Bounding The Coding Advantage of Combination Network Coding in Undirected Networks

Shreya Maheshwar; Zongpeng Li, *Member, IEEE*
Department of Computer Science
University of Calgary
{*smmahesh, zongpeng*}*@ucalgary.ca*

Baochun Li, *Senior Member, IEEE*
Deptartment of Electrical & Computer Engineering
University of Toronto
bli@eecg.toronto.edu

Abstract

We refer to network coding schemes in which information flows propagate along a combination network topology as combination network coding (CNC). CNC and its variations are the first network coding schemes studied in the literature, and so far still represent arguably the most important class of known structures where network coding is nontrivial. Our main goal in this paper is to seek a thorough understanding on the advantage of CNC in undirected networks, by proving a tight bound on its potential both in improving multicast throughput (the coding advantage) and in reducing multicast cost under a linear link flow cost model (the cost advantage). We prepare three results towards this goal. First, we show that the cost advantage of CNC is upper-bounded by $\frac{9}{8}$ under the uniform link cost setting. Second, we show that achieving a larger cost advantage is impossible by considering an arbitrary instead of uniform link cost configuration. Third, we show that in a given network topology, for any form of network coding, the coding advantage under arbitrary link capacity configurations is always upper-bounded by the cost advantage under arbitrary link cost configurations. Combining the three results together, we conclude that the potential for CNC to improve throughput and to reduce routing cost are both upper-bounded by a factor of $\frac{9}{8}$. The bound is tight since it is achieved in specific networks. This result can be viewed as a natural step towards improving the bound of 2 proved for the coding advantage of general multicast network coding.

Index terms: Multicast, Network Coding, Tree Packing, Combination Networks, Throughput-Critical Networks.

I. Introduction

While traditional information transmission in a data network is based on a store-and-forward style of routing, research on network coding has changed the landscape since its introduction about a decade ago [1]. With network coding, information flows can be 'mixed' at intermediate nodes in the network during the course of transmission. While the first *butterfly* network example used to motivate network coding shows an increase in multicast network capacity [1], it was later also discovered that network coding has other benefits such as reducing multicast cost, reducing energy consumption in wireless networks, improving transmission robustness and reducing the computational complexity of certain network routing problems (*e.g.*, see introduction in [2]).

In this paper, we study a special form of network coding referred to as *combination network coding (CNC)*, in which information flows propagate along a *combination network* topology $C_{n,k}$: an undirected graph containing one source, n relays, $\binom{n}{k}$ receivers, and each relay is connected to the source, each receiver is connected to a different set/combination of k receivers. Actually we can exchange the role between the source and one of the receivers without affecting the achievable multicast throughput, due to the following *source independent* property of multicast in undirected networks [3]: *in an undirected multicast network with given network topology, link capacities and terminal node set, the maximum achievable multicast throughput, either with or without network coding, is independent of the choice of the sender from the terminal node set.* An example application of CNC over a $C_{4,2}$ topology is illustrated in Fig. 1.

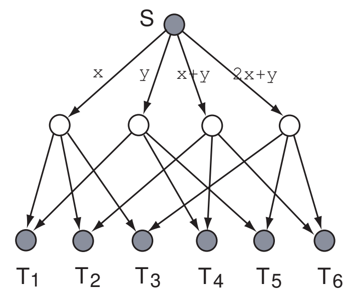

Fig. 1. An instance of CNC in the $C_{4,2}$ topology. Every link has unit capacity 1. x and y are two information flows of rate 1. $x+y$ and $2x+y$ are two flows encoded (over a finite field) from x and y, and each has a unit rate too. Each T_i can recover x and y from its received flows. A multicast throughput of 2 is achieved from S to $T_1, \ldots, T_6$. Note that the underlying network topology is undirected; the links directions shown are for the multicast flow of CNC.

An interesting observation is that, in the paradigm of wireline networks, many if not most known examples that demonstrate the advantage of network coding, either in improving throughput or in reducing cost, are of a structure identical or similar to that of CNC. We mention a few examples below. The butterfly network that is often used as a first example to motivate network coding (*e.g.*, [1], Fig.7; [4], Fig.1; [5], Fig.1) is isomorphic to $C_{3,2}$, only the source in $C_{3,2}$ is shifted to one of the receivers instead. Such a shift does not affect the multicast throughput in undirected networks [6], [3]. CNC over $C_{3,2}$ itself was also used in the first study of network coding ([1], Fig.8). CNC in $C_{4,2}$ ([7], Fig. 2) and $C_{6,3}$ ([4], Fig. 2; [7], Fig. 3) were used as examples to demonstrate the advantage of network coding in directed wireline networks. In a numerical comparison of network coding with tree packing in undirected networks, CNC was studied in a number of combination topologies including $C_{3,2}$, $C_{4,2}$, $C_{4,3}$, $C_{5,2}$ and $C_{5,3}$ ([8], Table II). The example that shows the largest known coding advantage of $\frac{8}{7}$ (the ratio of multicast cost with and without network coding is $\frac{8}{7}$) in undirected networks is constructed using $C_{3,2}$ as a gadget [9]. The general $C_{n,k}$ structure was also introduced and discussed ([10], Sec 4.3; [11], Sec. 4.1; [12]; [13]). In contrast, much less literature exists on other types of network coding schemes in wireline networks, which are fundamentally different than CNC. A thorough understanding of CNC therefore appears important for research of network coding in general.

We use two terms, *coding advantage* and *cost advantage*, to refer to the advantage of network coding in increasing throughput and in saving routing cost, respectively; more formal definitions of these two quantities can be found in

Sec. III. The coding advantage is the ratio of achievable throughput with network coding over that without network coding. The cost advantage is the minimum routing cost necessary for achieving a desired throughput without network coding over that with network coding. Our main topic of this paper is to study the coding advantage and the cost advantage achievable by CNC in undirected networks, and as a main result we prove a tight upper-bound on both of them. It is well known that the coding advantage of CNC is unbounded in directed networks; in fact, the examples used to show that the coding advantage of general network coding is unbounded in directed networks are CNC examples [7], [12], [4].

In a previous work [6], [3], we proved that the coding advantage is upper-bounded by a constant of 2, for a single multicast session in undirected networks. However, real-world values of the coding advantage observed are always close to 1, and much smaller than 2. The largest value known is $\frac{8}{7}$, achieved in a network pattern that grows to infinite size [9]. The largest value in small networks is $\frac{9}{8}$, observed in CNC over $C_{4,2}$ and $C_{4,3}$ topologies [14]. We previously tested the coding advantage in many randomly generated multicast networks, and the observed values are always 1. Closing the gap between the proven bound of 2 and the achievable value of $\frac{8}{7}$ is a natural open problem [15], which has proved to be hard for general network coding. In this paper, we restrict our attention to CNC, and prove that $\frac{9}{8}$ is a tight upper-bound for the coding advantage as well as the cost advantage of CNC, in networks with arbitrary link capacities and arbitrary link costs, respectively. We believe this is a reasonable approach to make progress towards closing the gap for general network coding, given the fact that little progress has been made in the literature in the general case, during the past six years.

Besides the tight bound of $\frac{9}{8}$ mentioned, our work in this paper is also motivated by the following questions, and answers or partially answers each of them: (1) For CNC with a $C_{n,k}$ flow topology, does the coding advantage grow or decrease as n and k become large? Is the trend monotonic? (2) Given n, for what values of k is the coding advantage large, or small? (3) The coding advantage observed for $C_{n,k}$ topologies with uniform link capacity [8] are the same as the cost advantage observed for $C_{n,k}$ with uniform link cost [16]. Is that a coincidence? (4) Can heterogeneous link capacity/cost lead to larger coding/cost advantage? (5) What is the relation between cost advantage and coding advantage for general network coding? (6) What properties of the $C_{n,k}$ topology class make it special for network coding?

A major challenge of analyzing the coding advantage directly lies in the complex nature of the multicast solutions without coding — multicast tree packing. A multicast solution with network coding is relatively simple, and is in the form of a multicast flow consisting of one flow rate per link. However, a multicast network in general, and especially a $C_{n,k}$ topology, may have a surprisingly large number of different multicast trees (*e.g.*, $C_{5,3}$ has $49,956,624$ different multicast trees [8]), and an optimal packing scheme may need to carefully select a substantial number of them for the optimal packing. Our technique of overcoming this difficulty is to avoid analyzing the coding advantage (and therefore the optimal packing scheme) directly. We analyze it indirectly through studying the cost advantage, where a single tree is always sufficient to form an optimal solution. Of course, we then need an extra step bridging the gap between coding advantage and cost advantage. We accomplish this within the primal-dual framework of linear optimization. It turns out that even a single minimum multicast tree is hard to determine, when link costs are independent and arbitrary. Consequently, bounding its cost relative to that of the optimal multicast flow with network coding is highly nontrivial. We overcome this difficulty by borrowing tools from another thread of research — throughput-critical multicast networks, which, interestingly, requires us to resort back to tree packing, but fortunately in a more manageable way.

In order to draw a conclusion on the advantage of CNC, we have derived three self-contained results using rather different techniques. The first two are about CNC in specific, and the third one is on general network coding. We first study the cost advantage of network coding in combination topologies with uniform link cost. The celebrated multicast rate feasibility condition for network coding in directed networks states that a multicast rate d is feasible if and only if there is no source-receiver cut with size smaller than d [1], [17]. Applying this condition to two different types of cuts in the network, we show that in the case of network coding, the minimum cost achieved with CNC is actually optimal, and can be represented as a closed-form function of n and k. We further describe a multicast tree construction procedure, and show that it generates minimum trees, whose size is also a closed-form function of n and k. Based on the CNC cost and tree cost functions, we derive a closed-form formula for the cost advantage of CNC in uniform-cost settings, and show it's tightly upper-bounded by $\frac{9}{8}$.

We next study CNC in the heterogeneous link cost setting. Here we first define the *throughput-critical* property of a multicast network: a network is throughput-critical if the maximum throughput achievable becomes infeasible with an arbitrarily small capacity reduction at any link. We show that any $C_{n,k}$ topology with uniform link capacity is throughput-critical with network coding or with tree packing. The case of tree packing is proved by building two classes of minimum multicast trees satisfying certain properties in top-layer versus bottom-layer link utilization, and using these trees with carefully selected tree flow rates to form a critical packing — a packing of trees that exactly saturates every link of the network. We show that when link costs become heterogeneous, one of the trees (although we do not know which) from the critical packing must be able to adapt to the changes better than CNC does, and consequently the cost advantage can not become larger. Together with result from the uniform-cost case, we obtain an upper-bound of $\frac{9}{8}$ for CNC under heterogeneous link capacities.

We then study the relation between cost advantage and coding advantage, and try to connect them for general network coding, including CNC as a special case. The first step is formulating both multicast with network coding and with tree packing into appropriate linear programs, and introduce dual variables that can be interpreted as link costs. Then Lagrange duals of the LPs are studied, and linear programming duality applied, to connect multicast throughput to multicast cost in each case. Connections from both cases are combined to obtain the following result: given a fixed network topology, for any link capacity vector c, there always exists a link cost vector w, such that the cost advantage under w is at least as high as the coding advantage under c. This implies that any upper-bound derived on the cost advantage under arbitrary link capacity configurations can also serve as an upper-bound to the coding advantage under arbitrary link capacity configurations.

Combining all three results above, we derive that both the cost advantage and the coding advantage of CNC are upper-bounded by $\frac{9}{8}$. This bound is tight in both cases since we know it can be achieved. We also propose a similar, yet subtly different and somewhat stronger statement on applying general network coding in combination networks as a conjecture, and show how the three results from this paper may be combined with a missing fourth piece to solve the puzzle. Finally, we discuss a few open questions on basic network coding research related to this paper.

We note that the undirected network model studied in this paper does not exactly model most computer networks. Nonetheless, the study of coding advantage in undirected networks can be justified by the following facts. (a) The coding advantage in directed networks is less interesting to study, since we know it is unbounded [4]. (b) The undirected network model is a simple conceivable network model of theoretical interest. It is a classic model in network flows and combinatorial optimization. (c) For the same total bandwidth between two nodes, utilizing them in one undirected link is more flexible than in two directed links with predefined bandwidth partitioning. As a result, future data communication networks may become closer to the undirected model. (d) The Internet core can be modeled as a balanced bi-directed network, with similar bandwidth available between neighboring routers in the two directions. Any bound proven on the coding advantage for general undirected networks implies a (larger) bound for balanced bi-directed networks [3]. (e) Wireless links exhibit a certain level of resemblance to undirected links. Communication between two wireless nodes are feasible in both directions, and the total throughput from both directions are together bounded by the channel capacity.

The rest of the paper is organized as follows. We first review previous research in Sec. II and describe the network model and notations in Sec. III. We study CNC with uniform-cost and hetero-cost in Sec. IV and Sec. V respectively, and study the connection between cost advantage and coding advantage in Sec. VI. In Sec. VII, we then combine the results together to draw a conclusion on the advantage of CNC, and further propose a slightly stronger conjecture and discuss a few related open problems. We conclude the paper in Sec. VIII.

II. Related Work

Ahlswede *et al.* [1] initiated the study of network coding. They show examples that demonstrate the benefit of network coding, in terms of throughput improvement. They also prove the fundamental result that, for a multicast transmission in a directed network, if a rate d can be achieved for each receiver independently, it can also be achieved for the entire session. Koetter and Médard [17] also derived this result for directed acyclic networks within an algebraic framework. They further extend the discussion to multiple transmissions, and examine the benefit of network coding in terms of robust networking.

Li *et al.* [18] show that linear codes suffice in achieving optimal throughput for a multicast transmission. The bound on the necessary base field size is first given by Koetter and Médard [17]. They show that for a multicast session with throughput r and number of receivers k, there exists a solution based on a finite field $GF(2^m)$, for some $m \leq \lceil \log_2(kr+1) \rceil$. This bound was then improved by Ho *et al.* to $m \leq \lceil \log_2(r+1) \rceil$ [19].

Li *et al.* [18] proposed the first code assignment algorithm, which performs an exponential number of vector independence tests. Jaggi *et al.* [4] observed that, computing a multicast flow first and then exploiting the flow structure dramatically simplifies the task, and designed a polynomial time code assignment algorithm accordingly. They also show that, in directed networks with integral routing, the coding advantage of CNC, and therefore general network coding, may increase proportionally as $\Omega(\log |V|)$, and therefore may be arbitrarily high.

We previously proved that the coding advantage is shown to be upper-bounded by 2, using graph theory techniques [6]. A simpler proof based on a different graph theory tool can be found in the book of Fragouli and Soljanin ([10], pp 73). Agarwal and Charikar also proved the same bound using the linear programming framework, by translating the coding advantage to the integrality gap of the bidirected cut relaxation of a minimum Steiner tree integer program [9]. Improving this bound to any value smaller than 2 has remained an open problem. Part of the motivation behind this work is to make progress towards that direction.

Chekuri *et al.* studied the coding advantage in directed networks, without symmetrical throughput requirement on the multicast receivers [20]. They demonstrate classes of networks where the coding advantage is bounded by 2, including networks with two unit sources, or networks with two receivers. They also construct a network pattern where the coding advantage grows at rate $\Theta(\sqrt{|V|})$. Chekuri *et al.* also discussed *minimal networks*, which is essentially the same as throughput-critical networks we define in this paper. Besides combination networks, they also studied network patterns which can be viewed as having two tiers of combinations of information flows.

Cannons *et al.* [21] and Dougherty *et al.* [22] also performed comparison studies between network capacity with network coding (coding capacity) and network capacity with routing only (routing capacity). In particular, they show that the network capacity is independent of the coding alphabet, and that while routing capacity of a network is always achievable, coding capacity is not.

Empirical comparisons of multicast throughput with and without network coding can be found in [23] and [8]. Both comparisons suggest that for random network topologies, the coding advantage is marginal at best. Other advantages are suggested instead, such as ease of management, robustness and ease for algorithm design.

The coding advantage in the case of multiple unicast sessions has also been examined in recent literature. While it is known that the coding advantage is larger than 1 for either directed networks or integral flows, it was conjectured by Harvey *et al.* [24] and by us [25] that network coding does not make a difference in undirected networks with fractional flows. This conjecture remains unsettled except for some special network cases [26], [27], [28]. Recently, Langberg and Médard took an interesting new perspective of analyzing the achievable throughput for multiple unicast sessions, by connecting it to achievable multiple multicast sessions in the same network [29]. We also note that variations of CNC have been used to study the coding advantage in unicast networks [30], [25].

Ngai and Yeung studied the coding advantage of network coding in combination networks, assuming the network is directed and each link has uniform capacity [12]. They derive a closed form representation for the coding advantage in directed uniform $C_{n,k}$ as $\frac{k(n-k+1)}{n}$, which tends to k as n grows, and is therefore unbounded, contrasting the tight upper-bound of $\frac{9}{8}$ proven in this work for the undirected case.

III. Network Model and Notations

We consider undirected multicast networks, and assume that a *multicast topology* G specifies the set of vertices V_G, the set of links E_G, as well as the multicast source S and m receivers $\{T_1, \ldots T_m\}$. Of particular interest in this paper are the *combination multicast topologies* $C_{n,k}$, as illustrated in Fig. 2. A $C(n,k)$ topology has three layers of nodes and two layers of links. A multicast source S lies at the top layer. At the middle layer, there are n relay nodes $\{R_1, \ldots, R_n\}$, each connected to the source, resulting in n top-layer links, denoted as E_{top}. $\binom{n}{k}$ receivers $\{T_1 \ldots T_{\binom{n}{k}}\}$ lie at the bottom layer, and each is connected to a different set of k relays, resulting in $k \cdot \binom{n}{k}$ bottom layer links, denoted as E_{bot}. Besides receiving information flows, each receiver may send out information as well.

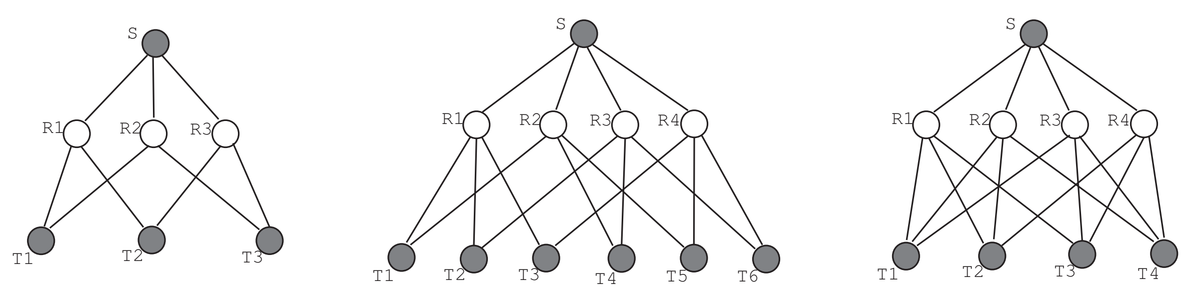

Fig. 2. Example combination multicast topologies: $C_{3,2}$, $C_{4,2}$ and $C_{4,3}$.

A multicast solution with network coding contains a *multicast flow* $f \in \mathcal{Q}_+^{A_G}$, where $A_G = \{\overrightarrow{uv}, \overrightarrow{vu} \,|uv \in E_G\}$, and $\mathcal{Q}_+$ denotes the set of positive rational numbers. A valid multicast flow f is the union of network flows from S to each of the receivers T_i, of some rate d. In other words, for each receiver T_i, let $f_i \in \mathcal{Q}_+^{E_G}$ be a network flow from S to T_i of end-to-end rate d. Then the union $f : f(\overrightarrow{uv}) = \max_i f_i(\overrightarrow{uv}), \forall \overrightarrow{uv} \in A_G$ forms a valid multicast flow of rate d.

A multicast scheme with network coding has two components: a flow routing component, as specified in the multicast flow vector f, and a code construction component, where one decides where and how coding is applied over the multicast flow. The celebrated result of Ahlswede *et al.* [1] on multicast rate feasibility reveals that for a single multicast session, these two components are separable in that routing can be solved independent of code construction. A *Combination network coding* (CNC) scheme is a network coding based multicast solution, whose underlying multicast flow exhibits a combination network topology, with uniform link flow rates. If we scale the desired multicast throughput d to 1, then $f(e) = 1/k$, for each link e in a combination topology $C_{n,k}$. CNC does not impose particular requirements on the code construction component, as long as the underlying multicast flow satisfies the definition above. In particular, both linear algebraic network coding and linear block network coding [31] can be used, where the source generates linearly independent information flows to feed into the relay nodes.

Without network coding, a certain multicast throughput d can be achieved by routing an information flow of rate d along a multicast tree t that connects all terminal nodes (source and receivers); the flow is replicated at branching points of the tree. We say that a *tree flow* $f(t)$ of rate d is established. In tree packing, multiple trees can be employed, and the total multicast throughput equals the summation of all tree flow rates. We denote the set of all multicast trees in G as $\mathcal{T}$. Compared to a directed network, an undirected network in general has substantially more multicast trees, which contributes towards the fact that the coding advantage can be arbitrarily large in directed networks but not undirected ones.

In max-throughput multicast in network (G, c), we target maximum multicast rate given a link capacity vector $c \in \mathcal{Q}_+^E$. We denote the maximum achievable throughput with network coding as $d^{NC}(G, c)$, and the maximum throughput without network coding (with multicast tree packing) as $d^{tree}(G, c)$. These two optimal quantities can be computed by LP (14) and LP (15) in Sec. VI, respectively. The coding advantage of the network (G, c) is $d^{NC}(G, c)/d^{tree}(G, c)$, and is always larger than or equal to 1.

In min-cost multicast in network (G, w), we target a multicast flow f that achieves a desired throughput d while minimizing the total cost $|f|_w = \sum_{e \in E_G} w(e)f(e)$ for a given link cost vector $w \in \mathcal{Q}_+^{E_G}$ ([32], LPs in Sec.VI.A.). Such a total flow cost depends on the multicast flow f, but not on the exact content of the information flows (the network code). In this work, we assume $d = 1$ by default for min-cost multicast. We also assume that link capacities are not a limiting factor in all min-cost multicast problems considered. For example, it is sufficient to assume that every link e has the same capacity as the desired throughput: $c(e) = d = 1, \forall e \in E_G$. We denote the minimum cost achievable with a network coding solution as $\min_{d^{NC}(f)=1} |f|_w$, and the min-cost achievable with tree packing as $\min_{t \in \mathcal{T}} |t|_w$, where $|t|_w = \sum_{e \in t} w(e)$. Here $d^{NC}(f)$ denotes the multicast throughput achieved by a multicast flow f. The cost advantage of the network is $\min_{t \in \mathcal{T}} |t|_w / \min_{d^{NC}(f)=1} |f|_w$, which is also always greater than or equal to 1, since any tree solution is just a special case of a network coding solution. Note that, while multiple trees are in general required for throughput maximization without network coding, a single tree (the minimum multicast tree) suffices to form the solution of min-cost multicast without network coding.

IV. CNC UNDER UNIFORM COST

Consider a CNC multicast flow over a $C_{n,k}$ topology, which can be either a stand-alone multicast network, or a sub-topology within a larger multicast network G. In this section, we focus on the uniform link cost case, where every link e in the $C_{n,k}$ topology has the same cost $w(e)$. The exact value of $w(e)$ does not affect the cost advantage and hence is not important; we assume $w(e) = 1, \forall e \in E_{C_{n,k}}$. We prove an upper-bound of $9/8$ for the cost advantage of network coding in this uniform-cost case. The proof consists of three steps. First, we analyze the cost achievable by network coding in Sec. IV-A, and show that CNC is actually the best network coding solution. Second, we analyze the cost achievable by a minimum multicast tree in Sec. IV-B. Third, based on results obtained from the first two steps, we derive in Sec. IV-C the cost advantage as a closed-form function of n and k, and prove that this function is upper-bounded by $9/8$.

A. Minimum Multicast Cost of Network Coding

Lemma 4.1. *The minimum multicast cost with network coding in a uniform-cost $C_{n,k}$ topology is $\binom{n}{k} + \frac{n}{k}$, and is achieved by CNC.*

Proof: We prove the lemma in two steps. First, we show that a cost of $\binom{n}{k} + \frac{n}{k}$ is possible with network coding, by constructing a concrete solution based on CNC. Second, we argue that a smaller cost is not possible with any network coding solution, by examining a number of cuts in the network.

Consider the CNC solution in a uniform $C_{n,k}$ with a multicast flow that orients all links 'downwards', *i.e.*, in the direction of either leaving the source (top-layer links) or entering the receivers (bottom layer links). It routes a flow of rate $1/k$ on each link. We can verify that, for each receiver T_i independently, we have a valid network flow from S to T_i of throughput 1, satisfying both flow conservation and link capacity constraints. By the fundamental multicast rate feasibility result of network coding [1], [17], we know that a specific code assignment exists for this multicast flow, such that each receiver simultaneously receives information at rate 1. Furthermore, the cost of such a CNC solution is $1 \times \frac{1}{k} \times n = \frac{n}{k}$ at the top layer and $1 \times \frac{1}{k} \times k\binom{n}{k} = \binom{n}{k}$ at the bottom layer, leading to a total cost of $\binom{n}{k} + \frac{n}{k}$.

We now show $\binom{n}{k} + \frac{n}{k}$ is actually the minimum cost possible. Link flows in a min-cost multicast solution can be separated into two groups: link flows that enter receivers, and link flows that enter relays. There should be no link flows entering the source in any min-cost multicast solution. As shown in Fig. 3 (bottom-left red circle), for each receiver T_i, all links adjacent to T_i constitutes a cut separating T_i from the rest of the network including the source. By the max-flow min-cut theorem, the sum of link flows entering T_i through this cut is at least 1. Since each link has unit cost of 1, the total cost for link flows entering each receiver is at least 1. There are $\binom{n}{k}$ receivers in total, and the sets of link flows entering different receivers are disjoint, hence the total cost of link flows entering all receivers is at least $\binom{n}{k}$.

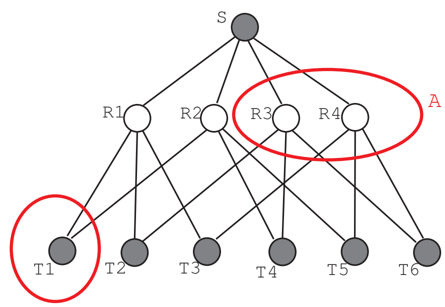

Fig. 3. Two types of cuts considered in the Proof of Lemma 4.1.

Now, consider the total cost for link flows that enter the set of relays, R. Assume, by way of contradiction, that such total cost is W for some $W < \frac{n}{k}$. The total rate of link flows entering R is also W. As shown in Fig. 3 (large red oval), let A be a subset of R with size k, and $f^{in}(A)$ be the total link flow rate entering A. Note that the average of $f^{in}(A)$ over all possible $A \subseteq R$ is kW/n. If $W < \frac{n}{k}$, then $kW/n < 1$, and consequently there must exist a particular $A^* \subseteq R$, $|A^*| = k$, and $f^{in}(A^*) < 1$. Due to the structure of $C_{n,k}$, there exists a receiver T_i such that A^* is exactly the neighbor set of T_i, and all links entering A^* forms a cut that separates T_i from the source. Then $f^{in}(A^*) < 1$ contradicts the fact that a multicast throughput 1 is achieved. $\square$

B. Minimum cost of Multicast Trees

Lemma 4.2. *The minimum cost of a multicast tree in a uniform-cost $C_{n,k}$ is $\binom{n}{k} + n - k + 1$.*

Proof: We first show that a cost of $\binom{n}{k} + n - k + 1$ is feasible, by constructing such a tree. Select $n - k + 1$ links arbitrarily from E_{top}. From E_{bot}, select one link for each of the $\binom{n}{k}$ receivers, to connect it to one of the relays that is already connected to the source. Such a relay exists, since only $k-1$ relays are not connected to the source, and each receiver is connected to k relays. Now every receiver is connected to the source, and we obtained a multicast tree with total cost of $\binom{n}{k} + n - k + 1$.

We next show that a cost of $\binom{n}{k} + n - k + 1$ is necessary. For any feasible multicast tree in uniform $C_{n,k}$, we can orient its links such that every link is in the direction of going downstream from the source to the receivers. There has to be at least one link entering each of the $\binom{n}{k}$ receivers. We also need at least $n - k + 1$ links entering the relays, otherwise there exists a set of k or more relays without an incoming link, and at least one receiver will be disconnected from the source. Since the set of links entering the receivers and the set entering the relays are disjoint, a valid multicast tree has at least $\binom{n}{k} + n - k + 1$ links. $\square$

C. Bounding the Cost Advantage

Theorem 4.1. *The cost advantage of network coding in a uniform-cost $C_{n,k}$ topology is*

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

and is tightly upper-bounded by $9/8$.

Proof: By Lemma 4.1 and Lemma 4.2, we can conclude that the cost advantage in a uniform $C_{n,k}$ is $\frac{\binom{n}{k}+n-k+1}{\binom{n}{k}+\frac{n}{k}}$. The rest of the proof is a mathematical analysis that shows $\frac{\binom{n}{k}+n-k+1}{\binom{n}{k}+\frac{n}{k}} \leq \frac{9}{8}$. We consider three cases: (a) $n \leq 16$, (b) $n > 16$ and $k = n-1$, and (c) $n > 16$ and $1 < k < n-1$.

Case (a), $n \leq 16$. We have numerically computed the function $\frac{\binom{n}{k}+n-k+1}{\binom{n}{k}+\frac{n}{k}}$ using a computer program, for all possible n and k such that $1 < k < n \leq 16$, the result is shown in Fig. 4. For $(n, k) = (4, 2)$ and $(n, k) = (4, 3)$, the function evaluates to $\frac{9}{8}$; for all other pairs of (n, k) computed, the function value is strictly less than $\frac{9}{8}$.

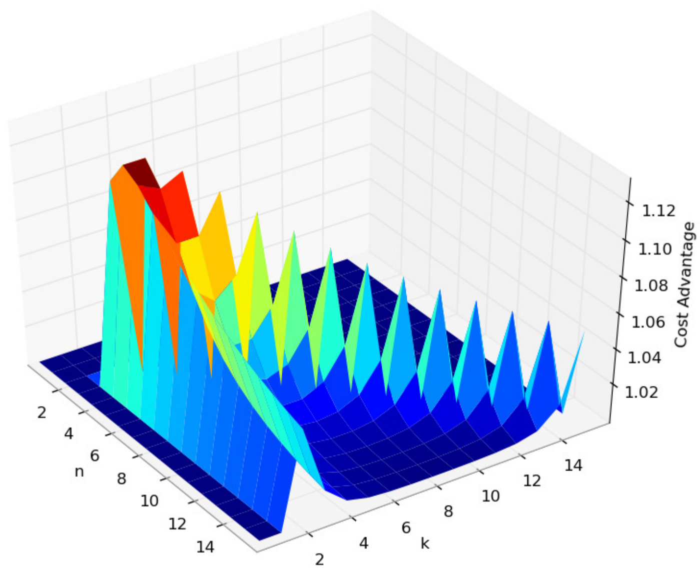

Fig. 4. Cost advantage for uniform $C_{n,k}$, $1 < k < n \leq 16$.

We next prepare for cases (b) and (c) by simplifying the inequality to be proven as follows:

$$\frac{\binom{n}{k} + n - k + 1}{\binom{n}{k} + \frac{n}{k}} \leq \frac{9}{8} \Longleftrightarrow \frac{n - k + 1 - \frac{n}{k}}{\binom{n}{k} + \frac{n}{k}} \leq \frac{1}{8} \Longleftrightarrow 8n - 8k - \frac{9n}{k} + 8 \leq \binom{n}{k} \tag{2}$$

Case (b), $n > 16$ and $k = n - 1$. We then have:

$$8n - 8k - \frac{9n}{k} + 8 = 8n - 8(n-1) - \frac{9n}{n-1} + 8 = 16 - \frac{9n}{n-1} \leq 16 < n = \binom{n}{k}. \tag{3}$$

Case (c), $n > 16$ and $1 < k < n-1$. We then have:

$$8n - 8k - \frac{9n}{k} + 8 \le 8n - 16 - \frac{9n}{k} + 8 = 8n - 8 - \frac{9n}{k} \le 8n - 8$$
$$= \frac{16(n-1)}{2} \le \frac{n(n-1)}{2} = \binom{n}{2} \le \binom{n}{k} \tag{4}$$

The bound $\frac{9}{8}$ is tight, since the cost advantage equals $\frac{9}{8}$ for $C_{4,2}$ and for $C_{4,3}$. $\square$

The cost advantage is in general decreasing for larger values of n and k, but not monotonically. For a given value of n, the cost advantage is large for k close to 1 or to n, and is small for k close to $\frac{n}{2}$. Theorem 4.1 implies the following bound on CNC:

Corollary 4.1. *The cost advantage of CNC applied to uniform-cost topologies is tightly upper-bounded by* $\frac{9}{8}$.

V. CNC Under Heterogeneous Cost

In this section, we show that in a multicast network with heterogenous link cost, CNC cannot achieve a larger cost advantage than in the uniform cost case. Consequently, the upper-bound of $\frac{9}{8}$ established in Sec. IV (Corollary 4.1) is still valid for CNC under heterogenous cost.

We now turn our attention to the case where the link cost vector w is heterogeneous, and each link e may have an independent cost $w(e) \in (0, \infty)$. We are concerned with the cost advantage, *i.e.*, the ratio of multicast cost without network coding over that with network coding, and scaling all link costs by a common factor does not affect such a ratio. Therefore, we assume without loss of generality that link costs have been normalized according to $\min_e w(e)$, such that $w(e) \in [1, \infty), \forall e$. Compared to the previous assumption of uniform link cost of 1, we essentially inflated each link cost independently by a factor of 1 or larger. Naturally, the minimum multicast cost can only increase, either with network coding or without network coding (with tree packing). Whether the cost advantage becomes larger or smaller depends on whether the network coding cost or the tree cost increases by a larger factor. If the tree cost inflates more, then a larger coding advantage is obtained than in the uniform case; if network coding cost inflates more instead, then a smaller cost advantage is obtained. We prove that the bound of $\frac{9}{8}$ holds for CNC with any link cost configuration w, by proving that the tree cost always has a smaller inflation factor. The logic here is as follows:

$$\frac{\min_{t\in\mathcal{T}} |t|_w}{\min_{t\in\mathcal{T}} |t|_{\vec{1}}} \le \frac{\min_{d^{NC}(f)=1} |f|_w}{\min_{d^{NC}(f)=1} |f|_{\vec{1}}} \Longrightarrow \frac{\min_{t\in\mathcal{T}} |t|_w}{\min_{d^{NC}(f)=1} |f|_w} \le \frac{\min_{t\in\mathcal{T}} |t|_{\vec{1}}}{\min_{d^{NC}(f)=1} |f|_{\vec{1}}} \tag{5}$$

Here $\min_{t\in\mathcal{T}} |t|_w$ and $\min_{t\in\mathcal{T}} |t|_{\vec{1}}$ are the minimum multicast tree costs under heterogeneous cost w and under uniform cost of 1, respectively; $\min_{d^{NC}(f)=1} |f|_w$ and $\min_{d^{NC}(f)=1} |f|_{\vec{1}}$ are the minimum multicast costs with network coding under heterogeneous cost w and under uniform cost of 1, respectively. After link costs changing to heterogeneous, the multicast cost in general increases both with and without network coding. We use the *cost inflation factor* to refer to the ratio of multicast cost after and before changing link costs to heterogeneous.

We first analyze the cost inflation factor of network coding. Since the coding solution is in the form of CNC, we have $f(e) = 1/k, \forall e \in E_{C_{n,k}}$, both with cost vector $\vec{1}$ (before cost inflation), and with cost vector w (after cost inflation). The total multicast cost before inflation is $\sum_e 1 \cdot f(e) = \frac{1}{k}|E_{C_{n,k}}|$. The multicast cost after inflation is $\sum_e w(e)f(e) = \frac{1}{k}\sum_e w(e)$. The inflation factor can therefore be derived as the average link cost of the $C_{n,k}$ flow topology (either as an independent network or as a sub-topology in a larger multicast network), *i.e.*,

$$\frac{\min_{d^{NC}(f)=1} |f|_w}{\min_{d^{NC}(f)=1} |f|_{\vec{1}}} = \frac{1}{|E_{C_{n,k}}|} \sum_{e \in E_{C_{n,k}}} w(e). \tag{6}$$

However, the cost inflation factor of the minimum multicast tree is less obvious, since we have no concrete knowledge of the topology of the new optimal tree under cost vector w. This new tree turns out to be elusive to analyze. Of course, to prove that the cost inflation of the minimum tree is smaller, one does not have to construct and bound the cost of the absolute minimum tree. If one can build a sub-optimal tree and still prove that its cost inflation from the optimal tree in the uniform-cost case is upper-bounded by the cost inflation of the network coding solution, then it immediately follows that the cost inflation of the minimum tree can only be even better. Nonetheless, even identifying such a sub-optimal but 'good' tree, and trying to upper-bound its cost, is challenging. This can be attributed in part to the fact that each link in the network can be independently assigned with an arbitrary cost, and therefore we can claim little on the topological properties of the new minimum tree. For example, we do not know how many links this new tree consists of.

Our solution towards the challenge above is motivated by another thread of research on multicast network coding, namely the study of *throughput-critical* multicast networks. In the rest of this section, we first introduce the concept of throughput-critical networks in Sec. V-A, and show that every $C_{n,k}$ network is throughput-critical under uniform link capacity. Then in Sec. V-B, we apply such a throughput-critical property of $C_{n,k}$ to argue that a 'good' multicast tree must exist in a heterogeneous cost $C_{n,k}$, without having to explicitly construct it.

A. The Throughput-Critical Property of the $C_{n,k}$ Topology

In graph theory, the study of graph coloring and graph chromatic numbers (the minimum number of colors required to properly color a graph) often focuses on a special class of graphs called *critical graphs* — a graph G is critical if the chromatic number of G is larger than that of G with any of its edge (assuming edge coloring) removed. Such graphs are canonical examples that help gain insight on what forces a graph to take a certain number of colors to color, with many irrelevant topological details filtered out. In parallel, we believe that the study of multicast throughput, and therefore the study of the multicast coding advantage, can benefit from the analysis of throughput-critical multicast networks, as defined below.

Definition. *A multicast network* (G, c) *with topology* G *and link capacity vector* c *is* throughput-critical *if the maximum throughput in* G*,* $d(G, c)$*, is not feasible with a capacity reduction of* ϵ *at any of its links, for any arbitrarily small constant* ϵ*.*

An arbitrary multicast network is rather unlikely to be throughput-critical, and in general contains redundant link capacity portions or even entire links whose removal does not jeopardize the achievable multicast throughput. Nonetheless, the $C_{n,k}$ network topology is special in that it is throughput-critical, either with or without network coding, under uniform link capacity.

Theorem 5.1. *A combination topology with uniform link capacity,* $(C_{n,k}, \vec{1})$*, is throughput-critical, assuming multicast with network coding.*

Proof: First, we note that $d^{NC}(C_{n,k}, \vec{1}) = k$. By orienting all links along the direction of either leaving the source or entering a receiver, we can verify that the max-flow from the source to each receiver is k. Furthermore, a cut that separates a receiver from the rest of the network has size of k, implying that a throughput of larger than k is impossible. Next, we need to show that the same throughput k is impossible, with a capacity reduction $\epsilon > 0$ at any link $e \in E_{C_{n,k}}$. There are two cases: (i) $e \in E_{bot}$ and (ii) $e \in E_{top}$.

Case (i), $e \in E_{bot}$. Let T_i be the receiver adjacent to e. The cut separating T_i from the rest of the network has size less than k, therefore a multicast rate of k is infeasible.

Case (ii), $e \in E_{top}$. In order for a multicast rate k to be feasible, we must orient all links in E_{bot} along the direction of entering the receivers. Otherwise, a similar cut as in Case (i) can be found, such that a receiver is separated to the rest of the network by a cut of size less than k. Let R_j be the relay adjacent to the link whose capacity is reduced to $1-\epsilon$. Let A be a subset of the relay nodes of size k, such that $R_j \in A$, and A is the neighbor set of a receiver T_l. Now, consider the cut separating $A \cup \{T_l\}$. The contribution from E_{top} to the size of this cut is $k - \epsilon$; the contribution from E_{bot} is 0, since all links in E_{bot} are oriented in the reverse direction. Therefore this cut has size smaller than k, making it impossible to achieve a multicast throughput of k. □

Theorem 5.2. *A combination topology with uniform link capacity,* $(C_{n,k}, \vec{1})$*, is throughput-critical, assuming multicast via tree packing.*

Proof: We prove this theorem by proving the following theorem, Theorem 5.3, which implies it. □

Definition. *A packing using a set of trees* $\mathcal{T}_P$ *with a flow vector* $f \in \mathcal{Q}_+^{\mathcal{T}_P}$ *for a multicast network* (G, c) *is* critical *if* f *exactly saturates every link in* G*:* $\sum_{t \in \mathcal{T}_\mathcal{P}: e \in t} f(t) = c(e), \forall e \in E_G$*.*

Theorem 5.3. *A combination topology with uniform link capacity,* $(C_{n,k}, \vec{1})$*, has a critical packing of multicast trees each with minimum number of links.*

Proof: We prove the theorem constructively, by building such a critical packing. Our packing is based on two disjoint subsets, $\mathcal{T}_A$ and $\mathcal{T}_B$, for two different classes of multicast trees described below.

We first prepare for the construction of $\mathcal{T}_A$ and $\mathcal{T}_B$ by showing that a minimum multicast tree with $\binom{n}{k}+n-k+1$ links can include any number of top layer links between 1 and $n-k+1$. Consider the following way of constructing a minimum tree. Select an arbitrary subset of relay nodes $R_{tree} \subseteq R$, such that $|R_{tree}| = n-k+1$. Connect every terminal node (source and receivers) to one of the relays in R_{tree}, in any way. We note the following facts. Fact A, since only $k-1$ relays are not included in R_{tree} and each receiver is connected to k different relays, it is possible to connect every receiver to one of the relays in R_{tree}. Fact B, every relay in R_{tree} has some receiver connected to it, since otherwise we have at least k relays with no receiver connection, and that contradicts the facts that every receiver is connected to a relay and that every set of k relays form the exclusive neighbor set for a receiver. Fact C, since relays are not directly connected and every terminal node has degree 1, the selected edges does not contain a cycle. From Facts A, B and C, we conclude that we have selected $\binom{n}{k}+1$ links that connect $\binom{n}{k}+n-k+2$ nodes (1 source, $n+k-1$ relays, $\binom{n}{k}$ receivers) without yielding cycles, which must be an $(n-k+1)$-component forest. Now, we can complete this forest into a tree of $\binom{n}{k}+n-k+1$ links by further selecting $n-k$ links to interconnect these $n-k+1$ components, and each of these last $n-k$ links can be freely selected from either E_{top} or E_{bot}. The end result is a minimum multicast tree with 1 to $n-k+1$ top layer links (inclusive).

We let $\mathcal{T}_A$ and $\mathcal{T}_B$ be the two extreme classes of minimum trees described above. The set $\mathcal{T}_A$ includes all possible multicast trees with $n-k+1$ links from top-layer links E_{top}, and $\binom{n}{k}$ links from bottom-layer links E_{bot}. The set $\mathcal{T}_B$ includes all possible multicast trees with 1 link from top-layer links E_{top}, and $\binom{n}{k}+n-k$ links from bottom-layer links E_{bot}.

We note that the $C_{n,k}$ topology is symmetrical in that, all top-layer links in E_{top} are topologically symmetrical among themselves, and so are all bottom-layer links in E_{bot}. Consequently, if we route a common flow rate f_A through every tree in $\mathcal{T}_A$, every top-layer link will have the same aggregated flow rate $\frac{n-k+1}{n}|\mathcal{T}_A|f_A$, and every bottom-layer link will have the same aggregated flow rate $\frac{\binom{n}{k}}{\binom{n}{k}k}|\mathcal{T}_A|f_A = \frac{1}{k}|\mathcal{T}_A|f_A$. Similarly, if we route a common flow rate of f_B through every tree in $\mathcal{T}_B$, every top-layer link will have the same aggregated flow rate $\frac{1}{n}|\mathcal{T}_B|f_B$, and every bottom-layer link will have the same aggregated flow rate $\frac{\binom{n}{k}+n-k}{k\binom{n}{k}}|\mathcal{T}_B|f_B$. Our target is to exactly saturate all links in the $C_{n,k}$ topology:

$$\begin{cases} \frac{n-k+1}{n}|\mathcal{T}_A|f_A + \frac{1}{n}|\mathcal{T}_B|f_B = 1 \\ \frac{1}{k}|\mathcal{T}_A|f_A + \frac{\binom{n}{k}+n-k}{k\binom{n}{k}}|\mathcal{T}_B|f_B = 1 \end{cases} \tag{7}$$

In the two equations above, n, k, $|\mathcal{T}_A|$ and $|\mathcal{T}_B|$ are constants, and f_A and f_B are the two variables. These two equations are both in the form of asserting a certain weighted average of f_A and f_B is of the same value 1. This equation group has positive solutions for f_A and f_B if and only if the two coefficient ratios $\frac{n-k+1}{n} : \frac{1}{k}$ and

$\frac{1}{n} : \frac{\binom{n}{k}+n-k}{k\binom{n}{k}}$ are either (a) both equal to 1, or (b) lying on different sides of point 1 on the number axis. We can verify that this is indeed the case, since $\frac{n-k+1}{n} : \frac{1}{k}$ is larger than 1, and $\frac{1}{n} : \frac{\binom{n}{k}+n-k}{k\binom{n}{k}} = k\binom{n}{k} : n(\binom{n}{k} + n - k) < k\binom{n}{k} : n\binom{n}{k} = k : n < 1$. This is actually how we wanted to construct the two sets of trees $\mathcal{T}_A$ and $\mathcal{T}_B$: we want their upper:down link utilization ratios to be larger than 1 and smaller than 1 respectively, so that it is possible to compute weights for the two sets to form a critical packing of the network $C_{n,k}$. The way that these two sets can be constructed to satisfy this condition is not unique. □

We now explain how Theorem 5.3 implies Theorem 5.2. Let $|t^*|$ be the minimum tree size, under uniform cost $\vec{1}$. A necessary condition for a multicast throughput with packing d_{tree} to be feasible is that the total available network capacity has to be at least the aggregated link capacities used in the trees:

$$\sum_e c(e) \geq d_{tree}|t^*|, \tag{8}$$

The condition above is also the easy direction of the iff condition later proven in Theorem 6.2. Theorem 5.3 shows that the maximum throughput $\frac{\sum_e c(e)}{|t^*|}$ is possible. Note that the total network capacity $\sum_e c(e)$ is tight to support this throughput. Any further link capacity deduction will break the necessary condition $\sum_e c(e) \geq d_{tree}|t^*|$, and make the throughput $\frac{\sum_e c(e)}{|t^*|}$ infeasible. Hence the network $(C_{n,k}, \vec{1})$ is throughput-critical, and Theorem 5.2 is true.

Theorem 5.3 implies the following Corollary on the coding advantage in a combination network $C_{n,k}$ with uniform link capacity. It shows that the coding advantage in a combination network with uniform link cost exactly equals the cost advantage in a combination network with uniform link cost.

Corollary 5.1. *The coding advantage in a uniform-capacity $C_{n,k}$ is $\frac{n-k+1+\binom{n}{k}}{\frac{n}{k}+\binom{n}{k}}$.*

Proof: First, the throughput with network coding in a uniform-capacity $C_{n,k}$ is k. Second, by Theorem 5.3, we further know that with tree packing, we can exactly utilize all available link capacities available in the network, in the optimal packing scheme. The network has $|E| = n + k\binom{n}{k}$ links each with unit capacity, each minimum tree in the packing has size $n - k + 1 + \binom{n}{k}$, therefore the throughput with tree packing is $\frac{n+k\binom{n}{k}}{n-k+1+\binom{n}{k}} = \frac{n-k+1+\binom{n}{k}}{\frac{n}{k}+\binom{n}{k}}$. □

B. Heterogeneous Cost Does Not Give CNC More Advantage

Theorem 5.4. *The cost advantage of CNC under heterogeneous link cost is upper-bounded by the cost advantage of CNC under uniform link cost.*

Proof: Consider CNC over a topology $C_{n,k}$ with heterogeneous cost w, and $w(e) \geq 1, \forall e \in E_{C_{n,k}}$. We prove the theorem by showing that there exists a multicast tree within this $C_{n,k}$, whose cost inflation compared to the optimal tree in the uniform-cost case ($w(e) = 1, \forall e \in E_{C_{n,k}}$) is no higher than the cost inflation of CNC. The key idea is to realize that one of the trees in the critical packing that we prove exist in Theorem 5.3 must satisfy this condition, although we do not know exactly which of them does.

Let $(\mathcal{T}^* = \mathcal{T}_\mathcal{A} \cup \mathcal{T}_\mathcal{B}, f)$ be a critical packing described in Theorem 5.3 for the $C_{n,k}$ topology. We then have the following equalities:

$$\sum_{t \in \mathcal{T}^*} f(t)|t|_{\vec{1}} = |E_{C_{n,k}}|, \tag{9}$$

$$\sum_{t \in \mathcal{T}^*} f(t)|t|_w = \sum_{e \in E_{C_{n,k}}} w(e). \tag{10}$$

Recall $|t|_{\vec{1}}$ is the cost of a tree t (the total cost of its edges) under uniform link cost of 1, and $|t|_w$ is the cost of t under heterogeneous cost w. Combining the two equations above, we can have:

$$\frac{\sum_{t \in \mathcal{T}^*} f(t)|t|_w}{\sum_{t \in \mathcal{T}^*} f(t)|t|_{\vec{1}}} = \frac{\sum_{e \in E_{C_{n,k}}} w(e)}{|E_{C_{n,k}}|} \tag{11}$$

The above equation can be interpreted as asserting that a certain weighted average of cost inflation for trees in $\mathcal{T}^*$ is equal to the cost inflation of CNC. It mathematically implies that:

$$\exists t' \in \mathcal{T}^*, s.t. : \frac{|t'|_w}{|t'|_{\vec{1}}} \leq \frac{\sum_{e \in E_{C_{n,k}}} w(e)}{|E_{C_{n,k}}|} \tag{12}$$

Finally, let t^* be the minimum tree in the $C_{n,k}$ topology under cost vector w. We have $|t^*|_w \leq |t'|_w$, because of the optimality of t^* by definition, under cost vector w. Therefore we conclude that:

$$\frac{|t^*|_w}{|t'|_{\vec{1}}} \leq \frac{|t'|_w}{|t'|_{\vec{1}}} \leq \frac{\sum_{e \in E_{C_{n,k}}} w(e)}{|E_{C_{n,k}}|}, \tag{13}$$

i.e., the cost inflation of the optimal tree can not be higher than that of CNC, and hence the cost advantage of CNC under heterogeneous cost can not be higher than the cost advantage of CNC under uniform cost. Here we note that t' is a minimum tree under uniform cost, due to the definition of $\mathcal{T}^*$. $\square$

Informally, the virtue of the proof of Theorem 5.4 can be interpreted as the following. A $C_{n,k}$ topology has a surprisingly large number of different multicast trees, and the number grows exponentially as n and k increases. For example, $C_{5,3}$, a topology with 16 nodes and 35 links, already has as many as $49,956,624$ different multicast trees [8]. Since the number of candidate trees is so large, when link cost becomes heterogeneous one can always flexibly pick a good tree to avoid the more expensive links well. In the end, heterogeneous cost is not as detrimental to trees as it is to network coding, and the heterogeneous cost case is not a better paradigm for demonstrating the advantage of CNC.

The converse of Theorem 5.4 is obviously true, since uniform link cost can be viewed as a special case of heterogenous link cost. Combining this with Theorem 4.1 and Theorem 5.4, we have the following corollary:

Corollary 5.2. *The cost advantage of CNC is tightly upper-bounded by $\frac{9}{8}$, under arbitrary link cost configurations.*

We conclude this section by pointing out that, if a multicast network (G, c, w) with both arbitrary link capacity c and arbitrary link cost w is considered, then the cost advantage of network coding is trivially unbounded, under either CNC or general network coding. One may connect the source to the set of receivers through two disjoint sub-networks, such that the first sub-network has enough capacity to achieve the desired throughput only with network coding, and the second sub-network has links arbitrarily more expensive than in the first.

VI. From Cost Advantage to Coding Advantage

In this section, we develop a general result that connects the two kinds of advantage of network coding: increasing multicast throughput and reducing multicast cost. This connection holds for a single multicast session in an undirected network, with either CNC or another form of network coding. The basic techniques employed are as follows. First, we formulate the max-throughput multicast problem, both with and without network coding, into appropriate linear programs. We then apply Lagrange duality theory to each LP, introducing link cost into the picture while at the same time connecting it to the multicast throughput. Finally, based on the individual relation from the case with network coding and the case without network coding, we conclude that cost advantage in general constitutes an upper-bound on coding advantage, in undirected multicast networks.

In deriving an alternative proof for the upper-bound of 2 on general network coding in undirected networks, Agarwal and Charikar [9] studied the connection between the coding advantage and the integrality gap of a natural Integer Programming formulation of the Steiner tree problem. Their result can also be directly used to perform the translation from cost advantage to coding advantage. Below we provide a different proof, based on partial LP duality instead of complete LP duality, to obtain exactly the result needed (only one of the two main inequality relations of Agarwal and Charikar is necessary for our translation). Besides theoretical analysis of the coding advantage, our technique can also be applied to design efficient and distributed solution algorithms for the optimal multicast problem with network coding [5].

A. Relating Network Coding Throughput to Network Coding Cost

In a previous work [8], [14], we formulated the max-throughput problem for multicast with network coding into a pair of primal and dual linear programs. By applying Lagrange relaxation techniques to the primal and to the dual, we derived a sufficient and necessary multicast rate feasibility condition for undirected networks and designed a distributed algorithm for achieving max-rate multicast, respectively. We will show that, with a different perspective for interpreting the dual variables, the necessary and sufficient condition leads to a connection between multicast throughput and multicast flow cost. We establish such a relation for network coding in this sub-section, and do the same for trees in Sec.VI-B. These relations will then become powerful tools for translating the obtained bound of cost advantage to a bound of coding advantage.

The following is the max-throughput multicast linear program, in the setting of undirected networks with network coding. Only the primal LP is of interest in this paper.

Maximize $\quad d^{NC}$ (14)

Subject to:

$$\begin{cases} d^{NC} \le f_i(\overrightarrow{T_iS}) & \forall T_i \in \mathcal{T} \\ f_i(\overrightarrow{uv}) \le c(\overrightarrow{uv}) & \forall T_i \in \mathcal{T}, \forall\, \overrightarrow{uv} \neq \overrightarrow{T_iS} \\ \sum_{v \in N(u)} (f_i(\overrightarrow{uv}) - f_i(\overrightarrow{vu})) = 0 & \forall T_i \in \mathcal{T}, \forall u \\ c(\overrightarrow{uv}) + c(\overrightarrow{vu}) \le c(uv) & \forall uv \neq T_iS \end{cases}$$

$$c(\overrightarrow{uv}), f_i(\overrightarrow{uv}), d^{NC} \ge 0 \qquad \forall T_i, \forall\, \overrightarrow{uv}$$

For the sake of compact LP representation, a conceptual link $\overrightarrow{T_iS}$ is introduced from each receiver T_i to S, with unlimited capacity. The vector f_i is used for a network flow from S to T_i, and $c(\overrightarrow{uv})$ is a variable that denotes the portion of link capacity $c(uv)$ between u and v that is oriented in the $\overrightarrow{uv}$ direction.

By relaxing the constraint $c(\overrightarrow{uv}) + c(\overrightarrow{vu}) \le c(uv)$, and studying the resulting Lagrange dual, the following if and only if condition for multicast rate feasibility is obtained [8], [14]. The notation is slightly modified to fit the context of this paper. The *only if* direction is a natural necessary condition, similar to the cut condition for network flow feasibility. The *if* direction is non-trivial, similar to the fact that having the cut condition satisfied at all cuts implies flow rate feasibility. This theorem provides a way to connect network coding cost $\min_{d^{NC}(f)=1} |f|_w$ with network coding throughput d, from the perspective of weighted link bandwidth demand and supply.

Theorem 6.1. [Li and Li, 2005] *A multicast rate d is feasible in an undirected multicast network (G, c) with network coding, if and only if for every link cost vector $w \in \mathcal{Q}_+^{E_G}$, $\frac{|G|_w}{\min_{d^{NC}(f)=1} |f|_w} \ge d$.*

Here $|G|_w = \sum_{e \in E_G} w(e)c(e)$ can be viewed as the cost-weighted grand capacity of the multicast network (G, c), and $\min_{d^{NC}(f)=1} |f|_w$ denotes the cost of the optimal multicast flow for achieving a unit throughput with network coding. It was shown that Theorem 6.1 subsumes the max-flow min-cut Theorem and Tutte-Nash-Williamson' Theorem as special cases ([14], Sec. V.C).

B. Relating Tree Throughput to Tree Cost

The problem of achieving maximum multicast throughput without network coding can be formulated into the following tree packing linear program, in a general undirected multicast network topology G.

Maximize $\quad \sum_{t \in \mathcal{T}} f(t)$ (15)

Subject to:

$$\sum_{t\in\mathcal{T}:e\in t} f(t) \leq c(e) \qquad \forall e \in E_G \qquad \longleftrightarrow w(e)$$

$$f(t) \geq 0 \qquad \forall t \in \mathcal{T}$$

By applying similar techniques as in the proof of Theorem 6.1, we can derive a parallel result for tree packing. We present details of the derivation below for completeness of the paper. We actually need a slightly stronger result with an extra claim that equality is possible.

Theorem 6.2. *A multicast rate d is feasible in an undirected multicast network (G, c) with tree packing, if and only if for every link cost vector $w \in \mathcal{Q}_+^{E_G}$, $\frac{|G|_w}{\min_{t\in\mathcal{T}}|t|_w} \geq d$. Furthermore, for the max-throughput d^*_{tree}, there exists a corresponding cost vector w^*_{tree} such that equality holds.*

Proof: We relax the constraint $\sum_{t\in\mathcal{T}:e\in t} f(t) \leq c(e)$ in the tree packing LP, and introduce a corresponding dual vector w, to translate the original LP into the following Lagrange dual:

Minimize $\quad L(w) \qquad (16)$

Subject to:

$$w(e) \geq 0, \qquad \forall e \in E_G,$$

where:

$$L(w) = \max_{f\geq 0}[\sum_{t\in\mathcal{T}} f(t) - \sum_{e\in E_G} w(e)(\sum_{t\in\mathcal{T}:e\in t} f(t) - c(e))]$$

Lagrange duality theory guarantees that the dual always has the same optimal solution as the primal LP does, *i.e.*:

$$d^*_{tree} = \min_{w\geq 0}\{\max_{f\geq 0}[\sum_{t\in\mathcal{T}} f(t) - \sum_{e\in E_G} w(e)(\sum_{t\in\mathcal{T}:e\in t} f(t) - c(e))]\}$$

$$= \min_{w\geq 0}\{\max_{f\geq 0}[\sum_{t\in\mathcal{T}} f(t)(1 - \sum_{t\in\mathcal{T}} |t|_w) + |G|_w]\} \qquad (17)$$

Note that when $\min_{t\in\mathcal{T}}|t|_w < 1$, the inner maximization above is unbounded, since we can scale f by an arbitrarily large number to get an arbitrarily large $\sum_{t\in\mathcal{T}} f(t)(1-\sum_{t\in\mathcal{T}}|t|_w)$. On the other hand, when $\min_{t\in\mathcal{T}}|t|_w \geq 1$, a zero flow vector f serves the best interest of the outer minimization. Therefore the equation above is equivalent to:

$$d^*_{tree} = \min_{w\geq 0, \min_{t\in\mathcal{T}}|t|_w \geq 1} |G|_w \qquad (18)$$

Further combining the equation above with the fact that both $|G|_w$ and $|t|_w$ grows proportionally with w, we finally obtain:

$$d^*_{tree} = \min_{w \geq 0} \frac{|G|_w}{\min_{t \in \mathcal{T}} |t|_w}, \tag{19}$$

and that concludes the proof. $\square$

C. Connecting Cost Advantage to Coding Advantage

We are now ready to present and prove the following relation between cost advantage and coding advantage, for general multicast network coding.

Theorem 6.3. *In any undirected multicast network topology G, for a given link capacity vector $c \in \mathcal{Q}_+^{E_G}$, there always exists a link cost vector $w \in \mathcal{Q}_+^{E_G}$, such that the cost advantage of network coding in (G, w) is at least as high as the coding advantage of network coding in (G, c).*

Proof: First, we apply Theorem 6.2 to (G, c). Let d^*_{tree} be the max-throughput with tree packing in (G, c), and let w^*_{tree} be the cost vector that satisfies (i) $\frac{|G|_{w^*_{tree}}}{\min_{t \in \mathcal{T}} |t|_{w^*_{tree}}} = d^*_{tree}$. Next, by Theorem 6.1, we further know that since d^{NC} is a feasible multicast rate with network coding, (ii) $\frac{|G|_{w^*_{tree}}}{\min_{d^{NC}(f)=1} |f|_{w^*_{tree}}} \geq d^*_{NC}$.

Combining (i) and (ii), we have:

$$d^*_{NC} \cdot \min_{d^{NC}(f)=1} |f|_{w^*_{tree}} \leq |G|_{w^*_{tree}} = d^*_{tree} \cdot \min_{t \in \mathcal{T}} |t|_{w^*_{tree}}, \tag{20}$$

and therefore:

$$\frac{d^*_{NC}}{d^*_{tree}} \leq \frac{\min_{t \in \mathcal{T}} |t|_{w^*_{tree}}}{\min_{d^{NC}(f)=1} |f|_{w^*_{tree}}}, \tag{21}$$

i.e., the cost advantage in (G, w^*_{tree}) is at least as high as the coding advantage in (G, c). $\square$

Theorem 6.3 directly implies the following corollary, which is sometimes more convenient to apply.

Corollary 6.1. *Given any undirected multicast network topology G, the maximum coding advantage achievable with arbitrary link capacity configuration is upper-bounded by the maximum cost advantage achievable with arbitrary link cost configuration.*

VII. Synthesizing the Results, and Beyond

We have now presented three results on the advantage of network coding, in Sec. IV, Sec. V and Sec. VI, respectively. The first two of them are on CNC in specific, and the third one on network coding in general. These results may each be of independent interest. Nonetheless, our studies of them have shared common motivation in that they are all required to answer some of the questions on the advantage of network coding. In this section, we try to combine them to work together. Sec. VII-A contains a straightforward combination of the three results, leading to a bound on the coding advantage of CNC in general undirected multicast networks with arbitrary link capacities. Sec. VII-B further proposes a similar, yet subtly different statement as a conjecture, and explains how the three results from this paper, plus a missing fourth piece, may work together to solve that puzzle. Sec. **??** contains a list of related open problems.

A. Bounding The Advantage of CNC

Theorem 7.1 below combines the results obtained in Sections IV, V and VI, for establishing an upper-bound on the coding advantage of CNC.

Theorem 7.1. *In any undirected multicast network topology G, with arbitrary link capacity vector c, CNC can improve multicast throughput by a factor of at most $\frac{9}{8}$, and the bound is tight.*

Proof: The three results listed above together imply that $\frac{9}{8}$ is an upper-bound for coding advantage of CNC. The fact that this bound is tight has been verified in $(C_{4,2}, \overrightarrow{1})$, where max throughput with network coding is 2, and $\frac{16}{9}$ without coding, and in $(C_{4,3}, \overrightarrow{1})$, where the max-throughput is 3 with network coding, and $\frac{8}{3}$ without coding [8]. □

We can now conclude that the advantage of CNC, in either increasing multicast throughput or saving multicast cost, is upper-bounded by a factor of $\frac{9}{8}$, in the undirected network setting.

B. A Stronger Conjecture

In a combination network $C_{n,k}$ with heterogeneous link cost, it is possible that the multicast flow of network coding is not exactly in the form of CNC, but a variation of it, with some (expensive) links unused. Fig. 5 shows a simple example.

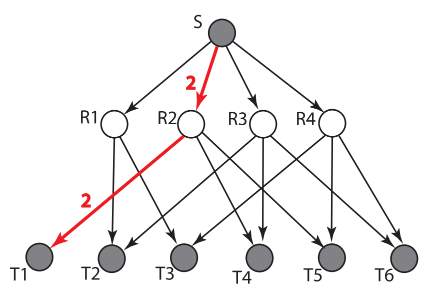

Fig. 5. Variation of $C_{4,2}$ with cost advantage of $\frac{18}{17}$, link directions shown are for the multicast flow with network coding. Link between R_1 and T_1 has high cost and is avoided in the multicast flow. All other links have uniform cost. The multicast flow in a network coding solution routes double amount of flows on links from S to R_2 then to T_1 than on all other links. Network coding cost for throughput 1 is 8.5; minimum multicast tree has cost of 9.

This example does not yield a cost advantage of larger than $\frac{9}{8}$. Are there other variations that do? Is $\frac{9}{8}$ still a valid upper-bound if we consider the cost advantage of general network coding applied to a $C_{n,k}$ network, with heterogeneous cost? We think that the bound $\frac{9}{8}$ is still valid. In particular, we believe that the following proposition is true:

Conjecture 7.1. *Given an undirected multicast network $(C_{n,k}, w)$, where the min-cost multicast flow f_w^* has zero flow rates on some of the links, then there must exist a different cost vector w', such that: the new min-cost multicast flow $f_{w'}^*$ is of the CNC structure, and the cost advantage in $(C_{n,k}, w')$ is larger than that in $(C_{n,k}, w)$.*

Intuitively, we believe that utilizing all links in a $C_{n,k}$ as CNC does is the best way to construct network instances with a large cost advantage. Combining the first two results in Theorem 4.1 and Theorem 5.4 with Conjecture 7.1, we can derive that in any $C_{n,k}$ multicast topology under any cost vector w, the cost advantage of network coding is upper-bounded by $\frac{9}{8}$. Then further including the third result in Corollary 6.1, we obtain the following statement:

Conjecture 7.2. *In any undirected combination multicast network $(C_{n,k}, c)$, the coding advantage (of general network coding) is tightly upper-bounded by $\frac{9}{8}$.*

The difference between Theorem 7.1 and Conjecture 7.2 is essentially from the subtle difference between applying combination network coding in general networks and applying general network coding in combination networks. In the former, CNC can be applied for multiple times, at different parts of a multicast network. However, for each application, we have exact knowledge on the atomic multicast flow structure where network coding is used. Such knowledge helps the analysis of the coding advantage. In the latter, the entire multicast network is confined to a combination topology. Even so, the exact structure of the multicast flow under network coding is still uncertain, and that introduces challenges to the analysis of the coding advantage.

C. The Coding Advantage in Directed Combination Networks

We have focused on the coding advantage in undirected combination networks so far in this paper. We now take a look at the directed counterpart, and discuss the counter part of Conjecture 7.2 in directed combination networks. We discuss how the proof technique of translating cost advantage into coding advantage can be applied in that context, and point out the equivalence between the following two quantities: (i) the coding advantage in directed combination networks, and (ii) the integrality gap of a natural linear programming formulation of the uncapacitatied facility location problem, a classic and extensively-studied problem in operations research.

Max Coding Advantage in Directed Combination Networks

For a given undirected combination network $C_{n,k}$, we can obtain its directed version by orienting all links *downward*, *i.e.*, in the direction of leaving the source (top layer links) or entering the receivers (bottom layer links). Ngai and Yeung [12] studied the coding advantage in directed combination networks, with uniform link capacities. Their main result is the following:

Theorem 8.1. [Ngai and Yeung, 2004] *In a directed combination network $C_{n,k}$ with uniform link capacities, the coding advantage is $\frac{(n-k+1)k}{n}$, which turns to k as n grows to infinity.*

The proof technique of Ngai and Yeung is to analyze the achievable throughput directly, both with and without network coding. A substantial amount of their effort is devoted to analyzing the throughput without coding in a directed $C_{n,k}$. Such analysis is non-trivial but still manageable, thanks to the uniform link capacity assumption. However, in the most general case where each link can independently take an arbitrary capacity, a direct analysis on throughput with tree packing becomes much harder. We conjecture that the bound $\frac{(n-k+1)k}{n}$ is till valid in the general case of heterogeneous link cost:

Conjecture 8.1. *In a directed combination network $C_{n,k}$ with arbitrary link capacities, the coding advantage is tightly upper-bounded by $\frac{(n-k+1)k}{n}$.*

A possible proof of Conjecture 8.1 is based on the framework established in this paper, *i.e.*, to analyze the cost advantage in directed combination networks, and then translate that into coding advantage using linear programming duality. We note that a cost advantage of $\frac{(n-k+1)k}{n}$ is indeed achievable. Given a directed $C_{n,k}$, assume that each link has uniform capacity 1, the desired throughput d is 1. Each top layer link has cost 1 and each bottom layer link has infinitesimally small cost ϵ.

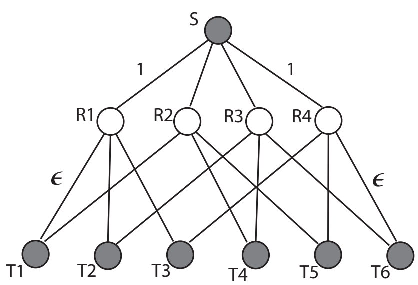

Fig. 6. Illustration of the cost assignment strategy where cost advantage of network coding is $\frac{(n-k+1)k}{n}$, in a directed $C_{4,2}$.

The total cost of multicast with network coding is $\frac{1}{k}(n + \binom{n}{k}k\epsilon)$, and the minimum multicast tree has cost of $n-k+1+\binom{n}{k}\epsilon$, leading to a cost advantage of $\frac{n-k+1+\binom{n}{k}\epsilon}{\frac{1}{k}(n+\binom{n}{k}k\epsilon)}$, and $\lim_{\epsilon\to 0} \frac{n-k+1+\binom{n}{k}\epsilon}{\frac{1}{k}(n+\binom{n}{k}k\epsilon)} = \frac{k(n-k+1)}{n}$. The remaining part of the proof, the more challenging part, is to show that such a cost vector is optimal in yielding a maximum cost advantage possible.

Connection to The Facility Location Problem

It is known that in general multicast networks, the coding advantage corresponds to the integrality gap of a natural Integer Programming formulation of the Steiner tree problem [9], [3]. We now further show the interesting result that, the coding advantage in a directed combination network corresponds to exactly the integrality gap of a natural formulation of the facility location problem, another classic problem in operations research.

The facility location problem [33] studies the optimal decision making on opening facilities at potential locations, and connecting customers to opened facilities, with the goal of minimizing the total cost of facility opening and customer connection. In the non-capacitated non-metric version of the problem, each customer needs to connect to one facility, and the cost of connecting a customer to a facility is arbitrary and does not need to satisfy metric distance constraints.

Let w_i denote the link cost for $\overrightarrow{SR_i}$, and w_{ij} denote the link cost for $\overrightarrow{R_iT_j}$. Let x_i and f_{ij} be binary variables indicating whether links $\overrightarrow{SR_i}$ and $\overrightarrow{R_iT_j}$ belong to the multicast tree, respectively. Then the minimum multicast cost without network coding in a directed combination network can be formulated into the following Integer Program:

$$\text{Minimize} \quad \sum_{R_i} x_i w_i + \sum_{R_i} \sum_{T_j} w_{ij} f_{ij} \tag{14}$$

Subject to:

$$\begin{cases} \sum_{R_i:\overrightarrow{R_iT_j}\in E} f_{ij} = 1 & \forall T_j \\ f_{ij} \le x_i & \forall R_i, \forall T_j \end{cases}$$

$$x_i, f_{i,j} \in \{0,1\}, \forall R_i, \forall T_j.$$

The IP above is a standard facility location IP. In particular, we can view each R_i as a facility, x_i represents whether R_i is opened. Each T_j can be viewed as a customer, and w_{ij} represents the cost of connecting customer T_j to facility R_i. The objective function minimizes the total facility opening and customer connection cost, and the constraints guarantee that each customer is connected to an opened facility. Furthermore, we note that the LP relaxation of this IP, obtained by replacing the discrete value requirement $x_i, f_{ij} \in \{0,1\}$ with $0 \le x_i, f_{i,j} \le 1$, we arrive at the LP for min-cost multicast with network coding, in the same combination network. Here x_i and $f_{i,j}$ are viewed as information flow rates from the top and bottom layers of the network, respectively.

D. Some Open Problems in Combinatorial Network Coding

We list below a few open problems that are related to the work in this paper. They are all basic network coding problems with a combinatorial flavor, which Ralf Koetter (along with co-authors) has referred to as *combinatorial network coding* in recent years (*e.g.*, in [34]).

P1. In the undirected network model with fractional routing, prove a less-than-2 upper-bound on the coding advantage for a single multicast session. We believe that the best known bound of 2 [6], [3], [9], [10] is loose, and the tight bound should be closer to 1 than to 2. This is a rather important problem, but probably also hard ([15], Conclusion).

P2. Since the best known upper-bound of coding advantage is $\frac{9}{8}$ for CNC and is 2 for general network coding, does there exist another type of network coding schemes that are based on a fundamentally different structure than CNC, and hopefully provide larger values of the coding advantage?

P3. If the answer to P2 is 'no', could it be possible that the tight upper-bound in P1 is actually very close to the current best known lower-bound of $\frac{8}{7}$ [9]? In that case, solving optimal multicast (max-rate or min-cost) with network coding, *e.g.*, using the linear optimization framework [35], [5], becomes an interesting perspective of designing polynomial-time approximation algorithms for Steiner tree problems in graphs (Steiner tree packing, minimum Steiner trees). In fact, researchers studying the Steiner tree problem had this hunch before the concept of network coding was in wide spread [36]. A very recent break through [15] that improves the approximation ratio for Steiner trees from 1.55 to 1.386 is indeed made using linear programming tools related to the suggested approach here.

P4. The approximation algorithm mentioned in P3 computes a scalar value (throughput or cost), based on a multicast flow. While the value may be proven to be a good approximation to that for the Steiner tree problem, one also needs to translate the optimal multicast flow with network coding into a tree solution (a packing scheme or a single tree), for the solution to be complete. The translation should be computationally efficient, and the size of the packing or the cost of the tree from the translation should be nicely bounded. How?

P5. The largest known value of coding advantage in planar networks, networks that can be drawn in a 2D plane without crossing links, is $\frac{9}{8}$, achieved in a uniform-capacity $C_{3,2}$. Is that indeed the largest coding advantage for all planar networks?

P6. In a canonical combination network $C_{n,k}$, there is actually no need for "network" coding — no coding is necessary at relay nodes, and encoding at the source alone is always sufficient. By Wu's result that no coding is necessary at links entering terminals [37], this is further true for all multicast networks with no pairs of relays directly connected by an edge, or *quasi-bipartite networks*. On one hand, being quasi-bipartite seems positive for achieving a large coding advantage — known examples with non-trivial coding advantages are quasi-bipartite. On the other hand, being quasi-bipartite seems negative — no 'network' coding is really necessary, the full potential of network coding may not have been fully realized. What role does relay-to-relay links play in a multicast network? Are they necessary for constructing the multicast examples achieving the largest coding advantage?

VIII. CONCLUSION

In this paper, we focused on CNC, a special form of network coding that frequently appears in the literature and is relatively well studied. Our main result is to draw a conclusion on the power of CNC in the undirected network setting, by proving that its potential both in increasing achievable throughput and in reducing routing cost are tightly upper-bounded by a factor of $\frac{9}{8}$. In order to derive this conclusion, we have developed three relatively self-contained results along the way, each of which may be of independent interest as well. The first states that cost advantage of CNC is upper-bounded by $\frac{9}{8}$ under uniform link cost. The second states that cost advantage of CNC under hetero-cost can be no larger than under uniform-cost. The third states that for general network coding, CNC or not, cost advantage always upper-bounds coding advantage. We also discussed a few conjectures and related open problems on basic network coding research with a combinatorial nature.

REFERENCES

[1] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network Information Flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] C. Fragouli, J.-Y. L. Boudec, and J. Widmer, "Network Coding: An Instant Primer," *ACM SIGCOMM Computer Communications Review*, vol. 36, no. 1, pp. 63–68, 2006.

[3] Z. Li, B. Li, and L. C. Lau, "A Constant Bound on Throughput Improvement of Multicast Network Coding in Undirected Networks," *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 997–1015, March 2009.

[4] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 1973–1982, June 2005.

[5] Z. Li and B. Li, "Efficient and Distributed Computation of Maximum Multicast Rates," in *Proceedings of IEEE INFOCOM*, 2005.

[6] ——, "Network Coding in Undirected Networks," in *Proc. of the 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004.

[7] P. Sanders, S. Egner, and L. Tolhuizen, "Polynomial Time Algorithm for Network Information Flow," in *Proceedings of the 15th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, 2003.

[8] Z. Li, B. Li, D. Jiang, and L. C. Lau, "On Achieving Optimal Throughput with Network Coding," in *Proceedings of IEEE INFOCOM*, 2005.

[9] A. Agarwal and M. Charikar, "On the Advantage of Network Coding for Improving Network Throughput," in *IEEE Information Theory Workshop*, October 2004.

[10] C. Fragouli and E. Soljanin, *Network Coding Fundamentals*. Boston, MA: NOW Publishers, 2007.

[11] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory*. Boston, MA: NOW Publishers, 2006.

[12] C. K. Ngai and R. W. Yeung, "Network Coding Gain of Combination Networks," in *IEEE Information Theory Workshop*, October 2004.

[13] M. Xiao, M. Médard, and T. Aulin, "A Binary Coding Approach for Combination Networks and General Erasure Networks," in *Proceedings of ISIT*, 2007.

[14] Z. Li, B. Li, and L. C. Lau, "On Achieving Optimal Multicast Throughput in Undirected Networks," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2467–2485, June 2006.

[15] J. Byrka, F. Grandoni, T. Rothvoss, and L. Sanità, "An Improved LP-Based Approximation for Steiner Tree," in *Proc. of the 42nd ACM Symposium on the Theory of Computing (STOC)*, 2010.

[16] A. Smith, B. Evans, Z. Li, and B. Li, " The Cost Advantage of Network Coding in Uniform Combinatorial Networks," in *Proc. of the First IEEE Workshop on Wireless Network Coding (WiNC)*, 2008.

[17] R. Koetter and M. Médard, "An Algebraic Approach to Network Coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.

[18] S. Y. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, February 2003.

[19] T. Ho, D. R. Karger, M. Médard, and R. Koetter, "Network Coding From a Network Flow Perspective," in *Proceedings of ISIT*, 2003.

[20] C. Chekuri, C. Fragouli, and E. Soljanin, "On Average Throughput and Alphabet Size in Network Coding," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2410–2424, June 2006.

[21] J. Cannons, R. Dougherty, C. Freiling, and K. Zeger, “Network Routing Capacity,” in *Proceedings of ISIT*, 2005.

[22] R. Dougherty, C. Freiling, and K. Zeger, “Unachievability of Network Coding Capacity,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2365–2372, June 2006.

[23] Y. Wu, P. A. Chou, and K. Jain, “A Comparison of Network Coding and Tree Packing,” in *Proceedings of ISIT*, 2004.

[24] N. J. A. Harvey, R. D. Kleinberg, and A. R. Lehman, “Comparing Network Coding with Multicommodity Flow for the k-pairs Communication Problem,” Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology, http://theory.csail.mit.edu/ nickh/Publications/Gaps/TR964.pdf, Tech. Rep., November 2004.

[25] Z. Li and B. Li, “Network Coding: The Case of Multiple Unicast Sessions,” in *42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

[26] N. J. A. Harvey, R. Kleinberg, and A. R. Lehman, “On the Capacity of Information Networks,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2345–2364, June 2006.

[27] N. J. A. Harvey and R. Kleinberg, “Tighter Cut-Based Bounds for k-pairs Communication Problems,” in *43rd Annual Allerton Conference on Communication, Control, and Computing*, 2005.

[28] K. Jain, V. V. Vazirani, and G. Yuval, “On the Capacity of Multiple Unicast Sessions in Undirected Graphs,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2805–2809, June 2006.

[29] M. Langberg and M. Médard, “On the Multiple Unicast Network Coding Conjecture,” in *47th Annual Allerton Conference on Communication, Control, and Computing*, 2009.

[30] J. Price and T. Javidi, “Network Coding for Resource Redistribution in a Unicast Network,” in *45th Annual Allerton Conference on Communication, Control, and Computing*, 2007.

[31] S. Jaggi, M. Effros, T. Ho, and M. Médard, “On Linear Network Coding,” in *42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

[32] Z. Li, “Min-Cost Multicast of Selfish Information Flows,” in *Proceedings of IEEE INFOCOM*, 2007.

[33] D. B. Shmoys, E. Tardos, and K. Aardal, “Approximation Algorithms for Facility Location Problems,” in *Proc. of the 29th ACM Symposium on the Theory of Computing (STOC)*, 1997.

[34] R. Koetter, M. Effros, and M. Médard, “On A Theory of Network Equivalence,” in *The IEEE Information Theory Workshop on Networking and Information Theory*, 2009.

[35] D. S. Lun, N. Ratnakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, E. Ahmed, and F. Zhao, “Minimum-cost Multicast over Coded Packet Networks,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2608–2623, 2006.

[36] K. Jain, M. Mahdian, and M. R. Salavatipour, “On the Bidirected Cut Relaxation for the Metric Steiner Tree Problem,” in *Proceedings of the 6th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1999.

[37] Y. Wu, “A Unification of Network Coding and Tree Packing (Routing) Theorems,” in *42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

PLACE
PHOTO
HERE

Shreya Maheshwar received her B.Sc. degree in Computer Science and Engineering from Visvesvaraya Technological University, India in 2004, and her M.Sc. degree in Computer Science from the University of Calgary in 2011. Shreya was a Software Engineer with Sasken Communication Technologies Ltd during 2004-2006, and was a Senior Software Engineer with Sasken during 2007-2008. Her research focuses on the analysis of the benefits of network coding in a number of network models, including in particular combination networks.

PLACE
PHOTO
HERE

Zongpeng Li received his B.Engr. degree in Computer Science and Technology from Tsinghua University (Beijing) in 1999, his M.S. degree in Computer Science from University of Toronto in 2001, and his Ph.D. degree in Electrical and Computer Engineering from University of Toronto in 2005. He has been working at the Department of Computer Science in the University of Calgary since August 2005, where he is now an Associate Professor. His research interests are in computer networks, particularly in network optimization, multicast algorithm design, network game theory and network coding. Zongpeng was named an Edward S. Rogers Sr. Scholar in 2004, won the Alberta Ingenuity New Faculty Award in 2007, was nominated for the Alfred P. Sloan Research Fellow in 2007, and received the Best Paper Award at the Ninth Passive and Active Measurement Conference (PAM) in 2008. He is a member of the IEEE and a member of the ACM.

PLACE
PHOTO
HERE

Baochun Li received the B.Engr. degree from the Department of Computer Science and Technology, Tsinghua University, China, in 1995 and the M.S. and Ph.D. degrees from the Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, in 1997 and 2000. Since 2000, he has been with the Department of Electrical and Computer Engineering at the University of Toronto, where he is currently a Professor. He holds the Nortel Networks Junior Chair in Network Architecture and Services from October 2003 to June 2005, and the Bell Canada Endowed Chair in Computer Engineering since August 2005. His research interests include large-scale multimedia systems, cloud computing, peer-to-peer networks, applications of network coding, and wireless networks. Dr. Li was the recipient of the IEEE Communications Society Leonard G. Abraham Award in the Field of Communications Systems in 2000. In 2009, he was a recipient of the Multimedia Communications Best Paper Award from the IEEE Communications Society, and a recipient of the University of Toronto McLean Award. He is a member of ACM and a senior member of IEEE.