# A GRAPH THEORETICAL APPROACH TO NETWORK ENCODING COMPLEXITY

LI XU*, WEIPING SHANG†, AND GUANGYUE HAN‡

**Key words.** Menger's theorem, max-flow min-cut theorem, network coding

**AMS subject classifications.** 68P30, 68R10, 90C27

**EXTENDED ABSTRACT.** Let $G(V, E)$ denote an acyclic directed graph, where $V$ denotes the set of all the vertices (or points) in $G$ and $E$ denotes the set of all the edges in $G$. In this paper, a *path* in $G$ is treated as a set of concatenated edges. For $k$ paths $\beta_1, \beta_2, \dots, \beta_k$ in $G(V, E)$, we say these paths *merge* [2] at an edge $e \in E$ if

1. $e \in \cap_{i=1}^{k} \beta_i$,
2. there are at least two distinct edges $f, g \in E$ such that $f, g$ are immediately ahead of $e$ on some $\beta_i, \beta_j$, $i \neq j$, respectively.

We call the maximal subpath that starts with $e$ and that is shared by all $\beta_i$'s *merged subpath* (or simply *merging*) by all $\beta_i$'s at $e$; see Figure 0.1 for a quick example.

For any two vertices $u, v \in V$, we call any set consisting of the maximum number of pairwise edge-disjoint directed paths from $u$ to $v$ a set of *Menger's paths* from $u$ to $v$. By Menger's theorem [4], the cardinality of Menger's paths from $u$ to $v$ is equal to the min-cut between $u$ and $v$.

Assume that $G(V, E)$ has $l$ sources $S_1, S_2, \dots, S_l$ and $l$ distinct sinks $R_1, R_2, \dots, R_l$. For $i = 1, 2, \dots, l$, let $c_i$ denote the min-cut between $S_i$ and $R_i$, and let $\alpha_i = \{\alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,c_i}\}$ denote a set of Menger's paths from $S_i$ to $R_i$. We are interested in the number of mergings among paths from different $\alpha_i$'s, denoted by $|G|_{\mathcal{M}}(\alpha_1, \alpha_2, \dots, \alpha_l)$. In this paper we will count number of mergings **without** multiplicity: all the mergings at the same edge $e$ will be counted as one merging at $e$.

The motivation for considerations of the number of mergings is more or less obvious in transportation networks: mergings among different groups of transportation paths can cause congestions, which may either decrease the whole network throughput or incur unnecessary cost. The connection between the number of mergings and the encoding complexity in computer networks, however, is a bit more subtle, which can be best illustrated by the following three examples in network coding theory.

The first example is the famous "butterfly network" [8]. As depicted in Figure 0.2(a), for the purpose of transmitting messages $a, b$ simultaneously from the sender $S$ to the receivers $R_1, R_2$, network encoding has to be done at node $C$. Another way to interpret the necessity of network coding at $C$ (for the simultaneous transmission to $R_1$ and $R_2$) is as follows: If the transmission to $R_2$ is ignored, Menger's paths $S \to A \to R_1$ and $S \to B \to C \to D \to R_1$ can be used to transmit messages $a, b$ from $S$ to $R_1$; if the transmission to $R_1$ is ignored, Menger's paths $S \to B \to R_2$ and $S \to A \to C \to D \to R_2$ to transmit messages $a, b$ to $R_2$. For the simultaneous transmission to $R_1$ and $R_2$, merging by these two groups of Menger's paths at $C \to D$ becomes a "bottle neck", therefore network coding at $C$ is required to avoid possible congestions.

*The University of Hong Kong (xuli@hku.hk)
†Zhengzhou University (shangwp@zzu.edu.cn)
‡The University of Hong Kong (ghan@hku.hk)

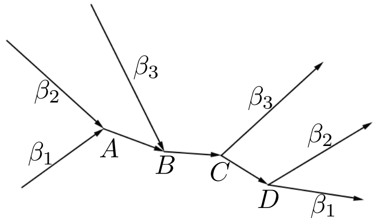


FIG. 0.1. *Paths $\beta_1, \beta_2$ merge at edge $A \to B$ and at merged subpath (or merging) $A \to B \to C \to D$, and paths $\beta_1, \beta_2, \beta_3$ merge at edge $B \to C$.*

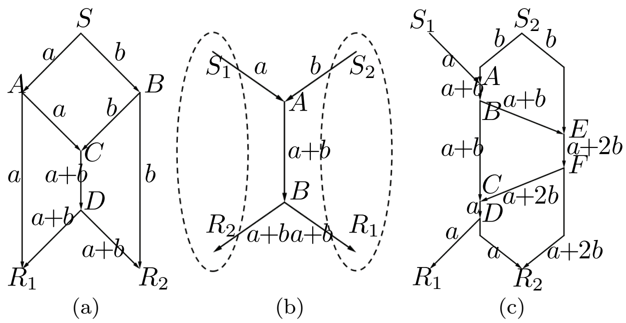


FIG. 0.2. *(a) Network coding on the butterfly network (b) Network coding on a variant of the butterfly network (c) Network coding on two sessions of unicast*

The second example is a variant of the classical butterfly network (see Example 17.2 of [8]) with two senders and two receivers, where the sender $S_1$ is attached to the receiver $R_2$ and the sender $S_2$ is attached to the receiver $R_1$. As depicted in Figure 0.2(b), the two parties wish to send messages $a, b$ to each other through the network. Similarly as in the first example, the edge $A \to B$ is where the Menger's paths $S_1 \to A \to B \to R_1$ and $S_2 \to A \to B \to R_2$ merge with each other, which is a bottle neck for the simultaneous transmission of messages $a, b$. The simultaneous transmission is achievable if upon receiving the messages $a$ and $b$, network encoding is performed at the node $A$ and the newly derived message $a+b$ is sent over the channel $AB$.

The third example is concerned with two sessions of unicast in a network [5]. As shown in Figure 0.2(c), the sender $S_1$ is to transmit message $a$ to the receiver $R_1$ using Menger's path $S_1 \to A \to B \to E \to F \to C \to D \to R_1$. And the sender $S_2$ is to transmit message $b$ to the receiver $R_2$ using two Menger's paths $S_2 \to A \to B \to C \to D \to R_2$ and $S_2 \to E \to F \to R_2$. Since mergings $A \to B$, $C \to D$ and $E \to F$ become bottle necks for simultaneous transmission of messages $a$ and $b$, network coding at these bottle necks, as shown in Figure 0.2(c), is performed to ensure the simultaneous message transmission.

Generally speaking, for a network with multiple groups of Menger's paths, each of which is used to transmit a set of messages to a particular sink, network encoding is needed at mergings by different groups of Menger's paths. As a result, the number of mergings is the number of network encoding operations required in the network. So, we are interested in the number of mergings among different groups of Menger's paths in such networks.

For the case when all sources in $G$ are in fact identical, $M^*(G)$ is defined as the minimum of $|G|_{\mathcal{M}}(\alpha_1, \alpha_2, \ldots, \alpha_l)$ over all possible Menger's path sets $\alpha_i$'s, and $\mathcal{M}^*(c_1, c_2, \ldots, c_l)$ is defined as the supremum of $M^*(G)$ over all possible choices of such $G$. It is clear that $M^*(G)$ is the least number of network encoding operations required for a given $G$, and $\mathcal{M}^*(c_1, c_2, \ldots, c_l)$ is the largest such number among all such $G$. As for $\mathcal{M}^*$, the authors of [1] use the idea of "subtree decomposition" to first prove that

$$\mathcal{M}^*(\underbrace{2, 2, \ldots, 2}_{l}) = l - 1.$$

It was first shown in [3] that $\mathcal{M}^*(c_1, c_2)$ is finite for all $c_1, c_2$ (see Theorem 22 in [3]), and subsequently $\mathcal{M}^*(c_1, c_2, \ldots, c_l)$ is finite for all $c_1, c_2, \ldots, c_l$.

For the case when all sources are distinct, $M(G)$ is defined as the minimum of $|G|_{\mathcal{M}}(\alpha_1, \alpha_2, \ldots, \alpha_l)$ over all possible Menger's path sets $\alpha_i$'s, and $\mathcal{M}(c_1, c_2, \ldots, c_l)$ is defined as the supremum of $M(G)$ over all possible choices of such $G$. Again, the encoding ideas for the second example can be easily generalized to networks, where each receiver is attached to all senders except its associated one. It is clear that the number of mergings is an tight upper bound for the number of network encoding operations required. For networks with several unicast sessions, in [5], an upper bound for the encoding complexity of a network with two unicast sessions is given. It is easy to see that for networks with multiple unicast sessions (straightforward generalizations of the third example), $\mathcal{M}$ with appropriate parameters can serve as an upper bound on network encoding complexity. It was first conjectured that $\mathcal{M}(c_1, c_2, \ldots, c_l)$ is finite in [6]. More specifically the authors proved that (see Lemma 10 in [6]) if $\mathcal{M}(c_1, c_2)$ is finite for all $c_1, c_2$, then $\mathcal{M}(c_1, c_2, \ldots, c_l)$ is finite as well. Here, we remark that we have rephrased the work in [1, 3, 6], since all of them are done using very different languages from ours.

In [2], we have shown that for any $c_1, c_2, \ldots, c_l$, $\mathcal{M}^*(c_1, c_2, \ldots, c_l)$, $\mathcal{M}(c_1, c_2, \ldots, c_l)$ are both finite, and we further studied the behaviors of $\mathcal{M}^*, \mathcal{M}$ as functions of the min-cuts. In this paper, further continuing the work in [2], we compute exact values of and give tighter bounds on $\mathcal{M}^*$ and $\mathcal{M}$ for certain parameters. The explicit results in this manuscript are omitted due to the space limit, and we refer to [7] for the full paper.

## REFERENCES


[1] C. Fragouli and E. Soljanin, "Information Flow Decomposition for Network Coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, Mar. 2006, pp. 829-848.

[2] G. Han, *Menger's Paths with Minimum Mergings*, arXiv: 0805.4059.

[3] M. Langberg, A. Sprintson and J. Bruck, "The Encoding Complexity of Network Coding," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, Jun. 2006, pp. 2386-2397.

[4] K. Menger, "Zur allgemeinen Kurventhoerie," *Fundamenta Math.*, vol. 10, 1927, pp. 96-115.

[5] W. Song, K. Cai, R. Feng and R. Wang, *Encoding Complexity for Intersession Network Coding with Two Simple Multicast Sessions*, arXiv: 1007.2928.

[6] A. Tavory, M. Feder and D. Ron, "Bounds on Linear Codes for Network Multicast," *Electron. Colloq. on Computational Complexity, Rep.* 33, 2003.

[7] L. Xu, W. Shang and G. Han, *A Graph Theoretical Approach to Network Encoding Complexity*, arXiv: 1202.0747

[8] R. W. Yeung, *Information theory and network coding* (*Information Technology: Transmission, Processing and Storage*), Springer-Verlag, 2008.