On Capacity of Deterministic Relay Networks

Mohammad A. (Amir) Khojastepour
NEC Laboratories America
4 Independence Way, Suite 200
Princeton, NJ, 08540
Email: amir@nec-labs.com

Alireza Keshavarz-Haddad
School of Engineering
Shiraz University
Zand ave., Shiraz, Iran
Email: alireza@rice.edu

***Abstract*—In this paper, we study network coding for a single multicast session in wireless networks based on "deterministic channel model" introduced recently in [2], [3]. We propose a network coding scheme for acyclic deterministic wireless network which can achieve the well-known cutset upper bound for such networks. While the original approach in [3] first proves the achievability of cutset upper bound for the layered networks and then extends the result to acyclic networks, we provide a short and direct proof for acyclic networks which trivially includes the layered networks.**

I. Introduction

In recent years network coding has become an important research topic in network information theory. It has been shown that network coding can help to improve the throughput, energy consumption, delay and some other performance metrics of communication networks. Network coding was first introduced in the seminal paper by Ahlswede et al. [1] in which it was proved that the maximum flow capacity of a single multicast session which is equal to the minimum cutset bound can be achieved using network coding in wired networks with directional links. Later, [4] and [5] showed constructively that the linear (random) network codes can achieve the minimum cutset bound of a single multicast session as well. Recently, a deterministic approach to study wireless networks was introduced in [2]. This model incorporates both broadcast and interference challenges in the wireless network. However, by removing the randomness in classic wireless channel models, this model makes the challenging problem of network coding for wireless network analytically tractable. For example, the problem of maximum flow capacity of a single multicast session in wireless networks was studied in [3] where it was shown that similar to the results for wired networks [1], the minimum cutset bound can be achieved. In [3], a random linear network coding scheme is proposed to show the achievability of the cutset upper bound in wireless networks with directional links. This result has been proved in two steps by first considering *layered* wireless networks and then by unfolding *acyclic* wireless networks over the time and using the results of layered networks. In this paper, we take new approach for solving this problem where we directly deal with a general acyclic network. Both the presented coding scheme in this paper and the one studied in [3] are based on large matrices with i.i.d. random components. However, these coding schemes are slightly different and we provide much shorter and direct proof for general acyclic networks.

This paper has been organized as follows. In Section II, we introduce the network model and notations. We describe our network coding scheme for a single multicast session in Section III. In Section IV, we show that the proposed coding scheme can asymptotically achieve the minimum cutset bound of the multicast session. Finally, we conclude the paper in Section V. In Appendix, we provide a proof for a mathematical lemma which is needed in our analysis.

II. Network Model and Notations

We consider a wireless network as a directed graph where each node can transmit the same message into *all* its outgoing links and receives the *superposition* of the signals arrived from the incoming links. We adopt the deterministic channel model in [2], [3] to model how the superposition is performed at each node. Notice that in this model the nodes can simultaneously transmit and receive data, so there is no concept of scheduling for transmissions at different nodes.

We assume that the network contains $1+N$ nodes, where one node is the source of a single multicast session and the rest of the nodes are relays or terminals (destinations) of the session. We use a universal index k for every node Φ_k where $k = 0, 1, \ldots, N$. We assume the output signal from node Φ_k at time t is a column vector $\mathbf{y}_t^k = [y_{t,1}^k, y_{t,2}^k, \ldots, y_{t,q}^k]^\dagger$ of size q where each element is a value in Galois field $\mathbb{F}(p^n)$ for some prime number p where p^n is the size of the finite field. Here $\dagger$ is used to denote the matrix transpose operation. Each link from the node Φ_l to Φ_m in the network is denoted by its transfer function $\mathbf{G}_l^m$ which is a $q \times q$ matrix with the entries in $\mathbb{F}(p^n)$. The output of the link at time t is $\mathbf{G}_l^m \mathbf{y}_t^l$. The received vector or input at the node Φ_m is the column vector $\mathbf{x}_t^m = [x_{t,1}^m, x_{t,2}^m, \ldots, x_{t,q}^m]^\dagger$ which is the superposition of the outputs of the links arriving at node Φ_m defined in $\mathbb{F}(p^n)$ on component-by-component basis, i.e.,

$$\mathbf{x}_t^m = \sum_{l \in \mathcal{I}_m} \mathbf{G}_l^m \mathbf{y}_t^l \tag{1}$$

where $\mathcal{I}_m$ is the set of all nodes for which an outgoing link goes to node Φ_m.

If we stack together the received vectors at multiple nodes $\Phi_k, k \in \mathcal{B} = \{j_1 \ldots j_b\}$, assuming they are characterized by the output at the nodes $\Phi_k, k \in \mathcal{A} = \{i_1 \ldots i_a\}$, then the transfer function is given by

$$\begin{bmatrix} \mathbf{x}_t^{j_1} \\ \mathbf{x}_t^{j_2} \\ \vdots \\ \mathbf{x}_t^{j_b} \end{bmatrix} = \begin{bmatrix} \mathbf{G}_{i_1}^{j_1} & \mathbf{G}_{i_2}^{j_1} & \dots & \mathbf{G}_{i_a}^{j_1} \\ \mathbf{G}_{i_1}^{j_2} & \mathbf{G}_{i_2}^{j_2} & \dots & \mathbf{G}_{i_a}^{j_2} \\ \vdots & & & \vdots \\ \mathbf{G}_{i_1}^{j_b} & \mathbf{G}_{i_2}^{j_b} & \dots & \mathbf{G}_{i_a}^{j_b} \end{bmatrix} \begin{bmatrix} \mathbf{y}_t^{i_1} \\ \mathbf{y}_t^{i_2} \\ \vdots \\ \mathbf{y}_t^{i_a} \end{bmatrix} = \mathbf{G}_{\mathcal{A}}^{\mathcal{B}} \begin{bmatrix} \mathbf{y}_t^{i_1} \\ \mathbf{y}_t^{i_2} \\ \vdots \\ \mathbf{y}_t^{i_a} \end{bmatrix} \quad (2)$$

We will use the notation of $\mathbf{G}_{\mathcal{A}}^{\mathcal{B}}$ at several place in our analysis.

III. Proposed linear network coding scheme for a single multicast session

Consider a coding as follows. We employ a block of KT channel uses which is divided into K frame of T channel uses each. Each node Φ_k in the network waits to get the m^{th} frame, transforms it to a new frame by multiplying its received frame into a matrix $\mathbf{F}_m^k$ of size $qT \times qT$ and sends it during the next T channel uses, i.e., $(m+1)^{\text{th}}$ frame.

The entries of the matrix $\mathbf{F}_m^k$ are i.i.d. uniformly distributed random numbers in $\mathbb{F}(p^n)$ and are independently generated for different node k and frame index m. However, note that the same matrix $\mathbf{F}_m^k$ is used for fixed k and m at different blocks.

The coding is done at each node over T vectors. After receiving T vectors $\mathbf{x}_t^k$ for $t = T(m-1)+1, \dots, T(m-1)+T$ in m^{th} frame, the node Φ_k performs a linear operation $\mathbf{F}_m^k$ on the Tq received symbols to find Tq new symbols. Then, it maps these new symbols into T column vectors $\mathbf{y}_t^k$ that are sent in the next time frame at $t = Tm+1, \dots, Tm+T$. To ease the notation, we denote the input and output of node Φ_k at frame m by $\mathbf{X}_m^k = [\mathbf{x}_{T(m-1)+1}^k, \dots, \mathbf{x}_{T(m-1)+T}^k]^\dagger$ and $\mathbf{Y}_m^k = [\mathbf{y}_{T(m-1)+1}^k, \dots, \mathbf{y}_{T(m-1)+T}^k]^\dagger$. Then,

$$\mathbf{Y}_{m+1}^k = \mathbf{F}_m^k \mathbf{X}_m^k \quad (3)$$

Let L denote the length of the longest path in the network. In our coding scheme, the source sets the last L frames of each block to zero and chooses the first $K-L$ frames (information codewords) from a vector space $V_{\text{in}} \subseteq \mathbb{F}(p^n)^{qT}$. In the next section, we will explain how the codewords are created.

We point out that in fact the coding scheme is a linear mapping; each terminal receives some linear combinations of $qT(K-L)$ symbols which are sent by the source at each block during KT uses of channel, and then it computes all symbols by an inverse linear mapping.

IV. Maximum throughput of the proposed coding scheme

In this section we demonstrate that our proposed coding scheme can achieve the minimum cutset bound for the single multicast session asymptotically as $T, K \to \infty$.

Without lack of generality, assume that Φ_0 is the source and Φ_N is an arbitrary terminal of the multicast session. We denote the set of all cutsets between the source and this destination by $\Lambda = \{\Omega : \Omega \subset \{0, 1, \dots, N\}, 0 \in \Omega, N \in \Omega^c\}$ and define minimum cutset rank between the source and the destination as $r_{\min} = \min_{\Omega \in \Lambda}\{\mathbf{rank}(\mathbf{G}_{\Omega}^{\Omega^c})\}$. If the multicast session contains more than one terminal, then we define $r_{\min}$ as the minimum value among the minimum cutset ranks of different terminals.

For the proof, we first study the null space of the linear mapping between the source and the destination which corresponds to this event: $\{\forall m \le K \ \mathbf{X}_m^N = \mathbf{0}\}$. We compute the probability of this event under the assumption that the entries of matrices $\mathbf{F}_m^k$ are i.i.d with uniformly random distribution. For analysis purposes, we divide the event into smaller events as following

$$\mathbb{P}[\forall m \le K \ \mathbf{X}_m^N = \mathbf{0}] = \sum_{\eta=L}^{K} \mathbb{P}[\forall m \le K \ \mathbf{X}_m^N = \mathbf{0}, \exists \text{ exactly } \eta \text{ frame } m : \mathbf{X}_m^0 = \mathbf{0}] \quad (4)$$

Notice that η starts from L here since in our coding scheme the last L frames of each block is set to zero at the source.

Consider a block of channel uses. Define $\Omega_m = \{k : \mathbf{X}_m^k \neq \mathbf{0}\}$ for every $m = 1, 2, \dots, K$. Based on the construction of the coding scheme, we can easily show that $\Omega_1, \Omega_2, \dots, \Omega_K$ is a Markov chain. Interestingly, based on the properties of Markov chain, we can break $\mathbb{P}[(\Omega_1, \dots, \Omega_K)]$ into state transition probabilities. Thus,

$$\begin{aligned} \mathbb{P}&[(\Omega_1, \dots, \Omega_K)] \\ &= \mathbb{P}[(\Omega_1, \dots, \Omega_{K-1})]\mathbb{P}[\Omega_K|(\Omega_1, \dots, \Omega_{K-1})] \\ &= \mathbb{P}[(\Omega_1, \dots, \Omega_{K-1})]\mathbb{P}[\Omega_K|\Omega_{K-1}] \\ &= \dots \\ &= \mathbb{P}[\Omega_1]\mathbb{P}[\Omega_2|\Omega_1] \dots \mathbb{P}[\Omega_K|\Omega_{K-1}] \\ &= \mathbb{P}[\Omega_1|\Omega_K]\mathbb{P}[\Omega_2|\Omega_1] \dots \mathbb{P}[\Omega_K|\Omega_{K-1}] \end{aligned} \quad (5)$$

The latter equation $\mathbb{P}[\Omega_1] = \mathbb{P}[\Omega_1|\Omega_K]$ results from the construction of coding scheme; since $\mathbf{X}_{K-L+1}^0 = \dots = \mathbf{X}_K^0 = \mathbf{0}$ the output values of all nodes will be zero at frame K, hence $\{\Omega_1\}$ and $\{\Omega_1|\Omega_K\}$ events are equivalent.

Now, we consider an arbitrary term m in the product $\prod_{m=1}^{K} \mathbb{P}[\Omega_{m+1}|\Omega_m]$. Define a column vector $\mathbf{Y}_{\Omega_m}$ by stacking all the output column vectors $\mathbf{Y}_{m+1}^k$ of the nodes k in Ω_m and $\mathbf{X}_{\Omega_{m+1}^c}$ by stacking all the input column vectors $\mathbf{X}_{m+1}^k$ in Ω_{m+1}^c. Then consider the corresponding linear transformation $I_T \otimes \mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c}$ from $\mathbf{Y}_{\Omega_m}$ to $\mathbf{X}_{\Omega_{m+1}^c}$ which satisfies $\mathbf{X}_{\Omega_{m+1}^c} = (I_T \otimes \mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})\mathbf{Y}_{\Omega_m}$. Here $\otimes$ is the Kronecker matrix product. Pay attention that $\mathbf{X}_m^0$ is generated independent from the state of the Markov chain. Also, the encoding function $\mathbf{F}_m^k$ has i.i.d. uniformly distributed entries on $\mathbb{F}(p^n)$, thus, the entries of $\mathbf{Y}_{m+1}^k$ are i.i.d and uniformly distributed. In addition, note that $\mathbf{rank}(I_T \otimes \mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c}) = T\mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})$. Therefore, we have

$$\begin{aligned} \mathbb{P}[\Omega_{m+1}|\Omega_m] &\le \mathbb{P}[\mathbf{X}_{m+1}^0 = \mathbf{U}_{m+1}] \cdot \mathbb{P}[\mathbf{X}_{\Omega_{m+1}^c} = \mathbf{0}|\mathbf{X}_{\Omega_m} \neq \mathbf{0}] \\ &= \mathbb{P}[\mathbf{X}_{m+1}^0 = \mathbf{U}_{m+1}] \cdot p^{-nT\mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})} \end{aligned} \quad (6)$$

where $\mathbf{U}_{m+1}$ in the generated value by the source[1] for frame $m+1$ which is chosen randomly from $\mathbf{V}_{\text{in}}$ and d_{in} denotes dimension of $\mathbf{V}_{\text{in}}$. Note that the latter equality is an approximation which is valid for large T. The first part of the above product can be computed as

$$\mathbb{P}[\mathbf{X}_m^0 = \mathbf{U}_m] = \begin{cases} p^{-nd_{\text{in}}}, & m \leq K-L \;\&\; \mathbf{U}_m = 0 \\ (1-p^{-nd_{\text{in}}}), & m \leq K-L \;\&\; \mathbf{U}_m \neq 0 \\ 1, & m > K-L \;\&\; \mathbf{U}_m = 0 \end{cases} \tag{7}$$

Note that if $\Omega_m = \Omega_{m+1}$ and $\mathbf{X}_m^0 \neq \mathbf{0}$, then $\mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})$ is the rank of a cutset in the original network and it is greater than or equal to $r_{\min}$. However, if $\Omega_m \neq \Omega_{m+1}$, then $\mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})$ is not necessarily related to any cutset in the network. Luckily, here we have a loop of states in the Markov chain with the length of K, i.e. $\Omega_1, \Omega_2, \ldots, \Omega_K, \Omega_1$, which helps us to find a lower bound. In the exponent of the probability function of occurrence of these sequence of states we have $\sum_{m=1}^{K} \mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c})$. By applying Lemma 1 (see Appendix) we are able to find a lower bound as following

$$\begin{aligned} \sum_{m=1}^{K} \mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c}) &\geq \sum_{i=1}^{K} \mathbf{rank}(\mathbf{G}_{[\Omega_1,\ldots,\Omega_K]_i}^{[\Omega_1^c,\ldots,\Omega_K^c]_{K-i+1}}) \\ &= \sum_{i=1}^{K} \mathbf{rank}(\mathbf{G}_{[\Omega_1,\ldots,\Omega_K]_i}^{([\Omega_1,\ldots,\Omega_K]_i)^c}) \end{aligned} \tag{8}$$

where $[\Omega_1, \ldots, \Omega_K]_i = \bigcup_{j_1,\ldots,j_i} (\bigcap_{l=1}^{i} \Omega_{j_l})$ is the union set of the intersection of all possible i subsets of $\Omega_1, \Omega_2, \ldots, \Omega_K$ and the latter equation (8) is obtained using De Morgan's laws. The inequality shows that for K transition between the states $\Omega_1, \ldots, \Omega_K, \Omega_1$ the sum of the ranks are greater than the sum of the ranks for K cutset of the original network defined by $[\Omega_1, \ldots, \Omega_K]_i$, $1 \leq i \leq K$.

It is easy to demonstrate based on the definitions that $[\Omega_1, \ldots, \Omega_K]_i$ includes the source node for every $i \leq K - \eta$. Each of these sets which contain the source node corresponds to a cutset between the source and the destination. Thus,

$$\sum_{m=1}^{K} \mathbf{rank}(\mathbf{G}_{\Omega_m}^{\Omega_{m+1}^c}) \geq r_{\min}(K-\eta) \tag{9}$$

Now based on equations (5) to (9), we have

$$\mathbb{P}[(\Omega_1, \ldots, \Omega_K), \eta] \leq p^{-nd_{\text{in}}(\eta-L)} p^{-nTr_{\min}(K-\eta)} \tag{10}$$

[1]To be rigorous, here we exclude Φ_0 from Ω_{m+1}^c for the cases that $\mathbf{X}_{m+1}^0 = \mathbf{0}$. We do not put this in equations to keep the notation simple.

Therefore,

$$\begin{aligned} &\mathbb{P}[\{\forall m \leq K, \mathbf{X}_m^N = \mathbf{0}\}] \\ &\leq \sum_{\eta=L}^{K} \sum_{(\Omega_1,\ldots,\Omega_K)} \mathbb{P}[(\Omega_1, \ldots, \Omega_K), \\ &\qquad \text{and } \mathbf{X}_m^0 = \mathbf{0} \text{ for exactly } \eta \text{ frames}] \\ &\leq \sum_{\eta=L}^{K} \binom{K-L}{\eta-L} p^{-nd_{\text{in}}(\eta-L)} p^{-nTr_{\min}(K-\eta)} \\ &= (p^{-nd_{\text{in}}} + p^{-nTr_{\min}})^{K-L} \end{aligned} \tag{11}$$

This equation shows that the dimension of the null space our linear mapping which maps $(K-L)$ independent symbols of the source in a block to the destination has at most dimension of $\log_{p^n}(p^{n(K-L)d_{\text{in}}}(p^{-nd_{\text{in}}} + p^{-nTr_{\min}})^{K-L}) = \frac{K-L}{n} \log_p(1 + p^{nd_{\text{in}} - nTr_{\min}})$.

For building the codewords, we define a space V_{in} with dimension of $d_{\text{in}} = Tr_{\min}$. Next, we exclude the null space of different terminals of multicast session from V_{in}. Since there are at most N terminals in the network, the dimension of remaining subspace $\tilde{V}_{\text{in}} \subset V_{\text{in}}$ will be at least $T(K-L)r_{\min} - N(K-L)/n \log_p(2)$. By choosing an arbitrary basis of $\tilde{V}_{\text{in}}$ as codeword for sending data we can achieve the following rate:

$$R = \frac{n \log_2(p)[T(K-L)r_{\min} - N(K-L)/n \log_p(2)]}{KT} \tag{12}$$

Clearly, as $T, K \to \infty$, $R \to r_{\min} n \log_2(p)$. ∎

Note that the equation (12) shows that the parameters T and K of the coding scheme need to be set large enough according to the size (N) and topology (L) of the network. Remarkably, we can find networks which need arbitrary large T and K for achieving the minimum cutset bound of a single multicast session.

V. Conclusion

In this paper, we proposed a network coding scheme for wireless networks modeled by "deterministic channel model" [2], [3]. We proved that our coding scheme can asymptotically achieve the minimum cutset upper bound for a single multicast session in acyclic wireless networks. Based on Markov chain properties of the underlying wireless network model, we provided a direct proof for achievability which directly deals with general acyclic networks and is shorter and simpler than the proof in existing work [3].

References

[1] R. Ahlswede, N. Cai, S. Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, 2000.

[2] Salman Avestimehr, Suhas Diggavi, and David Tse. A deterministic approach to wireless relay networks. In *ISIT*, 2007.

[3] Salman Avestimehr, Suhas Diggavi, and David Tse. Wireless network information flow. In *Allerton*, 2007.

[4] R. Koetter and M. Medard. An algebraic approach to network coding. *IEEE/ACM Transactions on Network*, 11(5):782–795, 2003.

[5] S. Y. R. Li, R. W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, 49(2):371–381, 2003.

APPENDIX

Lemma 1: Let $\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D}$ be four subsets of the nodes in the network. Then,

$$\mathbf{rank}(\mathbf{G}_{\mathcal{A}}^{\mathcal{B}}) + \mathbf{rank}(\mathbf{G}_{\mathcal{C}}^{\mathcal{D}}) \geq \mathbf{rank}(\mathbf{G}_{\mathcal{A}\cap\mathcal{C}}^{\mathcal{B}\cup\mathcal{D}}) + \mathbf{rank}(\mathbf{G}_{\mathcal{A}\cup\mathcal{C}}^{\mathcal{B}\cap\mathcal{D}})$$

This property can be generalized as follows. Let $\{\Omega_1,\Omega_2,\dots,\Omega_l\}$ and $\{\Theta_1,\Theta_2,\dots,\Theta_l\}$ be two sets each of them contain l subsets of the nodes in the network. Then,

$$\sum_{i=1}^{l}\mathbf{rank}(\mathbf{G}_{\Omega_i}^{\Theta_i}) \geq \sum_{i=1}^{l}\mathbf{rank}(\mathbf{G}_{[\Omega_1,\dots,\Omega_l]_i}^{[\Theta_1,\dots,\Theta_l]_{l-i+1}}) \qquad (13)$$

Proof of Lemma 1: To ease the notation in our proof, we define $\mathcal{U}_1 = \mathcal{A}\setminus(\mathcal{A}\cap\mathcal{C})$, $\mathcal{U}_2 = \mathcal{C}\setminus(\mathcal{A}\cap\mathcal{C})$, $\mathcal{U}_3 = \mathcal{A}\cap\mathcal{C}$, and $\mathcal{V}_1 = \mathcal{B}\setminus(\mathcal{B}\cap\mathcal{D})$, $\mathcal{V}_2 = \mathcal{D}\setminus(\mathcal{B}\cap\mathcal{D})$, $\mathcal{V}_3 = \mathcal{B}\cap\mathcal{D}$.

First we show the following rank inequality

$$\begin{aligned}\mathbf{rank}(\mathbf{G}_{\mathcal{A}}^{\mathcal{B}}) &= \mathbf{rank}(\begin{pmatrix}\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_1} & \mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3}\\ \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1} & \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}\end{pmatrix})\\ &\geq \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})\\ &\quad +\mathbf{rank}((\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})\end{aligned}$$

where $\mathbf{G}\setminus\mathbf{G}'$ is defined as the subspace of the span vector space of the columns of $[\mathbf{G}\ \mathbf{G}']$ which is orthogonal to the span vector space of the columns of $\mathbf{G}'$. Clearly, $\mathbf{rank}(\mathbf{G}\setminus\mathbf{G}') = \mathbf{rank}([\mathbf{G}\ \mathbf{G}']) - \mathbf{rank}(\mathbf{G}')$. For proving the above inequality, we note that the number of linearly independent rows (or columns) of $[\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\ \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}]$ is equal to $\mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})$. On the other hand there are at least $\mathbf{rank}((\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})$ linearly independent rows in $[\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_1}\ \mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3}]$ (rows in $\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3}$) which cannot be written as linear combination of the rows of $[\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\ \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}]$ (the rows of $\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}$). This shows that $\mathbf{G}_{\mathcal{A}}^{\mathcal{B}}$ has at least $\mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}((\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})$ linearly independent rows.

Similarly we have

$$\begin{aligned}\mathbf{rank}(\mathbf{G}_{\mathcal{C}}^{\mathcal{D}}) &= \mathbf{rank}(\begin{pmatrix}\mathbf{G}_{\mathcal{U}_2}^{\mathcal{V}_2} & \mathbf{G}_{\mathcal{U}_2}^{\mathcal{V}_3}\\ \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_2} & \mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}\end{pmatrix})\\ &\geq \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_2}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})\\ &\quad +\mathbf{rank}((\mathbf{G}_{\mathcal{U}_2}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})\end{aligned}$$

Finally, we complete the proof by using matrix rank inequality $\mathbf{rank}(\mathbf{G}) + \mathbf{rank}(\mathbf{G}') \geq \mathbf{rank}([\mathbf{G}\ \mathbf{G}'])$.

$$\begin{aligned}&\mathbf{rank}(\mathbf{G}_{\mathcal{A}}^{\mathcal{B}}) + \mathbf{rank}(\mathbf{G}_{\mathcal{C}}^{\mathcal{D}})\\ \geq\ & [\mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_1}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3}) + \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_2}\setminus\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})]\\ &+[\mathbf{rank}((\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger}) + \mathbf{rank}((\mathbf{G}_{\mathcal{U}_1}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})\\ &\quad+\mathbf{rank}((\mathbf{G}_{\mathcal{U}_2}^{\mathcal{V}_3})^{\dagger}\setminus(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3})^{\dagger})]\\ \geq\ & \mathbf{rank}(\mathbf{G}_{\mathcal{U}_3}^{\mathcal{V}_3\cup\mathcal{V}_1\cup\mathcal{V}_2}) + \mathbf{rank}((\mathbf{G}_{\mathcal{U}_3\cup\mathcal{U}_1\cup\mathcal{U}_2}^{\mathcal{V}_3})^{\dagger})\\ =\ & \mathbf{rank}(\mathbf{G}_{\mathcal{A}\cap\mathcal{C}}^{\mathcal{B}\cup\mathcal{D}}) + \mathbf{rank}(\mathbf{G}_{\mathcal{A}\cup\mathcal{C}}^{\mathcal{B}\cap\mathcal{D}})\end{aligned}$$

The second part of lemma can be proved easily by induction on the number of sets (l) and using the first part. ■