Tribes Is Hard in the Message Passing Model*

Arkadev Chattopadhyay[1] and Sagnik Mukhopadhyay[2]

1 **Tata Institute of Fundamental Research, Mumbai India**
`arkadev.c@tifr.res.in`

2 **Tata Institute of Fundamental Research, Mumbai India**
`sagnik@tifr.res.in`

Abstract

We consider the point-to-point message passing model of communication in which there are k processors with individual private inputs, each n-bit long. Each processor is located at the node of an underlying undirected graph and has access to private random coins. An edge of the graph is a private channel of communication between its endpoints. The processors have to compute a given function of all their inputs by communicating along these channels. While this model has been widely used in distributed computing, strong lower bounds on the amount of communication needed to compute simple functions have just begun to appear.

In this work, we prove a tight lower bound of $\Omega(kn)$ on the communication needed for computing the Tribes function, when the underlying graph is a star of $k+1$ nodes that has k leaves with inputs and a center with no input. A lower bound on this topology easily implies comparable bounds for others. Our lower bounds are obtained by building upon the recent information theoretic techniques of Braverman et al. ([4], FOCS'13) and combining it with the earlier work of Jayram, Kumar and Sivakumar ([10], STOC'03). This approach yields information complexity bounds that are of independent interest.

1998 ACM Subject Classification F.1.1 Models of computation, E.4.8 Coding and Information theory, F.2.2. Analysis of algorithms and Problem Complexity

Keywords and phrases communication complexity, Tribes, information complexity, direct-sum

Digital Object Identifier 10.4230/LIPIcs.STACS.2015.224

1 Introduction

The classical model of 2-party communication was introduced in the seminal work of Yao[18], motivated by problems of distributed computing. This model has proved to be of fundamental importance (see the book by Kushilevitz and Nisan [13]) and forms the core of the vibrant subject of communication complexity. It is fair to say that the wide applicability of this model to different areas of computer science cannot be over-emphasized.

However, a commonly encountered situation in distributed computing is one where there are multiple processors, each holding a private input, that are connected by an underlying communication graph. An edge of the graph corresponds to a private channel of communication between the endpoints. There are k processors located on distinct nodes of the graph that want to compute a function of their joint inputs. In such a networked scenario, a very natural question is to understand how much total communication is needed

* A. Chattopadhyay is partially supported by a Ramanujan Fellowship of the DST and S. Mukhopadhyay is supported by a TCS Fellowship.

to get the function computed. The classical 2-party model is just a special case where the graph is an edge connecting two processors.

Among others, this model has also been called the Number-in-hand multiparty point-to-point message passing model of communication. Apart from distributed computing, this model is used in secure multiparty computation. The study of the communication cost in the model was most likely introduced by Dolev and Feder [6] and further worked on by Duris and Rolim [8]. These early works focused on deterministic communication. There has been renewed interest in the model because it arguably better captures many of today's networks that is studied in various distributed models: models for map-reduce [11, 9], massively parallel model for computing conjunctive queries [3, 12], distributed models of learning [1] and in core distributed computing [7]. However, there were no known systematic techniques of proving lower bounds on the cost of randomized communication protocols that exploited the *non-broadcast* nature of the *private channels* of communication in the model. Recently, there has been a flurry of work developing new techniques for proving lower bounds on communication. Phillips, Verbin and Zhang [14] introduced the method of symmetrization to prove strong bounds for a variety of functions. Their technique was further developed in the works of Woodruff and Zhang [15, 16, 17].

All these works considered the co-ordinator model, a special case, that was introduced in the early work of [6]. In the co-ordinator model, the underlying graph has the star topology with $k+1$ nodes. There are k leaves, each holding an n-bit input. Each of the k leaf-nodes is connected to the center of the star. The node at the center has no input and is called the co-ordinator. The following two simple observations about the model will be relevant for this work: every function can be trivially computed using $O(nk)$ bits of communication by having each of the k players send their inputs to the co-ordinator who then outputs the answer. It is also easily observed that the co-ordinator model can simulate a communication protocol on an arbitrary topology having k nodes with at most a $\log k$ factor blow-up in the total communication cost.

A key lesson learnt from our experience with the classical 2-party model is that an excellent indicator of our understanding of a model is our ability to prove lower bounds for the widely known Set-Disjointness problem in the model. Indeed, as surveyed in [5], several new and fundamental lower bound techniques have emerged from efforts to prove lower bounds for this function. Further, the lower bound for Set-Disjointness, is what drives many of the applications of communication complexity to other domains. While the symmetrization technique of Phillips et.al and its refinements by Woodruff and Zhang proved several lower bounds, no strong lower bounds for Set-Disjointness were known until recently in the k-processor co-ordinator model. In this setting, the relevant definition of Set-Disjointness is the natural generalization of its 2-party definition: view the n-bit inputs of the k processors as a $k \times n$ Boolean matrix where the ith row corresponds to the Processor i's input. The Set-Disjointness function outputs 1 iff there exists a column of this matrix that has no zeroes.

In an important development, Braverman et al. [4] proved a tight $\Omega(kn)$ lower bound for Set-Disjointness in the co-ordinator model. Their approach is to build up new information complexity tools for this model that is a significant generalization of the 2-party technique of Bar-Yossef et al. [2]. In this work, we further develop this information complexity method for the co-ordinator model by considering another natural and important function, known as $\mathrm{Tribes}_{m,\ell}$. In this function, the n-bit input to each processor is grouped into m blocks, each of length ℓ. Thus, the overall $k \times n$ input matrix splits up into m sub-matrices $A_1, \ldots, A_m$, each of dimension $k \times \ell$. Tribes outputs 1 iff the Set-Disjointness function outputs 1 on each sub-matrix A_i. This obviously imparts a direct-sum flavor to the problem of determining the complexity of the Tribes function in the following sense: a naive protocol will solve Tribes by simultaneously running an optimal protocol for Set-Disjointness on each of the m instances $A_1, \ldots, A_m$. Is this strategy optimal?

This question was answered in the affirmative for the 2-party model by Jayram, Kumar and Sivakumar [10] when they proved an $\Omega(n)$ lower bound on the randomized communication complexity of the Tribes function. Their work delicately extended the information theoretic tools of Bar-Yossef et.al [2]. Interestingly, it also exhibited the power of the information complexity approach. There was no other known technique to establish a tight lower bound on the Tribes function[1].

In this work, we show that the naive strategy for solving Tribes is optimal also in the co-ordinator model:

▶ **Theorem 1.** *In the k-processor co-ordinator model, every bounded error randomized protocol solving the $\mathsf{Tribes}_{m,n}$ function, has communication cost $\Omega\big(m\ell k\big)$, for every $k \geq 2$.*

We prove this by extending and simplifying the information complexity approach of [4] and the earlier work of [10]. It is worth noting that our bounds in Theorem 1 hold for all values of k. In particular, this also yields a lower bound for Set-Disjointness for all values of k. The earlier bound of Braverman et al. only worked if $k = \Omega(\log n)$.

2 Overview & Comparison with Previous Work

We first provide a quick overview of our techniques and contributions. We follow this up with a more detailed description, elaborating on the main steps of the argument.

Brief Summary: Recall that the $\mathsf{Tribes}_{m,n}$ function can be written as an m-fold AND of Disj_n instances. One possible way to show that $\mathsf{Tribes}_{m,n}$ is hard in message-passing model is to show that any protocol evaluating $\mathsf{Tribes}_{m,n}$ must evaluate all the the Disj_n instances. This suffices to argue that $\mathsf{Tribes}_{m,n}$ is m times as hard as Disj_n. By now it is well known that information complexity provides a convenient framework to realize such direct sum arguments. In order to do so, one needs to define a distribution on inputs that is entirely supported on the ones of the m Set-Disjointness instances of Tribes. This was the general strategy of Jayram et al. [10] in the 2-party context. However, the first problem one encounters is to define an appropriate hard distribution and a right notion of information cost such that Disjointness has high information cost of $\Omega(k\ell)$ under that distribution *in the co-ordinator model.* This turns out to be a delicate and involved step. Various natural information costs do not work as observed by Phillips et al. [14]. Here, we are helped by the work of Braverman et al. [4]. They come up with an appropriate distribution τ and an information cost measure IC^0. However, we face some problems in using them. The first is that τ happens to be (almost) entirely supported on the zeroes of Set-Disjointness. Taking ideas from [10], we modify τ to get a distribution μ supported exclusively on the ones of Set-Disjointness. Roughly speaking, to sample from μ, we first sample from τ and then pick a random column of the sampled input and force it to all ones. Intuitively, the idea is that the all ones column is being well hidden at a random spot. The intuition is, if τ was hard, μ should also remain hard. It turns

[1] This is not surprising. Two other successful techniques, the discrepancy and the corruption method, both yield lower bounds on the non-deterministic complexity. On the other hand, Tribes and its complement, on n-bit inputs, both have only $\sqrt{n}$ non-deterministic complexity.

out that we prove the hardness of μ directly from scratch. To do so we appropriately modify the information cost measure IC^0 to IC so that it yields high information complexity under μ. Here, we use an idea of [10].

However, proving that IC is high for protocols when inputs are sampled according to μ raises new technical challenges. The first challenge is to prove a direct sum result on the information complexity of protocols as measured by IC. We do not know how to do that. Here we borrow ideas from [10] to introduce a new information measure, $\mathsf{PIC}(f)$ which is a lower bound on $\mathsf{IC}(f)$ and will be explained in relevant section. We show that $\mathsf{PIC}\big(\mathsf{Disj}_n\big)$ is at least $\Omega(\ell \cdot \mathsf{PIC}\big(\mathsf{Disj}_2\big))$. Implementing this step is a novelty of this work. The final challenge is to prove that $\mathsf{IC}\big(\mathsf{Disj}_2\big)$ is $\Omega(k)$. We again do that by first simplifying some of the lemmas of [4] and extending them using some ideas from the work of [10].

More Detailed Account: Among the many possible ways to define information cost of a protocol, the definition we work with stems from the inherent structure of the communication model. As evident from the previous discussion, in the model of communication we are interested in, the co-ordinator can see the whole transcript of the protocol but cannot see the inputs. On the other hand, the processors can only see a local view of the transcript - the message that is passed to them and the message they send - along with their respective inputs. From the point of view of the co-ordinator, who has no input, the information revealed by the transcript about the input can be expressed by $\mathbb{I}[X : \Pi(X)]$. This is small for the protocol where the co-ordinator goes around probing each player on each coordinate to see whether any player has 0 in it and gives up once she finds such a player.(We call it Protocol A). It is not hard to see that the information cost can only be as high as $O(n \log k)$ for protocol A. A relevant information cost measure from the point of view of processor i is $\mathbb{I}[X^{-i} : \Pi^i(X) \mid X^i]$ which measures how much information processor i learns about other inputs from the transcript. It turns out that this information cost is also very small for the protocol where all the processors send their respective inputs to the co-ordinator (We call this protocol as protocol B). Here $\mathbb{I}[X^{-i} : \Pi^i(X) \mid X^i]$ is 0 for all i. What is worth noticing is that in both protocols, if we consider the sum of the two information costs, i.e., $\mathbb{I}[X : \Pi(X)] + \sum_i \mathbb{I}[X^{-i} : \Pi^i(X) \mid X^i]$, it is $\Omega(nk)$ which is the kind of bound we are aiming for.

This cost trade-off was first observed in [14] but they were unable to prove a lower bound for Disj_n in this model of communication. Braverman et al [4] solved this problem by coming up with the following notion of information complexity. Let $(\mathbf{X}, \mathbf{M}, \mathbf{Z})$ be distributed jointly according to some distribution τ. The information cost of a protocol Π with respect to τ is defined as, $\mathsf{IC}^0_\tau(\Pi) = \sum_{i\in[k]} \Big[\underset{\tau}{\mathbb{I}}[\mathbf{X}^i : \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}] + \underset{\tau}{\mathbb{I}}[\mathbf{M} : \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z}]\Big]$.

Conditioning on the auxiliary random variables $\mathbf{M}$ and $\mathbf{Z}$ serves the following purpose: Even though the distribution τ is a non-product distribution, it can be thought of as a convex combination of product distributions, one for each specific values of $\mathbf{M}$ and $\mathbf{Z}$. It is well-known by now that such convex combination facilitates proving direct-sum like result.

The desired properties of the distribution τ are as follows. First, the distribution should have enough entropy to make it hard for the players to encode their inputs cheaply and send it across to the co-ordinator. Such an encoding is attempted in protocol B. Second, the distribution should be supported on inputs which have only a few 0's in each column of Disj_n. This makes sure that the co-ordinator has to probe $\Omega(k)$ processors in each column before he finds a 0 in that column. This attempt of probing was undertaken by the co-ordinator in protocol A. The first property can be individually satisfied by setting each processor's input to be 0 or 1 with equal probability in each column. The second property can also be individually satisfied by taking a random processor for each column and giving it a 0 and giving 1 to rest of the processors as their inputs. Let $\mathbf{Z}_j$ denote the processor whose bit was fixed to 0 in column j. The hard distribution for Disj_n is a convex combination of these two distributions. The way it is done is by setting a Bernoulli random variable $\mathbf{M}_j$ for each of the column j which acts as a switch, i.e., if $\mathbf{M}_j = 0$ the input to the column j is sampled from the first distribution, otherwise it is sampled from the second distribution. $\mathbf{M}_j$ takes value 0 with probability 2/3. We define $\mathbf{M} = \langle \mathbf{M}_1, \ldots, \mathbf{M}_\ell \rangle$ and $\mathbf{Z} = \langle \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell \rangle$.

At this point it is interesting to go back to the definition of IC^0 and try to see the implication of each term in the definition. For the coordinator, $\sum_i \underset{\tau}{\mathbb{I}}[\mathbf{X}^i : \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}]$ represents the amount of information revealed about the inputs of the processors by the transcript. For convenience, we can assume that $\mathbf{M}$ is with co-ordinator. We can do this without loss of generality as the co-ordinator can sample $O(\log k)$ inputs from column j and conclude the value of $\mathbf{M}_j$ from it, for any j. This amount of communication is okay for us as we are trying to show a lower bound of $\Omega(nk)$. However note that we cannot assume that the processors have the knowledge of $\mathbf{M}$. Had that been the situation, the processors would have employed protocol A or protocol B in column j depending on the value of the $\mathbf{M}_j$. The value of $\mathbb{I}[\mathbf{X}^i : \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}]$, in this protocol, would have been small. So we need to make sure that we charge the processors for their effort to know the value of $\mathbf{M}$. This is taken care by the second term in the definition of IC^0 i.e., $\underset{\mu}{\mathbb{I}}[\mathbf{M} : \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z}]$. Braverman et al. [4] used this notion of information complexity to achieve the $\Omega(\ell k)$ lower bound for the information cost of Disj_n with respect to the hard distribution.

As mentioned before, we, however, need the hard distribution ζ for $\mathsf{Tribes}_{m,n}$ to be entirely supported on 1s of Disj_n. But the distribution τ described above is supported on 0s of Disj_n. Here we borrow ideas from [10] and design a distribution μ by selecting a random column for the Disj_n instances and planting an all 1 input in it. We denote the random co-ordinate by $\mathbf{W}$. It is easy to verify that μ is a distribution supported in 1s of Disj_n. We set the hard distribution for $\mathsf{Tribes}_{m,n}$ to be an m-fold product distribution $\zeta = \mu^m$ denoted by the random variables $\langle \bar{\mathbf{X}}, \bar{\mathbf{M}}, \bar{\mathbf{Z}}, \bar{\mathbf{W}} \rangle$. It is to be noted that a correct protocol should work well for all inputs, not necessarily for the inputs coming from the distribution ζ. This property will be crucially used in later part of the proof. The modification of the input distribution from τ to μ and subsequently to ζ calls for changing the definition of the information complexity to suit our purpose. We define information complexity as follows which we will use in this paper.

► **Definition 2.** Let $(\bar{\mathbf{X}}, \bar{\mathbf{M}}, \bar{\mathbf{Z}}, \bar{\mathbf{W}})$ be distributed jointly according to ζ. The information cost of a protocol Π with k processors in NIH point-to-point coordinator model with respect to ζ is defined as,

$$\mathsf{IC}_\zeta(\Pi) = \sum_{i \in [k]} \left[\underset{\zeta}{\mathbb{I}}[\bar{\mathbf{X}}^i : \Pi^i(\bar{\mathbf{X}}) \mid \bar{\mathbf{M}}, \bar{\mathbf{Z}}, \bar{\mathbf{W}}] + \underset{\zeta}{\mathbb{I}}[\bar{\mathbf{M}} : \Pi^i(\bar{\mathbf{X}}) \mid \bar{\mathbf{X}}^i, \bar{\mathbf{Z}}, \bar{\mathbf{W}}] \right]. \tag{1}$$

For a function $f : \mathbf{X} \to \mathcal{R}$, the information complexity of the function is defined as, $\mathsf{IC}_{\zeta,\delta}(f) = \inf_\Pi \mathsf{IC}_\mu(\Pi)$, where the infimum is taken over all δ-error protocol Π for f.

By doing this, we are able to bound the information complexity of $\mathsf{Tribes}_{m,n}$ as m-times that of Disj_n. Although non-trivial, this step can be accomplished by exploiting the proof techniques used in [4]. The next step is to bound the information complexity of Disj_n, which turns out to be difficult for two reasons. First, the distribution μ is no more a 0 distribution for Disj_n. We get around this by defining a new information complexity measure, - which we call as partial information complexity - to show that the partial information complexity of Disj_n on distribution μ is at least $(\ell-1)$-times that of Disj_2. This is one of the main technical contributions of our paper. See Section 4.1 for details. The second hurdle we face is bounding the information complexity of Disj_2. Here we combine ideas from [10, 4] to conclude that the partial information complexity of Disj_2 is $\Omega(k)$. This is the second main technical contribution of this paper, which is explained in Section 4.2. Finally we give a simple argument in Section 5 to show that $\mathsf{IC}_\zeta(\Pi)$ lower bounds the communication cost of Π where Π is any correct protocol for $\mathsf{Tribes}_{m,n}$.

3 Preliminaries

Communication complexity. In this work, we are mainly interested in multiparty communication *number-in-hand* model. In this model of computation, the input is distributed between k players $P_1, \cdots, P_k$ who jointly wish to compute a function f on the combined input by communication with each other.

We work with randomized protocol where the players have access to private coins. (Though it might seem like that the public coin protocol can yield better upper bound, it can be noted that all the proofs can be modified to give the same result for public coin model.) The standard notion of private coin randomized communication complexity is adopted here, where we look at the worst-case communication of the protocol when the protocol is allowed to make only δ error (bounded away from 1/2) on each input. Here the probability is taken over the private coin tosses of the players. For more details, readers are referred to [13].

Information theory. We will quickly go through the information theoretic definitions and facts we need. For a random variable X taking value in the sample space Ω according to the distribution $p(\dot{})$, the entropy of X, denoted as $\mathcal{H}(X)$, is defined as $\mathcal{H}(X) = \sum_{x\in\Omega} \Pr[X = x] \log \frac{1}{\Pr[X=x]} = \mathbf{E}_x \left[\log \frac{1}{p(x)}\right]$.

For two random variables X and Y, the conditional entropy of X given Y is defined as $\mathcal{H}(X|Y) = \mathbf{E}_{x,y} \left[\log \frac{1}{p(x|y)}\right]$.

Informally, the entropy of a random variable measures the uncertainty associated with it. Conditioning on another random variable, i.e., knowing the value that another random variable takes can only decrease the uncertainty of the former one. This notion is captured in the following fact that $\mathcal{H}(X|Y) \le \mathcal{H}(X)$ where the equality is achieved when X is independent of Y. Given two random variables X and Y with joint distribution $p(x, y)$ we can talk about how much information one random variable reveals about the other random variable. The mutual information, as it is called, between X and Y is defined as $\mathbb{I}[X : Y] = \mathcal{H}(X) - \mathcal{H}(X|Y)$.

It is to be noted that the mutual information is a symmetric quantity, though it might not be obvious from the definition itself. From the previous discussion, it is easy to see that the mutual information is a non-negative quantity. As before, we can also define conditional mutual information as $\mathbb{I}[X : Y|Z] = \mathcal{H}(X|Z) - \mathcal{H}(X|Y, Z)$.

The following chain rule of mutual information will be crucially used in our proof.

$$\mathbb{I}[X_1, \ldots, X_n : Y] = \sum_{i\in[n]} \mathbb{I}[X_i : Y|X_{i-1}, \ldots, X_1]. \tag{2}$$

It is to be noted that the chain rule of mutual information will also work when conditioned on random variable Z.

▶ Remark. Consider a permutation $\sigma : [n] \to [n]$. The following observation will be useful in our proof.

$$\mathbb{I}[X_1, \dots, X_n : Y] = \sum_{i \in [n]} \mathbb{I}[X_{\sigma(i)} : Y | X_{\sigma(i-1)}, \dots, X_{\sigma(1)}]. \tag{3}$$

We will use the following lemma regarding mutual information.

▶ **Lemma 3.** *Consider random variables A, B, C and D. If A is independent of B given D then,*

$$\mathbb{I}[A : B, C \mid D] = \mathbb{I}[A : C \mid B, D], \tag{4}$$

and

$$\mathbb{I}[A : C \mid B, D] \geq \mathbb{I}[A : C \mid D]. \tag{5}$$

4 Lower Bound for $\mathsf{Tribes}_{m,n}$ in Message Passing Model

Here, in the first subsection, we will show two direct-sum results. In the first step we bound the information complexity of $\mathsf{Tribes}_{m,n}$ in terms of that of Disj_n. It is to be noted that the proof technique of [2] falls short of proving any lower bound on the information complexity measure we have defined - mainly because of the fact the information complexity measure consists of sum two different mutual information terms for each processor, and it is not clear that one can come up with lower bounds for both the terms simultaneously. This problem has already been attended to in [4] and the proof we present here resembles the proof technique used by them. For completeness we include the proof in this paper. In the second step, we will bound the information complexity of Disj_n in terms of Disj_2. This step is more difficult and a straight-forward application of the direct-sum argument of [4] will not work. First we use ideas from [10] to define partial information complexity measure which is more convenient to work with. Then we come up with a novel direct-sum argument for partial information complexity measure.

In Section 4.2, we show that the information complexity of Disj_2 is $\Omega(k)$. We manage to show this by combining ideas from [4, 10].

4.1 Direct Sum

In this section we prove that the information cost of computing $\mathsf{Tribes}_{m,n}$ is m times the information cost of computing Disj_n. The proof is almost the same proof as in [4] where the authors have used a direct sum theorem to show that the information cost of computing Disj_n is ℓ times the information cost of computing k-bit AND_k. Before going into details we need the following definitions which we will borrow from [10].

Consider $f : \mathcal{D}^m \to \mathcal{R}$ can be written as $f(X) = g(h(X_1), \dots, h(X_m))$ where $X = \langle X_1, \dots, X_m \rangle$, $X_i \in \mathcal{D}$ and $h : \mathcal{D} \to \mathcal{R}$. In other words, f is g-decomposable with primitive h.

▶ **Definition 4** (Collapsing distribution)**.** We call $X \in \mathcal{D}^m$ be a collapsing input for f if for any $i \in [m]$ and $y \in \mathcal{D}$, we have $f(X(i, y)) = h(y)$. Any distribution ζ supported entirely on collapsing inputs on f is called a collapsing distribution of f.

▶ **Definition 5** (Projection)**.** Given a distribution ν specified by random variable $(D_1, \dots, D_k)$ and a subset S of $[k]$, we call the projection of ν on $(D_i)_{i \in S}$, denoted as $\nu \downarrow_{(D_i)_{i \in S}}$, the marginal distribution of $(D_i)_{i \in S}$ induced by ν.

The proof is by reduction: we will show that given a protocol Π for $\mathsf{Tribes}_{m,n}$ and a collapsing distribution $\mu = \zeta_l^m$, we can construct a protocol Π' for Disj_n such that it computes Disj_n with the same error probability as that of Π and the information complexity of Π is m times that of Disj_n.

▶ **Theorem 6.** *Let $\mu = \zeta_\ell^m$ be a collapsing distribution for $\mathsf{Tribes}_{m,n}$ partitioned by $\mathbf{M}, \mathbf{Z}$ and $\mathbf{W}$ as described before. Then*

$$\mathsf{IC}_\mu(\mathsf{Tribes}_{m,n}) \geq m.\mathsf{IC}_{\zeta_\ell}(\mathsf{Disj}_n). \tag{6}$$

As mentioned before, the proof of Theorem 6 works out nicely by adapting the proof techniques of [4] and is omitted in this version.

Now our goal is to connect the information cost of Disj_n under ζ_ℓ to information cost of AND_k. So a natural attempt is to prove a theorem like Theorem 6 for reduction from Disj_n to AND_k. Unfortunately this is not possible. Recall that $\mathsf{Disj}_n(X) = \bigvee_{i=1}^{\ell} \bigwedge_{j=1}^{k} X_i^j$. Hence for a collapsing distribution each of the AND_ks should evaluate to 0, which is not the case for the distribution ζ_ℓ.

Inspired by [10], we define the following measure of information cost, namely, partial information cost. Let Π be a protocol for Disj_n. The partial information cost of Π is defined as,

$$\mathsf{PIC}(\Pi) = \sum_{i=1}^{k} \left(\mathbb{I}\left[\mathbf{M}_{-\mathbf{W}} : \Pi^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{Z}, \mathbf{W}\right] + \mathbb{I}\left[\mathbf{X}^i_{-\mathbf{W}} : \Pi^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{Z}, \mathbf{W}\right] \right). \tag{7}$$

The random variable $\mathbf{M}_{-\mathbf{W}}$ denotes $\mathbf{M}$ with its $\mathbf{W}$-th coordinate removed. Similarly, $\mathbf{X}^i_{-\mathbf{W}}$ denotes $\mathbf{X}^i$ with its $\mathbf{W}$-th coordinate removed. The partial information complexity of Disj_n is the partial information cost of the best protocol computing Disj_n. It is easy to see that the partial information complexity of any function f lower bounds the information complexity of f.

We prove the following theorem.

▶ **Theorem 7.** *Let ζ_ℓ be the distribution over the inputs of Disj_n partitioned by $\mathbf{M}, \mathbf{Z}, \mathbf{W}$ as described before. Then*

$$\mathsf{PIC}_{\zeta_\ell}(\mathsf{Disj}_n) \geq (\ell - 1).\mathsf{PIC}_{\zeta_2}(\mathsf{Disj}_2). \tag{8}$$

Here we will show the following reduction analogous to our previous reduction from $\mathsf{Tribes}_{m,n}$ to Disj_n. Given a protocol Π' for Disj_n and distribution ζ_ℓ (as described in Section 2, we will come up with a protocol Π'' for Disj_2 such that the partial information cost of Π'' w.r.t. ζ_ℓ is $1/(\ell-1)$ times the partial information cost of Π' w.r.t ζ_2.

Let us describe the construction of the protocol Π''. On an input $u = \langle u_1, u_2 \rangle$ for Disj_2, the processors and the coordinator sample a $k \times \ell$ random matrix $\mathbf{X}(u)$ in the following way.

1. The coordinator samples $\mathbf{P}$ and $\mathbf{Q}$ uniformly at random from $[\ell]$ such that $\mathbf{P} < \mathbf{Q}$.
2. The coordinator samples $\mathbf{Z}_{-\{\mathbf{P},\mathbf{Q}\}} = (\mathbf{Z}_i)_{i \in [\ell] \setminus \mathbf{P},\mathbf{Q}}$, where each $\mathbf{Z}_i \in_R [k]$, and sends it to all the processors.
3. The coordinator samples a number $\mathbf{R}$ uniformly at random from $\{0, ..., \ell-2\}$ and then samples a subset $\mathbf{T} \subseteq [\ell] \setminus \{\mathbf{P}, \mathbf{Q}\}$ uniformly at random from all sets of size $\mathbf{R}$ that do not contain $\mathbf{P}, \mathbf{Q}$. Then the coordinator samples $\mathbf{M}_t \sim \mathsf{Bin}(1/3)$ for all $t \in \mathbf{T}$ and sends them to all the processors. The processors use their private randomness to sample $\mathbf{X}_t$ for each

column t in $\mathbf{T}$ in the following way: The input of the $\mathbf{Z}_t$-th processor is fixed to 0 in $\mathbf{X}_t$ and the other processors get 1 if $\mathbf{M}_t = 1$, otherwise, if $\mathbf{M}_t = 0$, they get 0 or 1 uniformly at random. We will call this input sampling procedure as IpSample.

4. For the rest of the columns, the coordinator samples the inputs according to IpSample and sends the requisite inputs to the respective processors.
5. The processors form the input $\mathbf{X} \equiv \mathbf{X}(u, \mathbf{P}, \mathbf{Q})$ (i.e., $\mathbf{X}_\mathbf{P} = u_1$ and $\mathbf{X}_\mathbf{Q} = u_2$) and run the protocol Π' for Disj_n with $\mathbf{X}$ as input.

▶ **Observation 8.** *Consider the tuple* $(\mathbf{U}, \mathbf{N}, \mathbf{V}, \mathbf{S})$ *distributed according to* ζ_2*. If* $\mathbf{U}$ *is given as input to protocol* Π''*, then* $(\mathbf{X}, \mathbf{M}, \mathbf{Z}, \mathbf{W})$ *is distributed according to* ζ_ℓ*, where* $\mathbf{W}$ *is the unique all 1's coordinate in* $\mathbf{X}$*. Here* $\mathbf{W} = \mathbf{P}$ *if* $\mathbf{V} = 1$ *and* $\mathbf{W} = \mathbf{Q}$ *if* $\mathbf{V} = 2$.

Next we prove the following lemma connecting the information cost of Π' for Disj_n and that of Π'' for Disj_2. This lemma implies the Theorem 7.

▶ **Lemma 9.**

$$\mathop{\mathbb{I}}_{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2}[\mathbf{U}^i_{-\mathbf{V}}, \Pi''^i(\mathbf{U}) \mid \mathbf{N}, \mathbf{V}, \mathbf{S}] \le \frac{1}{\ell - 1} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{W},\mathbf{Z})\sim\zeta_\ell}[\mathbf{X}^i_{-\mathbf{W}}, \Pi'^i(\mathbf{X}) \mid \mathbf{M}, \mathbf{W}, \mathbf{Z}], \quad (9)$$

and

$$\mathop{\mathbb{I}}_{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2}[\mathbf{N}_{-\mathbf{V}}, \Pi''^i(\mathbf{U}) \mid \mathbf{U}^i, \mathbf{V}, \mathbf{S}] \le \frac{1}{\ell - 1} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{W},\mathbf{Z})\sim\zeta_\ell}[\mathbf{M}_{-\mathbf{W}}, \Pi'^i(\mathbf{X}) \mid \mathbf{X}^i, \mathbf{W}, \mathbf{Z}]. \quad (10)$$

Proof. We consider the LHS of Equation (10). The view of processor i of the transcript of protocol Π'', denoted as $\Pi''^i(\mathbf{U})$, is given as follows.

$$\Pi''^i(\mathbf{U}) = \langle \mathbf{P}, \mathbf{Q}, \mathbf{Z}_{-\mathbf{P},\mathbf{Q}}, \mathbf{R}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{X}^i_{\bar{\mathbf{T}}\setminus\{\mathbf{P},\mathbf{Q}\}}, \Pi'(\mathbf{X}(\mathbf{P}, \mathbf{Q}, \mathbf{U})))\rangle. \quad (11)$$

So the LHS of Equation (10) can be written as

$$\begin{aligned}
&\mathop{\mathbb{I}}_{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2}[\mathbf{N}_{-\mathbf{V}} : \Pi''^i(\mathbf{U}) \mid \mathbf{U}^i, \mathbf{V}, \mathbf{S}] \\
&= \mathop{\mathbb{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2 \\ (\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}}} [\mathbf{N}_{-\mathbf{V}} : \mathbf{P}, \mathbf{Q}, \mathbf{Z}_{-\mathbf{P},\mathbf{Q}}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{R}, \mathbf{X}^i_{\bar{\mathbf{T}}\setminus\{\mathbf{P},\mathbf{Q}\}}, \Pi'^i(\mathbf{X}(\mathbf{P}, \mathbf{Q}, \mathbf{U})) | \mathbf{U}^i, \mathbf{V}, \mathbf{S}] \\
&= \mathop{\mathbb{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2 \\ (\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}}} [\mathbf{N}_{-\mathbf{V}} : \Pi'^i(\mathbf{X}(\mathbf{P}, \mathbf{Q}, \mathbf{U})) | \mathbf{P}, \mathbf{Q}, \mathbf{Z}_{-\mathbf{P},\mathbf{Q}}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{R}, \mathbf{X}^i_{\bar{\mathbf{T}}\setminus\{\mathbf{P},\mathbf{Q}\}}, \mathbf{U}^i, \mathbf{V}, \mathbf{S}] \\
&\hspace{30em}\text{[Lemma 3 eqn. (4)]} \\
&= \mathop{\mathbb{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2 \\ (\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}}} [\mathbf{N}_{-\mathbf{V}} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{Z}, \mathbf{V}, \mathbf{X}^i_{\bar{\mathbf{T}}}] \\
&\hspace{20em}\text{[Combining } (\mathbf{U}^i, \mathbf{X}^i_{\bar{\mathbf{T}}\setminus\{\mathbf{P},\mathbf{Q}\}} \text{ and } (\mathbf{Z}_{-\mathbf{P},\mathbf{Q}}, \mathbf{S})] \\
&\le \mathop{\mathbb{I}}_{\substack{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2 \\ (\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}}} [\mathbf{N}_{-\mathbf{V}} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{Z}, \mathbf{V}, \mathbf{X}^i] \\
&\hspace{22em}\text{[Lemma 3 eqn. (5), } \mathbf{X}^i_\mathbf{S} \text{ ind. of } \mathbf{N}_{-\mathbf{V}}]
\end{aligned}$$

[V takes value in 1 and 2 uniformly at random. Hence we can write it as follows.]

$$\begin{aligned}
&= \frac{1}{2} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}} [\mathbf{M}_\mathbf{P} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i] \\
&\quad + \frac{1}{2} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}} [\mathbf{M}_\mathbf{Q} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M}_\mathbf{T}, \mathbf{Z}, \mathbf{V} = 1, \mathbf{X}^i].
\end{aligned} \quad (12)$$

Consider the first mutual information term.

$$\begin{aligned}
&\mathbb{I}[\mathbf{M_P} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M_T}, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i] \\
&= \frac{2}{\ell(\ell-1)} \sum_{p<q} \mathbb{I}[\mathbf{M}_p : \Pi'^i(\mathbf{X}) \mid \mathbf{P} = p, \mathbf{Q} = q, \mathbf{R}, \mathbf{T}, \mathbf{M_T}, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i] \\
&= \frac{2}{\ell(\ell-1)^2} \sum_{p<q} \sum_{r=0}^{\ell-2} \sum_{t:|t|=r} \Pr[\mathbf{T} = t] \mathbb{I}[\mathbf{M}_p : \Pi'^i(\mathbf{X}) \mid p, q, r, t, \mathbf{M}_t, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i] \\
&= \frac{2}{\ell(\ell-1)^2} \sum_{p<q} \sum_{r=0}^{\ell-2} \sum_{t:|t|=r} \frac{(\ell-r-2)!r!}{(\ell-2)!} \mathbb{I}[\mathbf{M}_p : \Pi'^i(\mathbf{X}) \mid p, q, r, t, \mathbf{M}_t, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i].
\end{aligned}$$

We can safely drop the conditioning $\mathbf{P} = p, \mathbf{R} = r, \mathbf{T} = t$ and $\mathbf{V} = 2$ in the following way. It is easy to see $\mathbf{R} = r, \mathbf{T} = t$ is implied by $\mathbf{M}_t$. $\mathbf{M}_p$ implies $\mathbf{P} = p$. Moreover, given (p, q), $\mathbf{V} = 2$ is equivalent to $\mathbf{W} = p$. So we can write,

$$\begin{aligned}
&\mathbb{I}[\mathbf{M_P} : \Pi'^i(\mathbf{X}) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M_T}, \mathbf{Z}, \mathbf{V} = 2, \mathbf{X}^i] \\
&= \frac{2}{(\ell-1)!\ell(\ell-1)} \sum_{q} \sum_{p:p<q} \sum_{r=0}^{\ell-2} \sum_{t:|t|=r} ((\ell-r-2)!r!) \mathbb{I}[\mathbf{M}_p : \Pi'^i(X) \mid \mathbf{W} = q, \mathbf{M}_t, \mathbf{Z}, \mathbf{X}^i].
\end{aligned} \tag{13}$$

Similarly, the second mutual information term of Equation (12) term can be written in the following way.

$$\begin{aligned}
&\mathbb{I}[\mathbf{M_Q} : \Pi'^i(X) \mid \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{T}, \mathbf{M_T}, \mathbf{Z}, \mathbf{V} = 1, \mathbf{X}^i] \\
&= \frac{2}{(\ell-1)!\ell(\ell-1)} \sum_{q} \sum_{p:p<q} \sum_{r=0}^{\ell-2} \sum_{t:|t|=r} ((\ell-r-2)!r!) \mathbb{I}[\mathbf{M}_q : \Pi'^i(\mathbf{X}) \mid \mathbf{W} = p, \mathbf{M}_t, \mathbf{Z}, \mathbf{X}^i].
\end{aligned} \tag{14}$$

Combining Equation (12), (13), (14), we get,

$$\begin{aligned}
&\mathbb{I}_{(\mathbf{U},\mathbf{N},\mathbf{V},\mathbf{S})\sim\zeta_2}[\mathbf{N}_{-\mathbf{V}}, \Pi'^i(\mathbf{U}) \mid \mathbf{U}^i, \mathbf{V}, \mathbf{S}] \\
&\leq \frac{1}{(\ell-1)!\ell(\ell-1)} \sum_{q'} \sum_{p':p'\neq q'} \sum_{r=0}^{\ell-2} \sum_{t:|t|=r} ((\ell-r-2)!r!) \mathbb{I}[\mathbf{M}_{p'} : \Pi'^i(X) | \mathbf{W} = q', \mathbf{M}_t, \mathbf{Z}, \mathbf{X}^i]
\end{aligned}$$

The number of permutations of $[\ell]\backslash q$ where the $r+1$th element is p' and the first r elements constitute the set t is $(\ell-r-2)!r!$. Hence we can write the previous summation as follows,

$$\begin{aligned}
&= \frac{1}{(\ell-1)!\ell(\ell-1)} \sum_{q'} \sum_{\sigma\in\mathcal{S}_{[\ell]\backslash q'}} \sum_{i\in[\ell]\backslash q'} \mathbb{I}[\mathbf{M}_{\sigma(i)} : \Pi'^i(\mathbf{X}) \mid \mathbf{M}_{\{\sigma(1),\ldots,\sigma(i-1)\}}, \mathbf{Z}, \mathbf{W} = q', \mathbf{X}^i] \\
&= \frac{1}{(\ell-1)!\ell(\ell-1)} \sum_{q'} \sum_{\sigma\in\mathcal{S}_{[\ell]\backslash q'}} \mathbb{I}[\mathbf{M}_{-q'} : \Pi'^i(\mathbf{X}) \mid \mathbf{Z}, \mathbf{W} = q', \mathbf{X}^i] \\
&\qquad\qquad \text{[Using chain rule of information, Eq. (3)]} \\
&= \frac{1}{\ell-1} \sum_{q'} \frac{1}{\ell} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z})\sim\zeta_\ell\downarrow\mathbf{X},\mathbf{M},\mathbf{Z}} [\mathbf{M}_{-q'} : \Pi'^i(\mathbf{X}) \mid \mathbf{Z}, \mathbf{W} = q', \mathbf{X}^i] \\
&= \frac{1}{\ell-1} \mathop{\mathbb{I}}_{(\mathbf{X},\mathbf{M},\mathbf{Z},\mathbf{W})\sim\zeta_\ell} [\mathbf{M}_{-\mathbf{W}} : \Pi'^i(\mathbf{X}) \mid \mathbf{Z}, \mathbf{W}, \mathbf{X}^i].
\end{aligned} \tag{15}$$

Equation (9) can be proved in the similar way and therefore omitted. ◀

4.2 Lower Bounding Disj_2

In this section we prove the following.

▶ **Theorem 10.** $\mathsf{IC}(\mathsf{Disj}_2) \geq \mathsf{PIC}(\mathsf{Disj}_2) \geq \Omega(1)$.

This, combined with Theorem 7 and Theorem 6 will imply a $\Omega(m\ell k)$ lower bound on the switched information complexity of $\mathsf{Tribes}_{m,n}$ which is the lower bound on $R_\delta(\mathsf{Tribes}_{m,n})$ we aimed for.

Notation. By $\bar{e}$ we mean the all 1 vector of size k. By $\bar{e}_{i,j}$, we mean the boolean vector of size k where all entries are 1 except the entries in index i and j. Similarly, $\bar{e}_i$ is the boolean vector where all entries are 1 except that of index i. $\Pi[i, x, m, z; \bar{e}_i]$ implies the transcript of the protocol Π on the following Disj_2 instance: the input of the first column comes from the distribution specified by $\mathbf{M} = m$, $\mathbf{Z} = z$ and $\mathbf{X}_1^i = x$ and the input of the second column is $\bar{e}_i$. Abusing notation slightly, $\Pi^i[x, m, z; \bar{e}_i]$ represents processor-i's view of the transcript $\Pi[i, x, m, z; \bar{e}_i]$.

Hellinger distance. For probability distributions P and Q supported on a sample space Ω, the Hellinger distance between P and Q, denoted as $h(P, Q)$, is defined as, $h(P, Q) = \frac{1}{\sqrt{2}}||\sqrt{P} - \sqrt{Q}||_2 = 1 - F(P, Q)$, where $F(P, Q) = \sum_{\omega\in\Omega} \sqrt{P(\omega)Q(\omega)}$ is also known as Bhattacharya coefficient. Below we will state a fact (without proof) about Hellinger distance.

▶ **Fact 11** ([2]). *Let Π be a δ-error protocol for function f. For inputs x and y such that $f(x) \neq f(y)$, we have,*

$$h(\Pi(x), \Pi(y)) \geq \frac{1-\delta}{\sqrt{2}}. \tag{16}$$

The following lemmas are generalization of their two-party analogues.

▶ **Lemma 12** (k-party cut-paste). *For any randomized protocol Π computing $f : X^k \to \{0, 1\}$ and for any $x, y \in X^k$ and for some i and j,*

$$h(\Pi(x_i x_j x_{-i,j}, y_i y_j y_{-i,j})) = h(\Pi(x_i y_j x_{-i,j}, y_i x_j y_{-i,j})). \tag{17}$$

▶ **Lemma 13** (Pythagorean). *For any randomized protocol Π and for any input $x, y \in X^k$ and for some i and j,*

$$2h^2(\Pi(x_i x_j x_{-i,j}, y_i y_j y_{-i,j})) \geq h^2(\Pi(x_i x_j x_{-i,j}, x_i y_j y_{-i,j})) + h^2(\Pi(y_i x_j x_{-i,j}, y_i y_j y_{-i,j})). \tag{18}$$

Following structural properties are generalizations of analogous properties shown in [4]. Simpler proofs will be included in full version.

▶ **Lemma 14** (Diagonal). *For $i \neq j$*

$$h^2(\Pi^i[0, 0, j; \bar{e}], \Pi^i[1, 1, z; \bar{e}]) \geq \frac{1}{2} h^2(\Pi^i(\bar{e}_{i,j}; \bar{e}), \Pi^i(\bar{e}_j; \bar{e})). \tag{19}$$

▶ **Lemma 15** (Global-to-local). *For $i \neq j$*

$$h^2(\Pi[i, 0, 0, z; \bar{e}], \Pi[i, 1, 0, z; \bar{e}]) = h(\Pi^i[0, 0, z; \bar{e}], \Pi^i[1, 0, z; \bar{e}]), \tag{20}$$

and

$$h(\Pi(\bar{e}_{i,j}; \bar{e}), \Pi(\bar{e}_i; \bar{e}) = h(\Pi^i(\bar{e}_{i,j}; \bar{e}), \Pi^i(\bar{e}_i; \bar{e})). \tag{21}$$

Now we are ready to prove the partial information cost of Disj_2 is $\Omega(k)$. We consider processor i and fix a value $j \neq i$.

▶ **Claim 16** ([4]).

$$(1)\ \mathbb{I}[\mathbf{M}_{-\mathbf{W}} : \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = j, \mathbf{W} = 2] \geq \frac{2}{3} h^2(\Pi^i[1,0,j;\bar{e}], \Pi^i[1,1,j;\bar{e}]), \tag{22}$$

$$(2)\ \mathbb{I}[\mathbf{X}^i_{-\mathbf{W}} : \Pi^i \mid \mathbf{M}, \mathbf{Z} = j, \mathbf{W} = 2] \geq \frac{2}{3} h^2(\Pi^i[0,0,j;\bar{e}], \Pi^i[1,0,j;\bar{e}]), \tag{23}$$

$$(3)\ \mathbb{I}[\mathbf{M}_{-\mathbf{W}} : \Pi^i \mid \mathbf{X}^i, \mathbf{Z} = j, \mathbf{W} = 1] \geq \frac{2}{3} h^2(\Pi^i[\bar{e};1,0,j], \Pi^i[\bar{e};1,1,j]), \tag{24}$$

$$(4)\ \mathbb{I}[\mathbf{X}^i_{-\mathbf{W}} : \Pi^i \mid \mathbf{M}, \mathbf{Z} = j, \mathbf{W} = 1] \geq \frac{2}{3} h^2(\Pi^i[\bar{e};0,0,j], \Pi^i[\bar{e};1,0,j]). \tag{25}$$

Using Cauchy-Schwarz and triangle inequality, we can write the following.

$$\begin{aligned}
&\sum_i \mathbb{I}[\mathbf{M}_{-W} : \Pi^i \mid \mathbf{X}^i, \mathbf{Z}, \mathbf{W}] + \mathbb{I}[\mathbf{X}^i_{-\mathbf{W}}; \Pi^i \mid \mathbf{M}, \mathbf{Z}, \mathbf{W}] \\
&\quad \geq \frac{1}{3k} \sum_i \sum_{j:i \neq j} \left[h^2(\Pi^i[1,1,j;\bar{e}], \Pi^i[0,0,j;\bar{e}]) + h^2(\Pi^i[\bar{e};1,1,j], \Pi^i[\bar{e};0,0,j])\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad [\text{Claim } 16] \\
&\quad \geq \frac{1}{6k} \sum_i \sum_{j:i \neq j} \left[h^2(\Pi^i(\bar{e}_{i,j}.\bar{e}), \Pi^i(\bar{e}_j.\bar{e})) + h^2(\Pi^i(\bar{e}\bar{e}_{i,j}), \Pi^i(\bar{e}\bar{e}_j)\right]. \qquad [\text{Lemma } 14] \\
&\quad \geq \frac{1}{6k} \sum_i \sum_{j:i \neq j} \left[h^2(\Pi(\bar{e}_{i,j}.\bar{e}), \Pi(\bar{e}_j.\bar{e})) + h^2(\Pi(\bar{e}\bar{e}_{i,j}), \Pi(\bar{e}\bar{e}_j))\right] \qquad [\text{Lemma } 15] \\
&\quad \geq \frac{1}{24k} \sum_{i \neq j} \Big[[h^2(\Pi(\bar{e}_i.\bar{e}), \Pi(\bar{e}_j.\bar{e}))] + [h^2(\Pi(\bar{e}\bar{e}_i), \Pi(\bar{e}\bar{e}_j))]\Big] \quad [\text{Recounting \& Tr. ineq.}] \\
&\quad = \frac{1}{24k} \sum_{i \neq j} \Big[[h^2(\Pi(\bar{e}.\bar{e}), \Pi(\bar{e}_{i,j}.\bar{e}))] + [h^2(\Pi(\bar{e}\bar{e}), \Pi(\bar{e}\bar{e}_{i,j}))]\Big] \qquad [\text{Lemma } 12] \\
&\quad \geq \frac{1}{48k} \sum_{i \neq j} [h^2(\Pi(\bar{e}.\bar{e}_{i,j}), \Pi(\bar{e}_{i,j}.\bar{e}))] \qquad [\text{Cauchy-Schwarz \& triangle inequality}] \\
&\quad \geq \frac{1}{96k} \sum_{i \neq j} \Big[[h^2(\Pi(\bar{e}.\bar{e}_{i,j}), \Pi(\bar{e}_j.\bar{e}_i)) + h^2(\Pi(\bar{e}_{i,j}\bar{e}), \Pi(\bar{e}_i.\bar{e}_j))]\Big] \qquad [\text{Lemma } 13] \\
&\quad = \frac{k-1}{384}(1-\delta)^2 = \Omega(k). \qquad [\text{Fact } 11\ ]
\end{aligned}$$

5 Putting Everything Together

In this section we show randomized communication complexity of any function f is lower bounded by the information complexity of f.

▶ **Theorem 17.** *For any distribution μ over the inputs,*

$$R_\epsilon(\mathsf{Tribes}_{m,n}) = \Omega(\mathsf{IC}_\mu(\mathsf{Tribes}_{m,n})). \tag{26}$$

This follows from the fact that the expected length of any instantaneous q-ary code for a random variable X is at least $\mathcal{H}(X)/\log q$. We omit the proof for space constraint. Using Theorem 6, 7, 10 and 17, it is not hard to see that Theorem 1 follows.

References

1 Maria-Florina Balcan, Avrim Blum, Shai Fine, and Yishay Mansour. Distributed learning, communication complexity and privacy. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland*, pages 26.1–26.22. JMLR.org, 2012.

2 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, pages 209–218. IEEE Computer Society, 2002.

3 Paul Beame, Paraschos Koutris, and Dan Suciu. Communication steps for parallel query processing. In Richard Hull and Wenfei Fan, editors, *Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA - June 22 - 27, 2013*, pages 273–284. ACM, 2013.

4 Mark Braverman, Faith Ellen, Rotem Oshman, Toniann Pitassi, and Vinod Vaikuntanathan. A tight bound for set disjointness in the message-passing model. In *FOCS*, pages 668–677. IEEE Computer Society, 2013.

5 Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *SIGACT News*, 41(3):59–85, 2010.

6 Danny Dolev and Tomás Feder. Determinism vs. nondeterminism in multiparty communication complexity. *SIAM J. Comput.*, 21(5):889–895, 1992.

7 Andrew Drucker, Fabian Kuhn, and Rotem Oshman. The communication complexity of distributed task allocation. In Darek Kowalski and Alessandro Panconesi, editors, *ACM Symposium on Principles of Distributed Computing, PODC '12, Funchal, Madeira, Portugal, July 16-18, 2012*, pages 67–76. ACM, 2012.

8 Pavol Duris and José D. P. Rolim. Lower bounds on the multiparty communication complexity. *J. Comput. Syst. Sci.*, 56(1):90–95, 1998.

9 Michael T. Goodrich, Nodari Sitchinava, and Qin Zhang. Sorting, searching, and simulation in the mapreduce framework. In Takao Asano, Shin-Ichi Nakano, Yoshio Okamoto, and Osamu Watanabe, editors, *Algorithms and Computation - 22nd International Symposium, ISAAC 2011, Yokohama, Japan, December 5-8, 2011. Proceedings*, volume 7074 of *Lecture Notes in Computer Science*, pages 374–383. Springer, 2011.

10 T. S. Jayram, Ravi Kumar, and D. Sivakumar. Two applications of information complexity. In Lawrence L. Larmore and Michel X. Goemans, editors, *STOC*, pages 673–682. ACM, 2003.

11 Howard J. Karloff, Siddharth Suri, and Sergei Vassilvitskii. A model of computation for mapreduce. In Moses Charikar, editor, *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 938–948. SIAM, 2010.

12 Paraschos Koutris and Dan Suciu. Parallel evaluation of conjunctive queries. In Maurizio Lenzerini and Thomas Schwentick, editors, *Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2011, June 12-16, 2011, Athens, Greece*, pages 223–234. ACM, 2011.

13 Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

14 Jeff M. Phillips, Elad Verbin, and Qin Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 486–501. SIAM, 2012.

15 David P. Woodruff and Qin Zhang. Tight bounds for distributed functional monitoring. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on*

Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012, pages 941–960. ACM, 2012.

16 David P. Woodruff and Qin Zhang. When distributed computation is communication expensive. In Yehuda Afek, editor, *Distributed Computing - 27th International Symposium, DISC 2013, Jerusalem, Israel, October 14-18, 2013. Proceedings*, volume 8205 of *Lecture Notes in Computer Science*, pages 16–30. Springer, 2013.

17 David P. Woodruff and Qin Zhang. An optimal lower bound for distinct elements in the message passing model. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 718–733. SIAM, 2014.

18 Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In Michael J. Fischer, Richard A. DeMillo, Nancy A. Lynch, Walter A. Burkhard, and Alfred V. Aho, editors, *Proceedings of the 11h Annual ACM Symposium on Theory of Computing, April 30 - May 2, 1979, Atlanta, Georgia, USA*, pages 209–213. ACM, 1979.