

# Recent Results on the Algorithmic Complexity of Network Coding

Michael Langberg  
Computer Science Division  
The Open University of Israel  
Raanana, 43107, Israel  
mikel@openu.ac.il

Alex Sprintson  
Department of Electrical and Computer Engineering  
Texas A&M University  
College Station, Texas, 77843  
spalex@ece.tamu.edu

**Abstract**—Given a network  $G(V, E)$ , a set of capacity constraints, and a set of communication requirements, how hard is it to find a network code (if such exists) that satisfies these requirements? How can we construct efficient network codes (e.g., over small alphabets)? How can we efficiently allocate resources in coding networks? This tutorial will address these questions in detail and present the current state of the art, open problems, and promising research directions.

In the first part of the tutorial we present an overview of the best known algorithms for construction of efficient network codes and resource allocation in coding networks. This part will focus primarily on multicast connections.

In the second part, we discuss the computational complexity of the general network coding problem, focusing on non-multicast scenarios. Specifically, we consider the problem of characterizing the rate regions of general network coding instances. Here not much is known. Namely, for scalar linear codes this problem is NP-Hard. However, for more general coding schemes, no such hardness results are known and the complexity of deciding whether a given rate vector lies in the capacity region is completely open. We will present recent results in this intriguing research area that touch both combinatorial graph theory and the study of entropic functions.

## I. INTRODUCTION

In the network coding paradigm, internal nodes of the network may mix the information content of the received packets before forwarding them. This mixing (or encoding) of information has been extensively studied over the last decade (see e.g., [1], [2], [3], [4], [5]). Design and analysis of efficient and capacity-achieving algorithms as well as complexity classification of the network coding problems are two of the most fundamental issues in the theory of network coding. While the advantages of network coding in the *multicast* setting are currently well understood, this is far from being the case in the context of *general* network coding. In particular, determining the capacity of a general network coding instance is a long standing open problem (see e.g., [6], [7]).

The goal of the tutorial is to provide an overview of the recent results in network coding algorithms and the computational complexity of the network coding problems. The first part of the tutorial will focus on the efficient algorithms for network code construction,

minimizing the complexity of network operation, and optimization of the total amount of consumed network resources. In the second part, we will discuss the computational complexity of the general network coding problem, focusing on non-multicast scenarios. We will present recent results in this intriguing research area that touch both combinatorial graph theory and the study of entropic functions. In this summary we will give only a high level description of the presented results. In particular, many of the terms that appear in *italics* will be defined rigorously during the talk.

## II. MULTICAST CONNECTIONS

An instance of multicast network coding problem includes a communication network  $G(V, E)$ , a source node  $s$ , and a set  $T$  of  $k$  terminals. We assume that the data is transmitted in packets such that each edge of the network can transmit one packet per channel use. The goal is to maximize the number of packets,  $h$ , delivered from the source node  $s$  to all terminals in  $T$ , under the constraint that each edge  $e \in E$  can be used exactly once. Note that all edges are assumed to be of unit capacity, but this assumption does not imply any loss of generality, as edges of larger capacity can be represented by multiple parallel edges of unit capacity.

It was shown in [1] that the capacity of the multicast networks, i.e., the maximum number of packets that can be delivered from the source  $s$  to the terminals in  $T$  is equal to the minimum total capacity of a cut that separates the source from a terminal. The maximum rate can be achieved by using *linear network codes* [2]. With linear network codes, all packets are elements of a finite field  $GF(q)$  and each packet transmitted over edge  $e = (v, u)$  is a linear combination of the packets received by the tail node  $v$  of  $e$ .

**Minimal networks.** We say that a multicast network over a graph  $G(V, E)$  with a source node  $s$  and a set  $T$  of destination nodes is *minimal* if removal of any edge  $e \in E$  would result in a decrease of its capacity. Clearly, a minimal network can be constructed through a simple greedy algorithm that iteratively removes redundant edges. Reference [8] shows an efficient algorithm

for construction of minimal networks. Minimal coding networks have several specific combinatorial properties that can be used for designing more efficient network coding algorithms. In particular, it was shown in [9] that the number of *encoding nodes* in a minimal acyclic coding network is bounded by  $O(h^3 k^2)$ .

**Randomized algorithms.** It was shown in [10] that a feasible network code can be efficiently constructed by choosing the *encoding functions*  $\beta_e$  for each  $e \in E$  at random. The success probability of the randomized scheme depends on the number of encoding edges in the network and the size of the underlying finite field. In particular, for  $q > k$  a random assignment of encoding function is feasible with probability at least  $(1 - \frac{k}{q})^\eta$ , where  $\eta$  is the total number of *encoding coefficients*. Note that the probability of finding a feasible network code can be amplified by repeating the random selection until a feasible solution is found. In a minimal coding network, the number of *encoding coefficients* is bounded by  $O(h^3 k^2)$ , hence the probability of success is higher.

**Deterministic algorithms.** In [4], Jaggi et al. proposed a deterministic polynomial-time algorithm for finding feasible network codes in multicast networks over a finite field of size  $O(k)$ . The algorithm visits each node of the network in a topological order and assigns the encoding coefficients for each outgoing edge  $e$  of node  $v$  in a way that preserves a certain invariant condition for each terminal  $t \in T$ . The computational complexity of this algorithm is  $O(|E|kh + |V|k^2 h^2(k + h))$ . Reference [8] presents an algorithm that has a running time of  $O(|E|kh + |V|k^2 h^2 + h^4 k^3(k + h))$ . This algorithm first constructs a minimal coding network and then applies the algorithm due to [4] on the resulting graph.

**Encoding complexity.** The important consideration in the design of network coding schemes is to minimize the total amount of computation performed by a network. The amount of computation performed by the network depends on the number of encoding nodes (i.e., nodes that generate new packets as opposed to nodes that just forward the received packets) and the size of the underlying field. It was shown in [9] that determining the minimal number of encoding nodes in a directed network is an NP-complete problem. Reference [9] also shows that the number of encoding nodes in an acyclic network is bounded by  $O(h^3 k^2)$ . It also presents an acyclic multicast network that requires as many as  $\Omega(h^2 k)$  encoding nodes. The number of encoding nodes in a multicast network depends on the size  $B$  of its *minimum feedback edge set*. In [9] it was shown that the number of encoding nodes in a network with cycles is bounded by  $(2B + 1)h^3 k^2$ .

**Field size.** The size of the finite field  $GF(q)$  employed by the network coding scheme is an important parameter that determines the amount of computation performed at each network node. In practical schemes an extension

field  $GF(2^\ell)$  of  $GF(2)$  is employed, i.e., every packet is a sequence of  $\ell$  bits. The addition operation in  $GF(2^\ell)$  can be implemented in  $O(\ell)$  time, while the multiplication typically requires  $O(\ell^2)$  time (the multiplication can be done in  $O(\ell)$  time using a look-up table, but such a table requires  $\Omega(q^2)$  storage space). The problem of finding the minimum required field size is NP-complete even for the special case of  $h = 2$  [11]. For  $h = 2$ , the field of size  $\lfloor \sqrt{2k - \frac{7}{4}} + \frac{1}{2} \rfloor$  is sufficient, and, for some networks, necessary [11], [12]. In the general case, a field of size  $k$  is sufficient for a network with  $k$  terminals for any  $h > 2$  [4].

**Low Complexity Encoding.** Reference [13] proposes a distributed randomized design of network codes for multicast networks with run time coding complexity of  $O(\ell)$  at source and internal nodes, where  $\ell$  is the number of bits in each transmitted packet. The proposed codes are referred to as *permute-and-add-network codes*. The main idea of this approach is that each encoding node first applies a random permutation on each of the incoming packets, and then generates a new packet by bitwise exclusive OR (XOR) operation on the resulting packets. The decoding complexity of this approach is  $O(\ell^2)$ .

**Cost minimization.** A important practical problem is to minimize the total amount of network resources allocated for a multicast connection. In such settings, we are given a graph  $G(V, E)$ , each edge  $e \in E$  is associated with a cost  $c_e$  of reserving this edge, a number of packets  $h$ , a source  $s$  and a set of terminals  $T$ . The goal is to find a minimum cost subgraph  $\hat{G}(\hat{V}, \hat{E})$  of  $G(V, E)$  that has a sufficient capacity to transmit  $h$  packets from  $s$  to each terminal  $t \in T$ . Note that if we restrict  $h$  to be one, the problem is equivalent to the well-known Steiner tree problem. The problem has several variations, all of them are NP-complete. Note that if  $G(V, E)$  is a directed graph, then the results on the directed Steiner trees [14] imply that it is hard to approximate the problem within a factor better than  $\ln k$ . For undirected graphs with no upper bounds on edge capacities (i.e., in the case in which an unrestricted number of copies of an edge can be used), the problem is related (but is not identical) to the network design problems [15], hence similar techniques can be used. Several techniques for cost minimization have been presented in [16].

### III. GENERAL NETWORK CODING

In this part of the talk we will discuss various results addressing the computational difficulty in deciding whether a given instance to the general network coding problem is feasible.

An instance of the general network coding problem includes a communication network  $G(V, E)$ , a set of source nodes  $S = \{s_i\}$ , and a set of terminal nodes  $T = \{t_j\}$ , and a set of source/terminal requirements  $\{(s_i, t_j)\}$  (implying that terminal  $t_j$  is interested in the

information available at source  $s_i$ ). We assume that the data is transmitted in packets such that each edge of the network can transmit one packet per channel use. Broadly speaking, the goal is to maximize the amount of information transmitted from source nodes  $S$  to the terminal nodes  $T$ .

**Scalar Linear Coding.** Informally, with *scalar linear codes* each packet cannot be split and needs to be sent throughout the network in one piece. This is in contrast to vector linear network codes, in which each packet can split into a vector of smaller size packets. In [11] it was proven that determining the *scalar linear capacity* of an instance  $I$  to the network coding problem is  $\mathcal{NP}$ -hard. The proof uses a reduction from the well known 3-SAT problem.

**Index Coding.** The Index Coding problem has a simple and elegant formulation [17], [18]. The problem captures many important aspects of the more general network coding problem.

The connection between the Index Coding and Network Coding problems was studied in [19] and [20]. In particular, these references use the connection to the Index Coding problem to show several properties of the Network Coding problem, such as hardness of finding an approximation solution for scalar linear codes and for a certain class of vector linear codes.

**Vector Linear Coding.** In a *vector linear network code*, each packet can be split into  $k$  smaller size packets, where  $k$  is said to be the *dimension* of the vector-linear code. References [21], [22] show that vector linear codes outperform scalar linear codes in terms of obtainable rate. For constant values of  $k$ , reference [19] shows the hardness of finding a *vector linear* solution to a general instance to the network coding problem which achieves (or even *approximately* achieves) *capacity*. In particular, [19] shows that for any constant  $\alpha < 1$  and  $k$  if one can efficiently find a *vector linear* solution of *dimension*  $k$  of rate  $\alpha$  times the *capacity* of the instance, then one can efficiently color 3-colorable graphs with a constant number of colors. The latter is a well known open problem (see e.g., [23]).

**Entropic functions.** Finally we turn to discuss the connection between the capacity of general network coding instances and the family  $\bar{\Gamma}^*$  of *entropic* functions [24], [25]. The characterization of  $\bar{\Gamma}^*$  is an intriguing open problem. Reference [26] shows that determining the *capacity region* for the general network coding problem implies a characterization of  $\bar{\Gamma}^*$ .

#### REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network Information Flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.
- [2] S.-Y. R. Li, R. W. Yeung, and N. Cai. Linear Network Coding. *IEEE Transactions on Information Theory*, 49(2):371 – 381, 2003.
- [3] R. Koetter and M. Medard. An Algebraic Approach to Network Coding. *IEEE/ACM Transactions on Networking*, 11(5):782 – 795, 2003.
- [4] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. Tolhuizen. Polynomial Time Algorithms for Multicast Network Code Construction. *IEEE Transactions on Information Theory*, 51(6):1973–1982, June 2005.
- [5] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong. A Random Linear Network Coding Approach to Multicast. *IEEE Transactions on Information Theory*, 52(10):4413–4430, 2006.
- [6] T. Cover and J. Thomas. *Elements of Information Theory*, second edition. John Wiley and Sons, 2006.
- [7] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang. *Network Coding Theory*. Now Publishers Inc, 2006.
- [8] M. Langberg, A. Sprintson, and J. Bruck. Network Coding: A Computational Perspective. *Information Theory, IEEE Transactions on*, 55(1):147–157, Jan. 2009.
- [9] M. Langberg, A. Sprintson, and J. Bruck. The Encoding Complexity of Network Coding. *Information Theory, IEEE Transactions on*, 52(6):2386–2397, June 2006.
- [10] T. Ho, R. Koetter, M. Medard, D. Karger, and M. Effros. The Benefits of Coding over Routing in a Randomized Setting. In *Proceedings of the IEEE International Symposium on Information Theory*, 2003.
- [11] A. Lehman and E. Lehman. Complexity Classification of Network Information Flow Problems. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '04)*, pages 142–150, Philadelphia, PA, USA, 2004.
- [12] C. Fragouli and E. Soljanin. Information Flow Decomposition for Network Coding. *Information Theory, IEEE Transactions on*, 52(3):829–848, March 2006.
- [13] S. Jaggi, Y. Cassuto, and M. Effros. Low Complexity Encoding for Network Codes. pages 40–44, July 2006.
- [14] M. Charikar, C. Chekuri, T. Cheung, Z. Dai, A. Goel, S. Guha, and M. Li. Approximation Algorithms for Directed Steiner Problems. *Journal of Algorithms*, 33(1):73 – 91, 1999.
- [15] K. Jain. A Factor 2 Approximation Algorithm for the Generalized Steiner Network Problem. pages 448–457, Nov 1998.
- [16] D. Lun, M. Médard, T. Ho, and R. Koetter. Network Coding with a Cost Criterion. In *Proceedings of International Symposium in Information Theory and its Applications*, Parma, Italy, October 2004.
- [17] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol. Index Coding with Side Information. In *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 197–206, 2006.
- [18] Y. Birk and T. Kol. Coding on Demand by an Informed Source (iscod) for Efficient Broadcast of Different Supplemental Data to Caching Clients. *Information Theory, IEEE Transactions on*, 52(6):2825–2830, June 2006.
- [19] M. Langberg and A. Sprintson. On the Hardness of Approximating the Network Coding Capacity. In *To appear in the proceedings of ISIT 2008*, Toronto, Canada, 2008.
- [20] S. Y. El Rouayheb, A. Sprintson, and C. Georgiades. On the Relation Between the Index Coding and the Network Coding Problems. In *proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2008.
- [21] M. Médard, M. Effros, T. Ho, and D. Karger. On Coding for Non-Multicast Networks. In *Proceedings of the 41st Allerton Conference on Communication, Control and Computing*, Monticello, IL, October 2003.
- [22] R. Dougherty, C. Freiling, and K. Zeger. Networks, Matroids, and Non-Shannon Information Inequalities. *IEEE Transactions on Information Theory*, 53(6):1949–1969, 2007.
- [23] U. Feige, M. Langberg, and G. Schechtman. Graphs with Tiny Vector Chromatic Numbers and Huge Chromatic Numbers. *SIAM Journal on Computing*, 33(6):1338–1368, 2004.
- [24] R. W. Yeung. A First Course in Information Theory. *Springer (Kluwer Academic/Plenum Publishers)*, 2002.
- [25] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang. Network Coding Theory. *Foundations and Trends in Communications and Information Theory*. Now Publishers, 2006.
- [26] T. Chan and A. Grant. Dualities Between Entropy Functions and Network Codes. *IEEE Transactions on Information Theory*, 54(10):4470–4487, 2008.