1} contains $2p(p-1)$ points with exactly $p-1$ preimages
- S_{3p-2} contains p points with exactly $3p-2$ preimages.

Therefore, the size of the image is $\frac{p}{2}(p^3 + p^2 - p + 1)$. Moreover, ψ_p has Rényi entropy of order $\alpha \neq 1$ given by

$$H_\alpha(\psi_p) = \frac{1}{1-\alpha}\log_p\left(3(p-1)^2 + (p-1)^2(p-3)2^{\alpha-1} + 2(p-1)(p-1)^\alpha + (3p-2)^\alpha\right) + \frac{1-4\alpha}{1-\alpha}$$

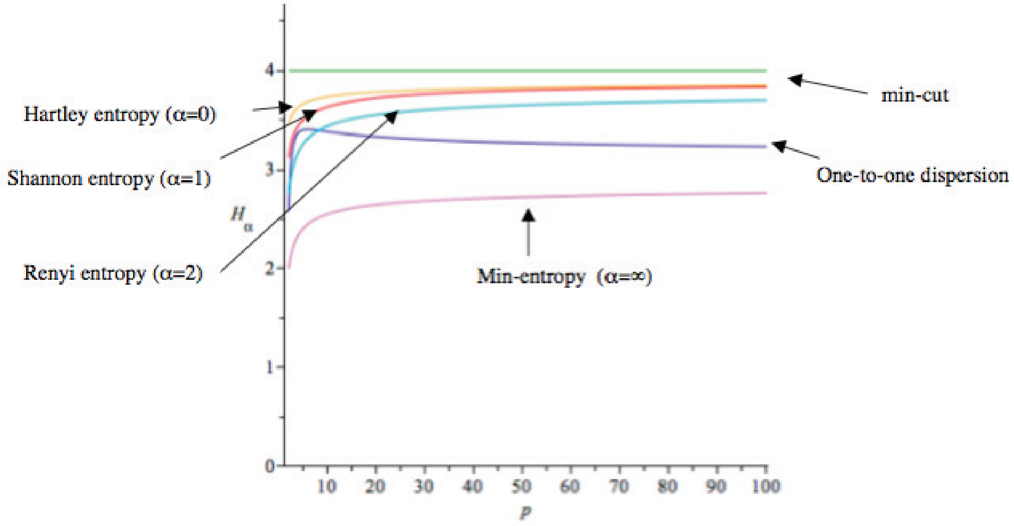


Figure 9: Dispersion and Rényi entropy of the interpretation ψ_p in Eq. (3) (p prime)

The limit of the Rényi entropy (including the case where $\alpha = 1$) when the size of the alphabet tends to infinity can be shown to be given by

$$\lim_{p\to\infty} H_\alpha(\psi_p) = \begin{cases} 4 & \text{if } 0 \le \alpha \le 2 \\ \frac{3\alpha-2}{\alpha-1} & \text{if } 2 \le \alpha < \infty \\ 3 & \text{if } \alpha = \infty \end{cases}$$

This shows that the bound $\frac{(2k-1)\alpha-k}{\alpha-1}$ is a matching upper and lower bound for $k = 2$.

7. Network Coding solvability and term sets

We now consider multi-user communications; our approach is to convert multi-user channels into single-user channels. A many-to-many cast is defined as follows.

Definition 10 (Many-to-many cast). *A multi-user communication problem instance (also referred to as a* many-to-many cast*) is a tuple* (V, E, S, U, A)*, where*

- $G = (V, E)$ *is an acyclic directed graph, where the vertices* $v_1, v_2, \dots, v_{|V|}$ *of* V *are sorted such that* $(v_i, v_j) \in E$ *only if* $i < j$.
- $S = \{s_1, s_2, \dots, s_k\} \subseteq V$ *is the set of sources, which are nodes with in-degree* 0*. Without loss of generality,* $s_i = v_i$ *for* $1 \le i \le k$.

- *$U = \{r_1, r_2, \dots, r_m\} \subseteq V$ is the set of users (receivers), which are nodes with out-degree 0. Without loss, $r_j = v_{|V|-m+j}$ for $1 \leq j \leq m$.*
- *A is an alphabet of size $|A| \geq 2$. Each source s_i sends a distinct message $a_i \in A$, and each user r_j requests $\{a_1, a_2, \dots, a_k\}$.*

Each vertex $v_k \in V \backslash S$ can manipulate the data it receives and transmit a function f_k of its inputs onto all its out-edges. A solution for the many-to-many cast instance is a choice of the functions at the intermediate nodes such that all the users' demands can be satisfied at the same time.

A communication problem can be equivalently defined with terms. Each user obtains a term built on the variables sent by the sources, where the function symbols represent the operations made by the intermediate nodes in $V \backslash (S \cup U)$. More formally, to each source s_i we associate the variable x_i and we denote $X = \{x_1, x_2, \dots, x_k\}$. Each user r_j requests all the variables in X. We then associate the function symbol f_l to all intermediate nodes $v_l \in V \backslash (S \cup U)$ and each vertex v_l is recursively assigned the term $u_l = f_l(u_{l_1}, u_{l_2}, \dots, u_{l_d})$, where $\{v_{l_1}, v_{l_2}, \dots, v_{l_d}\}$ is the in-neighborhood of v_l. We denote the in-neighborhood of the user r_j as $\{v_{j,1}, v_{j,2}, \dots, v_{j,p_j}\}$. Note that using this notation, it is possible that $v_{j,i} = v_{j',i'}$ for distinct j, j' and i, i'. However, this will not affect our definitions below. We finally associate the term $t_{j,i}$ to the vertex $v_{j,i}$. Thus, user j is satisfied if and only if it can recover X from the term set $\Gamma_j = \{t_{j,i}\}_{i=1}^{p_j}$ obtained from its in-neighborhood, or equivalently if and only if Γ_j has dispersion of k.

We now introduce a larger term set $\bar{\Gamma}$ which encapsulates the many-to-many cast solvability problem. For any $1 \leq j \leq m$, let $X^j = \{x_1^j, x_2^j, \dots, x_k^j\}$ be a new set of variables, and let Γ^j be the term set obtained from Γ_j by replacing the variable x_i by x_i^j for all $1 \leq i \leq k$; then let

$$\bar{\Gamma} := \bigcup_{j=1}^{m} \Gamma^j.$$

Proposition 11. *A many-to-many cast instance is solvable over A if and only if $\gamma_{\text{one}}(\Gamma_j, |A|) = \gamma(\Gamma_j, |A|) = k$ for all $1 \leq j \leq m$ or equivalently,*

$$\gamma_{\text{one}}(\bar{\Gamma}, |A|) = \gamma(\bar{\Gamma}, |A|) = km.$$

Proof. It is clear that any interpretation for Γ_j can be viewed as an interpretation for Γ^j, and that its Γ_j-dispersion (also one-to-one dispersion and Rényi entropy) is equal to its Γ^j-dispersion. Since the graphs G_{Γ^j} are components of $G_{\bar{\Gamma}}$, the latter has min-cut $\rho = \sum_{j=1}^{m} \rho_j$ and for any interpretation ψ for $\bar{\Gamma}$, we have $\gamma(\psi) = \sum_{j=1}^{m} \gamma(\psi^j)$, where ψ^j is the corresponding interpretation for Γ^j. We then conclude that there exists an interpretation for $\bar{\Gamma}$ over A with dispersion ρ if and only if there exist m interpretations for Γ_j with dispersions ρ_j for all $1 \leq j \leq m$.

We make two remarks about Proposition 11. Firstly, suppose a term is shared by the min-cut for two different receivers, so that the same set of functions is used for both. For the purpoose of the aggregate min-cut for $\bar{\Gamma}$, they are actually considered two different terms, since the variables have been diversified in $\bar{\Gamma}$. Secondly, we remark that Proposition 11 covers the case of a single source of k messages can be multicast to many different receivers if all individual min-cut is at least k (the seminal result on Network Coding in [2]). However, in order to asymptotically reach the min-cut of $\bar{\Gamma}$, one only needs to use dynamic routing, a novel technique that is not considered in [2].

Moreover, applying the max-flow min-cut theorems for the term set $\bar{\Gamma}$, we obtain the multi-user max-flow min-cut theorem below.

Theorem 5 (Multi-user max-flow min-cut theorem). *Let $\Gamma_1, \Gamma_2, \dots, \Gamma_m$ be term sets with respective min-cuts $\rho_1, \rho_2, \dots, \rho_m$. Then for any $\epsilon > 0$, there exists n_0 such that for all A with $|A| \geq n_0$, there is an interpretation ψ for $\bar{\Gamma}$ such that $\gamma_{\text{one}}(\psi^j) \geq \rho_j - \epsilon$ for all $1 \leq j \leq m$. Also, for any $0 \leq \alpha < 1$, there exists n_α such that $H_\alpha(\psi^j) \geq \rho_j - \epsilon$ for all $|A| \geq n_\alpha$.*

The multi-user max-flow min-cut theorem then shows that if the min-cut between the sources and each user is equal to k, then the multi-user instance is asymptotically solvable. Note that the equivalence in Proposition 11 together with Theorem 5 indicate that if each user can be asymptotically satisfied individually, then all the users' demands can asymptotically be satisfied simultaneously.

The following two examples are consequences of Proposition 11.

Example 13. *In [1], Yeung and Zhang made a simple observation that lay the foundation for network coding. The authors considered a situation where two users communicate via a satellite. User X wants to send a message $x \in A$ to Y, while Y at the same time wants to send a message $y \in A$ to X.*

The satellite communication problem is equivalent to the communication problem in Figure 10 a), referred to as the butterfly network. In Figure 10 a), the function symbols are affected to the vertices, accordingly to Definition 10; however the problem is sometimes represented in the literature by Figure 10 b), where the function symbols are assigned to the edges. A solution over an alphabet A is a function $\bar{f} : A^2 \to A$ with the property that there exist decoding functions $\bar{d}_1, \bar{d}_2 : A^2 \to A$ such that $b = \bar{d}_1\big(b, \bar{f}(a,b)\big)$ and $a = \bar{d}_2\big(a, \bar{f}(a,b)\big)$ for any $(a,b) \in A^2$.

By Proposition 11, the satellite problem (and the butterfly communication problem) is mathematically equivalent to the following problem on terms.

Problem 1 (Equivalent to the satellite communication problem). *Construct a function $\bar{f} : A^2 \to A$ with Γ-dispersion equal to 4, where $\Gamma := \{x_1, f(x_1,x_2), f(x_3,x_4), x_4\}$.*

This term set corresponds to the graph with vertex set

$$V = \Gamma_{\text{sub}} = \{x_1, x_2, x_3, x_4, f(x_1,x_2), f(x_3,x_4)\},$$

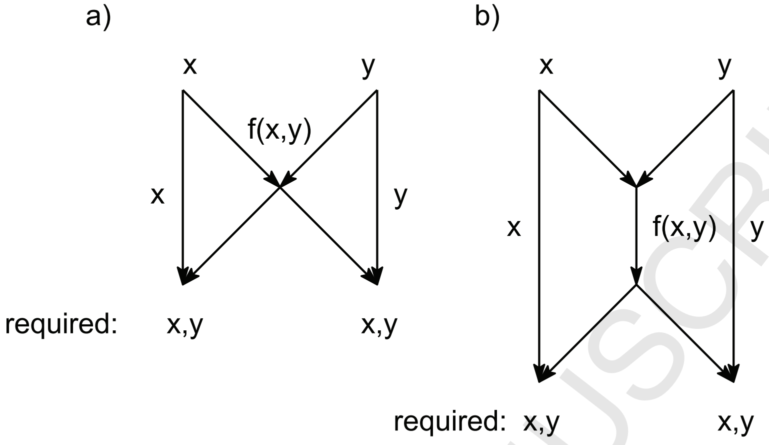


Figure 10: Butterfly network.

source set $S = \{x_1, x_2, x_3, x_4\}$ and target set $T = \Gamma$. The edge set $E \subseteq V \times V$ is given by

$$E = \big\{ \big(x_1, f(x_1, x_2)\big), \big(x_2, f(x_1, x_2)\big), \big(x_3, f(x_3, x_4)\big), \big(x_4, f(x_3, x_4)\big) \big\} .$$

The graph G_Γ is displayed in Figure 11.

Problem 1 can be viewed as a communication problem between a single source that transmits a message $(a_1, a_2, a_3, a_4) \in A^4$ and a user who receives a message of the form $\big(a_1, \bar{f}(a_1, a_2), \bar{f}(a_3, a_4), a_4\big) \in A^4$.

Example 14 (Distributed storage). *Assume that we want to store two messages $x, y \in A$ at four locations. The messages x and y are stored at two of the locations. At the two remaining locations two messages $f(x, y) \in A$ and $g(x, y) \in A$ are stored. The problem is to select the coding functions $f : A^2 \to A$ and $g : A^2 \to A$ such that it is always is possible to reconstruct x and y from accessing only two of the four locations.*

This type of problem has already been studied in the literature [17] as part of network coding, as well as an application of error correcting codes. The actual problem can be shown to be equivalent to the existence of two orthogonal Latin squares of order $|A|$ [18]. This problem was first posed by Euler around 1780 and was eventually completely solved in 1960, where it was shown in [19] that there exist orthogonal Latin squares of any order except of order 2 and order 6. The distributed storage problem is mathematically equivalent to the following problem.

Problem 2 (Equivalent to the distributed storage problem). *Construct two functions $\bar{f}, \bar{g} : A^2 \to A$ with Γ-dispersion equal to 10, where*

$$\Gamma := \{x_1, f(x_1, y_1), y_2, f(x_2, y_2), x_3, g(x_3, y_3), y_4, g(x_4, y_4), f(x_5, y_5), g(x_5, y_5)\}.$$

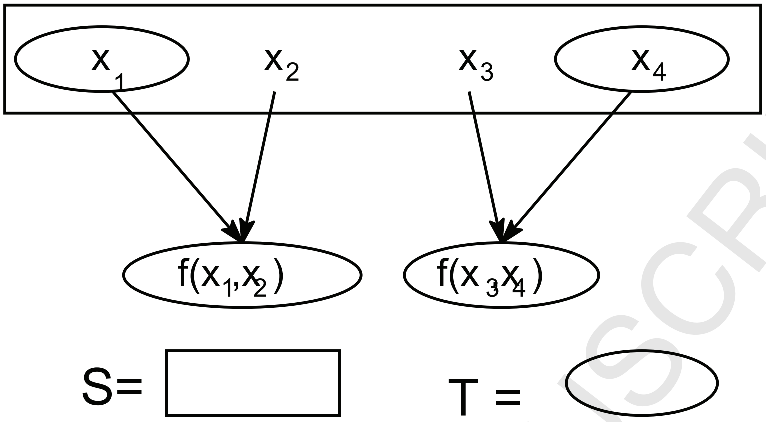


Figure 11: Graph G_Γ for the butterfly network.

Proof. We can view the storage problem as a many-to-many cast instance with two sources sending x and y and six users obtaining $\{x, y\}$, $\{x, f(x, y)\}$, $\{y, f(x, y)\}$, $\{x, g(x, y)\}$, $\{y, g(x, y)\}$, and $\{f(x, y), g(x, y)\}$ respectively. The demands of the first user are trivially satisfied, so only the last five need to be considered. Applying the transformations above and Proposition 11, we obtain the desired term set.

Problem 2 can be viewed as a communication problem between a single source that transmits a message $\mathbf{a} \in A^{10}$ and a user who receives a message $\psi(\mathbf{a}) \in A^{10}$.

We can give a network coding interpretation of the diversified term set. When the term set is diversified, then we can select a different solution for each user independently. Therefore, we have full dispersion in the diversified case if and only if each request can be satisfied individually.

Example 15. *Consider the term set* $\Gamma = \{x_1, f(x_1, x_2), x_4, f(x_3, x_4)\}$ *corresponding to the satellite communication problem. The diversified term set can be written as* $\Gamma^{\text{div}} = \{x_1, f(x_1, x_2), x_4, g(x_3, x_4)\}$. *Since the coding functions* f *and* g *are not required to be identical, it is easier to find coding functions that achieve the maximal dispersion of* 4 *in the case of* Γ^{div}. *One possible choice of coding functions that achieve dispersion of* 4 *is to let* $f(x_1, x_2) = x_2$ *and let* $g(x_3, x_4) = x_3$. *However, this choice does not correspond to any real-life situation for the original satellite communication problem. In fact, diversifying the term set is equivalent to considering the case where the satellite has access to two independent channels to the stations on which it can send a different message (one constructed by the function* f*, the other by the function* g*).*

The max-flow min-cut theorem for the dispersion then shows that if all demands can be satisfied individually, then for all large enough alphabets, all demands can be "nearly" satisfied at once. Theorems 1 and 2 quantify this statement in terms of a small loss in one-to-one dispersion. However, this small loss may be critical when network coding is considered. Indeed let Γ be the term set associated to a given multi-user communication problem, and let ψ be the induced function of an interpretation for Γ. Then, the one-to-one pre-image of ψ may contain very few points $(\mathbf{a}, \mathbf{a}, \ldots, \mathbf{a}) \in A^{km}$, where $\mathbf{a} \in A^k$. However, only these points make sense for the multi-user communication, as the variable x_i^j is merely an artificial variable representing x_i for all j. Consider the butterfly network for example. Then all the points in the one-to-one pre-image of dynamic routing satisfy $x_1 \neq x_3$, which does not correspond to any real-life situation for the satellite communication problem.

8. Dynamic networks

8.1. Theorem for multi-user dynamic networks

In the analysis of dynamic communication networks it is natural to take into account that networks change over time. Potential network changes including link failures, point failures and noisy channels can be modeled by ideas vaguely akin to Kripke's *possible world semantics* from logic and philosophy [20]. The idea is to consider a collection of *possible worlds* which each could become *the actual world* as time progresses.

A world is not only a representation of the network at a given time, but an expansion over a number of time slots. This is a generalization of the butterfly network, which can be viewed as an expansion over time of the satellite communication problem. As the model is discrete there are only a finite a number $|W|$ of possible worlds. We can think of each node in a world as a network node at a certain time slot. A point in each world thus represents a node at a given time slot, and as such there is a link from each node to its successor in time. This link is not a communication link, but represents the data transformation at the node in the two time slots. Communications might be instantaneous i.e. connect different nodes in the same time slot. The resulting network is acyclic. Formally we define

Definition 11 (Dynamic network). *Let $U, W,$ and T be finite sets. A* dynamic multiuser network *is a collection $\tilde{\Gamma} := \{\Gamma_{u,w,t} : u \in U, w \in W, t \in T\}$ of term sets indexed by a user/receiver $u \in U$, a world $w \in W$, and a time-slot $t \in T$. Each $\Gamma_{u,w,t}$ consists of a collection of terms as well as a collection of variables required. If each $\Gamma_{u,w,t}$ requires all the variables the network is a* dynamic many-to-many cast (or multi-cast) network

Assume all coding functions that occur in terms in $\tilde{\Gamma}$ have been given interpretations. Then we can associate a dispersion to each set term set $\Gamma_{u,w,t}$. For each set $\Gamma_{u,w,t}$ we associate a variable $\gamma_{u,w,t}$ which for each choice of coding functions denotes the dispersion of the term set $\Gamma_{u,w,t}$. Typically the same function symbol might occur in terms sets in multiple worlds. In general a coding function might be a good choice for some of the worlds, while it might be a bad choice for other worlds.

In order to take into account the fact that some worlds may be more likely than others, we consider the general case where each user $u \in U$ is assigned a *utility demand* D_u. This is of the form $D_u \equiv F_u(\gamma_u) > \mathrm{dem}_u$, where dem_u is a real number and the utility function F_u is a real-valued, non-decreasing, and continuous function in the variables $\gamma_u = \{\gamma_{u,w,t} : w \in W, t \in T\}$ representing the received dispersions. We say that the demand D_u is satisfied *locally* if it can be satisfied when all other user demands are disregarded. Conversely, we say that $D_1, D_2, \ldots, D_m$ are satisfied *globally* if they can all be satisfied by the same interpretation.

Theorem 6 (Dynamic multi-user theorem). *In a dynamic many-to-many cast where the demand D_u of each user $1 \leq u \leq m$ can be satisfied locally, the demands $D_1, D_2, \ldots, D_m$ can be satisfied globally. The same holds for Rényi entropy demands with $\alpha < 1$ and one-to-one dispersion demands.*

Proof. Suppose that each $F_u(\gamma_u) > \mathrm{dem}_u$ with $1 \leq u \leq m$ can be achieved locally. If we select $\delta > 0$ such that $\delta < \min_u\{F_u(\gamma_u) - \mathrm{dem}_u\}$, then in fact $F_u(\gamma_u) > \mathrm{dem}_u + \delta$ with $1 \leq u \leq m$ can be achieved locally. Assume that $F_u(\gamma_u) > \mathrm{dem}_u + \delta$ is achieved (locally) by the dispersions $\gamma_u = \{\gamma_{u,w,t} : w \in W, t \in T\}$. Since F_u is continuous, there exists $\epsilon > 0$ such that if the dispersion γ'_u of $\{\Gamma_{u,w,t} : w \in W, t \in T\}$ has $|\gamma'_{u,w,t} - \gamma_{u,w,t}| < \epsilon$, then $|F_u(\gamma'_u) - F_u(\gamma_u)| < \delta$. According to the multi-user max-flow min-cut theorem, for each $\epsilon > 0$ there exists an interpretation (over a sufficiently large alphabet) which globally achieves the dispersions $\gamma_{u,w,t} - \epsilon$, $u \in U, w \in W, t \in T$. Thus there exist coding functions such that $F_u(\gamma_u) > \mathrm{dem}_u$ for $1 \leq u \leq m$.

We remark that the demands could also be expressed as $F_u(\gamma_u) \geq \mathrm{dem}_u$. In that case, the dynamic multi-user theorem indicates that the demands can asymptotically be achieved globally if they can be asymptotically achieved locally. Example 16 below shows how the utility function can cover a broad family of performance measures.

Example 16. *Let $\tilde{\Gamma}$ consist of term sets $\Gamma_{u,w,t}$ where $u \in U, w \in W$, and $t \in T$. Assume world w occurs with probability p_w and the user u has assigned a weight ω_t proportional to the utility of the dispersion achieved in time slot t. Then the utility for each user $u \in U$ can be defined as*

$$F_u(\gamma_u) = \sum_{w \in W, t \in T} p_w \omega_t \gamma_{u,w,t}$$

which is a continuous function in the variables γ_u. Asymptotically, the maximal utility achievable for user u is given by

$$\sum_{w \in W, t \in T} p_w \omega_t \rho_{u,w,t}$$

where $\rho_{u,w,t}$ denotes the min-cut of $\Gamma_{u,w,t}$.

The example above can be viewed as an asymptotic generalization of the network coding theorem, which only considers one time slot and one possible world. A proper generalization of that result can be obtained by considering linear coding functions over a a diversified term set for a multi-user communication problem.

8.2. Clairvoyance and term equations

In a dynamic network, various unpredictable network changes (e.g. link failures) might happen in various of the possible worlds. A realistic choice of coding functions cannot look into the future and take into account which link might fail during transmission. Nonetheless, we define clairvoyance as the case where each node "knows" in which of the possible worlds the network is. We can define the formally as follows:

Definition 12 (Clairvoyant coding). *The clairvoyant version $\tilde{\Gamma}^{\text{clair}}$ of $\tilde{\Gamma}$ is defined as the collection of term sets where function symbols have been diversified so function symbols in different worlds are distinct, e.g. each function symbol the occurs in term sets with index $w \in W$ is assigned an (additional) index w. An assignment of functions symbols to $\tilde{\Gamma}^{\text{clair}}$ is said to consist of* clairvoyant coding functions *for $\tilde{\Gamma}$.*

As another application of the multi-user dispersion theorem we obtain that clairvoyance does not improve the performance of a network in terms of dispersion. The proof of Proposition 12 below is based on arguments similar to those used to prove the theorem for the dispersion.

Proposition 12 (Clairvoyance does not increase dispersion). *If in a dynamic many-to-many network the users demands $D_1, D_2, \ldots, D_r$ can be satisfied in $\tilde{\Gamma}^{\text{clair}}$ then they can be satisfied in $\tilde{\Gamma}$.*

It should be noticed that this result for the dispersion and the Rényi entropy is very much in the spirit of diversity coding and random linear network coding and is thus not surprising. Indeed, diversity coding intuitively deals with link failures and noisy channels by mixing the inputs and transmitting a large number of independent messages. Clairvoyance is thus rendered useless. However, our result about the one-to-one dispersion is remarkable, for–as seen in Section 4.1–a high one-to-one dispersion involves a controlled non-linear mixing, which contradicts the philosophy of diversity coding.

We finish this section by revealing that a model based on term equations can take into account the fact that each user not only requires a high dispersion, but also a certain number of specific messages. Recall from Section 5 that if a user receives the terms $\{t_1, t_2, \ldots, t_r\}$, we associate the decoding functions $\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_s$ where we require that $\bar{d}_i(\psi(\mathbf{a})) = a_i$ for all $1 \leq i \leq s$. This can be more succinctly expressed in the term equation $\tau_i = x_i$, where τ is the term defined as $\tau_i = d_i(t_1, t_2, \ldots, t_r)$.

In the case of message demands, then clairvoyance can clearly make a difference, as seen in Example 17 below.

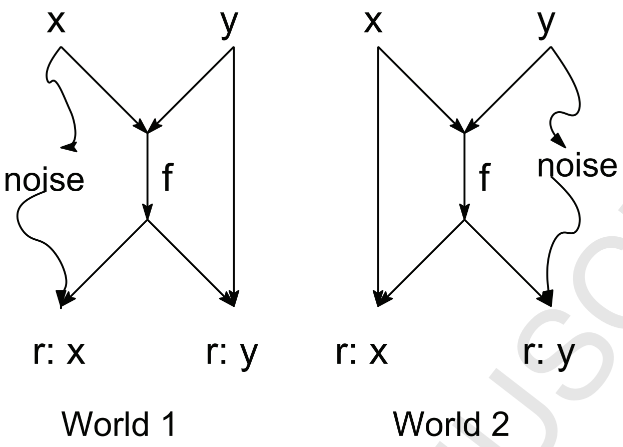


Figure 12: Dynamic network with message demands.

Example 17. *Consider the dynamic network depicted in Figure 12, where one of two links is always contaminated with pure noise. Remark that this is not the butterfly network in Figure 10. Without using clairvoyance, the message demands of the destinations can be expressed as term equations*

$$\begin{aligned} d_1(\text{noise}_1, f(x,y)) &= x \\ d_2(y, f(x,y)) &= y \\ d_3(x, f(x,y)) &= x \\ d_4(\text{noise}_2, f(x,y)) &= y. \end{aligned}$$

Here we assumed that the decoding functions can distinguish the messages x and y from noise, which is why we can apply different decoding functions to each of the four potential decoding situations. It is clear that the message demands of both users cannot be satisfied globally without clairvoyance. On the other hand, using clairvoyance, the problem is turned into the following set of term equations:

$$\begin{aligned} d_1(\text{noise}_1, f_1(x,y)) &= x \\ d_2(y, f_1(x,y)) &= y \\ d_3(x, f_2(x,y)) &= x \\ d_4(\text{noise}_2, f_2(x,y)) &= y. \end{aligned}$$

Thus, letting $\bar{f}_1(x,y) = x$ for world 1 and $\bar{f}_2(x,y) = y$ for world 2 solves the communication problem by an appropriate choice of decoding functions $\bar{d}_1, \bar{d}_2, \bar{d}_3$, and $\bar{d}_4$.

In brief, clairvoyance does not help in many-to-many casting where the users have dispersion demands. But not surprisingly in general–when different users have different message demands–clairvoyance can greatly increase the performance of the network.

9. Conclusion

There is an extensive literature for dealing with the logistics and scheduling in traditional commodity networks. The theories are very diverse ranging from linear programming, algorithms for transport of "discrete" goods, game theory, traffic flow theory, network exchange theory, economic network theory, packet switching, and queuing theory. It is less obvious that transport of digital information has a cost and that scheduling traffic of data is beneficial. Maybe this is why it historically was very late that people have begun (mainly in the field of Network Coding) to develop theories that cover transport of digital information in communication networks.

In this paper we developed a general theory for transport of digital information in relay networks. To summarise, we

- considered relay networks and explained how such networks can be used to model dynamic communication networks,
- introduced a formalism (the term model) that made it possible to handle communication problems in a graph-free approach,
- showed that the approximately throughput of any single-sender single-receiver relay network is given my its minimal term-cut,
- showed that any question about solvability of a given dynamic multi-user communication problem, can be restated as a question about solvability of a single-sender single-receiver problem in a specific relay network.

[1] R. Yeung, Z. Zhang, Distributed source coding for satellite communications, IEEE Trans. Info. Theory 45 (3) (1999) 1111–1120.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, R. W. Yeung, Network information flow, IEEE Trans. Info. Theory 46 (4) (2000) 1204–1216.

[3] R. Dougherty, C. Freiling, K. Zeger, Linearity and solvability in multicast networks, IEEE Transactions on Information Theory 50 (10) (2004) 2243–2256.

[4] R. Dougherty, C. Freiling, K. Zeger, Networks, matroids, and non-Shannon information inequalities, IEEE Transactions on Information Theory 53 (6) (2007) 1949–1969.

[5] R. Dougherty, C. Freiling, K. Zeger, Unachievability of network coding capacity, IEEE Transactions on Information Theory 52 (6) (2006) 2365–2372.

[6] S. Riis, Information flows, graphs and their guessing numbers, The Electronic Journal of Combinatorics 14 (R44) (2007) 1–17.

[7] T. Ho, M. Médard, R. Kötter, D. R. Karger, M. Effros, J. Shi, B. Leong, A random linear network coding approach to multicast, IEEE Trans. Info. Theory 52 (10) (2006) 4413–4430.

[8] A. Rényi, On measures of information and entropy, in: Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability, 1961, p. 547–561.

[9] S.-Y. R. Li, R. W. Yeung, N. Cai, Linear network coding, IEEE Trans. Info. Theory 49 (2) (2003) 371–381.

[10] R. Kötter, M. Médard, An algebraic approach to network coding, IEEE/ACM Transactions on Networking 11 (5) (2003) 782–795.

[11] S. Riis, Linear versus non-linear boolean functions in network flow, in: Proc. CISS, Princeton, NJ, 2004.

[12] R. Dougherty, C. Freiling, K. Zeger, Insufficiency of linear coding in network information flow, IEEE Transactions on Information Theory 51 (8) (2005) 2745–2759.

[13] J. Barwise, Handbook of Mathematical Logic, 2nd Edition, Studies in Logic and the Foundations of Mathematics, Amsterdam: North-Holland, 1977.

[14] K. Menger, Zur allgemeinen kurventhorie, Fundamenta Mathematicae 10 (1927) 95–115.

[15] E. Dinic, A solution of the max-flow problem with the polynomial estimation, Soviet Mathematics Doklady 11 (1970) 1277–1280.

[16] P. Keevash, O. Sisask, private communication.

[17] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, K. Ramchandran, Network coding for distributed storage systems, IEEE Trans. Info. Theory 56 (9) (2010) 4539–4551.

[18] S. Riis, R. Ahlswede, Problems in network coding and error-correcting codes, in: Proc. NetCod, Riva del Garda, Italy, 2005.

[19] R. Bose, S. Shrikhande, E. Parker, Further results on the construction of mutually orthogonal latin squares and the falsity of Euler's conjecture, Canadian Journal of Mathematics 12 (1960) 189–203.

[20] S. Kripke, A completeness theorem in modal logic, The Journal of Symbolic Logic 24 (1959) 1–15.