Electronic Colloquium on Computational Complexity, Report No. 74 (2014)

Topology Matters in Communication

Arkadev Chattopadhyay * Jaikumar Radhakrishnan † Atri Rudra ‡

May 14, 2014

Abstract

We provide the first communication lower bounds that are sensitive to the network topology for computing natural and simple functions by point to point message passing protocols for the 'Number in Hand' model. All previous lower bounds were either for the broadcast model or assumed full connectivity of the network. As a special case, we obtain bounds of the form $\Omega(k^2n)$ on the randomized communication complexity of many simple functions for a broad class of networks having k distributed nodes and each holding an n-bit input string. The best previous bounds were of the form $\Omega(kn)$. The main tool that we use for deriving our bounds is a new connection with the theory of metric embeddings. This enables us to prove a variety of results that include the following: A distributed XOR lemma; a tight bound (discarding poly-log factors) on the randomized complexity of Element Distinctness that answers a question of Phillips, Verbin and Zhang (SODA'12, [PVZ12]) and new lower bounds for composed functions that were also left open in the work of Phillips et al. [PVZ12]. Finally, these bounds yield new topology-dependent bounds for several natural graph problems considered by Woodruff and Zhang (DISC'13, [WZ13]).

*School of Technology and Computer Science, Tata Institute of Fundamental Research, email: arkadev.c@tifr.res.in.

†School of Technology and Computer Science, Tata Institute of Fundamental Research, email: jaikumar.radhakrishnan@gmail.com.

‡Department of Computer Science and Engineering, University at Buffalo, SUNY, email: atri@buffalo.edu.
Research supported in part by NSF grant CCF-0844796.

ISSN 1433-8092

1 Introduction

Multi-party communication complexity was introduced in the work of Chandra, Furst and Lipton [CFL83] where k players have inputs $X_1, \ldots, X_k \in \{0,1\}^n$ and the k players want to compute some common boolean function $f : (\{0,1\}^n)^k \to \{0,1\}$ with the goal of minimizing the total communication between the k players. We assume that each player can only look at her own input, i.e. we follow the so called Number in Hand (NIH) model.[1] The NIH multi-party model seems to have been first considered by Dolev and Feder [DF89]. The case for $k = 2$ is the standard two-party communication complexity introduced by Yao [Yao79]. Both two party and multi-party communication complexity has numerous applications: see e.g. the excellent book on this topic [KN97].

The generalization to the multi-party communication complexity model has to decide on various modes of communication:

1. Whether the communication is broadcast (i.e. everyone sees message sent by a player) or point to point (messages have a single sender and single receiver);

2. If the communication is point to point, how are the player communication channels connected, i.e. what is the structure of the underlying graph topology G?

For various reasons, the original model was with broadcast communication except for the early work of Duris and Rolim [DR98] who proved lower bounds on the deterministic and non-deterministic communication complexity in the point to point model over the complete graph. Recently, there has been a surge of interest in the point to point model [PVZ12, WZ12, WZ13, BEO$^+$13]. This is because the point to point model arguably better captures many of the modern day networks and has been studied in the many distributed models: e.g. the BSP model of Valiant [Val90], models for MapReduce [KSV10, GSZ11], massively parallel models to compute conjunctive queries [BKS13, KS11], distributed models for learning [BBFM12, IPSV12] and in the core distributed computing literature [DKO12]. The recent surge in interest in this model is also in part motivated by proving lower bounds for the distributed functional monitoring framework (see e.g. the recent survey [Cor13]), which generalizes the popular data streaming model [Mut05]. However, all of the recent work assumes that the underlying topology is fully connected[2]. In our opinion this is a strict restriction since in many situations assuming full connectivity would be too strong an assumption. Indeed in areas like sensor networks, researchers have considered the effects of network topology with some success [HK12] for simple topologies like trees.

The following is the motivating question for our work (which was also mentioned as an interesting direction to pursue in [BEO$^+$13]):

[1] The paper of Chandra, Furst and Lipton [CFL83] considered the model where each player gets to see everyone else's input (Number on Forehead or NOF model). To the best of our knowledge, the first paper with non-trivial randomized lower bounds in the NIH model was the work of Alon, Matias and Szegedy [AMS99].

[2] It is worthwhile to note that the effect of network topology on the cost of communication has been analyzed to quite an extent when the networks are *dynamic* in the context of distributed computing (see for example the recent survey of Kuhn and Oshman [KO11]). In this work in contrast, we are mainly concerned with static networks of arbitrary topology as embodied in the NIH model.

Can we prove lower bounds on multi-party communication complexity for the NIH point to point communication that are sensitive to the topology of the connections between the player?

To see how the network topology can make a big difference in the total communication cost, let us consider the trivial algorithm that can compute any function f: all players send their input to one designated player. If the topology G is the complete graph (or has constant diameter), then the trivial algorithm uses up $O(kn)$ communication. Now consider the case when G is the line graph. In this case it is best for all players to send their input to the "middle" node. However, note that in this case the total communication is $\Omega(k^2n)$.[3] (For general graphs, the total communication is bounded by the objective function of the 1-median problem where the distances are the shortest path distance in G.) Thus, ignoring the topology (as the current works do) could result in bounds that are sub-optimal by a $\Theta(k)$ factor. Our interest is in identifying situations where we can recover this extra $\Theta(k)$ factor in our lower bounds and in general match the bound of the trivial algorithm for any topology.

Our Contributions. Our main contribution is the first set of lower bounds for the NIH point to point communication model that are sensitive to the network topology.[4] We present a general framework to prove lower bounds for general topologies. Our framework is able to generalize many of the existing lower bound results for the complete graph topology and uses a new connection to the theory of metric embeddings. To the best of our knowledge this is the first work to apply results from metric embeddings to prove lower bounds on communication complexity. We would like to clarify that while none of our proofs are technically difficult by themselves, most of the tools we use are quite non-trivial. We believe our main contribution is more conceptual: we identify certain key components and show how to combine them to obtain topology dependent lower bounds. We also believe that our framework is fairly general and should be widely applicable. As a partial justification of this belief, we extend many known results on the complete graph to topology dependent, essentially tight, lower bounds for general graphs.

A natural function to start proving strong lower bounds is for the set disjointness problem (i.e. we want to compute $\vee_{i=1}^{n}\left(\wedge_{j=1}^{k}X_j(i)\right)$). Set disjointness is the "canonical" problem for two party communication complexity whose hardness implies lower bounds for myriads of problems in diverse models (see for example the survey [CP10]). It was also recently shown by Braverman et al. [BEO$^+$13] that for the k-party set disjointness problem on the complete graph, the total communication is $\Omega(kn)$. However, it is not too hard to see that for *any* topology, the intersection of the k sets (and in particular the set disjointness problem) as well as the union of the k sets can be computed with $O(kn)$ total communication.[5] This implies

[3] The two end point players have a total communication of kn, the next pair has total communication of $(k-2)n$ and so on.

[4] We note here that 2-party communication complexity lower bounds easily prove lower bounds of the form $\Omega(d \cdot n)$, where d is the diameter of G. In this work, we present bounds of the form $\Omega(kdn)$ for some situations.

[5] Consider a spanning tree of the underlying graph G and compute the intersection/union as one goes through all the nodes in the spanning tree say in pre-order traversal. It is easy to check that the total communication over each edge is $O(n)$ and that each edge needs to communicate only twice.

that existing reductions (e.g. those in [WZ13] for graph problems) from set disjointness (and related problems) cannot be used to prove topology dependent better lower bounds.

Thus, we need a problem where the players do need to send all their information to one player. Towards this end, consider the following problem that we call Element-Distinctness: the players want to decide if $X_i \neq X_j$ for every $i \neq j \in [k]$. If we allow randomization, the trivial algorithm on the line graph takes $\tilde{O}(k^2)$ amounts of communication.[6] In this case it does seem that all the pairs need to be compared and hence it seems like that the trivial algorithm is indeed optimal. We show that this is indeed the case for the Element-Distinctness problem as well as a bunch of other problems for all graph topology.

Our Results. We show that for all of the following problems, the trivial algorithm is optimal (up to $\operatorname{poly}\log k$ factors unless mentioned otherwise) for any network topology:

- Element-Distinctness: Output 1 if and only if $X_i \neq X_j$ (as vectors) for every $i \neq j \in [k]$. This answers a question of Philips et al. [PVZ12] who asked for the communication complexity of this function just for the case of G being a complete graph. In fact, [PVZ12] considered Element-Distinctness to be a variant of k-Equality in which players output 1 if and only if $X_i = X_j$ for every $i \neq j \in [k]$. They seemed to suggest that Element-Distinctness and Equality have the same complexity. We show that while for the complete graph they indeed have the same complexity, for general topologies the complexities of the two problems are entirely different.

- We prove the following XOR lemma. Consider any partition of $[k]$ for even k into two disjoint set S and $\bar{S}$ and a bijection ρ between S and $\bar{S}$. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Then computing the function XOR-$f \equiv \oplus_{i \in S} f(X_i, X_{\rho(i)})$ cannot be done better than the following trivial algorithm up to $\tilde{O}(\sqrt{k})$ factor: every pair $(i, \rho(i))$ for $i \in S$ computes $f(X_i, X_{\rho(i)})$ using the best two party communication protocol for f and then, say, players in S compute the final output bit. For certain functions we can improve the gap from $\tilde{O}(\sqrt{k})$ to just $\operatorname{poly}\log k$ factors. This extends the XOR lemma of Barak et al. [BBCR10] from the 2-party setting to the general multi-party setting[7]. XOR lemmas are of general interest in computer science.

- It is natural to consider what happens if one replaces the XOR function with OR or AND. Woodruff and Zhang [WZ14] showed that OR-f has communication complexity $\Omega\big(kR_\epsilon(f)\big)$ for the complete graph. While our techniques recover this result for complete graphs, we observe that such a generic OR/AND-lemma cannot have a topology dependent extension to general graphs. However, for some specific functions of natural interest, we prove topology dependent tight bounds. These include OR-Equality, OR-Disjointness and AND-Disjointness, also known as the tribes function. Besides being interesting by themselves, these results are also useful in proving lower bounds for several graph problems described next.

[6] There is no linear dependence on n since, the parties can just send fingerprints of their input to the designated player.

[7] We use the result of Barak et al. to prove ours.

- We extend the lower bounds on graph problems considered in [WZ13] to the general topology case. In particular, in these problems the k players get k subgraphs of some graph H (edges can be duplicated) and the players want to solve one of the following five problems: determining (i) the degree of a vertex in H, (ii) if H is acyclic; (iii) if H is triangle-free; (iv) if H is bipartite and (v) if H is connected. In all these cases we show that the trivial algorithm of all players sending their subgraphs to a designated player is the best possible up to $\operatorname{poly}\log k$ factors. Our reductions for (ii)-(v) are different from those in [WZ13] as the hard problem in [WZ13] can be solved with $O(kn)$ communication for any topology.

In Section 4.5, we present some other results on composed functions that showcase the generality of our techniques. While composed functions are a natural and important class of functions that have been widely studied in communication complexity, their study in the context of point to point communication model was suggested in the recent work of Phillips et al. [PVZ12].

Our Techniques. We now present an overview of our proof techniques. As mentioned earlier, each of our steps is technically simple and it is the combination of non-trivial results that seems crucial to prove our stronger results. We believe that the technical simplicity provides for easy and wide applicability of our techniques to many problems. Later, we will also show that our techniques generalize most of the existing techniques used to prove lower bounds in the special case when G is completely connected.

As usual for proving randomized lower bounds, we will prove a distributional lower bound for the problem. One way such lower bounds for 2-party problems are obtained is by proving a discrepancy/corruption bound for two dimensional rectangles/sub-matrices. For k players this would generalize to analyzing k-dimensional tensors that seem very challenging for large k. A more tractable option seems to try reducing the k-player problem to a hard 2-player problem by finding a convenient cut in the graph, where we give inputs on each side of the cut to a specific player. There are several obvious difficulties that come up when trying to pursue this option. We next sketch them and broadly describe how we get around these difficulties.

Note that we cannot work with just a single cut to get better than $\Omega(kn)$ bounds. This is because each player in the reduced 2-party problem across such a cut gets $O(kn)$-length inputs. Thus, we have to work with a family of cuts such that across each (or most) of them we have a fairly hard 2-party problem. Optimistically, one may hope then that these complexities can be added up to take us beyond the kn barrier. There are two things to take care of immediately before one can try implementing this idea. The global distribution on k players' inputs have to be chosen such that across every (or at least most) cut it becomes a hard distribution for the 2-party instance. Even if that happens, why are we allowed to add the distributional complexities of the various problems across cuts? Usually, the (μ, ϵ)-distributional complexity of a function f, denoted by $D_{\mu,\epsilon}(f)$, is defined as the *worst-case cost* of the best deterministic protocol that errs with probability ϵ when inputs are sampled from μ. But then, we cannot add the costs of these various problems across cuts because the worst-case of each individual problem may not give rise to a globally consistent input.

We get around this problem by considering the notion of expected cost for the 2-party

problem. Using linearity of expectation, one can now add the costs of the various 2-party problems. However, for technical reasons, we have to consider ϵ-error randomized expected cost wrt μ, i.e. a protocol that like a true randomized protocol errs with small error on *every* input but we measure its cost *only* w.r.t a distribution μ. This is a simple but subtle fix to the problem.

To illustrate our idea with a concrete and simple example, let us consider the Element-Distinctness problem on the line graph. Consider the following distribution on $X_1, \ldots, X_k$: randomly pick them to be k distinct values. Note that by linearity of expectation, the total expected communication is the sum of the expected communication over each edge. Consider any edge e. Let us assume e is such that there are $i \leq k/2$ players to the "left" of e and $k - i$ players to the "right" of e. Then note that any ϵ-error protocol for Element-Distinctness is solving the set disjointness problem among the sets $\{X_1, \ldots, X_i\}$ and $\{X_{i+1}, \ldots, X_k\}$ for every input with high probability. Ignoring the size of the domain of these values, this implies an $\Omega(i)$ lower bound on the communication on e. This is because our initial distribution is chosen so that the induced distribution for the 2-party set disjointness problem is such that every ϵ-error protocol has high expected cost w.r.t this distribution. Now just summing up the expected cost on each edge gives us an $\Omega(k^2)$ lower bound on the total expected communication, which was our aim.

The above argument crucially used the fact that the topology is a line graph. In particular, in the above argument when we considered an edge e, we basically used the fact that this induces a *cut* on the players (which in turn induces a two-party set disjointness problem). The more crucial aspect that might have been swept under the rug was the following fact: if one considers the set of all $k - 1$ cuts, then each pair (i, j) is cut exactly $|i - j|$ times, which is the same as the distance between player i and j on the line graph. Moreover, each edge e appears in precisely one cut which ensures that the summing up of expected costs is a valid counting of the expected total cost.

It turns out that for general graphs, we just need to find a set of cuts that has the property that every pair of players is separated as many times as (up to some slack) the shortest path distance between them in G. Further, to generalize the sum of expectation argument, we also need to ensure that every edge in G is not separated by many cuts. This is where the theory of metric embeddings plays a role. It turns out that one can find such cuts by known results on embedding metrics into ℓ_1 metric. (For those unfamiliar with metric embeddings, the connection is not that surprising since embeddings into ℓ_1 and cuts have a very close relationship.) In fact, for technical reasons we need the embedding to have a third property but that is also satisfied by known embeddings (e.g. the one due to Bourgain).

Once we have the above cut technology in place, we then need to select a global distribution of inputs such that the corresponding 2-party problems across cuts are hard wrt the expected cost measure over induced distributions across cuts. In most cases, we are able to appeal to a known 2-party result to finish off the argument. For instance, in the case of Element-Distinctness, the corresponding 2-party problem is k-set disjointness. For the XOR lemma, the induced problem exactly is the 2-party XOR problem lemma and we apply the result of Barak et al. [BBCR13].

Connections to Related Work. Finally, we put our techniques in the context of existing techniques used to argue lower bounds for the case when G is fully connected. In particular, we will argue that our techniques essentially generalize many of the existing techniques.

The first lower bounds for the message-passing NIH model seems to be due to Duris and Rolim [DR98]. They also considered the complete graph topology (co-ordinator model) and their bounds were for deterministic and non-deterministic complexity. In particular it uses a generalization of 2-party fooling set argument that does not seem to apply to bounded error randomized protocols. Very recently, the symmetrization technique was introduced by Phillips et al. [PVZ12] and was further developed by Woodruff and Zhang [WZ12, WZ13, WZ14]. At a very high level, the core idea in symmetrization is as follows. First we consider the case when G is a star graph of diameter 2 with a co-ordinator node at the center. Prototypical hard problems to consider are functions of the form $\vee_{i=1}^{k} f(X_i, Y)$, where the center gets Y and the k leaves of G get $X_1, \ldots, X_k$. If ν is a hard distribution for (the 2-party) function f, then the trick is to define a hard distribution μ on $X_1, \ldots, X_k, Y$ such that for every $i \in [k]$ the effective distribution on (X_i, Y) is ν. Then the argument, slightly re-phrased in our language, proceeds as follows: pick a random cut among the k cuts corresponding to the k edges. Then by definition of μ the induced 2-party problem across each cut is f and hence, the communication complexity is $\Omega(R(f))$, where $R(f)$ is the randomized two-party communication complexity of f. Then we note that since the cut was picked completely at random, and the distribution μ is symmetric with respect to the leaf-nodes, the communication across such a random cut in expectation is $\Theta(1/k)$ of the total communication, which leads to an overall $\Omega(k \cdot R(f))$ lower bound on the total communication. By contrast, our technique does not need this symmetric property though our use of linearity of expectation seems similar. Indeed in Section 4.4, we show how to recover the lower bound on the OR of f from [WZ13] using our techniques. Note that the cuts in a star-graph are all similar, as all leaves are symmetric with respect to the prototypical example. As identified by the authors [PVZ12] themselves, this property seems to be lost even for star graphs when the inputs held by leaf-nodes are not symmetric with respect to the function that players want to compute. For general graph topology, there might be very little symmetry left. In particular, in our technique, the cuts obtained are arbitrary with no guarantee of symmetry. Nevertheless, our technique seems flexible enough to handle such cases.

One technique that we cannot (yet) handle with ours is the result of Braverman et al. [BEO$^+$13] that proves a lower bound of $\Omega(kn)$ on the set disjointness problem. It is an interesting open question to see if we can port the techniques of Braverman et al. to our setting.

2 Preliminaries

Let f be any k-variate boolean function of the form $f : (\{0,1\}^n)^k \to \{0,1\}$ where each input X_i takes value in $\{0,1\}^n$. Let $G \equiv \big(V, E\big)$ be a graph with k vertices, i.e. $V = \{1, \ldots, k\}$. In the message passing game on the graph G for function f, there are k players of unbounded computational power. Player i is at vertex i of G and has access to only input X_i. This distribution of inputs is often called the 'Number in Hand' (NIH) model[8]. The players want

[8]There is another important multiparty communication model called the Number on Forehead model where Player i sees every input *except* X_i. We do not consider this model at all in this work.

to compute f collaboratively according to a unanimously agreed upon communication protocol according to which players send and receive messages to and from other players. In any such protocol, Player i is allowed to communicate with Player j if and only if their vertices are neighbors in the graph G. Like in the standard two party communication game, what (and to whom) Player i communicates at any round, only depends on input X_i and the messages received by Player i from other players until that round. Further, in contrast to the *broadcast* model, the message sent by Player i to j is only received by Player j and no other player. At termination of the protocol, each player should know the value $f(X_1, \ldots, X_k)$. The cost of an execution of the protocol is the total number of bits communicated on all the edges in all the rounds. Just as in standard two-party communication complexity, the protocols can be deterministic or randomized. All randomized protocols considered in this paper are public in the sense that players without communication share all public coin tosses. This is the most powerful model of randomness and thus lower bounds for this model imply lower bounds for weaker models.

We also need two notions of the cost of a protocol, the worst-case and the average-case/expected cost with respect to a distribution over the input. For any fixed $\epsilon < 1/2$, a randomized protocol makes ϵ error if on every input the probability (over the random coin tosses) of the protocol giving the wrong answer is at most ϵ. The worst-case cost of such a protocol Π, over the coin tosses and the inputs, is denoted by $\mathrm{Cost}\big(\Pi\big)$. The randomized ϵ-error message passing complexity of a function f for a graph G, denoted by $R_{\epsilon,G}\big(f\big)$, is the worst-case cost of the best ϵ-error protocol. Protocols with $\epsilon = 0$ are identified by a special term, called zero-error protocols. The zero-error complexity of a function f, denoted by $R_0\big(f\big)$, is the worst-case expected cost of the best randomized protocol computing f with no error[9].

Given a distribution μ over $(\{0,1\}^n)^k$, the expected cost of a protocol Π, denoted by $\mathrm{ECost}_\mu\big(\Pi\big)$ is the expectation of its cost over both the internal random coin tosses of the protocol and the distribution μ. The μ-expected ϵ-error complexity of a function f over G, denoted by $ER_{\mu,\epsilon,G}\big(f\big)$, is the expected cost of the best ϵ-error protocol over graph G for computing f. Naturally, $R_{\mu,0}\big(f\big)$, denotes the μ-expected zero-error complexity of f.

Let $G \equiv \big(V, E\big)$ be a graph. A cut C is a partition of its set of vertices, V, into two parts A, B. A pair of vertices $u, v \in V$ are separated by cut C if they lie in two different parts of the cut. The set of all pairs of vertices separated by C is denoted by $M(C)$. An edge in E is a cut-edge if its endpoints are separated by the cut. The set of cut-edges of C is denoted by $E(C)$. Given vertices u, v in graph G, we will use $d_G(u, v)$ (or just $d(u, v)$ when G is clear from the context) to denote the length of the shortest path between u and v in G. Throughout this paper, the underlying network graph G will be a connected graph.

Let f be any k-party problem associated with the graph G, where $k = |V(G)|$ and μ a distribution on the inputs to f. For any edge $e \in E(G)$, let $\mathrm{ECost}_\mu\big(\Pi, e\big)$ denote the expected total number of bits sent over e in both directions. Then, for any protocol Π and cut C of G, let $\mathrm{ECost}_\mu\big(\Pi, C\big)$ denote the expected total communication across C: $\mathrm{ECost}_\mu\big(\Pi, C\big) \equiv \sum_{e \in E(C)} \mathrm{ECost}_\mu\big(\Pi, e\big)$. Let $\mathcal{C} \equiv \{C_1, \ldots, C_t\}$ be a set of cuts. Define the expected cost of Π over $\mathcal{C}$ as $\mathrm{ECost}_\mu\big(\Pi, \mathcal{C}\big) \equiv \sum_{i=1}^{t} \mathrm{ECost}_\mu\big(\Pi, C_i\big)$.

We state below a simple but useful consequence of the linearity of expectation.

[9]Note here the expectation is only over the internal coin tosses of a protocol.

Observation 2.1 *Let $\mathcal{C}$ be a set of cuts of G such that any edge e of G appears as a cut edge in at most m cuts in $\mathcal{C}$. Then, $ECost_{\mu}(\Pi) \geq \frac{ECost_{\mu}(\Pi,\mathcal{C})}{m}$.*

We will need the following results from basic 2-party communication complexity where the graph of communication is just an edge connecting Player 1 and 2, often called Alice and Bob respectively. In general, k-party Set-Disjointness, denoted by k-DISJ, is defined as the following function: there is some universe $[N]$ and Player i gets a subset X_i of the universe. The function outputs 1 if and only if there is no element that appears in each X_i. The game in which the players are promised that the input sets further satisfy the condition $|X_i| = \ell$ and there is at most one element that is common to all X_i, is denoted by k-UDISJ$_\ell$. Let $\mu[\ell]$ be the distribution defined in the following way[10] on the inputs of $2-$UDISJ$_\ell$: with probability $3/4$, you sample uniformly a pair of sets from the space of all pairs of non-intersecting sets, each of size ℓ, and with probability $1/4$ you sample uniformly a pair of sets that intersect precisely at one element. For notational convenience, we will often drop ℓ from $\mu[\ell]$ when the context makes the value of ℓ clear.

Theorem 2.2 (Razborov[Raz92]) *There exists some universal constants δ, β such that $D_{\mu,\delta}(2-UDISJ_\ell)$ is $\Omega(\beta\ell)$, provided the size of the universe is at least $4\ell+1$.*

The following result will be useful for us for proving lower bounds.

Lemma 2.3 *Let ν be a distribution on the inputs of f, where f is any 2-party function. Let ν_0 (ν_1) be the marginal distribution on the zeroes (ones) of f induced by ν. Then, $ER_{\nu_0,\epsilon'}(f) \geq (\epsilon - \epsilon') \cdot D_{\nu,\epsilon}(f)$, where we have assumed $\epsilon' < \epsilon$.*

Proof: Assume that there is an ϵ'-erring randomized protocol Π with expected cost w.r.t ν_0 (ν_1) being c, where $\epsilon' < \epsilon$. Consider the following new protocol Π': let $\epsilon_d = \epsilon - \epsilon'$. Π' runs Π until c/ϵ_d bits have been communicated or Π has halted. If Π has halted, Π' outputs the answer of Π. Otherwise, Π' halts and outputs 1 (0).

We claim the following is true: $\Pr_{r\sim R, x\sim\nu}\left[\Pi'(x,r) \neq f(x)\right] \leq \epsilon$ Let x be a zero (one) of f. Conditioned on this, x is being sampled from ν_0 (ν_1). Hence, applying Markov's inequality, with probability less than ϵ_d , Π does not output an answer within communicating c/ϵ_d bits. Hence,

$$\Pr_{r\sim R, x\sim\nu}\left[\Pi'(x,r) = 1 | f(x) = 0\right] < \epsilon_d.$$

Now consider the other case, where x is a one (zero) of f. Note that for *every* input Π makes error with probability at most ϵ'. However, when x is a one (zero) of f, Π' does not make an error if Π did not. Thus,

$$\Pr_{r\sim R, x\sim\nu}\left[\Pi'(x,r) = 1 \wedge f(x) = 1\right] < \epsilon'.$$

[10] Razborov [Raz92] describes this distribution in another equivalent way that is more convenient for his analysis.

Combining these two cases immediately gives us our claim (2) The worst-case cost of Π' is at most c/ϵ_d. By fixing the randomness r of this Π', we get a deterministic protocol Π'' of cost at most c/ϵ_c and that errs w.r.t ν at most with probability ϵ. Thus, $c \geq \epsilon_d \cdot D_{\nu,\epsilon}\big(f\big)$. $\square$

Combining Theorem 2.2 with Lemma 2.3, the following corollary easily follows: let $\mu_0[\ell]$ ($\mu_1[\ell]$) be the uniform distribution on pairs of disjoint (uniquely-intersecting) sets, each of size ℓ. When the context makes it clear, we drop ℓ from the notation.

Corollary 2.4 *For each fixed $\epsilon < 1/2$, there exists β such that $ER_{\mu_0,\epsilon}\big(2 - UDISJ_\ell\big)$ and $ER_{\mu_1,\epsilon}\big(2 - UDISJ_\ell\big)$ are both at least $\beta.\ell$, if the size of the universe is at least $4\ell + 1$.*

We derive the following direct but useful consequence of Lemma 2.3:

Corollary 2.5 *Let ν be a distribution on the inputs of f, where f is any 2-party function. Then, $ER_{\nu,\epsilon'}\big(f\big) \geq \big(\epsilon - \epsilon'\big) \cdot D_{\nu,\epsilon}\big(f\big)$, where we have assumed $\epsilon' < \epsilon$.*

Proof: Using Lemma 2.3, we know that for any ϵ'-error protocol Π for f, $\mathrm{ECost}_{\nu_i}\big(\Pi\big) \geq (\epsilon - \epsilon')D_{\nu,\epsilon}\big(f\big)$, for $i = \{0,1\}$, where ν_i is the marginal of ν supported on points at which f evaluates to i. The corollary follows. $\square$

Another function that is classical in communication complexity is Equality. We consider its natural k-party version, denoted by $k-\mathrm{EQ}$, which outputs 1 if and only if all of its k-many n-bit input strings are equal. While EQ is relatively easy for bounded-error protocols, the cost of zero-error protocols is large under the following distribution: let $S \subseteq \{0,1\}^n$ let $\mathcal{U}^k_{=,S}$ be the uniform distribution on k tuples of equal strings from S. When $S = \{0,1\}^n$, we drop S from subscript of $\mathcal{U}$. Whenever the value of k becomes obvious from the context, we drop the superscript of $\mathcal{U}$. We outline a proof of the following classical result for the sake of completeness.

Theorem 2.6 $R_{\mathcal{U}_{=,S}}\big(2\text{-}EQ\big) = \Omega\big(\log|S|\big)$.

Proof: This uses a simple fooling set argument. Let $\mu \equiv \mathcal{U}_{=,S}$. Let Π be a zero-error randomized protocol. Then, there exists a fixing a of the random coins of Π such that the deterministic protocol Π^a has worst-case cost at most $\mathrm{ECost}_\mu\big(\Pi\big)$. Since Π was a zero-error protocol for every input, Π^a makes no errors as well. Hence, a standard fooling set argument shows, for every input in the support of μ, Π^a must generate a unique transcript. Thus, the length of one of those transcripts must be at least $\log|S|$. $\square$

2.1 Information Theoretic Techniques

Information theory techniques have been increasingly used to prove lower bounds in communication complexity (see for example [BYJKS04, JKS03, BBCR13, BEO$^+$13]). We will also use some of these techniques here. We quickly recall the basic notions needed here. Let X, Y, Z be discrete random variables taking values in some (discrete) set $\mathcal{D}$. Then, the entropy of X, denoted by $H(X)$, is defined as follows:

$$H(X) = -\sum_{x \in \mathcal{D}} \Pr\big[X = x\big] \log\big(\Pr[X = x]\big).$$

Informally, the entropy of X measures the uncertainty associated with it. Given two random variables X, Y, knowing the value of one may reduce the uncertainty associated with the other. More formally, the conditional entropy $H\big(X \,|\, Y\big)$, is defined as follows: $H\big(X \,|\, Y\big) = \mathbb{E}_y H\big(X \,|\, Y = y\big)$, where $H\big(X \,|\, Y = y\big)$ is the entropy of the conditional distribution of X given $Y = y$. It can be shown that $H\big(X \,|\, Y\big) \leq H\big(X\big)$. The mutual information between X, Y, denoted by $I\big(X \,;\, Y\big)$, is defined as follows:

$$I\big(X \,;\, Y\big) = H\big(X\big) - H\big(X \,|\, Y\big).$$

It is a non-trivial and useful property that mutual information is non-negative and symmetric. One can also define the conditional mutual information $I\big(X \,;\, Y \,|\, Z\big)$ as follows:

$$I\big(X \,;\, Y \,|\, Z\big) = H\big(X \,|\, Z\big) - H\big(X \,|\, YZ\big).$$

Given a 2-party randomized communication protocol Π and some input distribution μ, its external μ-information cost, denoted by $\mathrm{IC}_\mu\big(\Pi\big)$, is defined as follows:

$$\mathrm{IC}_\mu\big(\Pi\big) = I_{(X,Y)\sim\mu}\big(X, Y \,;\, \Pi(X,Y)\big),$$

where $\Pi(X, Y)$ is the random transcript of the protocol.

Remark 1 *The transcript of an execution of a protocol will contain the concatenation of messages sent by each player along with the public random coin tosses of the protocol.*

The ϵ-error (external) information complexity of a function f wrt distribution μ, denoted by $\mathrm{IC}_{\mu,\epsilon}\big(f\big)$, is the μ-information cost of the best ϵ-error protocol for f. An input distribution μ is called product if it can be decomposed as the product of one distribution μ_X on Alice's input and that of another one μ_Y on Bob's input, i.e. $\mu(X, Y) = \mu_X(X)\mu_Y(Y)$. Product distributions are convenient to analyze. However, the kind of distribution μ that we will need to analyze will not always be product, but a convex combination of product distributions. Such combinations are convenient to express in terms of an auxiliary random variable D. In particular, μ may be a non-product distribution for (X, Y, D). However, the conditional distribution $\big(X, Y \,|\, D = d\big)$ will be product for every d. Towards analyzing the cost of protocols wrt such distributions, we need the slightly more general notion of conditional information cost of a protocol Π, denoted by $\mathrm{CIC}_\mu\big(\Pi\big)$, and defined as follows:

$$\mathrm{CIC}_\mu\big(\Pi\big) = I_\mu\big(X, Y \,;\, \Pi(X,Y) \,|\, D\big).$$

This gives rise, just as in the case of information cost, to the notion of the ϵ-error conditional information complexity of a function f wrt μ, denoted by $\mathrm{CIC}_{\mu,\epsilon}\big(f\big)$.

We will need a relationship to be established between the information complexity of a function and its expected bounded error randomized complexity. Towards that, let us recall a useful inequality that lower bounds the compressibility of a random variable by its entropy. The proof of this can be found in any standard text on information theory like [CT91].

Theorem 2.7 (Theorem 5.3.1 in [CT91]) *The expected length L of any instantaneous q-ary code for a random variable X satisfies the following:*

$$L \geq \frac{1}{\log q} H(X).$$

We are now ready to make the connection between the two notions of complexity of a function:

Theorem 2.8 *For any distribution μ over the inputs to a function f, and $\epsilon < 1$, the following is satisfied:*

$$ER_{\mu,\epsilon}\big(f\big) = \Omega\Big(IC_{\mu,\epsilon}\big(f\big)\Big).$$

Proof: For any 2-party protocol Π, let us write its transcript as $\Pi_{-}\big(X,Y\big)R$, where R is the public coin tosses of Π and $\Pi_{-}\big(X,Y\big)$ are the concatenation of messages sent by the protocol. Note that we may assume that either the protocol uses a prefix-free code over the binary alphabet or it uses a special character to delimit the messages sent by players to each other, making the encoding of transcript prefix-free over the alphabet of size 3. Using the chain rule of information, the information cost of Π can be re-written as follows:

$$\mathrm{IC}_{\mu}\big(\Pi\big) = I_{\mu}\big(X,Y\,;\,R\big) \,+\, I_{\mu}\big(X,Y\,;\,\Pi_{-}(X,Y)\,|\,R\big).$$

But, $I_{\mu}\big(X,Y\,;\,R\big) = 0$ as the public coin tosses are independent of X,Y. Hence, expanding the conditional information term, we have

$$\mathrm{IC}_{\mu}\big(\Pi\big) = \mathbb{E}_{R}\big[\, I_{\mu}\big(X,Y\,;\,\Pi_{-}(X,Y)\,|\,R=r\big)\big].$$

However, invoking Theorem 2.7, we get

$$\mathrm{ECost}_{\mu}\big(\Pi\,|\,R=r\big) \geq \frac{1}{\log 3} H\big(\Pi_{-}\,|\,R=r\big) \geq \frac{1}{\log 3} I_{\mu}\big(X,Y\,;\,\Pi_{-}(X,Y)\,|\,R=r\big).$$

Now the claimed bound of our theorem easily follows. □

Now we state some results about the information complexity of functions which we will use. Let (U,V,W) be a triple of random variables sampled from $\{0,1\}^3$ as follows: sample W uniformly at random from $\{0,1\}$. If $W=0$, fix $U=0$ and sample V at random from $\{0,1\}$. If $W=1$, fix $V=0$ and sample U at random. Call this distribution τ. Note that the conditional distribution $\big(U,V\,|\,W=i\big)$, denoted by τ_i, for any $i \in \{0,1\}$ is product. Let ν be the marginal distribution of (U,V). Let $(X,Y,D) \sim \tau^n \equiv_{\mathrm{def}} \eta$. Let μ be the marginal of η on (X,Y), which is the same as ν^n. Then, Bar-Yossef et al. obtained the following remarkable result:

Theorem 2.9 (Bar-Yossef et al. [BYJKS04])

$$IC_{\mu,\epsilon}\big(UDISJ_n\big) \geq CIC_{\eta,\epsilon}\big(UDISJ_n\big) \geq \frac{n}{4}\Big(1-2\sqrt{\epsilon}\Big).$$

The following simple corollary will be useful for us.

Corollary 2.10

$$ER_{\mu,\epsilon}\big(UDISJ_n\big) \geq \frac{n}{4}\bigg(1 - 2\sqrt{\epsilon}\bigg).$$

Proof: Easily follows by combining Theorem 2.9 and Theorem 2.8. □

We next consider another important function in communication complexity. This is the tribes function, defined as follows:

$$\text{TRIBES}_{m,n}\big(\overline{X}, \overline{Y}\big) \equiv_{\text{def}} \bigwedge_{i=1}^{m} \text{DISJ}_n\big(X_i, Y_i\big),$$

where $\overline{X} = (X_1, \ldots, X_m)$, $\overline{Y} = (Y_1, \ldots, Y_m)$ and each $X_i, Y_i \in \{0,1\}^n$. In a 2-party game for tribes, Alice gets $\overline{X}$ and Bob gets $\overline{Y}$. Let $\overline{S} = (S_1, \ldots, S_m)$ with each $S_i \in [n]$ and let $\overline{D} = (D_1, \ldots, D_m)$ with each $D_i \in \{0,1\}^n$. Each (X_i, Y_i, S_i, D_i) are i.i.d. random variables sampled from a distribution γ. We sample (U, V, S, D) from γ as follows: sample S uniformly at random from $[n]$ and D at random from $\{0,1\}^n$. For every $j \leq n$, do the following: if $j \neq S$, then sample $\big(U_j, V_j\big)$ from τ_ℓ where[11] $\ell = D_j$. If $j = S$, then set $\big(U_j, V_j\big) = (1,1)$. Note that the conditional distribution $\big(U, V \,|\, S = s, D = d\big)$ is product. The common marginal distribution for each $\big(X_i, Y_i\big)$ is denoted by ρ. This implies that $\big(\overline{X}, \overline{Y}\big)$ has distribution ρ^m. We state the following result of Jayram et al.

Theorem 2.11 (Jayram et al. [JKS03])

$$IC_{\rho^m,\epsilon}\big(TRIBES_{m,n}\big) \geq \frac{m(n-1)}{16}\bigg(1 - 2\sqrt{\epsilon}\bigg).$$

Just as before, we derive the following:

Corollary 2.12

$$ER_{\rho^m,\epsilon}\big(TRIBES_{m,n}\big) \geq \frac{m(n-1)}{16}\bigg(1 - 2\sqrt{\epsilon}\bigg).$$

Proof: Easily follows by combining Theorem 2.11 and Theorem 2.8. □

3 A set of special cuts

Using Bourgain's theorem of embedding any graph metric with low distortion, we can derive the following:

Theorem 3.1 (Key Tool) *Let G be any graph with k vertices. Then there exists a set of cuts $\mathcal{C}$ that satisfies the following properties:*

[11] Distributions τ_0 and τ_1 were described just before Theorem 2.9.

1. *Every pair of vertices u, v in G are separated by at least $\Omega(\log k \cdot d(u,v))$ many cuts in $\mathcal{C}$.*

2. *Each edge in G appears as a cut-edge in at most $O(\log^2 k)$ many cuts in $\mathcal{C}$.*

We will prove the above theorem by first connecting the question about cuts to the problem of embedding a graph into ℓ_1 space. In particular, an embedding with specific properties immediately implies the required set of cuts:

Lemma 3.2 *Let $G = (V, E)$ be a graph and $f : V \to \mathbb{R}^D$ be a map for some dimension D that has the following properties:*

(i) For every $u, v \in V$, we have that $\|f(u) - f(v)\|_1 \geq \alpha \cdot d(u,v)$;

(ii) For every edge $(u,v) \in E$, we have that $\|f(u) - f(v)\|_1 \leq \beta$; and

(iii) For any dimension $i \in [D]$, we have that the set $\{f(u)_i | u \in V\}$ is the set $\{0, 1, 2, \ldots, M\}$ for some integer M.

Then there exists a collections of cuts $\mathcal{C}$ such that

1. *Every pair of vertices u, v in G are separated by at least $\alpha \cdot d(u,v)$ many cuts in $\mathcal{C}$.*

2. *Each edge in G appears as a cut-edge in at most β many cuts in $\mathcal{C}$.*

We note that just conditions (i) and (ii) imply that the mapping f is an embedding with distortion β/α. We need property (iii) to construct the required set of cuts $\mathcal{C}$. Next, we note that Bourgain's embedding of graphs (and in particular, any metric) into ℓ_1 satisfies properties (i)-(iii) in Lemma 3.2.

Theorem 3.3 *For any graph G with k vertices, there exists a mapping f such that it satisfies properties (i)-(iii) with $\alpha = \Omega(\log k)$ and $\beta = O(\log^2 k)$.*

Note that Lemma 3.2 and Theorem 3.3 immediately implies Theorem 3.1. In the rest of the section, we will prove Lemma 3.2 and outline why Bourgain's embedding proves Theorem 3.3.

Proof of Lemma 3.2: Let f be a mapping that satisfies property (iii). Next we define the set of cuts $\mathcal{C}$ and show that the number of cuts that separate any pair of vertices $u, v \in V$ is exactly $\|f(u) - f(v)\|_1$. Properties (i) and (ii) will then complete the proof.

For every dimension $i \in [D]$, we will define a family of cuts $\mathcal{C}_i$ and the final set of cuts will be their union $\cup_{i \in [D]} \mathcal{C}_i$. Fix any $i \in [D]$. Let $\{f(u)_i | u \in V\} = \{0, 1, 2, \ldots, M\}$. Then for every $j \in [M]$ include the cut $C_{ij} = \left\{u | f(u)_i < j - \frac{1}{2}\right\}$ in $\mathcal{C}_i$. By property (iii), note that these cuts $\mathcal{C}_{ij}$ are distinct (for fixed i).

To complete the proof, we need to argue that for every $u, v \in V$ exactly $\|f(u) - f(v)\|_1$ many cuts in $\mathcal{C}$ separate u and v. Towards this end, note that for any fixed $i \in [D]$, the number of cuts in $\mathcal{C}_i$ that separate u and v is exactly $|f(u)_i - f(v)_i|$ (this just follows from the construction of $\mathcal{C}_i$ and property (iii)). Thus, the total number of cuts in $\mathcal{C}$ that separate u and v is exactly

$$\sum_{i=1}^{D} |f(u)_i - f(v)_i| = \|f(u) - f(v)\|_1,$$

as desired. □

Theorem 3.3 without the added property (iii) is the usual statement of Bourgain's theorem for ℓ_1 embedding. Next we sketch why Bourgain's construction also satisfies property (iii).

Proof of Theorem 3.3: Bourgain's map f is defined as follows. Pick $D = O(\log^2 k)$ random subsets of V. For a given $u \in V$ and coordinate $i \in [D]$ that corresponds to the random subset S, define $f(u)_i = d(u, S)$, i.e. the distance of u to the closest vertex in S. Since the graph G is unweighted and connected, it is easy to check that this construction would satisfy property (iii). Indeed, consider a graph G' where S is contracted into a "super vertex" s (and (s, u) is an edge if and only if (u, s') is an edge for some $s' \in S$). Now run BFS starting from s in G'. Note that $d(u, S)$ is the level of u in the corresponding BFS tree. Further, by the fact that G' is connected (since G was connected), the $f(u)_i$ will take values in the set $\{0, 1, \ldots, M\}$ where there are M levels in the BFS tree.

Property (i) and (ii) with $\beta = D$ and $\alpha = \Omega(\log k)$ for Bourgain's map is well-known. In particular, Lemma 6.3 in Lecture 3 from [RÖ06] proves (ii) and Theorem 6.4 in Lecture 3 from [RÖ06] proves property (i). □

We end with two remarks. First we note that property (iii) is a bit stronger than what we need. Indeed, we can relax the condition that the distinct values in any dimension be consecutive integers to the following: all the consecutive values are separated by $\Theta(1)$. The in proof of Lemma 3.2, we would have that the number of cuts separating u and v would be $\Theta(\|f(u) - f(v)\|_1)$. This only affects the constants and thus, Theorem 3.1 would still be true. We chose the stronger version because it makes the proof a bit simpler and the fact that Bourgain's embedding already satisfies this stronger property.

Second, one might wonder if one can bypass the embeddings connections but it is easy to check that a set of cuts as defined in Theorem 3.1 indeed defines an embedding of G into ℓ_1 and further, the ratio $\beta/\alpha = \Omega(\log k)$ is true since this lower bound on distortion into ℓ_1 holds for expander graphs (see e.g. Section 7 in Lecture 3 from [RÖ06]).

4 Application

4.1 Element Distinctness

We prove an almost tight result on the randomized complexity of the Element-Distinctness function. Recall that this function outputs 1 if and only if each of the k strings are distinct and no string is repeated.

For any graph G and a vertex v, let $\Delta(v) \equiv \sum_{u \in V(G)} d\big(u, v\big)$. We call vertex c a center of G if $\Delta(c) \leq \Delta(v)$, for every other vertex v. Let the diameter of G be denoted by $D(G)$.

Theorem 4.1 *Let G be any graph with k vertices and c a center. The bounded-error randomized k-party complexity of Element-Distinctness over G is $\Theta\big(\Delta(c)\big)$, ignoring poly-log$(k)$ factors. The zero-error randomized complexity of Element-Distinctness is $\Theta\big(D(G) \cdot n + \Delta(c)\big)$, again ignoring poly-log(k) factors.*

Proof: Let τ be the following distribution: randomly pick k distinct strings $Z_1, \ldots, Z_k$ from $\{0,1\}^n$. Randomly assign them to the k nodes of G so that each node gets exactly one string. For the first part of the theorem, we first show that $ER_{\epsilon,\tau}\big(\text{ELMT-DIST}\big)$ is $\Omega\big(\Delta(c)\big)$.

Using Theorem 3.1, we obtain our set of cuts $\mathcal{C}$. Let C_i be any cut in $\mathcal{C}$. Let Π be any randomized ϵ-error protocol over G for Element-Distinctness. The simple but useful claim is the following: let V_i^0 and V_i^1 be the two sets of vertices separated by cut C_i and $\ell_i = \min\big\{|V_i^0|, |V_i^1|\big\}$.

Claim 4.2 *There is an ϵ-error randomized 2-party protocol solving UDISJ$_{\ell_i}$ with expected cost w.r.t $\mu_0[\ell_i]$ at most ECost$_\tau\big(\Pi, C_i\big)$.*

Let us first show why this claim gives us our desired bound. The claim along with Corollary 2.4 immediately yields that

$$\text{ECost}_\tau\big(\Pi, C_i\big) \geq \beta \cdot \ell_i.$$

Note that $\ell_i \geq |V_i^0| \cdot |V_i^1|/k$. Hence,

$$\sum_{i=1}^{t} \text{ECost}_\tau\big(\Pi, C_i\big) \geq \beta \cdot \frac{1}{k} \sum_{i=1}^{t} |V_i^0||V_i^1|.$$

Observe that $|V_i^0||V_i^1|$ is exactly the number of pairs of vertices separated by cut C_i. Using property (1) of $\mathcal{C}$ from Theorem 3.1, we bound it further:

$$\sum_{i=1}^{t} \text{ECost}_\tau\big(\Pi, C_i\big) \geq \beta \cdot \frac{1}{k} \sum_{u,v \in V(G): u \neq v} d(u,v) = \frac{\beta}{k} \sum_{u} \Delta(u) \geq \beta \cdot \Delta(c).$$

Combining these with Observation 2.1 and Theorem 3.1, we get our bound as follows:

$$\text{ECost}_\tau\big(\Pi\big) \geq \text{ECost}_\tau\big(\Pi, \mathcal{C}\big)/O(\log k) \geq \beta\Delta(c)/O(\log k),$$

where β just depends on ϵ.

What remains to prove is Claim 4.2. W.l.o.g., let $|V_i^0| - |V_i^1| = m_i \geq 0$. Consider any fixed assignment a of distinct strings to the first m_i nodes of V_i^0. Let $\tau|a$ denote the conditional distribution of inputs on the other nodes. We derive a 2-party ϵ-error protocol Π_a for UDISJ$_{\ell_i}$, where sets have elements from a universe of size $2^n - m_i$. Alice and Bob, on getting two sets X, Y respectively from this universe with $|X| = |Y| = \ell_i$, simulate Π as follows: Alice naturally encodes her set X as ℓ_i strings, each n-bit long and Bob does the same with his set Y. Alice assigns the encoded elements of her set and the fixed assignment a to nodes in V_i^0 and Bob assigns his elements to nodes in V_i^1. Then they simulate Π in the natural way, with Alice (Bob) communicating bits to Bob (Alice) whenever in Π a message is communicated along a cut-edge from V_i^0 (V_i^1) to V_i^1 (V_i^0). Using properties of Π, it is easily verified that this is an ϵ-error protocol for UDISJ$_{\ell_i}$. Observe that $\tau|a$ is the same distribution as $\mu_0[\ell_i]$. Using Corollary 2.4, we immediately get $\text{ECost}_{\tau|a}\big(\Pi, C_i\big) = \text{ECost}_{\mu_0[\ell_i]}\big(\Pi_a\big) \geq \beta\ell_i$, for each fixed a. Thus, we conclude that $\text{ECost}_\tau\big(\Pi, C_i\big) \geq \beta\ell_i$, establishing Claim 4.2.

The upper bound for the first part of the theorem follows from the natural fingerprinting algorithm. Every player sends a fingerprint of its input to the vertex c using $O(\log k)$ sized

hashes. The player at c then looks all the $\binom{k}{2}$ pairs of players and checks if the hashes of the corresponding inputs are the same or not. Note that this solves the problem as long as the collision probability of the hashes is $O(1/k^2)$, which can be arranged to be true with $O(\log k)$ sized hashes.

For proving the bound on zero-error protocols, we need to consider another distribution. Let γ be the distribution on k inputs, each n-bit long, generated by the following sampling method: consider two vertices u, v in G such that $d(u, v) = D(G)$. Let $Z = \{z_1, \ldots, z_{k-2}\}$ be a set of $k-2$ distinct strings and set $S = \{0,1\}^n - Z$. Let M be the output of a random coin-toss. If $M = 1$, sample inputs from τ, else sample inputs from ν, where ν is given as follows: assign each vertex, other than u, v, one distinct string from Z and sample a string x at random from S and assign x to both X_u and X_v. Let Π be any zero-error protocol for Element-Distinctness. We next bound $\mathrm{ECost}_\gamma\big(\Pi\big)$: clearly,

$$\mathrm{ECost}_\gamma\big(\Pi\big) = \frac{1}{2}\big[\mathrm{ECost}_\tau\big(\Pi\big) + \mathrm{ECost}_\nu\big(\Pi\big)\big].$$

The first term is at least $\Omega\left(\Delta(c)/\log k\right)$ from the first part of the theorem. We now bound the second term on the RHS above. Using a standard breadth-first search tree, we generate $t = D(G)$ many cuts, $C_1, \ldots, C_t$, such that each C_i separates u and v and the set of cut-edges of C_i and C_j are disjoint, if $i \neq j$. Thus, $\mathrm{ECost}_\nu\big(\Pi\big) = \sum_i \mathrm{ECost}_\nu\big(\Pi, C_i\big)$. We claim that $\mathrm{ECost}_\nu\big(\Pi, C_i\big) \geq R_{\mathcal{U}_{=,S}}\big(2 - EQ\big)$. Given this claim, one immediately gets the desired bound on the RHS.

To prove the claim, consider the following protocol Π' for Alice and Bob to solve 2-EQ, given inputs $x, y \in S$: Alice and Bob simulate, according to Π, respectively the nodes on the sides of C_i that have vertex u and v. Alice assigns x to X_u and Bob assigns y to Y_v. Then the follow Π assuming the other vertices got fixed inputs from Z. Clearly, this solves correctly 2-EQ for $x, y \in S$. Further, it easily follows $\mathrm{ECost}_{U_{=,2}}\big(\Pi'\big) = \mathrm{ECost}_\nu\big(\Pi, C_i\big) = \Omega\big(\log|S|\big)$, where the last step uses Theorem 2.6. This completes the argument for the lower bound.

The upper bound for the second part of the theorem follows from the following modification of the protocol from the first part. In the first phase, each player sends a hash of size $O(\log nk)$ to the player at vertex c. Using the hashes the player at vertex c will check which pairs of player can safely ruled to have distinct inputs. Call a pair of players that cannot be ruled out to have distinct inputs after the first phase to be *surviving*. In the second phase, the player at vertex c consider all surviving pairs (i, j) in some arbitrary order. For each surviving pair (i, j), players i and j send their inputs to c. If $X_i = X_j$, then the protocol terminates with a 0. Otherwise the protocol moves to the next surviving pair. If all surviving pairs have distinct inputs then the algorithm terminates with a 1. It is easy to check that the protocol always terminates with the correct answer. To complete the proof we briefly argue the claimed communication upper bound. First we note that the protocol needs at most one pair (i, j) such that $X_i = X_j$ to send their inputs to c with total communication $O(D(G) \cdot n)$. Now for every fixed pair (i, j) such that $X_i \neq X_j$, the probability that it survives the first phase is $O(1/(nk^2))$. This implies that the expected communication between (i, j) and c in second phase is $O(1/k^2)$. Thus, the total communication in the second phase for all the pairs with distinct inputs is $O(1)$. The total communication for the first phase is $\tilde{O}(\Delta(c))$ from the same argument as in the upper bound

for the first part of the theorem. Adding up all the communication costs completes the proof. □

4.2 Distributed XOR Lemma

So far we have proved lower bounds that showed that the trivial algorithm of sending all inputs (or hashes) to one player was optimal. In this section, we will consider functions for which the trivial algorithm can potentially have a smaller communication cost. In particular, we will consider functions where players are paired up and one only needs to send information from one player to its matched player. (However, there is no difference for the worst-case pairing: see Lemma 5.1 for a formal statement.)

To be more precise, let M be a disjoint pairing[12] of vertices in $G = (V, E)$ with $|V| = k$ (where k is even) and let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a boolean function. Consider the function $f^{\oplus G,M} : (\{0,1\}^n)^k \to \{0,1\}$ defined as follows:

$$f^{\oplus G,M}\left((X_u)_{u\in V}\right) = \bigoplus_{(u,v)\in M} f(X_u, X_v),$$

where $\oplus$ denotes the XOR operator.

Here is the trivial algorithm to compute $f^{\oplus G,M}$: for each pair $(u, v) \in M$, the players corresponding to u and v run the best possible (say randomized) protocol to compute $f(X_u, X_v)$. Note that this would have communication cost at most a factor $D(M)$ times the optimal communication complexity of computing f, where

$$D(M) = \sum_{(u,v)\in M} d(u, v).$$

Next, we show that the above in general, is tight to within a $\tilde{O}(\sqrt{k})$ factor. Further, for functions f where the optimal lower bound on communication complexity of f can be proved via distributional complexity under a product distribution, the above trivial algorithm is tight with poly-logarithmic factors.

Theorem 4.3 *For every constant $\epsilon > 0$ and every binary function f the following are true:*

1. $$R_{\epsilon,G}\left(f^{\oplus G,M}\right) \geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{D(M)}{\sqrt{k} \cdot \log k \cdot \log(nk)}\right),$$

   where $R_\gamma(f)$ is the optimal 2-party communication complexity of f with randomized protocols that err with probability γ.

2. $$R_{\epsilon,G}\left(f^{\oplus G,M}\right) \geq \Omega\left(\left(R'_{2\epsilon}(f) - 2\right) \cdot \frac{D(M)}{\text{poly}\log(nk)}\right),$$

[12] We stress that the pairs $(u, v) \in M$ need not be an edge.

where $R'_\gamma(f)$ is the optimal 2-party communication complexity of f with randomized protocols that err with probability γ that can be proved via distributional complexity on product distributions.

In addition to Theorem 3.1, the main tool that we will use is the following slightly modified result of Barak et al. (see the remark following the theorem):

Theorem 4.4 (Barak et al. [BBCR13]) *For every real numbers $\alpha, \rho > 0$, integer $k \geq 1$, binary function f and a distribution ν on the inputs of f, we have:*

(i)

$$ER_{\nu^k,\rho}(f^{\oplus k}) \cdot \log\left(ER_{\nu^k,\rho}(f^{\oplus k})/\alpha\right) \geq \Omega\left((D_{\nu,\rho+\alpha}(f) - 2)\alpha\sqrt{k}\right),$$

where $f^{\oplus k}$ is the 2-party equivalent[13] of the $f^{\oplus G,M}$ and ν^k is the product distribution where one takes k independent samples from ν.

(ii) If ν itself is a product distribution on the input space of f, then

$$ER_{\nu^k,\rho}(f^{\oplus k}) \cdot \operatorname{poly}\log\left(ER_{\nu^k,\rho}(f^{\oplus k})/\alpha\right) \geq \Omega\left((D_{\nu,\rho+\alpha}(f) - 2)\alpha \cdot k\right),$$

where $f^{\oplus k}$ and ν^k are as defined above.

Remark 2 *The LHS in both (i) and (ii) are slight modifications of the original statements in Theorem 2.8 and Theorem 2.9 of [BBCR13]. Their LHS was $D_{\nu^k,\rho}(f^{\oplus k})$. Our modification immediately follows from Corollary 2.5 of this work.*

We now prove Theorem 4.3.

Proof of Theorem 4.3: By Yao's lemma, there exists a distribution ν on $\{0,1\}^n \times \{0,1\}^n$ such that

$$R_{2\epsilon}(f) = D_{\nu,2\epsilon}(f).$$

Define

$$\mu = \nu^k.$$

By definition, $R_{\epsilon,G}\left(f^{\oplus G,M}\right) \geq ER_{\mu,\epsilon,G}(f^{\oplus G,M})$. Thus, to prove the bounds in the theorem statement it suffices to just lower bound $ER_{\mu,\epsilon,G}(f^{\oplus G,M})$, which is what we will do next.

Let Π be a protocol that computes $f^{\oplus G,M}$ with error ϵ such that

$$\mathrm{ECost}_\mu(\Pi) = ER_{\mu,\epsilon,G}(f^{\oplus G,M}). \tag{1}$$

Let $\mathcal{C} = \{C_1, \ldots, C_t\}$ be the set of special cuts on G guaranteed by Theorem 3.1. Recall that $M(C_i)$ denotes the set of pairs of vertices separated by C_i. We first claim that

[13] That is, G and M are both a single edge. In particular, all of the $k/2$ pairs are across the same edge.

Claim 4.5 *For every $1 \leq i \leq t$:*

$$ECost_{\mu}(\Pi, C_i) \geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{\sqrt{|M(C_i) \cap M|}}{\log(kn)}\right).$$

We will defer the proof of the claim above and complete the rest of the proof. Consider the following sequence of relationships:

$$\sum_{i=1}^{t} \mathrm{ECost}_{\mu}(\Pi, C_i) \geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{\sum_{i=1}^{t} \sqrt{|M(C_i) \cap M|}}{\log(kn)}\right) \tag{2}$$

$$\geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{\sum_{i=1}^{t} |M(C_i) \cap M|}{\sqrt{k} \cdot \log(kn)}\right) \tag{3}$$

$$\geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{D(M) \cdot \log k}{\sqrt{k} \cdot \log(kn)}\right). \tag{4}$$

In the above (2) follows from Claim 4.5, (3) follows from the fact that $|M| = k/2$ and (4) follows from part (1) of Theorem 3.1.

Next, note that (4) along with Observation 2.1 and part (2) of Theorem 3.1 implies that

$$\mathrm{ECost}_{\mu}(\Pi) \geq \Omega\left((R_{2\epsilon}(f) - 2) \cdot \frac{D(M)}{\sqrt{k} \cdot \log k \cdot \log(kn)}\right).$$

The above along with (1) completes the proof of part (1), modulo the proof of Claim 4.5.

To complete the argument, we now prove Claim 4.5. The basic idea is to show that we can use Π to solve the 2-party problem of computing $f^{\oplus |M(C_i) \cap M|}$, where Alice gets one half of the inputs indexed by $E(C_i) \cap M$ and Bob gets the other half and then invoke part (i) of Theorem 4.4. Next we present the details.

For notational convenience define $k' = |M(C_i) \cap M|$. Let a be an arbitrary assignment to inputs that correspond to pairs of vertices on the same side of the cut C_i. Let $\mu|_a$ be the distribution conditioned on this choice. Note by the definition of μ, this is just $\nu^{k'}$. Given Π, Alice and Bob solve for $f^{\oplus k'}$ as follows. They assign inputs for vertices that lie on the same side of the cut C_i with the specific instances from a. For the rest of the vertices in G, they assign the corresponding inputs to $f^{\oplus k'}$. The rest of the simulation is the obvious one: whenever Π needs to communicate over an edge in $E(C_i)$, they just exchange the corresponding messages (other communication in Π can be done by Alice or Bob on their own). It is easy to check that this protocol is an ϵ-error protocol for $f^{\oplus k'}$ (or its complement given the constants in a) given that Π is an ϵ-error protocol. Hence, part (i) of Theorem 4.4 implies that

$$\mathrm{ECost}_{\mu|_a}(\Pi) \geq \Omega\left(\left(D_{\nu,2\epsilon}(f) - 2\right) \cdot \frac{\sqrt{|M(C_i) \cap M|}}{\log(kn)}\right),$$

where we have used the fact that we can WLOG assume that the total communication for Π is $O(k^2 n)$ (which corresponds to the trivial algorithm). Since the above holds for every fixed a, and $D_{\nu,2\epsilon}(f) = R_{2\epsilon}(f)$ by our choice of ν, Claim 4.5 follows.

The proof of part (2) is exactly the same as above except we use of part (ii) of Theorem 4.4 instead of part (i) and is omitted. □

4.3 Distributed AND/OR

In the previous section, we considered the distributed computation of the XOR of several instances of a function f. It is natural to investigate the distributed computation of OR (AND). More formally, let G be a graph with k vertices and M a disjoint pairing of its vertices as before. Then for any $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, we define $f^{\vee G,M}, f^{\wedge G,M} : (\{0,1\}^n)^k \to \{0,1\}$ as follows:

$$f^{\vee G,M}\left((X_u)_{u\in V}\right) = \bigvee_{(u,v)\in M} f(X_u, X_v),$$

$$f^{\wedge G,M}\left((X_u)_{u\in V}\right) = \bigwedge_{(u,v)\in M} f(X_u, X_v).$$

As before in the case of XOR, trivial protocols compute it in $O\big(D(M)\cdot R_\epsilon(f)\big)$ cost. It turns out for specific f, one can design much more efficient protocols as stated below:

Observation 4.6 *For any G, M, both $R_{\epsilon,G}\big(NEQ^{\vee G,M}\big)$ and $R_{\epsilon,G}\big(EQ^{\wedge G,M}\big)$ are $O\big(k\big)$.*

Proof: We point out a simple protocol Π for $\text{NEQ}^{\vee G,M}$ and omit the very similar argument for $\text{EQ}^{\wedge G,M}$. We divide the set of k players into two groups arbitrarily such that the two players of each pairing in M go to different groups. Call one group X and the other Y such that the ith player in X (input X_i) is paired with ith player in Y (input Y_i). Let T be a rooted spanning tree of G. In Π, players sample $k/2$ public random strings $r_1, \ldots, r_{k/2}$, each n-bit long. A player at a node waits for all players at its child-nodes in T to communicate messages to it. Each player sends a two-bit message (b_X, b_Y) to its parent in T, where b_X, b_Y are called the X and Y part of the message respectively. Player i in X (Y) computes $\langle r_i, X_i\rangle (\text{mod } 2)$ ($\langle r_i, Y_i\rangle (\text{mod } 2)$) and then xors this bit and the X (Y) part of all the messages from its child-nodes. This new bit is the X (Y) part of its message to its parent. The Y (X) part is the xor of all the Y (X) part of the messages from its child nodes.

The protocol Π begins from the players at the leaf-nodes of T and propagates towards the root player. Finally, the root player checks if the X-bit and Y-bit computed by it are equal. If they are not, Π outputs 1 otherwise it outputs 0. It is simple to verify Π is a 1-sided protocol (never errs if it says 1) and it errs with probability 1/2. □

Observation 4.6 essentially rules out a topology dependent lower bound in general for $f^{\vee G,M}$ and $f^{\wedge G,M}$, analogous to Theorem 4.3. However, we show that for some simple functions f, we do get tight topology-dependent bounds. Moreover, some of these will be useful for proving lower bounds in Section 5 on the message passing complexity of solving some natural graph problems.

We complement Observation 4.6 by proving below topology-dependent tight bounds for $\text{EQ}^{\vee G,M}$ (and thus $\text{NEQ}^{\wedge G,M}$).

Theorem 4.7 *For any G, $n \geq 2$ and a disjoint pairing M of the nodes of G, $R_{\epsilon,G}\big(EQ_n^{\vee G,M}\big)$ is $\tilde{\Theta}\big(D(M)\big)$, ignoring* $\text{poly}\log(k)$ *factors.*

Proof: The upper bound follows easily using fingerprinting[14]. Using public coins, one of the nodes of each pair in M sends an $O(\log k)$-bit fingerprint to its mate. The mate checks it its fingerprint matches with the sent one. Clearly, with probability less than, say, $\frac{1}{2k}$, a pair of nodes will not detect an inequality. Thus, in total, they communicate $O\big(D(M)\log k\big)$ bits.

We next prove the lower bound. As before, we naturally divide the players arbitrarily into two groups X and Y of equal size, with X_i, Y_i paired in M. Consider any three distinct strings $s_x^0, s_y^0, s^1 \in \{0,1\}^n$, which exist as $n \geq 2$. We consider the following distribution: let $D_1, \ldots, D_{k/2}$ be independent 0/1 valued random variables, each D_i has equal probability of being 1 and 0. If $D_i = 0$ ($D_i = 1$), we set $X_i = s_x^0$ ($Y_i = s_y^0$) and sample Y_i (X_i) at random from $\{s_y^0, s^1\}$ ($\{s_x^0, s^1\}$). Call ν the marginal of this distribution on (X_i, Y_i). Let Π be any ϵ-error protocol for $\mathrm{EQ}_n^{\vee G,M}$. We next show that $\mathrm{ECost}_{\nu^k}\big(\Pi\big) = \Omega\big(D(M)/\log(k)\big)$, which will establish the claimed lower bound of our theorem.

Let $\mathcal{C} = \{C_1, \ldots, C_t\}$ be the special family of cuts of G guaranteed by Theorem 3.1. We make the following claim:

Claim 4.8 *For each $1 \leq i \leq t$: $ECost_{\nu^k}\big(\Pi, C_i\big) \geq \Omega\big(|M(C_i) \cap M|\big)$.*

The proof of the lower bound on $\mathrm{ECost}_{\nu^k}\big(\Pi\big)$ using Claim 4.8 is very similar to the proof of Theorem 4.3 using Claim 4.5 as given in the previous section. Hence, we omit that and just prove Claim 4.8 below.

Let $\psi = \nu^k$ and let $m = |M(C_i) \cap M|$. Let a be an arbitrary assignment to inputs that correspond to pairs of vertices on the same side of the cut C_i such that a is consistent with assignments in the support of ψ. Let $\psi|_a$ be the distribution conditioned on this choice. Let Π_a be the protocol induced from Π with this fixing. We first derive an ϵ-error protocol Γ for DISJ_m, using Π. Alice and Bob replace each of their 0's by strings s_x^0 and s_y^0 respectively. They both replace their 1's by s^1. Then they simulate Π_a in the natural way. Alice and Bob communicate to each other only when Π_a communicates message across cut C_i. It is easy to verify Γ is an ϵ-error protocol for DISJ_m. On any execution its cost is the same as Π_a. Recall the hard distribution μ for DISJ from Theorem 2.9 and Corollary 2.10. It is easily verified that

$$\mathrm{ECost}_{\psi|_a}\big(\Pi_a\big) = \mathrm{ECost}_{\mu}\big(\Gamma\big) \geq \frac{m}{4}\left(1 - 2\sqrt{\epsilon}\right),$$

where the last inequality follows from Corollary 2.10. Observing that this is true for every a, we immediately establish Claim 4.8.

□

Of particular interest for applications, would be $\mathrm{DISJ}_n^{\vee G,M}$ and $\mathrm{DISJ}_n^{\wedge G,M}$, where the subscript n refers to the fact that each of the k players holds an n-bit string. We will show the following:

Theorem 4.9 *1.*

$$R_{\epsilon,G}\big(DISJ_n^{\vee G,M}\big) = \Omega\left(\frac{D(M)}{\mathrm{poly}\log(k)} \cdot n\right)$$

.

[14]There is a more careful protocol that can save an $O(\log k)$ factor over the one described here. In this version, we do not describe it for the sake of simplicity.

2.

$$R_{\epsilon,G}\big(DISJ_n^{\wedge G,M}\big) = \Omega\left(\frac{D(M)}{\text{poly}\log(k)} \cdot n\right)$$

.

Proof: The proof of the two bounds are very similar. The bound for the complexity of $\text{DISJ}_n^{\vee G,M}$ makes use of Corollary 2.10 and the bound for $\text{DISJ}_n^{\wedge G,M}$ makes use of Corollary 2.12 in almost identical fashion. We therefore only prove the bound for the latter and omit the former.

The basic argument of the proof follows the same outline as those for Theorems 4.3 and 4.7. We first select ρ^k as our global distribution on inputs, where ρ is defined as in Corollary 2.12. This choice is natural because $\text{DISJ}_n^{\wedge G,M}$ defines a distributed version of the function $\text{TRIBES}_{k,n}$. Assume Π is an ϵ-error k-party randomized protocol solving $\text{DISJ}_n^{\wedge G,M}$. We will bound $\text{ECost}_{\rho^k}\big(\Pi\big)$ by the quantity appearing in the RHS of item 2 in the theorem above. This will finish the argument.

As in the arguments before, Theorem 3.1 provides us with the family of cuts $\mathcal{C} \equiv \{C_1, \ldots, C_t\}$. Then, we make the following claim:

Claim 4.10 *For each* $1 \leq i \leq t$,

$$ECost_{\rho^k}\big(\Pi, C_i\big) \geq \Omega\left(\frac{\big|M(C_i) \cap M\big| \cdot (n-1)}{16}\big(1 - 2\sqrt{\epsilon}\big)\right).$$

Given this claim, the bound of the theorem follows in very similar way as Theorem 4.3 followed from Claim 4.5 before. We omit these, by now, routine steps. All that remains is to prove Claim 4.10, which we do next. Note that the argument below is quite similar, in fact slightly simpler, to the proof of Claim 4.8 above.

Let $\phi = \rho^k$ and let $m = |M(C_i) \cap M|$. Let a be an arbitrary assignment to inputs that correspond to pairs of vertices on the same side of the cut C_i such that a is consistent with assignments in the support of ϕ. Let $\phi|_a$ be the distribution conditioned on this choice. Let Π_a be the protocol induced from Π with this fixing. We first derive an ϵ-error protocol Γ for $\text{TRIBES}_{m,n}$, using Π_a. Alice and Bob, given an input instance of $\text{TRIBES}_{m,n}$ just simulate Π_a in the natural way. Alice and Bob communicate to each other only when Π_a communicates message across cut C_i. Observing that the partial assignment a does not fix the output of $\text{DISJ}_n^{\wedge G,M}$, it is easy to verify Γ is an ϵ-error protocol for $\text{TRIBES}_{m,n}$. On any execution its cost is the same as Π_a. Noting that $\phi|_a$ is the same as ρ^m, it follows that

$$\text{ECost}_{\phi|_a}\big(\Pi_a\big) = \text{ECost}_{\rho^m}\big(\Gamma\big) \geq \frac{m(n-1)}{16}\left(1 - 2\sqrt{\epsilon}\right),$$

where the last inequality follows from Corollary 2.12. Observing that this is true for every a, we immediately establish Claim 4.10. □

4.4 OR of f on Star Graph

In this section, we show that our technique can recover the lower bound of Phillips et al. [PVZ12] (and strengthened by Woodruff and Zhang in [WZ13]) on the following function:

$$f^{\vee,*}(X_1,\ldots,X_k,Y) = \vee_{i=1}^{k} f(X_i,Y),$$

where the underlying topology G is the star graph with k leaves with the center getting Y and the k leaves get $X_1,\ldots,X_k$ separately. The following result is known:

Theorem 4.11 ([PVZ12, WZ13]) *For any $\epsilon > 0$ and boolean function f, we have for the star graph G on k leaves:*

$$R_{\epsilon,G}(f^{\vee,*}) \geq \Omega\left(k \cdot R_{2\epsilon}(f)\right).$$

Proof: By Yao's lemma, there exists a distribution ν on $\{0,1\}^n \times \{0,1\}^n$ such that

$$R_{2\epsilon}(f) = D_{\nu,2\epsilon}(f).$$

Let ν_0 be the marginal distribution on inputs where f evaluates to 0. Define the global distribution μ on the $k+1$ inputs as follows. Pick (X_1,Y) according to ν_0. Then conditioned on the choice of Y, pick $X_2,\ldots,X_k$ according to the conditional distribution on ν_0.

By definition, $R_{\epsilon,G}(f^{\vee,*}) \geq ER_{\mu,\epsilon,G}(f^{\vee,*})$. Thus, we will lower bound the latter quantity. Let Π be a protocol that computes $f^{\vee,*}$ with error ϵ such that

$$\mathrm{ECost}_{\mu}(\Pi) = ER_{\epsilon,\mu,G}(f^{\vee,*}).$$

Let $\mathcal{C} = \{C_1,\ldots,C_k\}$ be the set of cuts defined by picking a leaf to be one side of the cut. We first claim that:

Claim 4.12 *For every $1 \leq i \leq k$:*

$$ECost_{\mu}(\Pi, C_i) \geq \Omega\left(R_{2\epsilon}(f)\right).$$

Since each cut separates each of the k edges of G exactly once, Observation 2.1 completes the proof.

The proof of Claim 4.12 follows exactly the same argument as for similar claims in earlier proofs so we just sketch the argument here. WLOG consider the cut C_1 and assume that player 1 (with X_1) is on one side of the cut. The idea is that using Π we construct an ϵ-error randomized 2-party protocol Γ for f. Then we show that $\mathrm{ECost}_{\mu}(\Pi, C_1) = \mathrm{ECost}_{\nu_0}(\Gamma)$. This will allow us to conclude the argument as Lemma 2.3 yields $\mathrm{ECost}_{\nu_0}(\Gamma) \geq D_{2\epsilon,\nu}(f) = R_{2\epsilon}(f)$. So, here is how we construct such a protocol Γ. Given inputs, X, Y, Bob samples $X_2,\ldots,X_k$ according to ν_0 conditioned on Y. Let a be such a sampling of the values of $X_2,\ldots,X_k$ that is in the support of μ. Alice sets $X_1 = X$. Then, they simulate Π across the cut as usual. As each (X_j,Y) is in the support of ν_0, for $j \geq 2$, the effective function across the cut is $f(X_1,Y)$. This shows that Γ is an ϵ-error randomized protocol for f. Further, note that when the inputs to Γ are sampled from ν_0, Γ simulates Π on an input sampled from μ. This establishes the second property of Γ and we are done.

□

4.5 Composed Functions

In this section, we demonstrate further applicability of our technique by considering composed functions. These functions have been widely studied in the communication complexity literature. It is convenient to consider the following $k \times n$ Boolean matrix M representing the inputs of the players: each row i of M is the input string of Player i. Let $f : \{0,1\}^n \to \{0,1\}$ and $g : \{0,1\}^k \to \{0,1\}$ be two boolean functions. Then, $f \circ g : \big(\{0,1\}^n\big)^k \to \{0,1\}$ is defined as follows: the input to $f \circ g$ is naturally viewed as a $k \times n$ matrix M. Then, $(f \circ g)(M) = f\big(g(M^1), g(M^2), \ldots, g(M^n)\big)$, where M^j is the k-bit string in the j-th column of M.

For instance, $\big(\mathrm{OR} \circ \mathrm{AND}\big)$ is the k-party set-disjointness problem. Very recently, Braverman et.al. [BEO$^+$13] proved $\Omega\big(kn\big)$ lower bounds for any network topology. This is also tight, independent of the topology for the following simple reason. Let T be a rooted spanning tree. Staring from the players at leaf-nodes, every Player u computes the intersection of sets held by players in the tree rooted at u. Player u forwards this result using n-bits to its parent-node in T. Thus every player in T forwards an n-bit string to its parent resulting in a deterministic zero-error protocol of cost $(k-1)n$.

The main question that we are interested in is what combination of inner and outer functions makes the composed function's complexity sensitive to the graph topology, i.e. the complexity grows superlinearly with k the number of nodes for some topology. Clearly if the inner function g over k variables itself can be computed, by randomized protocols in $O(k)$ communication over a topology, then for any outer function f, $\big(f \circ g\big)$ can be computed in $\tilde{O}(kn)$ cost. This rules out several functions like XOR, AND, Equality as inner functions.

On the other hand, we show that for hard inner functions like Inner Product or Set-Disjointness we indeed get topology sensitive complexity bounds. Let G be any graph on a set of k nodes $\{1, \ldots, k\}$, and M a disjoint pairing of its nodes, assuming k is even. Then, $\big(f_n \circ \mathrm{IP}_k\big)^{G,M}$ is defined as follows on any input $X_1, \ldots, X_k$: let $Y_j = \oplus_{(u,v)\in M}\big(X_u\big)_j \cdot \big(X_v\big)_j$, i.e. we apply the IP function, as defined by pairing M, on the jth column of $k \times n$ matrix A formed naturally from the inputs of the k players. Then,

$$\big(f_n \circ \mathrm{IP}_k\big)^{G,M}\big(X_1, \ldots, X_k\big) \equiv f\big(Y_1, \ldots, Y_n\big).$$

Theorem 4.13 *The k-party ϵ-error randomized complexity of both $\big(XOR_n \circ IP_k\big)^{G,M}$ and $\big(OR_n \circ IP_k\big)^{G,M}$ are $\tilde{\Theta}\big(D(M) \cdot n\big)$.*

Proof: Recall the distribution ν defined just before the statement of Theorem 2.9 in Section 2. We consider the following input distribution γ: let $\big((X_u)_j, (X_v)_j\big)$ be i.i.d according to ν, for $(u,v) \in M$ and $j \in [n]$. Note that $\gamma \equiv \nu^\ell$, where $\ell = \frac{k}{2}n$. We show that for every protocol Π computing $\big(f \circ \mathrm{IP}_k\big)^{G,M}$, where f is either XOR_n or OR_n, $\mathrm{ECost}_\gamma\big(\Pi\big) = \tilde{\Omega}\big(D(M) \cdot n\big)$ which will establish the theorem.

We make the following claim:

Claim 4.14 *Let C be any cut of G: $ECost_{\nu^\ell}\big(\Pi, C\big) = \Omega\big(|M(C) \cap M| \cdot n\big)$.*

The proof of the lower bound on $\mathrm{ECost}_{\nu^k}\big(\Pi\big)$ using Claim 4.14 and the set of special cuts of G using Theorem 3.1, is very similar to the proof of Theorem 4.3 using Claim 4.5 as given in Section 4.2. Hence, we omit that and just prove Claim 4.14 below.

We note that the following argument is quite similar to proof of Claim 4.8 and Claim 4.10. Let $\phi = \nu^\ell$ and let $m = |M(C) \cap M|$. Let a be an arbitrary assignment to inputs that correspond to pairs of vertices on the same side of the cut C such that a is consistent with assignments in the support of ϕ. Let $\phi|_a$ be the distribution conditioned on this choice. Let Π_a be the protocol induced from Π with this fixing. We first derive an ϵ-error protocol Γ for UDISJ_{mn}, using Π_a. Alice and Bob, given an input instance of UDISJ_{mn} just simulate Π_a in the natural way. They communicate to each other only when Π_a communicates message across cut C. Observing that the partial assignment a does not fix the output of $\big(f \circ \mathrm{IP}_k\big)^{G,M}$, it is easy to verify Γ is an ϵ-error protocol for UDISJ_{mn}. On any execution its cost is the same as Π_a. Noting that $\phi|_a$ is the same as ν^{mn}, it follows that

$$\mathrm{ECost}_{\phi|_a}\big(\Pi_a\big) = \mathrm{ECost}_{\nu^{mn}}\big(\Gamma\big) \geq \frac{mn}{4}\left(1 - 2\sqrt{\epsilon}\right),$$

where the last inequality follows from Corollary 2.10. Observing that this is true for every a, we immediately establish Claim 4.14. □

Similarly, define $\big(f_n \circ \mathrm{DISJ}_k\big)^{G,M}$ on inputs $X_1, \ldots, X_k$ as follows: let $Z_j = \vee_{(u,v)\in M}\big(X_u\big)_j \cdot \big(X_v\big)_j$, i.e. we apply the DISJ function, as specified by pairing M, on the jth column of $k \times n$ matrix A formed naturally from the inputs of the k players.

Theorem 4.15 *The k-party ϵ-error randomized complexity of $\big(f_n \circ DISJ_k\big)^{G,M}$ is $\tilde{\Theta}\big(D(M) \cdot n\big)$ when f_n is any of the following functions: XOR_n, OR_n.*

Proof: The proof is almost identical to that of the previous argument for Theorem 4.13. We choose the same global input distribution ν^ℓ as there, where $\ell = \frac{k}{2}n$. Then, just as there, for any cut C, and any bounded-error protocol Π computing $\big(f \circ \mathrm{DISJ}_k\big)^{G,M}$, where f is either XOR_n or OR_n, we claim the following:

Claim 4.16 $ECost_{\nu^\ell}\big(\Pi, C\big) = \Omega\big(|M(C) \cap M| \cdot n\big)$.

Both the proof of this claim and the application of the claim to prove our theorem is identical to the proof and use of Claim 4.14 for proving Theorem 4.13. □

5 Lower Bounds for Graph Problems

In this section, we will consider the case when the k players are trying to compute a function about a graph $H = (V, E)$ that is distributed among the k players. In particular, for player $i \in [k]$, we will denote its subgraph by H_i. We consider the graph based problems considered in [WZ13] and show that in all of them the trivial algorithm where all players send their input to one player is the best possible algorithm. In particular, these give topology dependent extensions

to the corresponding results in [WZ13]. In this section, we will not explicitly differentiate whether the edge sets of H_i are disjoint for every i or not. In what follows we will use m to denote the total number of edges in H_i ($i \in [k]$) with duplication.

We begin with a technical result that we will use multiple times:

Lemma 5.1 *For any graph G on k vertices for even k, let $\mathcal{M}(G)$ denote the set of all disjoint pairings of the k vertices. Then for any $k \geq 2$,*

$$\frac{1}{2} \cdot \Delta(c) \leq \max_{M \in \mathcal{M}(G)} D(M) \leq \Delta(c).$$

Proof: The upper bound easily follows from the triangle inequality. Indeed consider any pairing M. Then note that for any pair u and v, we have $d(u,v) \leq d(u,c) + d(v,c)$. Summing up both side over all pairs in M (and noting that when u or v is c then we only need one term in the RHS of the inequality) proves the upper bound.

For the lower bound, we first claim that there exists $k-1$ pairings $M_1, \ldots, M_{k-1}$ such that every pair of distinct vertices of G appears in exactly one M_j. (This follows from the fact that a complete graph on even number of vertices has a 1-factorization.) This implies that

$$\sum_{j=1}^{k-1} D(M_j) = \frac{1}{2} \sum_{u \neq v} d(u,v) \geq \frac{k}{2} \cdot \Delta(c),$$

where the equality follows from the definition of $M_1, \ldots, M_{k-1}$ while the inequality follows from the same argument used in proof of Theorem 4.1. Thus, by a counting argument there exists an M_j such that

$$D(M_j) \geq \frac{k}{2(k-1)} \cdot \Delta(c) \geq \frac{\Delta(c)}{2},$$

as desired. □

5.1 Reductions from Element-Distinctness

Degree. In this problem in addition to H_i each player also receives a vertex $v \in V$. Together they have to compute the degree of v. The following result follows from the same reduction as in [WZ13].

Theorem 5.2 *Solving the degree problem on graph G needs $\Omega\left(\Delta(c) \cdot \frac{m}{k \cdot \text{poly}\log k}\right)$ communication.*

Proof: We use the reduction from [WZ13]. Consider an instance of Element-distinctness problem where the k inputs are $X_1, \ldots, X_k \in \{0,1\}^n$ with $n = O(\log k)$. Now consider H as follows. $V = \{0,1\}^n \cup \{v\}$. Then player $i \in [k]$ gets the edge (v, X_i). Note that each player can construct H_i from its input X_i, that $m = k$ and that the degree of v is k if and only if all of $X_1, \ldots, X_k$ are distinct. The claimed lower bound then follows from Theorem 4.1. □

5.2 Reductions from OR-DISJ

In this section, we consider the following three problems.

Acyclicity. Given H_i to player $i \in [k]$, the players need to decide if H is acyclic or not.

Triangle-Freeness. Given H_i to player $i \in [k]$, the players have to decide if H has a triangle or not.

Bipartitiness. Given H_i to player $i \in [k]$, the players have to decide if H is bipartite or not.

We will show that for all of the problems above, the trivial algorithm is the best.

Theorem 5.3 *Each of the problems of acyclicity, triangle-freeness and bipartitenes on graph G needs $\Omega\left(\Delta(c) \cdot \frac{m}{k \cdot \log k}\right)$ communication.*

Proof: We will prove the claimed lower bounds by showing the claimed lower bound on the problem where the players have to decide if (i) H is a forest vs (ii) H has a triangle. (Note that solving any of acyclicity, triangle-freeness or bipartitenes will determine which of the two cases H falls in.)

Consider the OR-DISJ problem with pairing M, where we pick M so that it maximizes $D(M)$ (and hence, we can apply Lemma 5.1). For notational convenience let us assume that

$$M = \{(i, i') | i \in [k/2] \text{ and } i' = i + k/2\}.$$

Assume that for the $DISJ_n^{\vee G,M}$ problem, player $i \in [k]$ gets a set/vector $X_i \in \{0,1\}^n$. We now define the subgraphs H_i. To begin with we have

$$V = \cup_{j=1}^{k/2} U_j \cup \{w^1, \ldots, w^k\},$$

where

$$U_j = \{u_1^j, \ldots, u_n^j\}.$$

For every $p \in [k]$, H_p has the following edges: $(w^p, w^{p'})$ and $(w^p, u_j^{p'})$ for every $j \in [n]$ such that $X_p(j) = 1$, where $p' = p$ if $p \leq k/2$ and $p' = p - k/2$ otherwise. See Figure 5.2 for an illustration of this reduction.

Note that each player p can construct its subgraph just from its X_p and that H is in case (i) if $DISJ_n^{\vee G,M}(X_1, \ldots, X_k) = 0$ and is in case (ii) otherwise. Theorem 4.9 along with the lower bound in Lemma 5.1 completes the proof. □

5.3 Reductions from AND-DISJ

In this section, we prove lower bounds for the following connectivity problems.

Connectivity. Given H_i to player $i \in [k]$, the players need to decide if H is connected or not.

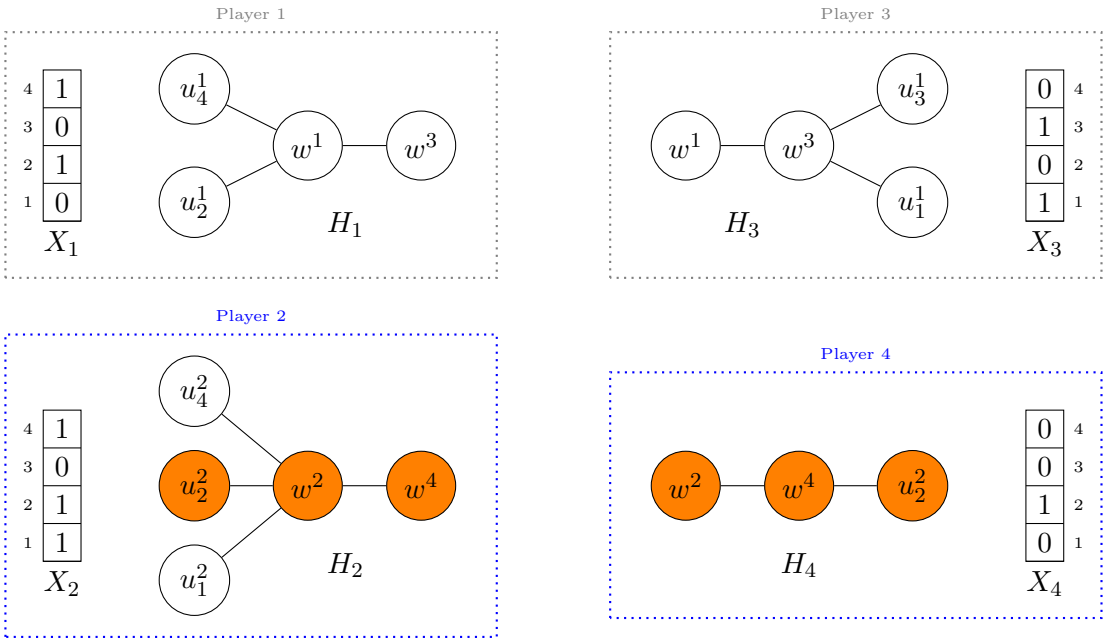

Figure 1: Illustration of the reduction in proof of Theorem 5.3 for $n = k = 4$. The pairing $M = (1,3),(2,4)$ is denoted by the paired players having the same colored boxes. In this example the overall graph $H = \cup_{i=1}^{4} H_i$ has a triangle and the three participating nodes are colored in orange.

Connected Components. Given H_i to player $i \in [k]$, the players need to compute the number of connected components in H.

Since a lower bound for the first problem implies a lower bound for the second problem, we only present a lower bound for the connectivity problem.

Theorem 5.4 *The problem of connectivity on graph G needs $\Omega\left(\Delta(c) \cdot \frac{m}{k \cdot \log k}\right)$ communication.*

Proof: We will reduce from AND-DISJ. As in proof of Theorem 5.3, we pick M so that it maximizes $D(M)$ (and hence, we can apply Lemma 5.1). For notational convenience let us assume that

$$M = \{(i, i') | i \in [k/2] \text{ and } i' = i + k/2\}.$$

Consider the AND-DISJ problem with pairing M, where player $i \in [k]$ gets the set $X_i \in \{0,1\}^n$. We now define the subgraphs H_i. To begin with we have

$$V = \cup_{j=1}^{k/2} U_j \cup \{\ell_1, \ldots, \ell_{k/2}, r\},$$

where

$$U_j = \{u_1^j, \ldots, u_n^j\}.$$

For every player $p \in [k]$, player p has the following edges in H_p. If $p \leq k/2$, then it has the edges (ℓ_j, u_j^p) for every j such that $X_p(j) = 1$. Otherwise it has the edges $(r, u_j^{p'})$ for every j such that $X_p(j) = 1$ where $p' = p - k/2$. See Figure 5.3 for an illustration of this reduction.

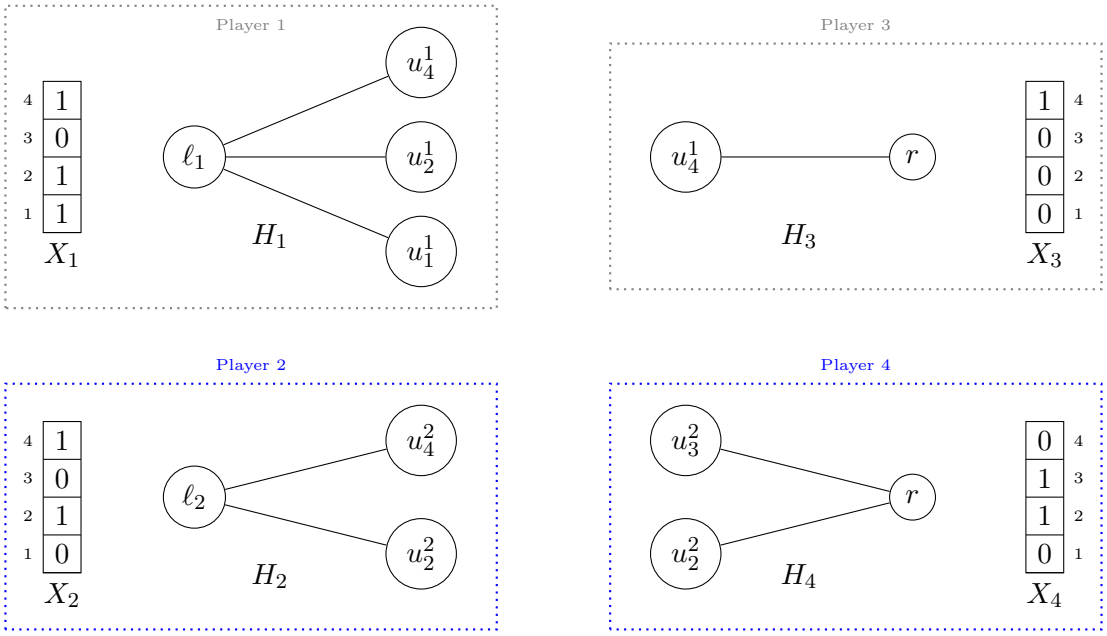

Figure 2: Illustration of the reduction in proof of Theorem 5.4 for $n = k = 4$. The pairing $M = (1,3),(2,4)$ is denoted by the paired players having the same colored boxes. In this example the overall graph $H = \cup_{i=1}^{4} H_i$ is connected.

It is easy to check that each player p can construct H_p just from X_p and that H is connected if and only if $DISJ_n^{\wedge G,M}(X_1, \ldots, X_k) = 1$. Theorem 4.9 along with the lower bound in Lemma 5.1 completes the proof. □

6 Conclusion and Open Questions

In this paper we obtained the first lower bounds in NIH point to point communication model that non-trivially depend on the underlying network topology. We presented a simple technique based on the theory of embeddings and the linearity of expectation that is able to prove our results. Our results include topology dependent lower bounds on the Element-Distinctness problem, an XOR lemma, various results on OR and AND of functions as well as natural graph problems. We also showed that our techniques are powerful enough to recover results from existing techniques.

Many questions still remain unresolved. The biggest question is arguably to determine the complexity of composed functions. We know that we cannot have non-trivial topology dependent lower bounds for every composed functions: broad sufficient conditions on the composed

functions that lead to non-trivial topology dependent lower bounds would be very interesting. Resolving this question even for the special case of the outer function being OR/AND seems non-trivial.

Another special case of this question would be fix the inner functions to be a hard function like DISJ and to consider outer functions that are lower degree polynomials over $\mathbb{F}_2$. We can handle the very special cases of linear outer functions but even handling quadratic polynomials seems challenging.

There is one aspect of our technique that we find intriguing. To explain this, note that a key step of the technique is to reduce the problem to a bunch of 2-player communication games played across a nice family of cuts. But in each of these 2-player communication games, Alice gets access to *all* inputs lying on one side of the cut and Bob gets similar access to *all* inputs lying on the other side. However, in the original k-player game, each player gets access to only his/her input. In other words, it seems natural to expect that our technique for proving lower bounds is still not able to capture all the bottleneck in the original point to point communication problem. What then explains the many tight lower bounds that the technique establishes nevertheless? Is it possible that for a certain class of functions f, one can formulate a set of conditions for induced 2-party games across cuts of a graph G that are sufficient (and necessary) to yield a genuine point to point k-party efficient communication protocol for f over topology G. This would be an interesting direction to investigate.

Acknowledgments

We thank Anupam Gupta for answering our newbie questions on embeddings. Thanks also to David Woodruff and Qin Zhang for answering our questions on their paper [WZ13].

References

[AMS99] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *J. Comput. Syst. Sci.*, 58(1):137–147, 1999.

[BBCR10] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[BBCR13] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. *SIAM J. Computing*, 42(3):1327–1363, 2013.

[BBFM12] Maria-Florina Balcan, Avrim Blum, Shai Fine, and Yishay Mansour. Distributed learning, communication complexity and privacy. In *COLT*, pages 26.1–26.22, 2012.

[BEO+13] M. Braverman, F. Ellen, R. Oshman, T. Pitassi, and V. Vaikuntanathan. A tight bound for set disjointness in the message-passing model. In *FOCS*, pages 668–677, 2013.

[BKS13] Paul Beame, Paraschos Koutris, and Dan Suciu. Communication steps for parallel query processing. In *PODS*, pages 273–284, 2013.

[BYJKS04] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, pages 94–99, New York, NY, USA, 1983. ACM.

[Cor13] Graham Cormode. The continuous distributed monitoring model. *SIGMOD Rec.*, 42(1):5–14, May 2013.

[CP10] A. Chattopadhyay and T. Pitassi. The story of set disjointness. *SIGACT News*, 41(3):59–85, 2010.

[CT91] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. J. Wiley and Sons, 1991.

[DF89] D. Dolev and T. Feder. Multiparty communication complexity. In *FOCS*, pages 428–433, 1989.

[DKO12] A. Drucker, F. Kuhn, and R. Oshman. The communication complexity of distributed task allocation. In *PODC*, pages 67–76, 2012.

[DR98] P. Duris and J.D.P. Rolim. Lower bounds on the multiparty communication complexity. *J.Comput.Syst.Sci.*, 56(1):90–95, 1998.

[GSZ11] Michael T. Goodrich, Nodari Sitchinava, and Qin Zhang. Sorting, searching, and simulation in the mapreduce framework. In *ISAAC*, pages 374–383, 2011.

[HK12] H.Kowshik and P.R. Kumar. Optimal function computation in directed and undirected graphs. *IEEE Transcations on Information Theory*, 58(6):3407–3418, 2012.

[IPSV12] Hal Daumé III, Jeff M. Phillips, Avishek Saha, and Suresh Venkatasubramanian. Protocols for learning classifiers on distributed data. In *AISTATS*, pages 282–290, 2012.

[JKS03] T.S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *STOC*, pages 673–682, 2003.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KO11] F. Kuhn and R. Oshman. Dynamic networks: models and algorithms. *SIGACT News*, 42(1):82–96, 2011.

[KS11] Paraschos Koutris and Dan Suciu. Parallel evaluation of conjunctive queries. In *PODS*, pages 223–234, 2011.

[KSV10] Howard J. Karloff, Siddharth Suri, and Sergei Vassilvitskii. A model of computation for mapreduce. In *SODA*, pages 938–948, 2010.

[Mut05] S. Muthukrishnan. Data streams: Algorithms and applications. *Foundations and Trends in Theoretical Computer Science*, 1(2), 2005.

[PVZ12] J. M. Phillips, E. Verbin, and Q. Zhang. Lower bounds for number-in-hand multiparty communication complexity, made easy. In *SODA*, pages 486–501, 2012.

[RÖ6] Harald Räcke. CMCS 39600: Theory of Metric Embeddings, 2006. Lecture Notes. Available at `http://ttic.uchicago.edu/ harry/teaching/teaching.html`.

[Raz92] A. Razborov. On the distributional complexity of Disjointness. *Theor.Comput.Sci.*, 106(2):385–390, 1992.

[Val90] Leslie G. Valiant. A bridging model for parallel computation. *Commun. ACM*, 33(8):103–111, 1990.

[WZ12] D. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. In *STOC*, pages 941–960, 2012.

[WZ13] D. Woodruff and Q. Zhang. When distributed computation is communication expensive. In *DISC*, pages 16–30, 2013.

[WZ14] D. Woodruff and Q. Zhang. An optimal lower bound for distinct elements in the message passing model. In *SODA*, pages 718–733, 2014.

[Yao79] A. C. C. Yao. Some complexity questions related to distributed computing. In *11th ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de