The Cost Advantage of Network Coding in Uniform Combinatorial Networks

Andrew Smith†, Bryce Evans†, Zongpeng Li†, Baochun Li‡
† Department of Computer Science, University of Calgary
‡ Department of Electrical and Computer Engineering, University of Toronto

***Abstract*—The coding advantage refers to the potential for network coding to improve end-to-end throughput or reduce routing cost. How large can the coding advantage be? We investigate this fundamental question in the classic undirected network model. After almost a decade of research in network coding, so far all known networks where such potential exists are based on a special class of topologies known as combinatorial networks. We try to prove a rather small upper-bound (close to 1) for the coding advantage for the class of combinatorial networks and its variations. Such a result, interestingly, will lead us to the following dilemma: either we are still missing the most appropriate network topologies for demonstrating the power of network coding, after a decade of research in network coding, or we have been ignoring a very effective perspective for efficiently approximating the minimum Steiner tree problem, after a few decades of research in Steiner trees. We elaborate on the above arguments and present the early stage results of our research: the coding advantage (in terms of routing cost) is upper-bounded by 1.125 in the class of uniform combinatorial networks.**

I. Introduction

Network Coding [1], [2] is a field of information and coding theory which allows the expansion of the function of individual nodes in a computer network beyond the standard operations used in routing. It does this by allowing nodes to perform encoding operations on incoming information flows, which provides more flexible solutions for propagating information throughout the network.

The benefits of network coding have been shown to be multi-fold. Among them two salient ones are increasing the end-to-end throughput and reducing the routing cost, especially for multicast transmissions or in wireless ad hoc networks. In the context of throughput, the coding advantage is defined as the ratio of the maximum end-to-end throughput with and without the use of network coding; in the context of routing cost, the coding advantage is defined as the ratio of the minimum cost necessary to achieve a certain throughput, without and with network coding. In the latter case we also refer to the coding advantage as the cost advantage of network coding. When the coding advantage is 1, network coding does not make a difference; when the coding advantage is large, network coding provides a substantial boost in throughput or a dramatic reduction in cost. In the classic undirected network model that we study here, the coding advantage is best shown when the communication session is in the form of one-to-many multicast. Previous research has shown that (a) for directed networks, the coding advantage is essentially unbounded and can grow at rate $\Theta(n)$, where n is the size of a multicast network [3]; (b) for undirected networks, the coding advantage is finite and its best upper-bound proven so far is 2 [4], [5]. Note that wireless ad hoc networks with uniform communication radius, *i.e.*, the unit-disk graphs, can be modelled using an undirected network since (i) reachability between neighboring nodes are always symmetrical, and (ii) data transmission in the two directions between a pair of neighbors share the total available channel capacity. A more accurate model would take into account details in wireless interference and the wireless broadcast advantage, though.

Although the best theoretical bound so far is 2, the largest value of the coding advantage observed in practice, including in both contrived and random networks, is only 1.125 or $\frac{9}{8}$ for finite networks, and 1.143 or $\frac{8}{7}$ for infinite networks. Therefore it is probable that the tight upper-bound of the coding advantage is much closer to 1 than to 2. In this paper, we prove that the cost advantage of network coding is tightly upper-bounded by 1.125 in the class of uniform combinatorial networks. This class of networks is particularly important since almost all known examples that shows a larger than 1 coding advantage belongs to it or can be reduced to its instances.

Definition A *Combinatorial Network* with parameters (n, k), denoted $C_{n,k}$, is an undirected graph $G = (V, E)$.

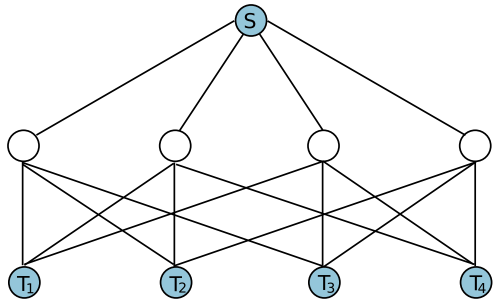

Fig. 1. An example of a combinatorial network, $C_{4,3}$. Here the coding advantage is 1.125, in terms of both throughput and cost.

$C_{n,k}$ consists of a sender, n relay nodes, and $\binom{n}{k}$ receivers. Each of the n relay nodes are connected to the sender directly, while each of the receivers is connected to a unique subset of size k of the relay nodes.

If further, all edges in E have uniform capacity and cost, we refer to the network as a *uniform combinatorial network*.

While the ultimate goal remains to provide an improved bound on coding advantage in general undirected networks, we initially restrict our attention to combinatorial networks for several reasons. First, most undirected network topologies exhibit a coding advantage of 1, which is equivalent to no improvement with the employment of network coding, while combinatorial networks generally exhibit a non-trivial coding advantage. Second, the largest coding advantage yet demonstrated was shown in two combinatorial networks, specifically $C_{4,3}$ (shown in Fig. 1) and $C_{4,2}$.

Furthermore, almost all[1] undirected topologies with non trivial coding advantage contain simple variants of combinatorial networks, suggesting that the structure of these networks is fundamentally related to network coding's potential to improve multicast routing solutions. For example, the celebrated butterfly network example (shown in Fig. 2), frequently cited and through which network coding is introduced to the public [1], has a topology that is isomorphic to $C_{3,2}$. The example topology depicted by Jaggi *et al.* [3] when showing that the coding advantage is unbounded in directed networks, is exactly $C_{6,3}$. The contrived topologies tested by Li *et al.* [6] for the coding advantage included $C_{3,2}$, $C_{4,2}$, $C_{4,3}$, $C_{5,2}$ and $C_{5,3}$. The reason that larger combinatorial

[1]The only exception we know of is the infinite network topology that leads to a coding advantage of 1.143, as mentioned earlier in the paper.

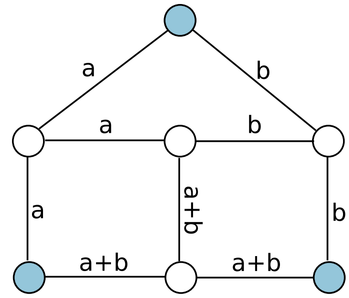

Fig. 2. The Butterfly Network, with a topology isomorphic to Fig. 3.

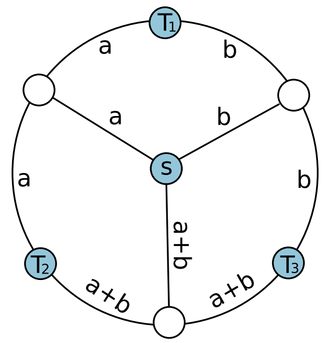

Fig. 3. The combinatorial network $C_{3,2}$, a possible drawing.

networks were not tested is computational: the general Steiner tree algorithm employed requires examining nearly 50 million different trees already for $C_{5,3}$. We develop a close-form formula for the coding advantage in this paper instead, by exploiting the specific structure of combinatorial networks, thus avoiding expensive computations.

We are working to prove that, for all combinatorial networks and their certain variations, the coding advantage is upper-bounded by a small number close to 1. Since proving a tight bound (in fact anything smaller than 2) for the coding advantage for general network topologies has proven hard, working on the combinatorial class seems to be a reasonable first step, especially that all known topologies for network coding to make a difference are variants of this class.

Below we discuss another important motivation for our research on the coding advantage, which is related to efficiently approximating the minimum Steiner tree and Steiner tree packing problems. Both variants of the Steiner tree problems have been shown to be NP-hard, and it is also known that an α-approximation exists for one of them in polynomial time if and only if the same is true for the other [7]. On the other hand, both multicast throughput and cost with network coding can be computed in polynomial time [6], [8]. Therefore, a bound of α on the coding advantage in terms of throughput or cost, directly implies the existence of a polynomial-time α-approximation algorithm for the Steiner tree packing and minimum Steiner tree problems, respectively. Note that the Steiner tree packing and minimum Steiner tree problems correspond to multicast routing without network coding. If we accomplish the proof of a small upper-bound α on the coding advantage for combinatorial networks and their variants, then the following question can be asked:

Does there exist another fundamentally different class of network topologies, for which a larger coding advantage can be observed?

If the answer is 'Yes', then that means after a decade of research in network coding, we are still missing the essence of it, in that the most appropriate networks for network coding to show its power have not been discovered and studied yet. If the answer is 'No', then that leads to a much better approximation algorithm for Steiner trees than the state-of-the-art 1.58-approximation [9], achieved after a few decades of research in the area of Steiner trees. We note that results we derive in this paper already represents an improvement in approximating the Steiner tree problems in uniform combinatorial networks: before this work, the best polynomial-time algorithm is a 1.27-approximation that works for all quasi-bipartite networks [9]; our results implies a 1.125-approximation instead.

We define our network model in the next Section, and present our early stage result in Sec. III: the cost advantage of network coding is tightly upper-bounded by 1.125 for uniform combinatorial networks. In Sec. IV, we describe how we plan to extend this bound from uniform combinatorial networks to general combinatorial networks with heterogeneous costs and capacities on links. Concluding remarks are in Sec. V.

II. Notation and Network Model

We denote the topology of a (multicast) network using an undirected graph $G = (V, E)$. $M = \{S, T_1, \ldots, T_k\} \subseteq V$ contains terminal nodes in the multicast group, in which S is the multicast sender. The desired multicast throughput rate is d. Each link $e \in E$ has a capacity $C(e)$ and a cost $w(e)$ that are both positive rational numbers. In this papers we focus on uniform combinatorial networks with uniform link capacities and costs, and assume that $C(e)$ and $w(e)$ are always 1 for all links. Without loss of generality, we also scale the throughput d to 1 (the important point is that link capacities are no less than the desired throughput and therefore will not become a limiting factor in routing).

Now, for a given multicast problem $\mathcal{P}$ defined upon a certain network configuration (topology, throughput, link costs and capacities), define $\pi_{\mathcal{P}}$ to be the minimum cost required to achieve the throughput without network coding, and define $\chi_{\mathcal{P}}$ to be the minimum cost required when employing network coding. Then, the *cost advantage* of network coding is given as $\pi_{\mathcal{P}}/\chi_{\mathcal{P}}$.

III. A Tight Upper-Bound for Uniform Combinatorial Networks

Now focus on a multicast problem $\mathcal{P}$ that is defined upon a uniform combinatorial network $C_{n,k}$, with desired multicast throughput 1 and uniform link cost of 1. We prove a tight upper-bound of 1.125 for the cost advantage of network coding.

Theorem III.1. *The cost advantage of network coding for the combinatorial network $C_{n,k}$ has a closed-form representation of*

$$\frac{\binom{n}{k} + n - k + 1}{\binom{n}{k} + \frac{n}{k}}$$

Proof: The theorem III.1 follows from lemmas III.2 and III.3, which prove the minimum cost of multicast in combinatorial networks with and without network coding respectively. ■

Lemma III.2. *The minimum multicast cost of rate 1 in a combinatorial network $C_{n,k}$ with network coding is $\binom{n}{k} + \frac{n}{k}$.*

We begin by proving the existence of a multicast solution of the given cost, followed by a proof of its minimality.

Proof of existence:

First note that the combinatorial network $C_{n,k}$ contains k links to each of the $\binom{n}{k}$ receivers and a single link to each of the n relays for a total of $k\binom{n}{k} + n$ links. Clearly, a flow assignment of $\frac{1}{k}$ on every link results in the desired throughput. In addition, every relay node has an incoming link with a flow of $\frac{1}{k}$ and can thus provide the necessary flow of $\frac{1}{k}$ on each of its outgoing edges. ■

Proof of minimality:

Consider the following statement of the MIN-CUT MAX-FLOW theorem.

> A flow rate d from node S to node T is achievable iff every cut between S and T has size at least d.

Further, note that a multicast rate d is only achievable if (and obviously only if) a unicast rate d is feasible from the sender to each receiver, in a directed network [1], [2]. This allows us to consider each flow from the sender node S to some receiver T_i independently. One cut that separates S from T_i is to remove all k links entering T_i. In order to sustain a throughput of 1, the sum of the flows assigned to these links must be at least 1. By applying the same reasoning to each of the $\binom{n}{k}$ receivers, the total flow on all links connecting the receivers must be $\binom{n}{k}$ (noticing that for each T_i there is no overlap of links included in the cut with any other receiving node).

Every subset of k relay nodes must also have an incoming flow that sums to 1, also by the MIN-CUT MAX-FLOW theorem. Because there are $\binom{n}{k}$ of these subsets, that would imply a minimum additional flow of $\binom{n}{k}$ if they shared no links. However, each relay node and its incoming links participates in $\binom{n-1}{k-1}$ of these subsets and so the flow on incoming links to that particular relay is shared by those subsets (over counted $\binom{n-1}{k-1}$ times). The resulting minimum flow implied by these cuts is thus

$$\frac{\binom{n}{k}}{\binom{n-1}{k-1}} = \frac{\frac{n!}{k!(n-k)!}}{\frac{(n-1)!}{(k-1)!((n-1)-(k-1))!}} = \frac{n!(k-1)!(n-k)!}{(n-1)!k!(n-k)!} = \frac{n}{k}$$

Therefore, by the MIN-CUT MAX-FLOW theorem, the minimum possible cost for unit flow in $C_{n,k}$ is $\binom{n}{k} + \frac{n}{k}$ ■

Lemma III.3. *The minimum multicast cost of rate* 1 *in a combinatorial network $C_{n,k}$ without network coding is* $\binom{n}{k} + n - k + 1$.

Proof: It is has been shown that the minimum cost of multicast without network coding is equivalent to the cost of the minimum Steiner tree. We demonstrate that in any combinatorial network, a Steiner tree of can be constructed with no fewer than $\binom{n}{k} + n - k + 1$ links.

To connect each of the $\binom{n}{k}$ receivers, each receiver must have an incoming link from a relay node. This gives the first $\binom{n}{k}$ links. We want to connect the minimum number of relay nodes while still having a connected relay available to each receiver. If we connect $r \leq n-k$ of the n relay nodes to the sender, the receivers that remain without connection are those that choose their k available links from the $n-r$ disconnected relays. By the uniform construction of the network, independent of which r nodes are connected, there remain $\binom{n-r}{k}$ disconnected receivers. Thus when $r = n-k$ there remains $\binom{k}{k} = 1$ receiver disconnected. Hence there must be at least $n-k+1$ relays connected to connect the Steiner tree to all receiver nodes. The total number of links in this minimum Steiner tree is $\binom{n}{k} + n - k + 1$. ■

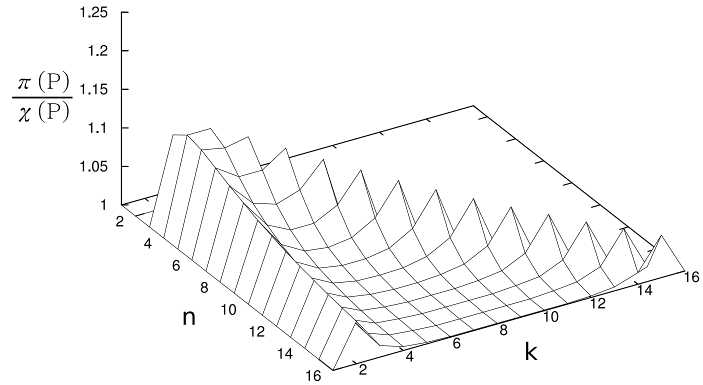

Fig. 4. Coding Advantage for Uniform Combinatorial Networks Base Case

So we now have a formula for computing coding advantage for any combinatorial network with unbounded capacity and uniform cost. The next result shows that this formula, and hence the coding advantage, is bounded inclusively by $\frac{9}{8}$.

Theorem III.4. *The network coding cost advantage for the combinatorial network $C_{n,k}$ is bounded inclusively by $\frac{9}{8}$.*

Proof:

Investigation of this formula for values of n and k up to 16 has shown no value for the coding advantage greater than $\frac{9}{8}$ [10]. This is demonstrated in Figure 4, which displays the Coding Advantage as a height map for values of n and k up to 16. This forms the base case for our induction on n for all $C_{n,k}$.

We now demonstrate that this upper bound holds for all values of n and k. First observe that a proof that the above formula is bounded by $\frac{9}{8}$ can be achieved by proving the following series of equivalent expressions:

$$\begin{aligned}
\frac{\binom{n}{k}+n-k+1}{\binom{n}{k}+\frac{n}{k}} &\leq \frac{9}{8} \\
\frac{\binom{n}{k}+n-k+1}{\binom{n}{k}+\frac{n}{k}} - \frac{\binom{n}{k}+\frac{n}{k}}{\binom{n}{k}+\frac{n}{k}} &\leq \frac{9}{8}-1 \\
\frac{n-k+1-\frac{n}{k}}{\binom{n}{k}+\frac{n}{k}} &\leq \frac{1}{8} \\
\frac{8(n-k+1-\frac{n}{k})}{\binom{n}{k}+\frac{n}{k}} &\leq 1 \\
8(n-k+1-\frac{n}{k}) &\leq \binom{n}{k}+\frac{n}{k} \\
8n-8k+8-\frac{8n}{k}-\frac{n}{k} &\leq \binom{n}{k} \\
8n-8k-\frac{9n}{k}+8 &\leq \binom{n}{k}
\end{aligned}$$

Now, treating separately the cases where $k = 1$ or $k = n-1$, we assume that $2 \leq k \leq n-2$. Beginning with the right hand side, notice that:

$$\binom{n}{k} \geq \binom{n}{2} = \binom{n}{n-2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}$$

Using this as a bound, we show that $8n - 8k - \frac{9n}{k} + 8 \leq \frac{n(n-1)}{2}$. Still working under the assumption that $2 \leq k \leq n-2$:

$$\begin{aligned}
8n-8k-\frac{9n}{k}+8 &\leq 8n-16-\frac{9n}{k}+8 \\
&= 8n-8-\frac{9n}{k} \\
&\leq 8n-8 \\
&= 8(n-1) \\
&= \frac{16(n-1)}{2} \\
&\leq \frac{n(n-1)}{2}, \forall n \geq 16
\end{aligned}$$

Thus, for $n \geq 16$ and for all $k, k \geq 2, k \leq n-2$ we have established a bound of $\frac{9}{8}$ for the formula.

For the case where $k = 1$ or $k = n-1$ we have the following:

$$\binom{n}{k} \geq \binom{n}{1} = \binom{n}{n-1} = n$$

So we show that $8n - 8k - \frac{9n}{k} + 8 \leq n$. For $k = 1$:

$$8n-8k-\frac{9n}{k}+8 = 8n-8-9n+8 = -n \leq n$$

and for $k = n-1$

$$8n-8k-\frac{9n}{k}+8 = 16-\frac{9n}{n-1} \leq 16$$

which shows that the bound also holds in this case when $n \geq 16$.

■

Hence we have the desired upper-bound of 1.125 for the class of uniform combinatorial networks. Since the value of 1.125 is achieved at both $C_{4,2}$ and $C_{4,3}$, we know the bound is tight.

IV. Next Steps

As discussed earlier in Sec. I, our eventual goal is to prove a small upper-bound for the coding advantage (for both throughput and cost) in all networks that are either combinatorial networks or their variations. A simple variation can be obtained, for example, by replacing a link in $C_{n,k}$ with two sequential links, each with the original capacity and cost. Such a variation is modelled with heterogeneous link costs — doubling the cost on the original link has the same effect. We have finished extending the proof in this paper so that the upper-bound of 1.125 still holds when relaxing the uniform link cost assumption; this result will be presented in an upcoming work. We are currently working on further generalizing the bound towards all combinatorial networks, with both heterogeneous link costs and heterogeneous link capacities. After that, we shall leverage the results obtained for cost advantage towards proving a similar bound on the throughput advantage.

V. Conclusions

At this point the result is proven for the subclass of uniform combinatorial networks where link capacities are non-limiting and the cost is equal across all links. We also expect to prove that having heterogeneous costs with or without link capacity constraints will still lead to combinatorial networks with coding advantage less than or equal to $\frac{9}{8}$.

The result in this paper improves upon the upper-bound established by Li et al. [6] for general undirected networks. While the class the bound is identified with is highly structured, it is one of the few topologies that demonstrate a non-trivial coding advantage. Hence boundaries on this class of structures may have farther reaching consequences regarding the maximum coding advantage possible for multicast in undirected networks.

Another motivation for our research on bounding the coding advantage, as discussed in the Introductions, has been related to the topic of designing efficient approximation algorithms for Steiner tree problems.

In future work, we plan to study whether the bound 1.125 is still valid once heterogeneous link capacities are introduced, and throughput d is large such that the finite link capacities becomes a non-trivial issue in multicast routing. We expect the answer to be ‘yes’. We also plan to translate the bound of 1.125 from cost reduction to improvement, for all combinatorial networks. The long-term direction after these, will be to determine the tight upper-bound for the coding advantage, for all general undirected networks.

REFERENCES

[1] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, “Network information flow,” *Information Theory, IEEE Transactions on*, vol. 46, no. 4, pp. 1204–1216, Jul 2000.

[2] R. Koetter and M. Medard, “Beyond routing: an algebraic approach to network coding,” *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 1, pp. 122–130 vol.1, 2002.

[3] 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*, vol. 51, no. 6, pp. 1973–1982, June 2005.

[4] Z. Li and B. Li, “Network Coding in Undirected Networks,” in *Proc. of the 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004.

[5] A. Agarwal and M. Charikar, “On the advantage of network coding for improving network throughput,” *Proc. 2004 IEEE Information Theory Workshop*, pp. 247–249, 2004.

[6] Z. Li, B. Li, D. Jiang, and L. C. Lau, “On achieving optimal throughput with network coding,” *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 3, pp. 2184–2194 vol. 3, 13-17 March 2005.

[7] K. Jain, M. Mahdian, and M. R. Salavatipour, “Packing Steiner Trees,” in *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2003.

[8] D. S. Lun, N. Ratnakar, R. Koetter, M. Médard, E. Ahmed, and H. Lee, “Achieving Minimum-cost Multicast: a Decentralized Approach Based on Network Coding,” in *Proceedings of IEEE INFOCOM*, 2005.

[9] C. Gröpl, S. Hougardy, T. Nierhoff, and H. J. Prömel, “Steiner trees in uniformly quasi-bipartite graphs,” *Inf. Process. Lett.*, vol. 83, no. 4, pp. 195–200, 2002.

[10] B. Evans and A. J. Smith, “The cost advantage of network coding,” December 2007, submitted to Dr. Zongpeng Li in fulfillment of the requirements of CPSC599.18.