Simultaneous communication in noisy channels

(draft : filename: simcom.tex)

AUTHORS

Abstract

A sender wishes to broadcast a message of length \( n \) over an alphabet of size \( k \) to \( r \) users, where each user \( i \), \( 1 \leq i \leq r \) should be able to receive one of possible \( m_i \) messages. The channel in which the message is sent has noise for each of the users (possibly different noise for different users), who cannot distinguish between some pairs of letters. The vector \( (m_1, m_2, \ldots, m_r)n \) is said to be feasible if length \( n \) encoding and decoding schemes exist by which every user is able to decode his message. A rate vector \((R_1, R_2, \ldots, R_r)\) is feasible if there exists a sequence of feasible vectors \( (m_1, m_2, \ldots, m_r)n \) such that \( (R_1, \ldots, R_r) = \lim_{n \to \infty} (\frac{\log_2{m_1}}{n}, \ldots, \frac{\log_2{m_r}}{n}) \).

In this work we investigate which rate vectors are feasible for several different scenarios and describe some of their properties. The maximum total rate, which is defined to be the sum of all users’ rates, is an additional parameter for which we give lower and upper bounds.

An interesting case we discuss is when each user has a partition of the letters into disjoint sets and for each transmitted letter, the user only knows the set which the letter belongs to. We give tight lower and upper bounds for this case and hence describe all feasible rate vectors. In particular, this generalizes the symmetric case where there are \( r = k \) users and each user \( i \) distinguishes only a single unique letter from all other letters. In such a case, the rate vector \((\log_2{k}/k, \ldots, \log_2{k}/k)\) whose total rate is \( \log_2{k} \), is feasible.

The proofs are based on the probabilistic method, tools from information theory and the relation to the Shannon capacity of a graph, which is the special case of the problem corresponding to a single user.

1 Introduction

A sender has to transmit messages to \( r \)-users, where the user \( i \) should be able to receive any one of \( m_i \) messages. To this end, the sender broadcasts a message of length \( n \) over an alphabet \( \Sigma \) of size \( k \). Each user \( i \) has a confusion graph \( G_i \) on the set of letters of \( \Sigma \), where two letters \( a, b \in \Sigma \) are connected if and only if user \( i \) cannot distinguish between \( a \) and \( b \). The sender and users can agree on a (deterministic) coding scheme. For each possible values \( a_i \) of the messages, \( 1 \leq a_i \leq m_i \), the scheme should enable the sender to transmit a string of length \( n \) over \( \Sigma \) so that each user \( i \) will be able to recover \( a_i \). The vector of a scheme for length \( n \) messages is \((m_1, m_2, \ldots, m_r)\). The rate vector of a sequence of schemes is the limit

\[
\lim_{n \to \infty} \left( \frac{\log_2{m_1}}{n}, \frac{\log_2{m_2}}{n}, \ldots, \frac{\log_2{m_r}}{n} \right),
\]

1
assuming the limit exists for this sequence. Our objective is to study which vectors and which rate vectors are feasible for a given set of confusion graphs $G_i$. This seems to be difficult even for relatively small cases, and reveals some interesting phenomena. Note that this problem generalizes the problem of computing the Shannon capacity of a graph which was first considered by Shannon in [5]. In the case of a single user (i.e. a single confusion graph $G$), the maximum feasible rate is precisely $\log_2 c(G)$ where $c(G)$ denotes the Shannon capacity of $G$. As this is true for any single user, the following upper bound follows.

Proposition 1 Given $r$ users whose confusion graphs are $G_1, G_2, \ldots, G_r$, a feasible rate vector $(R_1, R_2, \ldots, R_r)$ must satisfy $R_i \leq \log_2 c(G_i)$ for every $1 \leq i \leq r$, hence $\sum_{i=1}^{r} R_i \leq \sum_{i=1}^{r} \log_2 c(G_i)$.

Investigating the feasible rate vectors for a given set of confusion graphs raises an interesting question. What is the maximum capacity of the channel for all users together? The total capacity can be measured as the sum of rates for each user. This sum encapsulates the usability of the channel. Although in practice we have several users, their total rate cannot exceed the possible rate of a single user who shares all their information. Given a set of confusion graphs $G_1, \ldots, G_r$, let $G = \bigcap_{i=1}^{r} G_i$ be the graph over the same alphabet $\Sigma$, where $a, b \in \Sigma$ are connected in $G$ if and only if they are connected in $G_i$ for every $i$. The confusion graph $G$ represents the information all users have together and therefore can bound their total rate as follows.

Proposition 2 Given $r$ users whose confusion graphs are $G_1, G_2, \ldots, G_r$, any feasible rate vector $(R_1, R_2, \ldots, R_r)$ must satisfy $\sum_{i=1}^{r} R_i \leq \log_2 c(\bigcap_{i=1}^{r} G_i)$.

1.1 Our Results

For simplicity we omit all floor and ceiling signs whenever these are not crucial.

Throughout the paper we investigate the feasible rate vectors for several sets of confusion graphs. An important simple property of the feasible rate vectors is convexity. This property can be stated formally as follows.

Proposition 3 Let $G_1, G_2, \ldots, G_r$ be the confusion graphs for $r$ users. Given two feasible rate vectors $\overline{R}, \overline{R}'$ and $\alpha \in [0, 1]$, the rate vector $\alpha \overline{R} + (1 - \alpha) \overline{R}'$ is also feasible.

Proof: Since both $\overline{R}$ and $\overline{R}'$ are feasible rate vectors, each has some corresponding encoding scheme. Our new encoding scheme would be to use the encoding scheme corresponding to $\overline{R}$ in the first $\alpha n$ coordinates and the one corresponding to $\overline{R}'$ in the remaining $(1 - \alpha)n$ coordinates. The resulting rate vector is precisely $\alpha \overline{R} + (1 - \alpha) \overline{R}'$, as required. □

Corollary 4 Let $G_1, G_2, \ldots, G_r$ be the confusion graphs for $r$ users. Given $x_1, x_2, \ldots, x_r \in [0, 1]$ so that $\sum_{i=1}^{r} x_i = 1$, the rate vector $(x_1 \cdot \log_2 c(G_1), \ldots, x_r \cdot \log_2 c(G_r))$ is feasible.
Proof: For every $1 \leq i \leq r$, the rate vector consisting of rate $\log_2 c(G_i)$ for the $i$’th user and zero rate for all other users is trivially feasible. The result thus follows by Proposition 3. □

Let $\Sigma_k = \{\sigma_1, \sigma_2, \ldots, \sigma_k\}$ be an alphabet of size $k$. The first case we analyze is where each confusion graph is a disjoint union of cliques. Given user $i$, one can view his noise as receiving $y_i = f_i(x)$ whenever $x$ is transmitted, where $f_i(x) : \Sigma_k \mapsto \{1, 2, \ldots, \ell_i\}$ is the index of the clique which $x$ belongs to and $\ell_i$ is the number of cliques in user’s $i$ confusion graph. Note that we consider isolated vertices as cliques of size one and hence $f_i$ is well defined up to the order of the cliques.

Definition 1 Given a probability distribution $p$ over $\Sigma_k$ and a subset of the users $I \subseteq \{1, 2, \ldots, r\}$, let $H(p)(\{Y_i\}_{i \in I})$ be the binary entropy of the random variables $\{Y_i\}_{i \in I}$ where $Y_i = f_i(X)$ and $X$ is the random variable distributed over $\Sigma_k$ according to $p$.

For each subset of the users $I \subseteq \{1, 2, \ldots, r\}$, the alphabet $\Sigma_k$ can be partitioned into $s \leq k$ disjoint parts $(A_1, \ldots, A_s)$ according to what these users receive $(\{f_i(x)\}_{i \in I})$. These users cannot distinguish between different letters from the same part $A_j$, so their joint information when sending a letter $X$ from $\Sigma_k$ according to the probability distribution $p$ can be computed as

$$ H(p)(\{Y_i\}_{i \in I}) = - \sum_{1 \leq j \leq s} \Pr[X \in A_j] \cdot \log_2 \Pr[X \in A_j] . $$

Therefore in a sense that will be made precise later, when using only messages in which the letters are distributed according to some distribution $p$, we expect no subset $I$ of users to have total rate which exceeds $H(p)(\{Y_i\}_{i \in I})$.

The following theorem provides a full characterization of all feasible rate vectors in such a scenario.

Theorem 5 Let $G_1, G_2, \ldots, G_r$ be the confusion graphs for $r$ users over the alphabet $\Sigma_k = \{\sigma_1, \sigma_2, \ldots, \sigma_k\}$, where each confusion graph is a disjoint union of cliques. Using the notations and definitions above, a rate vector $(R_1, R_2, \ldots, R_r)$ is feasible if and only if there exists a probability distribution $p = (p_1, \ldots, p_k)$ over $\Sigma_k$ so that for every subset $I \subseteq \{1, 2, \ldots, r\}$ of the users,

$$ \sum_{i \in I} R_i \leq H(p)(\{Y_i\}_{i \in I}) . $$

An interesting special case of the above theorem is the symmetric dense scenario of $r = k$ users, where the confusion graph $G_i$ of user $i$ is a clique over $\Sigma_k - \{\sigma_i\}$. In other words, user $i$ only distinguishes the letter $\sigma_i$ from all other letters. When decoding, the relevant information for user $i$ is only the locations of $\sigma_i$ in the transmitted message. The following theorem is thus a corollary of Theorem 5.
Theorem 6 For every fixed \( k \geq 3 \), \((\log_2 k, \ldots, \log_2 k)\) is a feasible rate vector over the alphabet \( \Sigma_k = \{\sigma_1, \sigma_2, \ldots, \sigma_k\} \), when each confusion graph \( G_i \) is a clique on \( \Sigma_k - \{\sigma_i\} \).

Theorem 6 indicates the possible gain of encoding schemes for several users simultaneously. The total rate here is \( \log_2 k \), whereas using convexity with encoding schemes for single users cannot exceed a total rate of 1 (as this is the maximum rate for each single user; the Shannon capacity \( c(G_i) \) is precisely 2 for each \( 1 \leq i \leq k \)). However, in some cases there is no such gain. Several examples are discussed in what follows.

Let \( \Sigma_k = \{\sigma_1, \sigma_2, \ldots, \sigma_k\} \) be an alphabet of size \( k \) and let \( \Sigma_d = \{\sigma_1, \ldots, \sigma_d\} \) denote the set of the first \( d \) letters of \( \Sigma_k \), where \( 2 \leq d \leq k \). Consider the case where user 1 has a confusion graph

\[ G_1 = (\Sigma_k, \{ab \mid a \in \Sigma_k \land b \in \Sigma_k - \Sigma_d\}), \]

meaning the complete graph over \( \Sigma_k \) minus a clique over \( \Sigma_d \). The Shannon capacity of such graphs is easily shown to be \( c(G_1) = d \), hence the maximum rate of user 1 is at most \( \log_2 d \). The following results indicate, that for two different confusion graphs of user 2, nothing can be gained beyond convexity of single user encoding schemes. We need the following definition.

Definition 2 A rate vector is optimal if no user can increase his rate while the other user maintains the same rate.

Note that the total rate of such optimal rate vectors does not necessarily reach the maximum total rate possible.

Theorem 7 In the scenario described above for \( 2 \leq d \leq k \), when user 2 has the empty confusion graph \( G_2 = (\Sigma_k, \emptyset) \), the rate vectors \((\alpha \log_2 d, (1 - \alpha) \log_2 k)\) for \( \alpha \in [0, 1] \) are optimal.

Theorem 8 In the scenario described above for \( 2 \leq d \leq \frac{k + 1}{2} \), when user 2 has the complement confusion graph, that is

\[ G_2 = \overline{G_1} = (\Sigma_k, \{ab \mid a, b \in \Sigma_d\}), \]

the rate vectors \((\alpha \log_2 d, (1 - \alpha) \log_2 (k - d + 1))\) for \( \alpha \in [0, 1] \) are optimal.

The rest of the paper is organized as follows. In Section 2 we characterize the feasible rate vectors for the first scenario, where each confusion graph is a union of disjoint cliques, (Theorem 5), and demonstrate how combining encoding schemes for many users can sometimes outperform convexity (Theorems 6). Section 3 deals with the second family of confusion graphs described above in which convexity yields the optimal rate vectors (Theorems 7 and 8). Combining these results, one can characterize all feasible rate vectors for all scenarios involving two users and alphabet of size 3. In Section 4 we elaborate in more details on this analysis. The final Section 5 contains some concluding remarks and open problems.
2 Unions of disjoint cliques - outperforming convexity

We consider the case where the confusion graph of each user $i$ is a disjoint union of cliques. This case is of special interest as we can give a full description of all feasible rate vectors and it is deterministic in the sense that given the transmitted letter, we can transform it to the different symbols that each user receives. Moreover, choosing specific confusion graphs, it demonstrates how the maximum possible total rate can be achieved by combining schemes for many users (and only this way), even when the confusion graphs are nearly complete.

2.1 Proof of Theorem 5

Upper bound: Let $G_1, G_2, \ldots, G_r$ be a set of confusion graphs for $r$ users over the alphabet $\Sigma_k$, where each $G_i$ is a disjoint union of cliques, and let $(R_1, R_2, \ldots, R_r)$ be a feasible rate vector for this scenario. Suppose $k \geq 3$, as the case $k \leq 2$ is trivial. Let $(m_1, m_2, \ldots, m_r)(n)$ be a sequence of feasible vectors which converges to $(R_1, R_2, \ldots, R_r)$. Fix a large $n$ and fix a length $n$ scheme which realizes $(m_1, m_2, \ldots, m_r)$. The messages used in this scheme can be divided into families according to the number of appearances of each letter $\sigma_i$. As there are only $k$ letters and all the messages are of length $n$, there are at most $n^{k-1}$ different families. One of these families therefore contains at least $1/n^{k-1}$ of all possible messages $\prod_{i \in [r]} m_i$ (where $[r] = \{1, 2, \ldots, r\}$). Let $p = (p_1, p_2, \ldots, p_k)$ be the distribution of the letters of $\Sigma_k$ matching this largest family of messages.

Definition 3 Given a probability distribution $p$ over $\Sigma_k$ and a subset of the users $I \subseteq [r]$, let $N(p)(n; \{Y_i\}_{i \in I})$ denote the number of possible (different) messages for these users under the restriction that each message is originated in a length $n$ message over $\Sigma_k$ in which $\sigma_i$ appears $p_i n$ times.

Lemma 9

$$\frac{2^{H(p)(\{Y_i\}_{i \in I})n}}{n^k} \leq N(p)(n; \{Y_i\}_{i \in I}) \leq 2^{H(p)(\{Y_i\}_{i \in I})n}.$$ 

Proof: As stated before, given a subset of the users $I \subseteq [r]$, the alphabet $\Sigma_k$ can be partitioned into $s \leq k$ disjoint parts $(A_1, \ldots, A_s)$ according to $\{f_i(x)\}_{i \in I}$. The number of different messages for these users can be computed as

$$N(p)(n; \{Y_i\}_{i \in I}) = \binom{n}{p'_1 n, p'_2 n, \ldots, p'_s n}$$
Given some large $G$ where each $p_i$ is a disjoint union of cliques, and let $p = (p_1, \ldots, p_k)$ be a probability distribution over $\Sigma$. Fix $(R_1, R_2, \ldots, R_r)$ so that for any subset of the users $I \subseteq [r]$, $\sum_{i \in I} R_i \leq H(p)(\{Y_i\}_{i \in I})$. Given some large $n$, we set the number of messages $m_i$ for each user $i$ to be $m_i = \frac{2R_i n}{n^{k+2}}$ (which clearly satisfies $\lim_{n \to \infty} \frac{\log_2 m_i}{n} = \lim_{n \to \infty} \frac{R_i n^{-(k+2)} \log_2 n}{n} = R_i$). By Lemma 9, for every subset of the users $I \subseteq [r]$, 

$$\prod_{i \in I} m_i = \prod_{i \in I} \frac{2R_i n}{n^{k+2}} = \frac{2^{\sum_{i \in I} R_i n}}{n^{(k+2)|I|}} \leq \frac{2^{H(p)(\{Y_i\}_{i \in I}) n}}{n^{k+2}} \leq \frac{N(p)(n; \{Y_i\}_{i \in I})}{n^2}. \quad (1)$$

Our encoding scheme will use only messages in which the letters of $\Sigma$ are distributed according to $p$. For each user $i$ there are $N(p)(n; \{Y_i\})$ different messages that he can identify. We randomly
divide them into \( m_i \) families \( \mathcal{F}_{i,1}, \ldots, \mathcal{F}_{i,m_i} \), where each family represents a different message for user \( i \). When the message \( \overline{x} \in \Sigma_k^n \) is transmitted, user \( i \) receives \( \overline{y}_i = f_i(\overline{x}) = f_i(x_1) \cdots f_i(x_n) \) and decodes the message \( j \), the single \( j \in [m_i] \) for which \( \overline{y}_i \in \mathcal{F}_{i,j} \).

In order to complete the proof we show the described encoding scheme is valid with high probability. Such a scheme is valid if for every set of messages \( \{j_i \in [m_i]\}_{i \in [r]} \) there exists a message \( \overline{x} \) so that for every user \( i \in [r] \), \( \overline{y}_i \in \mathcal{F}_{i,j_i} \). Our proof uses the extended Janson inequality (c.f., e.g., [2], Chapter 8), which we describe here for completeness.

**Theorem 10 (Janson)** Let \( S \) be a set, for each \( s \in S \) let \( p_s \) be a real \( 0 \leq p_s \leq 1 \). Let \( R \) be a random subset of \( S \) obtained by selecting each \( s \in S \), randomly and independently, to lie in \( R \) with probability \( p_s \). Let \( A_i, i \in I \) be a family of subsets of \( S \). For each \( i \in I \), let \( B_i \) be the event that \( A_i \subset R \). Let \( \mu = \sum_{i \in I} \text{Prob}[B_i] \) be the expected number of events \( B_i \) that occur. Define \( \Delta = \sum \text{Prob}[B_i \wedge B_j] \), where the sum ranges over all ordered pairs \( i, j \in I, i \neq j \) such that \( A_i \cap A_j \neq \emptyset \). Then the probability that none of the events \( B_i \) occurs is at most \( e^{-\mu + \Delta/2} \). If the further assumption that \( \Delta > \mu \) holds, this probability can also be bounded by \( e^{-n^2/2\Delta} \).

Fix messages \( j_1, j_2, \ldots, j_r \) for the \( r \) users. Consider the \( r \)-uniform \( r \)-bipartite hypergraph whose classes of vertices are \( \mathcal{F}_{i,j_i}, 1 \leq i \leq r \). Each edge represents a consistent message, i.e. for every message \( \overline{x} \in \Sigma_k^n \) there exists an edge \( \{\overline{y}_i \in \mathcal{F}_{i,j_i}\}_{i \in [r]} \) if indeed for every \( i \in [r] \), \( \overline{y}_i = f_i(\overline{x}) \in \mathcal{F}_{i,j_i} \). Existence of some edge in the hypergraph indicates that this set of messages can be transmitted as required.

Using the notations of Theorem 10, our set \( S \) is the union of all possible messages \( f_i(\overline{x}) \) for each user \( i \) and for \( \overline{x} \in \Sigma_k^n \) which is distributed according to \( p \). The probability of each element is \( 1/m_i \) for the relevant user \( i \). The sets \( A_i \) represent all the consistent messages \( \{f_i(\overline{x})\}_{i \in [r]} \) where again, \( \overline{x} \in \Sigma_k^n \) and is distributed according to \( p \). The quantity \( \mu \) is exactly

\[
N(\mu)(n; \{Y_i\}_{i \in [r]}) \prod_{i=1}^r \frac{1}{m_i} \geq n^2,
\]

where the inequality follows from (1). The quantity \( \Delta \) is given by

\[
\Delta \leq N(\mu)(n; \{Y_i\}_{i \in [r]}) \prod_{i=1}^r \frac{1}{m_i} \sum_{\emptyset \neq I \subset [r]} \frac{N(\mu)(n; \{Y_i\}_{i \in [r]})}{N(\mu)(n; \{Y_i\}_{i \in [r]})} \prod_{i \notin I} \frac{1}{m_i}
\]

Indeed, there are \( N(\mu)(n; \{Y_i\}_{i \in [r]}) \) possible messages for all the users together, each with probability \( \prod_{i=1}^r \frac{1}{m_i} \). For every fixed message and a fixed subset of the users \( \emptyset \neq I \subset [r] \), there are by symmetry at most \( N(\mu)(n; \{Y_i\}_{i \in [r]})/N(\mu)(n; \{Y_i\}_{i \in [r]}) \) possible different ways to complete the message for the other users (which are not in \( I \)) given the choice for the users which are in \( I \), and the probability that the alternative extension of the message is available is \( \prod_{i \notin I} \frac{1}{m_i} \), giving the expression above.
Therefore,

{\Delta} \leq N(p)(n; \{Y_i\}_{i \in [r]}) \frac{1}{m_i} \sum_{\emptyset \neq I \subseteq [r]} N(p)(n; \{Y_i\}_{i \in I}) \prod_{i \notin I} \frac{1}{m_i}

= \mu \sum_{\emptyset \neq I \subseteq [r]} \frac{N(p)(n; \{Y_i\}_{i \in [r]})}{N(p)(n; \{Y_i\}_{i \in [r]})} \prod_{i \notin I} \frac{1}{m_i}

= \mu^2 \sum_{\emptyset \neq I \subseteq [r]} \prod_{i \notin I} \frac{1}{m_i} \leq \mu^2 \sum_{\emptyset \neq I \subseteq [r]} \frac{1}{n^2} = \frac{\mu^2}{n^2} (2^k - 2),

where the last inequality follows from (1). By Theorem 10 the probability that there exists no valid message as required is at most \(e^{\mu^2/2\Delta} \leq e^{-n^2/2k+1} < \frac{1}{2^{n^2}}\) (here we assumed \(\Delta > \mu\), if this is not the case, Theorem 10 yields an upper bound of \(e^{-\mu + \Delta/2} \leq e^{-\mu/2} < e^{-n^2/2} < \frac{1}{2^{n^2}}\). As there are \(\prod_{i=1}^r m_i \leq k^n\) distinct choices for \(j_1, \ldots, j_r\), the assertion of Theorem 5 follows by the union bound. \(\square\)

2.2 Proof of Theorem 6

Theorem 6 is a corollary of Theorem 5 for the symmetric dense case where there are \(r = k\) users, and each user \(i\) distinguishes a single letter from all other letters. This simple case demonstrates how the maximum rate can be achieved only by mixing the messages for all the users. Since each confusion graph \(G_i\) is a clique on \(\Sigma_k - \{\sigma_i\}\), the Shannon capacity of this graph is 2 and therefore the maximal rate for any single user is 1. However, the theorem shows that indeed a total rate of \(\log_2 k\) can be achieved, and obviously this is best possible.

Let \(k \geq 3\) be fixed. In order to prove the theorem, a probability distribution \(p\) is required so that for every subset of the users \(I \subseteq [r] = [k]\),

\[
\sum_{i \in I} R_i = |I| \frac{\log_2 k}{k} \leq H_p(\{Y_i\}_{i \in I}). \tag{2}
\]

Let us consider the uniform probability distribution \(p, p_i = 1/k\) for every letter \(\sigma_i\). When considering all users together, the random variables \(\{Y_i\}_{i \in [r]}\) indicate the exact letter \(x\) that was transmitted. Therefore, since the entropy is exactly \(\log_2 k\), \(H_p(\{Y_i\}_{i \in [r]}) = \log_2 k\) which indeed satisfies (2) as \(\sum_{i \in [r]} R_i = \frac{\log_2 k}{k} = \frac{\log_2 k}{k}\).

Given a subset of the users \(I \subset [r]\), the random variables \(\{Y_i\}_{i \in I}\) indicate which letter \(x\) was transmitted if \(x = \sigma_i\) for \(i \in I\) or alternatively, that some other letter was transmitted. Hence its binary entropy satisfies

\[
H_p(\{Y_i\}_{i \in I}) = |I| \frac{\log_2 k}{k} + \frac{k - |I|}{k} \log_2 \frac{k}{|I|} > |I| \frac{\log_2 k}{k} = \sum_{i \in I} R_i
\]

thus (2) holds for every subset of the users \(I \subseteq [r]\), as required. \(\square\)
3 A clique minus a clique - convexity is everything

We consider the case where the confusion graph $G_1$ of the first user is the complete graph on $k$ vertices minus a clique on $d$ vertices. We give an upper bound on the rate of the other user for both the empty confusion graph and for $\bar{G}_1$. In both cases, the results imply that optimal encoding can be achieved by convexity, that is nothing can be gained by encoding the messages together. In order to prove Theorems 7 and 8, we need the following lemmas.

Lemma 11 Given $a, b \in \mathbb{N}$ s.t. $2 \leq b \leq a$ and $x_1 \geq x_2 \geq \cdots \geq x_b \geq 0$, 

$$(a - b + 1)a^{\log_b x_b} + \sum_{i \in [b-1]} a^{\log_b x_i} \leq a^{\log_b \sum_{i \in [b]} x_i}.$$ 

Remark: Define here $a^{\log_b 0} = 0$ hence we allow $x_b$ to be 0.

Proof: When $x_1 = x_2 = \cdots = x_b$ equality holds as

$$(a - b + 1)a^{\log_b x_b} + \sum_{i \in [b-1]} a^{\log_b x_i} = a \cdot a^{\log_b x_b} = a^{1 + \log_b x_b} = a^{\log_b b \cdot x_b} = a^{\log_b \sum_{i \in [b]} x_i}. $$

In order to complete the proof, it suffices to show that the partial derivatives $\frac{\partial}{\partial x_i}$ are smaller on the left hand side than those on the right hand side for any $i \in [b - 1]$, regardless of the values $\{x_i\}$. Given a fixed $i \in [b - 1]$, the derivative of the left hand side is $\frac{\partial}{\partial x_i} a^{\log_b x_i} = \frac{\partial}{\partial x_i} x_i^{\log_b a} = \log_b a \cdot x_i^{\log_b a - 1}$. On the other hand, the derivative of the right hand side is $\log_b a \cdot (\sum_{i \in [b]} x_i)^{\log_b a - 1}$ which is at least as big. \[\] 

Lemma 12 Given $2 \leq d \leq k \in \mathbb{N}$ and a set $\mathcal{G} \subseteq \Sigma_k^n$, we define $\mathcal{G}' = \mathcal{G} \cap \Sigma_d^n$. If $\mathcal{G}$ is closed under replacing each $\sigma_i$ with $\sigma_j$ for any $i > d$ and $j \in [k]$, then either $|\mathcal{G}| = |\mathcal{G}'| = 0$ or

$$\log_k |\mathcal{G}| \leq \log_d |\mathcal{G}'|.$$ 

Proof: We apply induction on $n$. For $n = 1$, if $\mathcal{G}' = \mathcal{G}$ the inequality holds as $k \geq d$ (or both sets are empty). Otherwise there exists $\sigma_i \in \mathcal{G}$ for $i > d$, hence $|\mathcal{G}'| = |\Sigma| = k$ and $|\mathcal{G}'| = d$ for which equality holds.

Assuming the lemma holds for any $n' < n$ we prove it for $n$. Define $\mathcal{G}_i = \{g_1g_2 \cdots g_n-1 \mid g \in \mathcal{G} \land g_n = \sigma_i\}$ and $\mathcal{G}'_i = \mathcal{G}_i \cap \Sigma_d^n$ for every $i \in [k]$. Note that $\mathcal{G}'_i \subseteq \mathcal{G}'_j$ and hence $|\mathcal{G}'_i| \leq |\mathcal{G}'_j|$ for every $i > d$ and $j \in [d]$ (since $\mathcal{G}$ is closed under replacing $\sigma_i$ with $\sigma_j$ for $i > d$ and $j \in [k]$). In particular, this is true for $m \in [d]$ so that $|\mathcal{G}'_m| = \min_{j \in [d]} |\mathcal{G}'_j|$. Therefore, by the induction hypothesis and
Lemma 11 with \( a = k \) and \( b = d \),

\[
|\mathcal{G}| = \sum_{i \in [k]} |\mathcal{G}_i| \leq \sum_{i \in [k]} k^{\log_d |\mathcal{G}_i|} \\
\leq (k-d) k^{\log_d |\mathcal{G}_{m_i}|} + \sum_{i \in [d]} k^{\log_d |\mathcal{G}_i|} \\
\leq (k-d+1) k^{\log_d |\mathcal{G}_{m_i}|} + \sum_{i \in [d]-\{m\}} k^{\log_d |\mathcal{G}_i|} \\
\leq k^{\log_d \sum_{i \in [d]} |\mathcal{G}_i|} = k^{\log_d |\mathcal{G}|}
\]

completing the proof. \( \square \)

**Lemma 13** Given \( 2 \leq d \leq k \in \mathbb{N} \) s.t. \( d \leq \frac{k+1}{2} \) and a set \( \mathcal{G} \subseteq \Sigma_k^n \), define \( \mathcal{G}' = \mathcal{G} \cap \Sigma_d^n \) and \( \mathcal{G}'' = \{ f(g_1) f(g_2) \cdots f(g_n) \mid g \in \mathcal{G} \} \) where \( f(\sigma_i) = \sigma_{\max(i,d)} \). If \( \mathcal{G} \) is closed under replacing each \( \sigma_i \) with \( \sigma_j \) for any \( i > d \) and \( j \in [k] \), then either \( |\mathcal{G}'| = |\mathcal{G}''| = 0 \) or

\[
\log_{k-d+1} |\mathcal{G}''| \leq \log_d |\mathcal{G}'|.
\]

**Remark:** The restriction of \( d \) is required as one can easily find an example where \( d = \lceil \frac{k+1}{2} \rceil \) for which the lemma does not hold (such examples are given in Appendix A.1).

**Proof:** Again we apply induction on \( n \). For \( n = 1 \), if \( \mathcal{G}' = \mathcal{G} \) we have no \( \sigma_i \in \mathcal{G} \) for \( i > d \) and the inequality holds as \( |\mathcal{G}''| = 1 \) or both sets are empty. Otherwise there exists \( \sigma_i \in \mathcal{G} \) for \( i > d \), hence \( |\mathcal{G}''| = k-d+1 \) and \( |\mathcal{G}'| = d \) for which equality holds.

Assuming the lemma holds for any \( n' < n \) we prove it for \( n \). Extending the previous definitions of \( \mathcal{G}_i \) and \( \mathcal{G}'_i \), let \( \mathcal{G}''_i = \{ f(g_1) f(g_2) \cdots f(g_{n-1}) \mid g \in \mathcal{G}_i \} \) for every \( i \in [k] \). Note that \( \mathcal{G}''_i \subseteq \mathcal{G}''_j \) and hence \( |\mathcal{G}''_i| \leq |\mathcal{G}''_j| \) for every \( i > d \) and \( j \in [k] \) (since \( \mathcal{G} \) is closed under replacing \( \sigma_i \) with \( \sigma_j \) for \( i > d \) and \( j \in [k] \)). Similarly, \( |\mathcal{G}''_i| \leq |\bigcup_{j \in [d]} \mathcal{G}''_j| \) for all \( i > d \). Therefore,

\[
|\bigcup_{j \in [d]} \mathcal{G}''_j| \leq \sum_{j \in [d]} |\mathcal{G}''_j| = (d-1) \cdot |\bigcup_{j \in [d]} \mathcal{G}''_j| \\
\leq \sum_{j \in [d]} |\mathcal{G}''_j| - (d-1) \cdot \max_{i \in [k]-[d]} |\mathcal{G}''_i| \\
\leq \sum_{j \in [d]} |\mathcal{G}''_j| - \frac{d-1}{k-d} \sum_{i \in [k]-[d]} |\mathcal{G}''_i|.
\]

As before, \( |\mathcal{G}_i| \leq |\mathcal{G}''_m| \) for every \( i > d \) and \( m \in [d] \) so that \( |\mathcal{G}_m'| = \min_{j \in [d]} |\mathcal{G}_j'| \). Therefore, by the induction hypothesis

\[
\sum_{i \in [k]-[d]} |\mathcal{G}''_i| \leq \sum_{i \in [k]-[d]} (k-d+1)^{\log_d |\mathcal{G}_i|} \\
\leq (k-d)(k-d+1)^{\log_d |\mathcal{G}_m'|}.
\]
By Lemma 11 with \( a = k - d + 1 \) and \( b = d \) (which indeed satisfies \( 2 \leq b \leq a \) as \( d \leq (k + 1)/2 \)),

\[
|\mathcal{G}''| = \sum_{i \in [k]-[d]} |\mathcal{G}'_i| + \left| \bigcup_{j \in [d]} \mathcal{G}'_j \right|
\]

\[
\leq \frac{(k - d) - (d - 1)}{k - d} \sum_{i \in [k]-[d]} |\mathcal{G}'_i| + \sum_{j \in [d]} |\mathcal{G}'_j|
\]

\[
\leq \frac{k - 2d + 1}{k - d} (k - d)(k - d + 1)^{\log_d |\mathcal{G}'_m|} + \sum_{j \in [d]} (k - d + 1)^{\log_d |\mathcal{G}'_j|}
\]

\[
= (k - 2d + 2)(k - d + 1)^{\log_d |\mathcal{G}'_m|} + \sum_{j \in [d]-\{m\}} (k - d + 1)^{\log_d |\mathcal{G}'_j|}
\]

\[
\leq (k - d + 1)^{\log_d \sum_{j \in [d]} |\mathcal{G}'_j|} = (k - d + 1)^{\log_d |\mathcal{G}'|}
\]

completing the proof. \( \Box \)

**Proof of Theorem 7:** Let \( G_1, G_2 \) be the confusion graphs as defined in the theorem and assume the rate of the first user is \( \alpha \log_2 d \) for some \( \alpha \in [0, 1] \). The messages used can be divided into disjoint families \( \mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_{d^n} \) according to the message for the first user. Since the first user can only distinguish between the letters \( \sigma_i \) for \( i \in [d] \), we can and will assume each such family \( \mathcal{F}_a \) is closed under replacing \( \sigma_i \) with \( \sigma_j \) for \( i > d \) and \( j \in [k] \).

In order to prove this assumption is valid, it suffices to show user 1 can still distinguish between each of the families after these replacements. Let \( \mathcal{G}_a \) denote the family \( \mathcal{F}_a \) after replacing \( \sigma_i \) with \( \sigma_j \) for \( i > d \) and \( j \in [k] \). Assume by contradiction that there exist two families \( \mathcal{F}_a, \mathcal{F}_b \) and two vectors \( v_a \in \mathcal{G}_a, v_b \in \mathcal{G}_b \) so that user 1 can distinguish between \( \mathcal{F}_a \) and \( \mathcal{F}_b \) but cannot distinguish between \( v_a \) and \( v_b \). By the definition of \( G_1 \), for every coordinate \( i \in [n] \), either \( v_a[i] \not\in \Sigma_d \) or \( v_b[i] \not\in \Sigma_d \) or \( v_a[i] = v_b[i] \in \Sigma_d \) as otherwise user 1 would be able to distinguish between them (here \( v_x[i] \) denotes the \( i \)th letter in the vector \( v_x \)). Since \( v_a \in \mathcal{G}_a \) and \( v_b \in \mathcal{G}_b \), we know there exists \( u_a \in \mathcal{F}_a \) and \( u_b \in \mathcal{F}_b \) from which \( v_a \) and \( v_b \) can be derived by the replacements above. Therefore, for every coordinate \( i \in [n] \), either \( u_a[i] \not\in \Sigma_d \) or \( u_b[i] \not\in \Sigma_d \) or \( u_a[i] = u_b[i] \in \Sigma_d \). However, this is in contradiction to the fact that user 1 was able to distinguish between \( \mathcal{F}_a \) and \( \mathcal{F}_b \) as he cannot distinguish between \( u_a \) and \( u_b \).

Define \( \mathcal{F}'_a = \mathcal{F}_a \cap \Sigma_d \) for every \( \mathcal{F}_a \). Since these families are pairwise disjoint, by an averaging argument there exists some message \( a \) for which \( |\mathcal{F}'_a| \leq d^{-\alpha n} \cdot d^n = d^{(1-\alpha)n} \). By Lemma 12, for this specific message \( a \), \( |\mathcal{F}_a| \leq k^{(1-\alpha)n} \) which implies the rate of the second user is at most \( (1-\alpha) \log_2 k \). \( \Box \)

**Corollary 14** For the confusion graph \( G_1 \) as above and any confusion graphs \( G_2, \ldots, G_r \), a feasible rate vector \( (\alpha \log_2 d, R_2, \ldots, R_r) \) for \( \alpha \in [0, 1] \) must satisfy \( \sum_{i=2}^r R_i \leq (1-\alpha) \log_2 k \).

**Proof of Theorem 8:** Let \( G_1, G_2 \) be the confusion graphs as defined in the theorem. Note that the Shannon capacity \( c(G_2) \) is precisely \( k-d+1 \), hence these rate vectors are feasible by Corollary
4 (as is also easy to see directly). Assume the rate of the first user is $\alpha \log_2 d$ for some $\alpha \in [0, 1]$. The messages used can be divided into disjoint families $F_1, F_2, \ldots, F_{d^n}$ according to the message for the first user. Since the first user can only distinguish between the letters $\sigma_i$ for $i \in [d]$, we can and will assume, as in the proof of Theorem 7, that each such family $F_a$ is closed under replacing $\sigma_i$ with $\sigma_j$ for $i > d$ and $j \in [k]$.

Define $F'_a = F_a \cap \Sigma_d^n$ for every $F_a$. Since these families are pairwise disjoint, by an averaging argument there exists some message $a$ for which $|F'_a| \leq d^{-\alpha} n = d^{(1-\alpha)n}$. Given the first user should receive the message $a$, the second user has at most $|F''_a| = |\{f(g_1)f(g_2) \cdots f(g_n) \mid g \in F_a\}|$ different messages where $f(\sigma_i) = \sigma_{\max\{i,d\}}$ (as the second user can only distinguish the locations of $\sigma_i$ for $i \in [k] - [d]$ and all other letters are indistinguishable for him). By Lemma 13, for this specific message $a$, $|F''_a| \leq (k - d + 1)(1-\alpha)n$ which implies the rate of the second user is at most $(1 - \alpha) \log_2 (k - d + 1)$. □

**Corollary 15** For the confusion graph $G_1$ as above and any confusion graphs $G_2, \ldots, G_r \supseteq G_1$, a feasible rate vector $(\alpha \log_2 d, R_2, \ldots, R_r)$ for $\alpha \in [0, 1]$ must satisfy $\sum_{i=2}^r R_i \leq (1-\alpha) \log_2 (k - d + 1)$.

### 4 Two users, three letters - the complete story

Two users and three letters is the smallest possible example of non-trivial scenario. Having only two letters result in each user either knowing everything or knowing nothing and obviously having a single user coincides with the Shannon capacity question. These smallest scenarios however already contain some interesting cases which we analyze using our previous results.

Let $\Sigma = \{\sigma_0, \sigma_1, \sigma_2\}$ be our alphabet and let $G_1, G_2$ be the confusion graphs of the two users correspondingly. If one of the confusion graphs is the complete graph, again it coincides with the Shannon capacity of a single graph (for the non-complete confusion graph). As stated earlier, there is a strong connection between the feasible rate vectors and the Shannon capacity of graphs. In the cases we are about to analyze, we use the fact that the Shannon capacity of every graph on 3 vertices which is neither the empty graph nor the clique, is precisely 2 (each such graph is perfect, hence its Shannon capacity equals its independence number). By the symmetry between the users and the letters in the alphabet, it suffices to discuss only subset of the possible confusion graphs.

#### 4.1 Confusion graph with two edges

In this subsection we show that when the first confusion graph has two edges, then no scheme can outperform what follows from convexity.

**Proposition 16** Let $G_1 = (\Sigma, \{\sigma_0 \sigma_1, \sigma_0 \sigma_2\})$, meaning the first user only distinguishes between the letters $\sigma_1$ and $\sigma_2$. For every $G_2$, the optimal rate vectors are given by Corollary 4, i.e.

$$(\alpha \cdot \log_2 c(G_1), (1 - \alpha) \cdot \log_2 c(G_2))$$
for \(\alpha \in [0, 1]\) where \(\log_2 c(G_1) = 1\).

**Remark:** Note that this matches the case of a clique minus a clique for the parameters \(k = 3\) and \(d = 2\) as denoted in previous sections, but here we do not limit the confusion graph of the second user.

**Proof:** The proof is divided into two parts, according to the intersection between the edges of \(G_1\) and \(G_2\). In the first case where there is a non-empty intersection, the bound given by Proposition 2 yields a maximum total rate of \(\log_2 c(G_1 \cap G_2) = 1\). Therefore, one could not hope for finding a feasible rate vector which is not of this form (assuming \(G_2\) is not the complete graph, these are all optimal rate vectors as they have a total rate of 1).

Let us assume there is no intersection between the two confusion graphs, meaning either \(G_2\) is the empty graph or \(G_2 = \overline{G_1}\). These cases match Theorems 7 and 8 respectively which indeed yield the desired result. \(\square\)

### 4.2 The first confusion graph has a single edge

Throughout this section we denote \(H\) as the binary entropy function where given some probability distribution \(p_1, p_2, \ldots, p_k, q\) (where \(q = 1 - \sum_{i=1}^{k} p_i\)),

\[
H(p_1, p_2, \ldots, p_k, q) = H(p_1, p_2, \ldots, p_k) = -\sum_{i=1}^{k} p_i \log_2 p_i - q \log_2 q.
\]

**Proposition 17** Let \(G_1 = (\Sigma, \{\sigma_0\sigma_1\})\) and \(G_2 = (\Sigma, \{\sigma_0\sigma_2\})\). The following rate vectors are optimal:

- \((R_1, H(R_1))\) for \(R_1 \in [1/2, 2/3]\).
- \((R_1, \log_2 3 - R_1)\) for \(R_1 \in [2/3, \log_2 3 - 2/3]\).
- \((H(R_2), R_2)\) for \(R_2 \in [1/2, 2/3]\).

**Remark:** By the proposition a rate vector \((R_1, R_2)\) in the above case is feasible if and only if

1. \(R_1 \in [0, 1/2]\) and \(R_2 \in [0, 1]\) or
2. \(R_1 \in [1/2, 2/3]\) and \(R_2 \in [0, H(R_1)]\) or
3. \(R_1 \in [2/3, \log_2 3 - 2/3]\) and \(R_2 \in [0, \log_2 3 - R_1]\) or
4. \(R_1 \in [\log_2 3 - 2/3, 1]\) and \(R_2 \in [0, H^{-1}(R_1)]\).
**Proof:** The scenario described above is a special case of Theorem 5. Given a probability distribution \(p = (p_0, p_1, p_2)\), \(y_1\) is distributed \((p_0 + p_1, p_2)\), \(y_2\) is distributed \((p_0 + p_2, p_1)\) and \(\{y_1, y_2\}\) is distributed according to \(p\).

The uniform distribution \(p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\) yields that the rate vectors \((R_1, R_2)\) are feasible if \(R_1 + R_2 \leq \log_2 3\) and each \(R_i \leq H(2/3) = \log_2 3 - 2/3\). This matches the second case described in the theorem, which is obviously optimal as one cannot hope to exceed a total rate of \(\log_2 3\).

By symmetry, it suffices to analyze the first case of the theorem in order to complete the proof. Setting \(p_1\) to be some probability smaller than half bounds the rate of the second user by \(R_2 \leq H(p_1) = H(1 - p_1)\). Moreover, the total rate \(R_1 + R_2\) is bounded by \(H(\frac{1-p_1}{2}, \frac{1-p_2}{2}, p_1) = H(p_1) + (1-p_1)\). This shows that the rate vectors \((R_1, H(R_1))\) are feasible as \(R_1 = 1-p_1 \leq H(\frac{1-p_1}{2})\) for \(p_1 \in [0, 1/2]\) (indeed equality holds for \(p_1 = 0\) and since \(H'(x) = \log_2(1-x) - \log_2 x < 2\) for \(x \in [1/4, 1/2]\), or equivalently \(H'(\frac{1-p_1}{2}) > -1\) for \(p_1 \in [0, 1/2]\), this holds for every \(p_1 \in [0, 1/2]\) as well). Moreover, these rate vectors are also optimal as the bound for the total rate \(H(p_1) + (1-p_1)\) decreases while \(p_1\) increases in the section \([1/3, 1/2]\) (using \(H'(x) < 1\) for \(x \in [1/3, 1/2]\)).

**Remark:** Although the rate vector \((1, 1/2)\) is feasible, the vector \((2^n, 2)\) is not feasible. If user 1 needs to be able to receive \(2^n\) distinct messages, one of them has to be encoded by \((\sigma_2, \sigma_2, \cdots, \sigma_2)\). But this has to be transmitted independently of the message of the second user, showing there is no \((2^n, 2)\)-scheme.

**Proposition 18** Let \(G_1 = (\Sigma, \{\sigma_0 \sigma_1\})\) and \(G_2\) be the empty graph. The following rate vectors are optimal:

- \((R_1, \log_2 3 - R_1)\) for \(R_1 \in [0, H(2/3)]\).
- \((H(R_2), R_2)\) for \(R_2 \in [1/2, 2/3]\).

**Remark:** By the proposition a rate vector \((R_1, R_2)\) in the above case is feasible if and only if

1. \(R_1 \in [0, \log_2 3 - 2/3]\) and \(R_2 \in [0, \log_2 3 - R_1]\) or
2. \(R_1 \in [\log_2 3 - 2/3, 1]\) and \(R_2 \in [0, H^{-1}(R_1)]\).

**Proof:** Our problem is monotone in the following sense. Removing an edge from one of the confusion graphs can only improve the feasible rate vectors. Therefore in our case, the lower bound of \((H(R_2), R_2)\) for \(R_2 \in [1/2, 2/3]\) we achieved when both confusion graphs had a single edge can be applied here. Showing these rate vectors are also optimal will complete the proof as we have the trivial upper bound of \(\log_2 3\) on the total rate, and by combining convexity with the fact that the rate vectors \((H(2/3), 2/3)\) and \((0, \log_2 3)\) are feasible, we conclude that all other required rate vectors are achieved.

Our problem is a special case of Theorem 5. Note that in the proof of Proposition 17 we showed an upper bound for the rates \((R_1, H(R_1))\) which only depended on the second user. Assuming
the second user has rate of $H(R_1)$ already bounds the total rate by $H(R_1) + R_1$. Similarly in our case, assuming the first user has rate of $H(R_2)$ for $R_2 \in [1/2, 2/3]$ bounds the total rate of the two users together by $H(R_2) + R_2$. □

5 Conclusions and open problems

In this work we have introduced a notion of simultaneous communication in a noisy channel where the channel’s noise may differ for each of the users. The goal is to find, for a given set of confusion graphs which represent the noise for each of the users, which rate vectors (or alternatively vectors) are feasible. As in the Shannon capacity of a channel, we care about the average rate per letter when the length of the messages tends to infinity.

Our work demonstrates basic lower and upper bounds for the general case. A simple yet useful tool in understanding the feasible rate vectors is the convexity property which is described in Proposition 3. We saw several examples where convexity and basic encoding schemes (derived from the Shannon capacity of the confusion graphs) are optimal. On the other hand, there are examples where much more can be gained by mixing the encoding for several users. The case in which every graph is a disjoint union of cliques is fully understood, and so is the case of 2 users and alphabet of size 3. Many other cases remain open.

We conclude with several open problems it would be interesting to solve.

- The lower and upper bounds for the maximum total rate given in this paper apply combinatorial and probabilistic techniques. It would be interesting to find stronger bounds which possibly extend the algebraic and geometric bounds known for the Shannon capacity, such as the bounds given by Lovász in [4], Hamers [3] or the first author [1].

- In the non-symmetric case where we have a user whose confusion graph is a clique over $k$ letters minus a clique over $d$ letters and the other user’s confusion graph is its complement, it would be interesting to know if Theorem 8 still holds for $d > \frac{k+1}{2}$. Since Lemma 13 does not hold for such $d$, a different approach must be used.

- Determine a graph $G$ which brings the maximum total rate to minimum when using $G$ and $\overline{G}$ as the two confusion graphs for two users. In such a case, the upper bound of the Shannon capacity of the intersection (Proposition 2) does not help as the two graphs are disjoint. However, by Theorem 1.1 of [1], we know there exist graphs $G$ on $k$ vertices for which both $c(G)$ and $c(\overline{G})$ are at most $e^{O(\sqrt{\log k \log \log k})}$. For such a graph, the upper bound in Proposition 1 of the maximum total rate yields $O(\sqrt{\log k \log \log k})$ which is far less than the trivial upper bound of $\log_2 k$.

- The encoding schemes described throughout the paper use randomness and therefore are not given explicitly. As a result, the encoding and decoding schemes are not efficient. Finding
explicit and efficient encoding and decoding schemes for the scenarios described in the paper remains open.

References


A Appendix

A.1 An example in which Lemma 13 does not hold when $d > \frac{k+1}{2}$

Let $d = 3$, $k = 4$ and $n = 2$ where indeed $d > \frac{k+1}{2} = 2.5$. Define

$$\mathcal{G} = \{(\sigma_1, \sigma_1), (\sigma_1, \sigma_2), (\sigma_1, \sigma_3), (\sigma_1, \sigma_4), (\sigma_2, \sigma_1), (\sigma_3, \sigma_1), (\sigma_4, \sigma_1)\}$$

which can also be viewed as $\{(\sigma_1, \sigma_4), (\sigma_4, \sigma_1)\}$ after replacing $\sigma_4$ with every $\sigma \in \Sigma_4$ (as $\mathcal{G}$ has to be closed under these replacements). By the definitions of the lemma,

$$\mathcal{G}' = \{(\sigma_1, \sigma_1), (\sigma_1, \sigma_2), (\sigma_1, \sigma_3), (\sigma_2, \sigma_1), (\sigma_3, \sigma_1)\}, \quad \mathcal{G}'' = \{(\sigma_3, \sigma_3), (\sigma_3, \sigma_4), (\sigma_4, \sigma_4)\}$$

and therefore the lemma does not hold as

$$\log_{k-d+1} |\mathcal{G}''| = \log_2 3 > \log_3 5 = \log_d |\mathcal{G}'|.$$  

The same example can be used with larger parameters, for instance with $k = 100$ and $d = 51$ (which is the minimal $d$ for which $d > \frac{k+1}{2}$). With these parameters, $|\mathcal{G}'| = 2d - 1 = 101$ and $|\mathcal{G}''| = 2(k - d + 1) - 1 = 99$ and indeed $\log_{k-d+1} |\mathcal{G}''| = \log_{50} 99 > \log_{51} 101 = \log_d |\mathcal{G}'|$. 