Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality

Elchanan Mossel *   Ryan O'Donnell †   Oded Regev ‡
Jeffrey E. Steif §   Benny Sudakov ¶

Abstract

In this paper we study *non-interactive correlation distillation (NICD)*, a generalization of noise sensitivity previously considered in [5, 31, 39]. We extend the model to *NICD on trees*. In this model there is a fixed undirected tree with players at some of the nodes. One node is given a uniformly random string and this string is distributed throughout the network, with the edges of the tree acting as independent binary symmetric channels. The goal of the players is to agree on a shared random bit without communicating.

Our new contributions include the following:

- In the case of a k-leaf star graph (the model considered in [31]), we resolve the open question of whether the success probability must go

*Department of Statistics, U.C. Berkeley. mossel@stat.berkeley.edu. Supported by a Miller fellowship in CS and Statistics, U.C. Berkeley and by an Alfred P. Sloan fellowship.

†Institute for Advanced Study, Princeton, NJ. odonnell@ias.edu. Most of this work was done while the author was a student at Massachusetts Institute of Technology. This material is based upon work supported by the National Science Foundation under agreement No. CCR-0324906. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

‡Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Most of this work was done while the author was at the Institute for Advanced Study, Princeton, NJ. Work supported by an Alon Fellowship, ARO grant DAAD19-03-1-0082 and NSF grant CCR-9987845.

§Department of Mathematics, Chalmers University of Technology, 412 96 Gothenburg, Sweden. steif@math.chalmers.se. Supported in part by NSF grant DMS-0103841, the Swedish Research Council and the Göran Gustafsson Foundation (KVA).

¶Department of Mathematics, Princeton University, Princeton, NJ 08544, USA. bsudakov@math.princeton.edu. Research supported in part by NSF grant DMS-0106589, DMS-0355497, and by an Alfred P. Sloan fellowship.

to zero as $k \to \infty$. We show that this is indeed the case and provide matching upper and lower bounds on the asymptotically optimal rate (a slowly-decaying polynomial).

- In the case of the k-vertex path graph, we show that it is always optimal for all players to use the same 1-bit function.
- In the general case we show that all players should use monotone functions. We also show, somewhat surprisingly, that for certain trees it is better if not all players use the same function.

Our techniques include the use of the *reverse* Bonami-Beckner inequality. Although the usual Bonami-Beckner has been frequently used before, its reverse counterpart seems not to be well known. To demonstrate its strength, we use it to prove a new isoperimetric inequality for the discrete cube and a new result on the mixing of short random walks on the cube. Another tool that we need is a tight bound on the probability that a Markov chain stays inside certain sets; we prove a new theorem generalizing and strengthening previous such bounds [2, 3, 6]. On the probabilistic side, we use the "reflection principle" and the FKG and related inequalities in order to study the problem on general trees.

1 Introduction

1.1 Non-interactive correlation — the problem and previous work

Our main topic in this paper is the problem of *non-interactive correlation distillation (NICD)*, previously considered in [5, 31, 39]. In its most general form the problem involves k players who receive noisy copies of a uniformly random bit string of length n. The players wish to agree on a single random bit but are not allowed to communicate. The problem is to understand the extent to which the players can successfully distil the correlations in their strings into a shared random bit. This problem is relevant for cryptographic information reconciliation, random beacons in cryptography and security, and coding theory; see [39].

In its most basic form, the problem involves only two players; the first gets a uniformly random string x and the second gets a copy y in which each bit of x is flipped independently with probability ε. If the players try to agree on a shared bit by applying the same Boolean function f to their strings, they will fail with probability $\mathbf{P}[f(x) \neq f(y)]$. This quantity is known as the *noise sensitivity of f at ε*, and the study of noise sensitivity has played an important role in several areas of mathematics and computer science (e.g., inapproximability [26], learning theory [17, 30], hardness amplification [33], mixing of short random walks [27], percolation [10]; see also [34]). In [5], Alon, Maurer, and Wigderson showed that if the players want to use a balanced function f, no improvement over the naive strategy of letting $f(x) = x_1$ can be achieved.

The paper [31] generalized from the two-player problem NICD to a k-player problem, in which a uniformly random string x of length n is chosen, k players receive independent ε-corrupted copies, and they apply (possibly different) balanced Boolean functions to their strings, hoping that all output bits agree. This generalization is equivalent to studying high norms of the Bonami-Beckner operator applied to Boolean functions (i.e., $\|T_\rho f\|_k$); see Section 3 for definitions. The results in [31] include: optimal protocols involve all players using the same function; optimal functions are always monotone; for $k = 3$ the first-bit ('dictator') is best; for fixed ε and fixed n and $k \to \infty$, all players should use the majority function; and, for fixed n and k and $\varepsilon \to 0$ or $\varepsilon \to 1/2$ dictator is best.

Later, Yang [39] considered a different generalization of NICD, in which there are only two players but the corruption model is different from the "binary symmetric channel" noise considered previously. Yang showed that for certain more general noise models, it is still the case that the dictator function is optimal; he also showed an upper bound on the players' success rate in the erasure model.

1.2 NICD on trees; our results

In this paper we propose a natural generalization of the NICD models of [5, 31], extending to a tree topology. In our generalization we have a network in the form of a tree; k of the nodes have a 'player' located on them. One node broadcasts a truly random string of length n. The string follows the edges of the trees and eventually reaches all the nodes. Each edge of the tree independently introduces some noise, acting as a binary symmetric channel with some fixed crossover probability ε. Upon receiving their strings, each player applies a balanced Boolean function, producing one output bit. As usual, the goal of the players is to agree on a shared random bit without any further communication; the protocol is successful if all k parties output the same bit. (For formal definitions, see Section 2.) Note that the problem considered in [31] is just NICD on the star graph of $k+1$ nodes with the players at the k leaves.

We now describe our new results:

The k-leaf star graph: We first study the same k-player star problem considered in [31]. Although this paper found maximizing protocols in certain asymptotic scenarios for the parameters k, n, and ε, the authors left open what is arguably the most interesting setting: ε fixed, k growing arbitrarily large, and n unbounded in terms of ε and k. Although it is natural to guess that the success rate of the players must go to zero exponentially fast in terms of k, this turns out not to be the case; [31] notes that if all players apply the majority function (with n large enough) then they succeed with probability $\Omega(k^{-C(\varepsilon)})$ for some finite constant $C(\varepsilon)$ (the estimate [31] provides is not sharp). [31] left as a major open problem to prove that the success probability goes to 0 as $k \to \infty$.

In this paper we solve this problem. In Theorem 4.1 we show that the success probability must indeed go to zero as $k \to \infty$. Our upper bound is a slowly-decaying polynomial. Moreover, we provide a matching lower bound: this follows from a tight analysis of the majority protocol. The proof of our upper bound depends crucially on the reverse Bonami-Beckner inequality, an important tool that will be described later.

The k-vertex path graph: In the case of NICD on the path graph, we prove in Theorem 5.1 that in the optimal protocol all players should use the same 1-bit function. In order to prove this, we prove in Theorem 5.4, a new tight bound on the probability that a Markov chain stays inside certain sets. Our theorem generalizes and strengthens previous work [2, 3, 6].

Arbitrary trees: In this general case, we show in Theorem 6.3 that there always exists an optimal protocol in which all players use monotone functions. Our analysis uses methods of discrete symmetrization together with the FKG correlation inequality.

In Proposition 6.2 we show that for certain trees it is better if not all players use the same function. This might be somewhat surprising: after all, if all players wish to obtain the same result, won't they be better off using the same function? The intuitive reason the answer to this is negative can be explained by Figure 1: players on the path and players on the star each 'wish' to use a different function. Those on the star wish to use the majority function and those on the path wish to use a dictator function. Indeed, we will show that this strategy yields better success probability than any strategy in which all players use the same function.

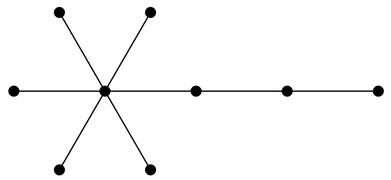

Figure 1: The graph T with $k_1 = 5$ and $k_2 = 3$

1.3 The reverse Bonami-Beckner inequality

Let us start by describing the original inequality (see Theorem 3.1), which considers an operator known as the Bonami-Beckner operator (see Section 3). It is easy to prove that this operator is contractive with respect to any norm. However, the strength in the Bonami-Beckner inequality is that it shows that this operator remains contractive from L_p to L_q for certain values of p and q with $q > p$. This is the reason it is often referred to as a hypercontractive inequality. The inequality was originally proved by Bonami in 1970 [12] and then independently by Beckner in 1973 [8]. It was first used to analyze discrete problems in a remarkable paper by Kahn, Kalai and Linial [27] where they considered the influence of variables on Boolean functions. The inequality has proved to be of great importance in the study of combinatorics of $\{0, 1\}^n$ [15, 16, 22], percolation and random graphs [38, 23, 10, 14] and many other applications [9, 4, 36, 7, 35, 18, 19, 28, 33].

Far less well-known is the fact that the Bonami-Beckner inequality admits a reversed form. This reversed form was first proved by Christer Borell [13] in 1982. Unlike the original inequality, the reverse inequality says that some low norm of the Bonami-Beckner operator applied to a non-negative function can be bounded *below* by some higher norm of the original function. Moreover, the norms involved in the reverse inequality are all at most 1 while the norms in the original inequality are all at least 1. Technically these should not be called norms since they do not satisfy the triangle inequality; nevertheless, we use this terminology.

We are not aware of any previous uses of the reverse Bonami-Beckner inequality for the study of discrete problems. The inequality seems very promising and we hope it will prove useful in the future. To demonstrate its strength, we provide two applications:

Isoperimetric inequality on the discrete cube: As a corollary of the reverse Bonami-Beckner inequality, we obtain in Theorem 3.4 a type of isoperimetric inequality on the discrete cube. It differs from the usual isoperimetric inequality in that the "neighborhood" structure is slightly different. Although it is a simple corollary, we believe that the isoperimetric inequality is interesting. It is also used later to give a sort of hitting time upper-bound for short random walks. In order to illustrate it, let us consider two subsets $S, T \subseteq \{-1,1\}^n$ each containing a constant fraction σ of the 2^n elements of the discrete cube. We now perform the following experiment: we choose a random element of S and flip each of its n coordinates with probability ε for some small ε. What is the probability that the resulting element is in T? Our isoperimetric inequality implies that it is at least some constant independent of n. For example, given any two sets with fractional size $1/3$, the probability that flipping each coordinate with probability .3 takes a random point chosen from the first set into the second set is at least $(1/3)^{1.4/.6} \approx 7.7\%$. We also show that our bound is close to tight. Namely, we analyze the above probability for diametrically opposed Hamming balls and show that it is close to our lower bound.

Short random walks: Our second application, Proposition 3.6, is to short random walks on the discrete cube. We point out however that this does not differ substantially from what was done in the previous paragraph. Consider the following scenario. We have two sets $S, T \subseteq \{-1,1\}^n$ of size at least $\sigma 2^n$ each. We start a walk from a random element of the set S and at each time step proceed with probability $1/2$ to one of its neighbors which we pick randomly. Let τn be the length of the random walk. What is the probability that the random walk terminates in T? If $\tau = C \log n$ for a large enough constant C then it is known that the random walk mixes and therefore we are guaranteed to be in T with probability roughly σ. However, what happens if τ is, say, 0.2? Notice that τn is then less than the diameter of the cube! For certain sets S, the random walk might have zero probability to reach certain vertices, but if σ is at least, say, a constant then there will be some nonzero probability of ending in T. We bound from below the probability that the walk ends in T by a function of σ and τ only. For example, for $\tau = 0.2$, we obtain a bound of roughly σ^{10}. The proof crucially depends on the reverse Bonami-Beckner inequality; to the best of our knowledge, known techniques, such as spectral methods, cannot yield a similar bound.

2 Preliminaries

We now formally define the problem of "non-interactive correlation distillation (NICD) on trees with the binary symmetric channel (BSC)." In general we have four parameters. The first is T, an undirected tree giving the geometry of the problem. Later the vertices of T will become labeled by binary strings, and the edges of T will be thought of as independent binary symmetric channels. The second parameter of the problem is $0 < \rho < 1$ which gives the *correlation* of bits on opposite sides of a channel. By this we mean that if a bit string $x \in \{-1,1\}^n$ passes through the channel producing the bit string $y \in \{-1,1\}^n$ then $\mathbf{E}[x_i y_i] = \rho$ independently for each i. We say that y is a ρ-correlated copy of x. We will also sometimes refer to $\varepsilon = \frac{1}{2} - \frac{1}{2}\rho \in (0, \frac{1}{2})$, which is the probability with which a bit gets flipped — i.e., the crossover probability of the channel. The third parameter of the problem is n, the number of bits in the string at every vertex of T. The fourth parameter of the problem is a subset of the vertex set of T, which we denote by S. We refer to the S as the set of *players*. Frequently S is simply all of $V(T)$, the vertices of T.

To summarize, an instance of the NICD on trees problem is parameterized by:

1. T, an undirected tree;
2. $\rho \in (0, 1)$, the correlation parameter;
3. $n \geq 1$, the string length; and,
4. $S \subseteq V(T)$, the set of players.

Given an instance, the following process happens. Some vertex u of T is given a uniformly random string $x^{(u)} \in \{-1,1\}^n$. Then this string is passed through the BSC edges of T so that every vertex of T becomes labeled by a random string in $\{-1,1\}^n$. It is easy to see that the choice of u does not matter, in the sense that the resulting joint probability distribution on strings for all vertices is the same regardless of u. Formally speaking, we have n independent copies of a "tree-indexed Markov chain;" or a "Markov chain on a tree" [24]. The index set is $V(T)$ and the probability measure $\mathbf{P}$ on $\alpha \in \{-1,1\}^{V(T)}$ is defined by

$$\mathbf{P}(\alpha) = \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2}\rho\right)^{A(\alpha)}\left(\frac{1}{2} - \frac{1}{2}\rho\right)^{B(\alpha)},$$

where $A(\alpha)$ is the number of pairs of neighbors where α agrees and $B(\alpha)$ is the number of pairs of neighbors where α disagrees.

Once the strings are distributed on the vertices of T, the player at the vertex $v \in S$ looks at the string $x^{(v)}$ and applies a (pre-selected) Boolean function $f_v \colon \{-1,1\}^n \to \{-1,1\}$. The goal of the players is to maximize the probability that the bits $f_v(x^{(v)})$ are identical for *all* $v \in S$. In order to rule out the trivial solutions of constant functions and to model the problem of flipping a shared random coin, we insist that all functions f_v be *balanced*; i.e., have equal probability of being -1 or 1. As noted in [31], this does not necessarily ensure that when all players agree on a bit it is conditionally equally likely to be -1 or 1; however, if the functions are in addition antisymmetric, this property does hold. We call a collection of balanced functions $(f_v)_{v\in S}$ a *protocol* for the players S, and we call this protocol *simple* if all of the functions are the same.

To conclude our notation, we write $\mathcal{P}(T,\rho,n,S,(f_v)_{v\in S})$ for the probability that the protocol succeeds – i.e., that all players output the same bit. When the protocol is simple we write merely $\mathcal{P}(T,\rho,n,S,f)$. Our goal is to study the maximum this probability can be over all choices of protocols. We denote by

$$\mathcal{M}(T,\rho,n,S) = \sup_{(f_v)_{v\in S}} \mathcal{P}(T,\rho,n,S,(f_v)_{v\in S}),$$

and define

$$\mathcal{M}(T,\rho,S) = \sup_{n} \mathcal{M}(T,\rho,n,S).$$

3 Reverse Bonami-Beckner and applications

In this section we recall the reverse Bonami-Beckner inequality and obtain as a corollary an isoperimetric inequality on the discrete cube. These results will be useful in analyzing the NICD problem on the star graph and we believe they are of independent interest. We also obtain a new result about the mixing of relatively short random walks on the discrete cube.

3.1 The reverse Bonami-Beckner inequality

We begin with a discussion of the Bonami-Beckner inequality. Recall the Bonami-Beckner operator T_ρ, a linear operator on the space of functions $\{-1,1\}^n \to \mathbb{R}$ defined by

$$(T_\rho f)(x) = \mathbf{E}[f(y)],$$

where y is a ρ-correlated copy of x. The usual Bonami-Beckner inequality, first proved by Bonami [12] and later independently by Beckner [8], is the following:

Theorem 3.1 *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ and $q \geq p \geq 1$. Then*

$$\|T_\rho f\|_q \leq \|f\|_p \qquad \textit{for all } 0 \leq \rho \leq (p-1)^{1/2}/(q-1)^{1/2}.$$

The reverse Bonami-Beckner inequality is the following:

Theorem 3.2 *Let $f\colon \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ be a nonnegative function and let $-\infty < q \leq p \leq 1$. Then*

$$\|T_\rho f\|_q \geq \|f\|_p \qquad \textit{for all } 0 \leq \rho \leq (1-p)^{1/2}/(1-q)^{1/2}. \tag{1}$$

Note that in this theorem we consider r-norms for $r \leq 1$. The case of $r = 0$ is a removable singularity: by $\|f\|_0$ we mean the geometric mean of f. Note also that since T_ρ is a convolution operator, it is positivity-improving for any $\rho < 1$; i.e., when f is nonnegative so too is $T_\rho f$, and if f is further not identically zero, then $T_\rho f$ is everywhere positive.

The reverse Bonami-Beckner theorem is proved in the same way the usual Bonami-Beckner theorem is proved; namely, one proves the inequality in the case of $n = 1$ by elementary means, and then observes that the inequality tensors. Since Borell's original proof may be too compact to be read by some, we provide an expanded version of it in Appendix A for completeness.

We will actually need the following "two-function" version of the reverse Bonami-Beckner inequality which follows easily from the reverse Bonami-Beckner inequality using the (reverse) Hölder inequality (see Appendix A):

Corollary 3.3 *Let $f, g\colon \{-1,1\}^n \to \mathbb{R}^{\geq 0}$ be nonnegative, let $x \in \{-1,1\}^n$ be chosen uniformly at random, and let y be a ρ-correlated copy of x. Then for $-\infty < p, q < 1$,*

$$\mathbf{E}[f(x)g(y)] \geq \|f\|_p \|g\|_q \qquad \textit{for all } 0 \leq \rho \leq (1-p)^{1/2}(1-q)^{1/2}. \tag{2}$$

3.2 A new isoperimetric inequality on the discrete cube

In this subsection we use the reverse Bonami-Beckner inequality to prove an isoperimetric inequality on the discrete cube. Let S and T be two subsets of $\{-1,1\}^n$. Suppose that $x \in \{-1,1\}^n$ is chosen uniformly at random and y is a ρ-correlated copy of x. We obtain the following theorem, which gives a lower bound on the probability that $x \in S$ and $y \in T$ as a function of $|S|/2^n$ and $|T|/2^n$ only.

Theorem 3.4 *Let $S, T \subseteq \{-1,1\}^n$ with $|S| = \exp(-s^2/2)2^n$ and $|T| = \exp(-t^2/2)2^n$. Let x be chosen uniformly at random from $\{-1,1\}^n$ and let y be a ρ-correlated copy of x. Then*

$$\mathbf{P}[x \in S, y \in T] \geq \exp\left(-\frac{1}{2}\frac{s^2 + 2\rho st + t^2}{1-\rho^2}\right). \tag{3}$$

Proof: Take f and g to be the 0-1 characteristic functions of S and T, respectively. Then by Corollary 3.3, for any choice of $p, q < 1$ with $(1-p)(1-q) = \rho^2$, we get

$$\mathbf{P}[x \in S, y \in T] = \mathbf{E}[f(x)g(y)] \geq \|f\|_p \|g\|_q = \exp(-s^2/2p)\exp(-t^2/2q). \tag{4}$$

Write $p = 1 - \rho r$, $q = 1 - \rho/r$ in (4), with $r > 0$. Maximizing the right-hand side as a function of r the best choice is $r = ((t/s) + \rho)/(1 + \rho(t/s))$ which yields in turn

$$p = 1 - \rho r = \tfrac{1-\rho^2}{1+\rho(t/s)}, \quad q = 1 - \rho/r = \tfrac{t}{s}\tfrac{1-\rho^2}{\rho+(t/s)}.$$

(Note that this depends only on the ratio of t and s.) Substituting this choice of r (and hence p and q) into (4) yields $\exp(-\frac{1}{2}\frac{s^2+2\rho st+t^2}{1-\rho^2})$, as claimed. ∎

We now obtain the following corollary of Theorem 3.4.

Corollary 3.5 *Let $S \subseteq \{-1,1\}^n$ have fractional size $\sigma \in [0,1]$, and let $T \subseteq \{-1,1\}^n$ have fractional size σ^α, for $\alpha \geq 0$. If x is chosen uniformly at random from S and y is a ρ-correlated copy of x, then the probability that y is in T is at least*

$$\sigma^{(\sqrt{\alpha}+\rho)^2/(1-\rho^2)}.$$

In particular, if $|S| = |T|$ then this probability is at least $\sigma^{(1+\rho)/(1-\rho)}$.

Proof: Choosing s and t so that $\sigma = \exp(-s^2/2)$ and $\sigma^\alpha = \exp(-t^2/2)$ we obtain

$$\begin{aligned} -\frac{1}{2}(s^2 + 2\rho st + t^2) &= \log\sigma - \rho\sqrt{-2\log\sigma}\sqrt{-2\alpha\log\sigma} + \alpha\log\sigma \\ &= \log\sigma(1 + 2\rho\sqrt{\alpha} + \alpha), \end{aligned}$$

and therefore

$$\exp\left(-\frac{1}{2}\frac{s^2+2\rho st+t^2}{1-\rho^2}\right) = \sigma^{(1+2\rho\sqrt{\alpha}+\alpha)/(1-\rho^2)}.$$

Theorem 3.4 therefore tells us that conditioned on starting in S, the probability of ending in T is at least

$$\sigma^{(1+2\rho\sqrt{\alpha}+\alpha)/(1-\rho^2)-1} = \sigma^{(\sqrt{\alpha}+\rho)^2/(1-\rho^2)}.$$

■

In Subsection 3.4 below we show that the isoperimetric inequality is almost tight. First, we prove a similar bound for random walks on the cube.

3.3 Short random walks on the discrete cube

We can also prove a result of a similar flavor about short random walks on the discrete cube:

Proposition 3.6 *Let $\tau > 0$ be arbitrary and let S and T be two subsets of $\{-1,1\}^n$. Let $\sigma \in [0,1]$ be the fractional size of S and let α be such that the fractional size of T is σ^α. Consider a standard random walk on the discrete cube that starts from a uniformly random vertex in S and walks for τn steps. Here by a standard random walk we mean that at each time step we do nothing with probability $1/2$ and we walk along the ith edge with probability $1/2n$. Let $p^{(\tau n)}(S,T)$ denote the probability that the walk ends in T. Then,*

$$p^{(\tau n)}(S,T) \geq \sigma^{\frac{(\sqrt{\alpha}+\exp(-\tau))^2}{1-\exp(-2\tau)}} - O\Big(\frac{\sigma^{(-1+\alpha)/2}}{\tau n}\Big).$$

In particular, when $|S| = |T| = \sigma 2^n$ then $p^{(\tau n)}(S,T) \geq \sigma^{\frac{1+\exp(-\tau)}{1-\exp(-\tau)}} - O(\frac{1}{\tau n})$.

The Laurent series of $\frac{1+e^{-\tau}}{1-e^{-\tau}}$ is $2/\tau + \tau/6 - O(\tau^3)$ so for $1/\log n \ll \tau \ll 1$ our bound is roughly $\sigma^{2/\tau}$.

For the proof we will first need a simple lemma:

Lemma 3.7 *For $y > 0$ and any $0 \leq x \leq y$,*

$$0 \leq e^{-x} - (1-x/y)^y \leq O(1/y).$$

Proof: The expression above can be written as

$$e^{-x} - e^{y \log(1-x/y)}.$$

We have $\log(1 - x/y) \leq -x/y$ and hence we obtain the first inequality. For the second inequality, notice that if $x \geq 0.1y$ then both expressions are of the form $e^{-\Omega(y)}$ which is certainly $O(1/y)$. On the other hand, if $0 \leq x < 0.1y$ then there is a constant c such that

$$\log(1 - x/y) \geq -x/y - cx^2/y^2.$$

The Mean Value Theorem implies that for $0 \leq a \leq b$, $e^{-a} - e^{-b} \leq e^{-a}(b - a)$. Hence,

$$e^{-x} - e^{y \log(1-x/y)} \leq e^{-x}(-y \log(1 - x/y) - x) \leq \frac{cx^2 e^{-x}}{y}.$$

The lemma now follows because $x^2 e^{-x}$ is uniformly bounded for $x \geq 0$. ■

We now prove Proposition 3.6. The proof uses Fourier analysis; for the required definitions see, e.g., [27].

Proof: Let x be a uniformly random point in $\{-1, 1\}^n$ and y a point generated by taking a random walk of length τn starting from x. Let f and g be the 0-1 indicator functions of S and T, respectively, and say $E[f] = \sigma$, $E[g] = \sigma^\alpha$. Then by writing f and g in their Fourier decomposition we obtain that

$$\sigma \cdot p^{(\tau n)}(S, T) = \mathbf{P}[x \in S, y \in T] = \mathbf{E}[f(x)g(y)] = \sum_{U,V} \hat{f}(U)\hat{g}(V)\mathbf{E}[x_U y_V]$$

where U and V range over all subsets of $\{1, \ldots, n\}$. Note that $\mathbf{E}[x_U y_V]$ is zero unless $U = V$. Therefore we may write

$$\begin{aligned}
& \sigma p^{(\tau n)}(S,T) \\
= \quad & \sum_U \hat{f}(U)\hat{g}(U)\mathbf{E}[(xy)_U] = \sum_U \hat{f}(U)\hat{g}(U)\Big(1 - \frac{|U|}{n}\Big)^{\tau n} \\
= \quad & \sum_U \hat{f}(U)\hat{g}(U)\exp(-\tau|U|) \\
+ \quad & \sum_U \hat{f}(U)\hat{g}(U)\Big[\Big(1 - \frac{|U|}{n}\Big)^{\tau n} - \exp(-\tau|U|)\Big] \\
= \quad & \langle f, T_{\exp(-\tau)} g\rangle + \sum_U \hat{f}(U)\hat{g}(U)\Big[\Big(1 - \frac{|U|}{n}\Big)^{\tau n} - \exp(-\tau|U|)\Big] \\
\geq \quad & \langle f, T_{\exp(-\tau)} g\rangle - \max_{|U|}\Big|\Big(1 - \frac{|U|}{n}\Big)^{\tau n} - \exp(-\tau|U|)\Big| \sum_U |\hat{f}(U)\hat{g}(U)|.
\end{aligned}$$

By Corollary 3.5,

$$\sigma^{-1}\langle f, T_{\exp(-\tau)} g\rangle \geq \sigma^{\frac{(\sqrt{\alpha}+\exp(-\tau))^2}{1-\exp(-2\tau)}}.$$

By Cauchy-Schwarz and Parseval's identity,

$$\sum_U |\hat{f}(U)\hat{g}(U)| \leq \|\hat{f}\|_2\|\hat{g}\|_2 = \|f\|_2\|g\|_2 = \sigma^{(1+\alpha)/2}.$$

In addition, from Lemma 3.7 with $x = \tau|U|$ and $y = \tau n$ we have that

$$\max_{|U|}\Big|\Big(1 - \frac{|U|}{n}\Big)^{\tau n} - \exp(-\tau|U|)\Big| = O\Big(\frac{1}{\tau n}\Big).$$

Hence,

$$p^{(\tau n)}(S,T) \geq \sigma^{\frac{(\sqrt{\alpha}+\exp(-\tau))^2}{1-\exp(-2\tau)}} - O\Big(\frac{\sigma^{(-1+\alpha)/2}}{\tau n}\Big).$$

■

3.4 Tightness of the isoperimetric inequality

We now show that Theorem 3.4 is almost tight. Suppose $x \in \{-1,1\}^n$ is chosen uniformly at random and y is a ρ-correlated copy of x. Let us begin by understanding more about how x and y are distributed. Define

$$\Sigma(\rho) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

and recall that the density function of the bivariate normal distribution $\phi_{\Sigma(\rho)} : \mathbb{R}^2 \to \mathbb{R}^{\geq 0}$ with mean 0 and covariance matrix $\Sigma(\rho)$, is given by

$$\begin{aligned}\phi_{\Sigma(\rho)}(x,y) &= (2\pi)^{-1}(1-\rho^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{x^2-2\rho xy+y^2}{1-\rho^2}\right) \\ &= (1-\rho^2)^{-\frac{1}{2}}\phi(x)\phi\left(\frac{y-\rho x}{(1-\rho^2)^{\frac{1}{2}}}\right).\end{aligned}$$

Here ϕ denotes the standard normal density function on $\mathbb{R}$, $\phi(x) = (2\pi)^{-1/2}e^{-x^2/2}$.

Proposition 3.8 *Let $x \in \{-1,1\}^n$ be chosen uniformly at random, and let y be a ρ-correlated copy of x. Let $X = n^{-1/2}\sum_{i=1}^n x_i$ and $Y = n^{-1/2}\sum_{i=1}^n y_i$. Then as $n \to \infty$, the pair of random variables (X,Y) approaches the distribution $\phi_{\Sigma(\rho)}$. As an error bound, we have that for any convex region $R \subseteq \mathbb{R}^2$,*

$$\left|\mathbf{P}\big[(X,Y) \in R\big] - \iint_R \phi_{\Sigma(\rho)}(x,y)\,dy\,dx\right| \leq O((1-\rho^2)^{-1/2}n^{-1/2}).$$

Proof: This follows from the Central Limit Theorem (see, e.g., [20]), noting that for each coordinate i, $\mathbf{E}[x_i^2] = \mathbf{E}[y_i^2] = 1$, $\mathbf{E}[x_iy_i] = \rho$. The Berry-Esséen-type error bound is proved in Sazonov [37, p. 10, Item 6]. ■

Using this proposition we can obtain the following result for two diametrically opposed Hamming balls.

Proposition 3.9 *Fix $s,t > 0$, and let $S, T \subseteq \{-1,1\}^n$ be diametrically opposed Hamming balls, with $S = \{x\colon \sum_i x_i \leq -sn^{1/2}\}$ and $T = \{x\colon \sum_i x_i \geq tn^{1/2}\}$. Let x be chosen uniformly at random from $\{-1,1\}^n$ and let y be a ρ-correlated copy of x. Then we have*

$$\lim_{n\to\infty} \mathbf{P}[x \in S, y \in T] \leq \frac{\sqrt{1-\rho^2}}{2\pi s(\rho s+t)} \exp\left(-\frac{1}{2}\frac{s^2+2\rho st+t^2}{1-\rho^2}\right).$$

Proof:

$$\begin{aligned}
& \lim_{n\to\infty} \mathbf{P}[x \in S, y \in T] \\
= & \int_s^\infty \int_t^\infty \phi_{\Sigma(-\rho)}(x,y)\, dy\, dx \quad \text{(By Lemma 3.8)} \\
\le & \int_s^\infty \int_t^\infty \frac{x(\rho x + y)}{s(\rho s + t)} \phi_{\Sigma(-\rho)}(x,y)\, dy\, dx \\
& \left(\text{Since } \frac{x(\rho x + y)}{s(\rho s + t)} \ge 1 \text{ on } x \ge s, y \ge t \right) \\
= & \frac{1}{\sqrt{1-\rho^2}} \int_s^\infty \int_t^\infty \frac{x(\rho x + y)}{s(\rho s + t)} \phi(x) \phi\left(\frac{y + \rho x}{\sqrt{1-\rho^2}} \right) dy\, dx \\
\le & \frac{1}{\sqrt{1-\rho^2}} \int_s^\infty \int_{\rho s + t}^\infty \frac{xz}{s(\rho s + t)} \phi(x) \phi\left(\frac{z}{\sqrt{1-\rho^2}} \right) dz\, dx \\
& \left(\text{Using } z = \rho x + y \text{ and noting } \frac{xz}{s(\rho s + t)} \ge 1 \text{ on } x \ge s, z \ge \rho s + t \right) \\
= & \frac{1}{s(\rho s + t)\sqrt{1-\rho^2}} \left(\int_s^\infty x\phi(x) dx \right) \left(\int_{\rho s + t}^\infty z\phi\left(\frac{z}{\sqrt{1-\rho^2}} \right) dz \right) \\
= & \frac{\sqrt{1-\rho^2}}{s(\rho s + t)} \phi(s) \phi\left(\frac{\rho s + t}{\sqrt{1-\rho^2}} \right) \\
= & \frac{\sqrt{1-\rho^2}}{2\pi s(\rho s + t)} \exp\left(-\frac{1}{2} \frac{s^2 + 2\rho s t + t^2}{1-\rho^2} \right).
\end{aligned}$$

The result follows. ■

By the Central Limit Theorem, the set S in the above statement satisfies (see [1, 26.2.12]),

$$\lim_{n\to\infty} |S| 2^{-n} = \frac{1}{\sqrt{2\pi}} \int_s^\infty e^{-x^2/2}\, dx \sim \exp(-s^2/2)/(\sqrt{2\pi} s).$$

For large s (i.e., small $|S|$) this is dominated by $\exp(-s^2/2)$. A similar statement holds for T. This shows that Theorem 3.4 is nearly tight.

4 The best asymptotic success rate in the k-star

In this section we consider the NICD problem on the star. Let Star_k denote the star graph on $k+1$ vertices and let S_k denote its k leaf vertices. We shall study the same problem considered in [31]; i.e., determining $\mathcal{M}(\mathrm{Star}_k, \rho, S_k)$. Note that it was shown in that paper that the best protocol in this case is always simple (i.e., all players should use the same function).

The following theorem determines rather accurately the asymptotics of $\mathcal{M}(\mathrm{Star}_k, \rho, S_k)$:

Theorem 4.1 *Fix $\rho \in (0, 1]$ and let $\nu = \nu(\rho) = \frac{1}{\rho^2} - 1$. Then for $k \to \infty$,*

$$\mathcal{M}(\mathrm{Star}_k, \rho, S_k) = \tilde{\Theta}\left(k^{-\nu}\right),$$

where $\tilde{\Theta}(\cdot)$ denotes asymptotics to within a subpolynomial ($k^{o(1)}$) factor. The lower bound is achieved asymptotically by the majority function MAJ_n with n sufficiently large.

Note that if the corruption probability is very small (i.e., ρ is close to 1), we obtain that the success rate only drops off as a very mild function of k. We first prove the upper bound.

Proof of upper bound: We know that all optimal protocols are simple, so assume all players use the same balanced function $f : \{-1, 1\}^n \to \{-1, 1\}$. Let $F_{-1} = f^{-1}(-1)$ and $F_1 = f^{-1}(1)$ be the sets where f obtains the values -1 and 1 respectively. The center of the star gets a uniformly random string x, and then independent ρ-correlated copies are given to the k leaf players. Let y denote a typical such copy. The probability that all players output -1 is thus $\mathbf{E}_x[\mathbf{P}[f(y) = -1|x]^k]$. We will show that this probability is $\tilde{O}(k^{-\nu})$. This complete the proof since we can replace f by $-f$ and get the same bound for the probability that all players output 1.

Suppose $\mathbf{E}_x[\mathbf{P}[f(y) = -1|x]^k] \geq 2\delta$ for some δ; we will show δ must be small. Define

$$S = \{x : \mathbf{P}[f(y) = -1 \mid x]^k \geq \delta\}.$$

By Markov's inequality we must have $|S| \geq \delta 2^n$. Now on one hand, by the definition of S,

$$\mathbf{P}[y \in F_1 \mid x \in S] \leq 1 - \delta^{1/k}. \tag{5}$$

On the other hand, applying Corollary 3.5 with $T = F_1$ and $\alpha \leq 1/\log_2(1/\delta) < 1/\log(1/\delta)$ (since $|F_1| = \frac{1}{2}2^n$), we get

$$\mathbf{P}[y \in F_1 \mid x \in S] \geq \delta^{(\log^{-1/2}(1/\delta)+\rho)^2/(1-\rho^2)}. \tag{6}$$

Combining (5) and (6) yields the desired upper bound on δ in terms of k, $\delta \leq k^{-\nu+o(1)}$ by the following calculations. We have

$$1 - \delta^{1/k} \geq \delta^{(\log^{-1/2}(1/\delta)+\rho)^2/(1-\rho^2)}.$$

We want to show that the above inequality cannot hold if

$$\delta \geq \left(\frac{e^{c\sqrt{\log k}}}{k}\right)^{\nu}, \tag{7}$$

where $c = c(\rho)$ is some constant. We will show that if δ satisfies (7) and c is sufficiently large then for all large k

$$\delta^{1/k} + \delta^{(\log^{-1/2}(1/\delta)+\rho)^2/(1-\rho^2)} > 1.$$

Note first that

$$\delta^{1/k} > \left(\frac{1}{k}\right)^{\frac{\nu}{k}} = \exp\left(-\frac{\nu \log k}{k}\right) > 1 - \frac{\nu \log k}{k}. \tag{8}$$

On the other hand,

$$\delta^{(\log^{-1/2}(1/\delta)+\rho)^2/(1-\rho^2)} = \delta^{-\log^{-1}\delta/(1-\rho^2)} \cdot \delta^{2\rho\log^{-1/2}(1/\delta)/(1-\rho^2)} \cdot \delta^{\rho^2/(1-\rho^2)}. \tag{9}$$

Note that

$$\delta^{\rho^2/(1-\rho^2)} = \delta^{1/\nu} \geq \frac{e^{c\sqrt{\log k}}}{k}$$

and

$$\begin{aligned}\delta^{2\rho\log^{-1/2}(1/\delta)/(1-\rho^2)} &= \exp\left(-\frac{2\rho}{1-\rho^2}\sqrt{\log(1/\delta)}\right)\\ &\geq \exp\left(-\frac{2\rho}{1-\rho^2}\sqrt{\nu\log k}\right).\end{aligned}$$

Finally,

$$\delta^{-\log^{-1}\delta/(1-\rho^2)} = \exp\left(-\frac{1}{1-\rho^2}\right).$$

Thus if $c = c(\rho)$ is sufficiently large then the left hand side of (9) is at least $\frac{\nu \log k}{k}$. This implies the desired contradiction by (7) and (8). ■

Proof of lower bound: We will analyze the protocol where all players use MAJ_n, similarly to the analysis of [31]. Our analysis here is more careful resulting in a tighter bound.

We begin by showing that the probability with which all players agree if they use MAJ_n, in the case of fixed k and $n \to \infty$, is:

$$\lim_{\substack{n\to\infty \\ n \text{ odd}}} \mathcal{P}(\mathrm{Star}_k, \rho, n, S_k, \mathsf{MAJ}_n) = 2\nu^{1/2}(2\pi)^{(\nu-1)/2} \int_0^1 t^k I(t)^{\nu-1}\, dt, \qquad (10)$$

where $I = \phi \circ \Phi^{-1}$ is the so-called Gaussian isoperimetric function, with $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $\Phi(x) = \int_{-\infty}^{x} \phi(t)dt$ the density and distribution functions of a standard normal random variable respectively.

Apply Proposition 3.8, with $X \sim N(0,1)$ representing $n^{-1/2}$ times the sum of the bits in the string at the star's center, and $Y|X \sim N(\rho X, 1-\rho^2)$ representing $n^{-1/2}$ times the sum of the bits in a typical leaf player's string. Thus as $n \to \infty$, the probability that all players output $+1$ when using MAJ_n is precisely

$$\int_{-\infty}^{\infty} \Phi\left(\frac{\rho\, x}{\sqrt{1-\rho^2}}\right)^k \phi(x)\, dx = \int_{-\infty}^{\infty} \Phi\left(\nu^{-1/2}x\right)^k \phi(x)\, dx.$$

Since MAJ_n is antisymmetric, the probability that all players agree on $+1$ is the same as the probability they all agree on -1. Making the change of variables $t = \Phi(\nu^{-1/2}x)$, $x = \nu^{1/2}\Phi^{-1}(t)$, $dx = \nu^{1/2}I(t)^{-1}\, dt$, we get

$$\begin{aligned}
\lim_{\substack{n\to\infty \\ n \text{ odd}}} \mathcal{P}(\mathrm{Star}_k, \rho, n, S_k, \mathsf{MAJ}_n) &= 2\nu^{1/2} \int_0^1 \frac{t^k \phi(\nu^{1/2}\Phi^{-1}(t))}{I(t)}\, dt \\
&= 2\nu^{1/2}(2\pi)^{(\nu-1)/2} \int_0^1 t^k I(t)^{\nu-1}\, dt,
\end{aligned}$$

as claimed.

We now estimate the integral in (10). It is known (see, e.g., [11]) that $I(s) \geq J(s(1-s))$, where $J(s) = s\sqrt{\ln(1/s)}$. We will forego the marginal improvements given by taking the logarithmic term and simply use the estimate $I(t) \geq t(1-t)$. We then get

$$\begin{aligned}
\int_0^1 t^k I(t)^{\nu-1}\, dt &\geq \int_0^1 t^k (t(1-t))^{\nu-1}\, dt \\
&= \frac{\Gamma(\nu)\Gamma(k+\nu)}{\Gamma(k+2\nu)} \qquad \text{([1, 6.2.1, 6.2.2])} \\
&\geq \Gamma(\nu)(k+2\nu)^{-\nu} \qquad \text{(Stirling approximation).}
\end{aligned}$$

Substituting this estimate into (10) we get

$$\lim_{n\to\infty} \mathcal{P}(\mathrm{Star}_k, \rho, n, S_k, \mathsf{MAJ}_n) \geq c(\nu)k^{-\nu},$$

where $c(\nu) > 0$ depends only on ρ, as desired. ■

We remark that in the upper bound above we have in effect proved the following theorem regarding high norms of the Bonami-Beckner operator applied to Boolean functions:

Theorem 4.2 *Let $f\colon \{-1,1\}^n \to \{0,1\}$ and suppose $\mathbf{E}[f] \leq 1/2$. Then for any fixed $\rho \in (0,1]$, as $k \to \infty$, $\|T_\rho f\|_k^k \leq k^{-\nu+o(1)}$, where $\nu = \frac{1}{\rho^2} - 1$.*

Since we are trying to bound a high norm of $T_\rho f$ knowing the norms of f, it would seem as though the usual Bonami-Beckner inequality would be effective. However this seems not to be the case: a straightforward application yields

$$\begin{aligned} \|T_\rho f\|_k &\leq \|f\|_{\rho^2(k-1)+1} = \mathbf{E}[f]^{1/(\rho^2(k-1)+1)} \\ \Rightarrow \qquad \|T_\rho f\|_k^k &\leq (1/2)^{k/(\rho^2(k-1)+1)} \approx (1/2)^{1/\rho^2}, \end{aligned}$$

only a constant upper bound.

5 The optimal protocol on the path

In this section we prove the following theorem which gives a complete solution to the NICD problem on a path. In this case, simple dictator protocols are the unique optimal protocols, and any other simple protocol is exponentially worse as a function of the number of players.

Theorem 5.1

- *Let $\mathrm{Path}_k = \{v_0, v_1, \ldots, v_k\}$ be the path graph of length k, and let S be any subset of Path_k of size at least two. Then simple dictator protocols are the unique optimal protocols for $\mathcal{P}(\mathrm{Path}_k, \rho, n, S, (f_v))$. In particular, if $S = \{v_{i_0}, \ldots, v_{i_\ell}\}$ where $i_0 < i_1 < \cdots < i_\ell$, then we have*

$$\mathcal{M}(\mathrm{Path}_k, \rho, S) = \prod_{j=1}^{\ell} \left(\frac{1}{2} + \frac{1}{2}\rho^{i_j - i_{j-1}}\right).$$

- *Moreover, for every ρ and n there exists $c = c(\rho, n) < 1$ such that if $S = \mathrm{Path}_k$ then for any simple protocol f which is not a dictator,*

$$\mathcal{P}(\mathrm{Path}_k, \rho, n, S, f) \leq \mathcal{P}(\mathrm{Path}_k, \rho, n, S, \mathcal{D})c^{|S|-1}$$

where $\mathcal{D}$ denotes the dictator function.

5.1 A bound on inhomogeneous Markov chains

A crucial component of the proof of Theorem 5.1 is a bound on the probability that a reversible Markov chain stays inside certain sets. In this subsection, we derive such a bound in a fairly general setting. Moreover, we exactly characterize the cases in which the bound is tight. This is a generalization of Theorem 9.2.7 in [6] and of results in [2, 3].

Let us first recall some basic facts concerning reversible Markov chains. Consider an irreducible Markov chain on a finite set S. We denote by $M = \big(m(x,y)\big)_{x,y\in S}$ the matrix of transition probabilities of this chain, where $m(x,y)$ is the probability to move in one step from x to y. We will always assume that M is ergodic (i.e., irreducible and aperiodic).

The rule of the chain can be expressed by the simple equation $\mu_1 = \mu_0 M$, where μ_0 is a starting distribution on S and μ_1 is the distribution obtained after one step of the Markov chain (we think of both as row vectors). By definition, $\sum_y m(x,y) = 1$. Therefore, the largest eigenvalue of M is 1 and a corresponding right eigenvector has all its coordinates equal to 1. Since M is ergodic, it has a unique (left and right) eigenvector corresponding to an eigenvalue with absolute value 1. We denote the unique right eigenvector $(1,\ldots,1)^t$ by $\mathbf{1}$. We denote by π the unique left eigenvector corresponding to the eigenvalue 1 whose coordinate sum is 1. π is the stationary distribution of the Markov chain. Since we are dealing with a Markov chain whose distribution π is not necessarily uniform it will be convenient to work in $L^2(S,\pi)$. In other words, for any two functions f and g on S we define the inner product $\langle f,g\rangle = \sum_{x\in S}\pi(x)f(x)g(x)$. The norm of f equals $\|f\|_2 = \sqrt{\langle f,f\rangle} = \sqrt{\sum_{x\in S}\pi(x)f^2(x)}$.

Definition 5.2 *A transition matrix $M = \big(m(x,y)\big)_{x,y\in S}$ for a Markov chain is reversible with respect to a probability distribution π on S if $\pi(x)m(x,y) = \pi(y)m(y,x)$ holds for all x, y in S.*

It is known that if M is reversible with respect to π, then π is the stationary distribution of M. Moreover, the corresponding operator taking $L^2(S,\pi)$ to itself defined by $Mf(x) = \sum_y m(x,y)f(y)$ is self-adjoint, i.e., $\langle Mf, g\rangle = \langle f, Mg\rangle$ for all f, g. Thus, it follows that M has a complete set of orthonormal (with respect to the inner product defined above) eigenvectors with real eigenvalues.

Definition 5.3 *If M is reversible with respect to π and $\lambda_1 \le \ldots \le \lambda_{r-1} \le \lambda_r = 1$ are the eigenvalues of M, then the* spectral gap *of M is defined to be $\delta = \min\big\{|-1-\lambda_1|, |1-\lambda_{r-1}|\big\}$.*

For transition matrices $M_1, M_2, \ldots$ on the same space S, we can consider the time-inhomogeneous Markov chain which at time 0 starts in some state (perhaps randomly) and then jumps using the matrices $M_1, M_2, \ldots$ in this order. In this way, M_i will govern the jump from time $i-1$ to time i. We write I_A for the indicator function of the set A and π_A for the function defined by $\pi_A(x) = I_A(x)\pi(x)$ for all x. Similarly, we define $\pi(A) = \sum_{x \in A} \pi(x)$. The following theorem provides a tight estimate on the probability that the inhomogeneous Markov chain stays inside certain specified sets.

Theorem 5.4 *Let $M_1, M_2, \ldots, M_k$ be ergodic transition matrices on the state space S, all of which are reversible with respect to the same probability measure π with full support. Let $\delta_i > 0$ be the spectral gap of matrix M_i and let $A_0, A_1, \ldots, A_k$ be nonempty subsets of S.*

- *If $\{X_i\}_{i=0}^k$ denotes the time-inhomogeneous Markov chain using the matrices $M_1, M_2, \ldots, M_k$ and starting according to distribution π, then $\mathbf{P}[X_i \in A_i \;\; \forall i = 0 \ldots k]$ is at most*

$$\sqrt{\pi(A_0)}\sqrt{\pi(A_k)} \prod_{i=1}^{k} \Big[1 - \delta_i \Big(1 - \sqrt{\pi(A_{i-1})}\sqrt{\pi(A_i)}\Big)\Big]. \qquad (11)$$

- *Suppose we further assume that for all i, $\delta_i < 1$ and that $\lambda_1^i > -1 + \delta_i$ (λ_1^i here is the smallest eigenvalue for the ith chain). Then equality in (11) holds if and only if all the sets A_i are the same set A and for all i the function $I_A - \pi(A)\mathbf{1}$ is an eigenfunction of M_i corresponding to the eigenvalue $1-\delta_i$.*

- *Finally, suppose even further that all the chains M_i are the same chain M. Then there exists a constant $c = c(M) < 1$ such that for all sets A for which strict inequality holds in (11) when each A_i is taken to be A, we have the stronger inequality*

$$\mathbf{P}[X_i \in A \;\; \forall i = 0 \ldots k] \leq c^k \pi(A) \prod_{i=1}^{k} \left[1 - \delta(1 - \pi(A))\right]$$

  for every k.

Remark: Notice that if all the sets A_i have π-measure at most $\sigma < 1$ and all the M_i's have spectral gap at least δ, then the upper bound in (11) is bounded above by

$$\sigma[\sigma + (1-\delta)(1-\sigma)]^k.$$

Hence, the above theorem generalizes Theorem 9.2.7 in [6] and strengthens the estimate from [3].

5.2 Proof of Theorem 5.1

If we look at the NICD process restricted to positions $x_{i_0}, x_{i_1}, \ldots, x_{i_\ell}$, we obtain a time-inhomogeneous Markov chain $\{X_j\}_{j=0}^{\ell}$ where X_0 is uniform on $\{-1,1\}^n$ and the ℓ transition operators are powers of the Bonami-Beckner operator,

$$T_\rho^{i_1-i_0}, T_\rho^{i_2-i_1}, \cdots, T_\rho^{i_\ell-i_{\ell-1}}.$$

Equivalently, these operators are $T_{\rho^{i_1-i_0}}$, $T_{\rho^{i_2-i_1}}$, $\ldots$, $T_{\rho^{i_\ell-i_{\ell-1}}}$. It is easy to see that the eigenvalues of T_ρ are $1 > \rho > \rho^2 > \cdots > \rho^n$ and therefore its spectral gap is $1-\rho$. Now a protocol for the $\ell+1$ players consists simply of $\ell+1$ subsets $A_0, \ldots, A_\ell$ of $\{-1,1\}^n$, where A_j is a set of strings in $\{-1,1\}^n$ on which the jth player outputs the bit 1. Thus, each A_j has size 2^{n-1}, and the success probability of this protocol is simply

$$\mathbf{P}[X_i \in A_i \ \ \forall i = 0 \ldots \ell] + \mathbf{P}[X_i \in \bar{A}_i \ \ \forall i = 0 \ldots \ell].$$

But by Theorem 5.4 each summand is bounded by

$$\frac{1}{2}\prod_{j=1}^{\ell}\left(\frac{1}{2} + \frac{\rho^{i_j - i_{j-1}}}{2}\right),$$

yielding our desired upper bound. It is easy to check that this is precisely the success probability of a simple dictator protocol.

To complete the proof of the first part it remains to show that every other protocol does strictly worse. By the second statement of Theorem 5.4 (and the fact that the simple dictator protocol achieves the upper bound in Theorem 5.4), we can first conclude that any optimal protocol is a simple protocol, i.e., all the sets A_j are identical. Let A be the set corresponding to any potentially optimal simple protocol. By Theorem 5.4 again the function $I_A - (|A|2^{-n})\mathbf{1} = I_A - \frac{1}{2}\mathbf{1}$ must be an eigenfunction of T_{ρ^r} for some r corresponding to its second largest eigenvalue ρ^r. This implies that $f = 2I_A - \mathbf{1}$ must be a balanced *linear* function, $f(x) = \sum_{|S|=1} \hat{f}(S) x_S$. It is well known (see, e.g., [32]) that the only such Boolean functions are dictators. This completes the proof of the first part. The second part of the theorem follows immediately from the third part of Theorem 5.4 ∎

5.3 Inhomogeneous Markov chains

In order to prove Theorem 5.4 we need a lemma that provides a bound for one step of the Markov chain.

Lemma 5.5 *Let M be an ergodic transition matrix for a Markov chain on the set S that is reversible with respect to the probability measure π and has spectral gap $\delta > 0$. Let A_1 and A_2 be two subsets of S and let P_1 and P_2 be the corresponding projection operators on $L^2(S, \pi)$ (i.e., $P_i f(x) = f(x) I_{A_i}(x)$ for every function f on S). Then*

$$\|P_1 M P_2\| \leq 1 - \delta \left(1 - \sqrt{\pi(A_1)}\sqrt{\pi(A_2)}\right),$$

where the norm on the left is the operator norm for operators from $L^2(S, \pi)$ into itself.

Further, suppose we assume that $\delta < 1$ and that $\lambda_1 > -1 + \delta$. Then equality holds above if and only if $A_1 = A_2$ and the function $I_{A_1} - \pi(A_1)\mathbf{1}$ is an eigenfunction of M corresponding to $1 - \delta$.

Proof: Let $e_1, \ldots, e_{r-1}, e_r = \mathbf{1}$ be an orthonormal basis of right eigenvectors of M with corresponding eigenvalues $\lambda_1 \leq \ldots \leq \lambda_{r-1} \leq \lambda_r = 1$. For a function f on S, denote by $\mathsf{supp}(f) = \{x \in S \mid f(x) \neq 0\}$. It is easy to see that $\|P_1 M P_2\|$ is given by

$$\sup\left\{|\langle f_1, M f_2\rangle| : \|f_1\|_2 = 1, \|f_2\|_2 = 1, \mathsf{supp}(f_1) \subseteq A_1, \mathsf{supp}(f_2) \subseteq A_2\right\}.$$

Given such f_1 and f_2, expand them as

$$f_1 = \sum_{i=1}^{r} u_i e_i, \quad f_2 = \sum_{i=1}^{r} v_i e_i$$

and observe that for $j = 1, 2$,

$$|\langle f_j, \mathbf{1}\rangle| = |\langle f_j, I_{A_j}\rangle| \leq \|f_j\|_2 \|I_{A_j}\|_2 = \sqrt{\pi(A_j)}. \tag{12}$$

But now by the orthonormality of the e_i's we have

$$\begin{aligned}
|\langle f_1, M f_2\rangle| &= \left|\sum_{i=1}^{r} \lambda_i u_i v_i\right| \sum_{i=1}^{r} |\lambda_i u_i v_i| && (13)\\
&\leq |\langle f_1, \mathbf{1}\rangle\langle f_2, \mathbf{1}\rangle| + (1 - \delta) \sum_{i \leq r-1} |u_i v_i| && (14)\\
&\leq |\langle f_1, \mathbf{1}\rangle\langle f_2, \mathbf{1}\rangle| + (1 - \delta)(1 - |\langle f_1, \mathbf{1}\rangle\langle f_2, \mathbf{1}\rangle|) && (15)\\
&\leq \sqrt{\pi(A_1)}\sqrt{\pi(A_2)} + (1 - \delta)\left(1 - \sqrt{\pi(A_1)}\sqrt{\pi(A_2)}\right) && (16)\\
&= 1 - \delta\left(1 - \sqrt{\pi(A_1)}\sqrt{\pi(A_2)}\right).
\end{aligned}$$

For the third inequality, we used that $\sum_i |u_i v_i| \leq 1$ which follows from f_1 and f_2 having norm 1.

As for the second part of the lemma, if equality holds then all the derived inequalities must be equalities. In particular, if (12) holds as an equality, it follows that for $j = 1, 2$, $f_j = \pm\big(1/\sqrt{\pi(A_j)}\big) I_{A_j}$. Since $\delta < 1$ is assumed, it follows from the third inequality in (13) that we must also have that $\sum_i |u_i v_i| = 1$ from which we can conclude that $|u_i| = |v_i|$ for all i. Since $-1 + \delta$ is not an eigenvalue, for the second inequality in (13) to hold we must have that the only nonzero u_i's (or v_i's) correspond to the eigenvalues 1 and $1 - \delta$. Next, for the first inequality in (13) to hold, we must have that $u = (u_1, \ldots, u_n) = \pm v = (v_1, \ldots, v_n)$ since λ_i can only be 1 or $1 - \delta$ and $|u_i| = |v_i|$ for each i. This gives us that $f_1 = \pm f_2$ and therefore $A_1 = A_2$.

Finally, we also get that $f_1 - \langle f_1, \mathbf{1}\rangle \mathbf{1}$ is an eigenfunction of M corresponding to the eigenvalue $1 - \delta$. To conclude the proof, note that if $A_1 = A_2$ and $I_{A_1} - \pi(A_1)\mathbf{1}$ is an eigenfunction of M corresponding to $1 - \delta$, then it is easy to see that when we take $f_1 = f_2 = I_{A_1} - \pi(A_1)\mathbf{1}$, all inequalities in our proof become equalities. ■

Proof of Theorem 5.4: Let P_i denote the projection onto A_i, as in Lemma 5.5. It is easy to see that

$$\mathbf{P}[X_i \in A_i \ \ \forall i = 0 \ldots k] = \pi_{A_0} P_0 M_1 P_1 M_2 \cdots P_{k-1} M_k P_k I_{A_k}.$$

Rewriting in terms of the inner product, this is equal to

$$\langle I_{A_0}, (P_0 M_1 P_1 M_2 \cdots P_{k-1} M_k P_k) I_{A_k} \rangle.$$

By Cauchy-Schwarz it is at most

$$\|I_{A_0}\|_2 \|I_{A_k}\|_2 \|P_0 M_1 P_1 M_2 \cdots P_{k-1} M_k P_k\|,$$

where the third factor is the norm of $P_0 M_1 P_1 M_2 \cdots P_{k-1} M_k P_k$ as an operator from $L^2(S, \pi)$ to itself. Since $P_i^2 = P_i$ (being a projection), this in turn is equal to

$$\sqrt{\pi(A_0)}\sqrt{\pi(A_k)} \|(P_0 M_1 P_1)(P_1 M_2 P_2) \cdots (P_{k-1} M_k P_k)\|.$$

By Lemma 5.5 we have that for all $i = 1, \ldots, k$

$$\|P_{i-1} M_\ell P_i\| \leq 1 - \delta_i \Big(1 - \sqrt{\pi(A_{i-1})}\sqrt{\pi(A_i)}\Big).$$

Hence

$$\Big\| \prod_{i=1}^{k} (P_{i-1} M P_i) \Big\| \le \prod_{i=1}^{k} \Big[1 - \delta_i \Big(1 - \sqrt{\pi(A_{i-1})} \sqrt{\pi(A_i)} \Big) \Big], \qquad (17)$$

and the first part of the theorem is complete.

For the second statement note that if we have equality, then we must also have equality for each of the norms $\|P_{i-1} M_i P_i\|$. This implies by Lemma 5.5 that all the sets A_i are the same and that $I_{A_i} - \pi(A_i)\mathbf{1}$ is in the $1 - \delta_i$ eigenspace of M_i for all i. For the converse, suppose on the other hand that $A_i = A$ for all i and $I_A - \pi(A)\mathbf{1}$ is in the $1 - \delta_i$ eigenspace of M_i. Note that

$$\begin{aligned} P_{i-1} M_i P_i I_A &= P_{i-1} M_i I_A = P_{i-1} M_i \big(\pi(A)\mathbf{1} + (I_A - \pi(A)\mathbf{1}) \big) \\ &= P_{i-1} \big(\pi(A)\mathbf{1} + (1 - \delta_i)(I_A - \pi(A)\mathbf{1}) \big) \\ &= \pi(A) I_A + (1 - \delta_i) I_A - (1 - \delta_i)\pi(A) I_A \\ &= \big(1 - \delta_i (1 - \pi(A) \big) I_A. \end{aligned}$$

Since $P_i^2 = P_i$, we can use induction to show that

$$\begin{aligned} \pi_{A_0} P_0 M_1 P_1 M_2 \cdots P_{k-1} M_k P_k I_{A_k} &= \pi_A \big[\prod_{i=1}^{k} (P_{i-1} M_i P_i) \big] I_A \\ &= \pi(A) \prod_{i=1}^{k} \big(1 - \delta_i (1 - \pi(A) \big), \end{aligned}$$

completing the proof of the second statement.

In order to prove the third statement, first note that if strict inequality holds in (11) when each A_i is taken to be A, then, by the second part of this result, the function $I_A - \pi(A)\mathbf{1}$ is not an eigenfunction of M corresponding to the eigenvalue $1 - \delta$. It then follows from Lemma 5.5 that $\|PMP\| < 1 - \delta(1 - \pi(A))$ where P is the corresponding projection onto A. The result now immediately follows from (17). ∎

6 NICD on general trees

In this section we give some results for the NICD problem on general trees. Theorem 1.3 in [31] stated that for the star graph where S is the set of leaves, the simple dictator protocols constitute all optimal protocols when $|S| = 2$ or $|S| = 3$. The proof of that result immediately leads to the following.

Theorem 6.1 *For any NICD instance (T, ρ, n, S) in which $|S| = 2$ or $|S| = 3$ the simple dictator protocols constitute all optimal protocols.*

6.1 Example with no simple optimal protocols

It appears that the problem of NICD in general is quite difficult. In particular, using Theorem 5.1 we show that there are instances for which there is no simple optimal protocol. Note the contrast with the case of stars, where it is proven in [31] that there is always a simple optimal protocol.

Proposition 6.2 *There exists an instance (T, ρ, n, S) for which there is no simple optimal protocol. In fact, given any ρ and any $n \geq 4$, there are integers k_1 and k_2, such that if T is a k_1-leaf star together with a path of length k_2 coming out of the center of the star (see Figure 1) and S is the full vertex set of T, then this instance has no simple optimal protocol.*

Proof: Fix ρ and $n \geq 4$. Recall that we write $\varepsilon = \frac{1}{2} - \frac{1}{2}\rho$ and let $\mathrm{Bin}(3, \varepsilon)$ be a binomially distributed random variable with parameters 3 and ε. As was observed in [31],

$$\mathcal{P}(\mathrm{Star}_k, \rho, n, S_k, \mathsf{MAJ}_3) \geq \frac{1}{8}\mathbf{P}[\mathrm{Bin}(3, \varepsilon) \leq 1]^k.$$

To see this, note that with probability $1/8$ the center of the star gets the string $(1, 1, 1)$. Since

$$\mathbf{P}[\mathrm{Bin}(3, \varepsilon) \leq 1] = (1 - \varepsilon)^2(1 + 2\varepsilon) > 1 - \varepsilon$$

for all $\varepsilon < 1/2$, we can pick k_1 large enough so that

$$\mathcal{P}(\mathrm{Star}_{k_1}, \rho, n, S_{k_1}, \mathsf{MAJ}_3) \geq 8(1 - \varepsilon)^{k_1}.$$

Next, by the last statement in Theorem 5.4, there exists $c_2 = c_2(\rho, n) > 1$ such that for all balanced non-dictator functions f on n bits

$$\mathcal{P}(\mathrm{Path}_k, \rho, n, \mathrm{Path}_k, \mathcal{D}) \geq \mathcal{P}(\mathrm{Path}_k, \rho, n, \mathrm{Path}_k, f)c_2^k.$$

Choose k_2 large enough so that

$$(1 - \varepsilon)^{k_1} c_2^{k_2} > 1.$$

Now let T be the graph consisting of a star with k_1 leaves and a path of length k_2 coming out of its center (see Figure 1), and let $S = V(T)$. We claim that the NICD instance (T, ρ, n, S) has no simple optimal protocol. We first observe that if it did, this protocol would have to be $\mathcal{D}$, i.e., $\mathcal{P}(T, \rho, n, S, f) < \mathcal{P}(T, \rho, n, S, \mathcal{D})$ for all simple protocols f which are not equivalent to dictator. This is because the quantity on the right is $(1-\varepsilon)^{k_1+k_2}$ and the quantity on the left is at most $\mathcal{P}(\mathrm{Path}_{k_2}, \rho, n, \mathrm{Path}_{k_2}, f)$ which in turn by definition of c_2 is at most $(1-\varepsilon)^{k_2}/c_2^{k_2}$. This is strictly less than $(1-\varepsilon)^{k_1+k_2}$ by the choice of k_2.

To complete the proof it remains to show that $\mathcal{D}$ is not an optimal protocol. Consider the protocol where k_2 vertices on the path (including the star's center) use the dictator $\mathcal{D}$ on the first bit and the k_1 leaves of the star use the protocol MAJ_3 on the last three out of n bits. Since $n \geq 4$, these vertices use completely independent bits from those that vertices on the path are using. We will show that this protocol, which we call f, does better than $\mathcal{D}$.

Let A be the event that all vertices on the path have their first bit being 1. Let B be the event that each of the k_1 leaf vertices of the star have 1 as the majority of their last 3 bits. Note that $P(A) = \frac{1}{2}(1-\varepsilon)^{k_2}$ and that, by definition of k_1, $P(B) \geq 4(1-\varepsilon)^{k_1}$. Now the protocol f succeeds if both A and B occur. Since A and B are independent (as distinct bits are used), f succeeds with probability at least $2(1-\varepsilon)^{k_2}(1-\varepsilon)^{k_1}$ which is twice the probability that the dictator protocol succeeds. ■

Remark: It was not necessary to use the last 3 bits for the k_1 vertices; we could have used the first 3 (and had $n = 3$). Then A and B would not be independent but it is easy to show (using the FKG inequality) that A and B would then be positively correlated which is all that is needed.

6.2 Optimal monotone protocols always exist

Next, we present some general statements about what optimal protocols must look like. Using discrete symmetrization together with the FKG inequality we prove the following theorem, which extends one of the results in [31] from the case of the star to the case of general trees.

Theorem 6.3 *For all NICD instances on trees, there is an optimal protocol in which all players use a monotone function.*

One of the tools that we need to prove Theorem 6.3 is the correlation inequality obtained by Fortuin et al. [21] which is usually called the FKG inequality. We first recall some basic definitions.

Let D be a finite linearly ordered set. Given two strings x, y in D^m we write $x \leq y$ iff $x_i \leq y_i$ for all indices $1 \leq i \leq m$. We denote by $x \vee y$ and $x \wedge y$ two strings whose ith coordinates are $\max(x_i, y_i)$ and $\min(x_i, y_i)$ respectively. A probability measure $\mu : D^m \to \mathbb{R}^{\geq 0}$ is called *log-supermodular* if

$$\mu(\eta)\mu(\delta) \leq \mu(\eta \vee \delta)\mu(\eta \wedge \delta) \tag{18}$$

for all $\eta, \delta \in D^m$. If μ satisfies (18) we will also say that μ satisfies the FKG lattice condition. A subset $A \subseteq D^m$ is *increasing* if whenever $x \in A$ and $x \leq y$ then also $y \in A$. Similarly, A is *decreasing* if $x \in A$ and $y \leq x$ imply that $y \in A$. Finally, the measure of A is $\mu(A) = \sum_{x \in A} \mu(x)$. The following well known fact is a special case of the FKG inequality.

Proposition 6.4 *Let $\mu : \{-1, 1\}^m \to \mathbb{R}^{\geq 0}$ be a log-supermodular probability measure on the discrete cube. If A and B are two increasing subsets of $\{-1, 1\}^m$ and C is a decreasing subset then*

$$\mu(A \cap B) \geq \mu(A) \cdot \mu(B) \quad \textit{and} \quad \mu(A \cap C) \leq \mu(A) \cdot \mu(C).$$

It is known that in order to prove that μ satisfies the FKG lattice condition, it suffices to check this for "smallest boxes" in the lattice, i.e., for η and δ that agree at all but two locations. Since we don't know a reference, for completeness we prove this here.

Lemma 6.5 *Let μ be a measure with full support. Then μ satisfies the FKG lattice condition (18) if and only if it satisfies (18) for all η and δ that agree at all but two locations.*

Proof: We will prove the non-trivial direction by induction on $d = d(\eta, \delta)$, the Hamming distance between η and δ. The cases where $d(\eta, \delta) \leq 2$ follow from the assumption. The proof will proceed by induction on d. Let $d = d(\eta, \delta) \geq 3$ and assume the claim holds for all smaller d. We can partition the set of coordinates into 3 subsets $I_=, I_{\{\eta > \delta\}}$ and $I_{\{\eta < \delta\}}$, where η and δ agree, where $\eta > \delta$ and where $\eta < \delta$ respectively. Without loss of generality $|I_{\{\eta > \delta\}}| \geq 2$. Let $i \in I_{\{\eta > \delta\}}$ and let η' be obtained from η by setting $\eta'_i = \delta_i$ and letting $\eta'_j = \eta_j$ otherwise. Then since $\eta' \wedge \delta = \eta \wedge \delta$,

$$\frac{\mu(\eta \wedge \delta)\mu(\eta \vee \delta)}{\mu(\eta)\mu(\delta)} = \left(\frac{\mu(\eta' \wedge \delta)\mu(\eta' \vee \delta)}{\mu(\eta')\mu(\delta)} \right) \times \left(\frac{\mu(\eta')\mu(\eta \vee \delta)}{\mu(\eta)\mu(\eta' \vee \delta)} \right).$$

The first factor is ≥ 1 by the induction hypothesis since $d(\delta, \eta') = d(\delta, \eta) - 1$. Note that $\eta' = \eta \wedge (\eta' \vee \delta)$, $\eta \vee \delta = \eta \vee (\eta' \vee \delta)$, and $d(\eta', \eta \vee \delta) = 1 + |I_{\{\eta < \delta\}}| < d$. Therefore by induction, the second term is also ≥ 1. ■

The above tools together with symmetrization now allow us to prove Theorem 6.3.

Proof of Theorem 6.3: The general strategy of the proof is a shifting technique together with using FKG to prove that this shifting improves things.

Recall that we have a tree T with m vertices, $0 < \rho < 1$, and a probability measure $\mathbf{P}$ on $\alpha \in \{-1, 1\}^{V(T)}$ which is defined by

$$\mathbf{P}(\alpha) = \tfrac{1}{2}(\tfrac{1}{2} + \tfrac{1}{2}\rho)^{A(\alpha)}(\tfrac{1}{2} - \tfrac{1}{2}\rho)^{B(\alpha)},$$

where $A(\alpha)$ is the number of pairs of neighbors where α agrees and $B(\alpha)$ is the number of pairs of neighbors where α disagrees. To use Proposition 6.4 we need to show that $\mathbf{P}$ is a log-supermodular probability measure.

Lemma 6.5 tells us that we need only check the FKG lattice condition for configurations that differ in only two sites. Note that (18) holds trivially if $\alpha \leq \beta$ or $\beta \leq \alpha$. Thus it suffices to consider the case where there are two vertices u, v of T on which α and β disagree and that $\alpha_v = \beta_u = 1$ and $\alpha_u = \beta_v = -1$. If these vertices are not neighbors then by definition of $\mathbf{P}$ we have that $\mathbf{P}(\alpha)\mathbf{P}(\beta) = \mathbf{P}(\alpha \vee \beta)\mathbf{P}(\alpha \wedge \beta)$. Similarly, if u is a neighbor of v in T, then one can easily check that

$$\frac{\mathbf{P}(\alpha)\mathbf{P}(\beta)}{\mathbf{P}(\alpha \vee \beta)\mathbf{P}(\alpha \wedge \beta)} = \left(\frac{1-\rho}{1+\rho}\right)^2 \leq 1.$$

Hence we conclude that measure $\mathbf{P}$ is log-supermodular.

Let $f_1, \ldots, f_k$ be the functions used by the parties at nodes $S = \{v_1, \ldots, v_k\}$. We will shift the functions in the sense of Kleitman's monotone "down-shifting" [29]. Namely, define functions $g_1, \ldots, g_k$ as follows: If

$$f_i(-1, x_2, \ldots, x_n) = f_i(1, x_2, \ldots, x_n)$$

then we set

$$\begin{aligned} g_i(-1, x_2, \ldots, x_n) &= g_i(1, x_2, \ldots, x_n) = f_i(-1, x_2, \ldots, x_n) \\ &= f_i(1, x_2, \ldots, x_n). \end{aligned}$$

Otherwise, we set $g_i(-1, x_2, \ldots, x_n) = -1$ and $g_i(1, x_2, \ldots, x_n) = 1$. We claim that the agreement probability for the g_i's is at least the agreement probability for the f_i's. Repeating this argument for all bit locations will prove that there exists an optimal protocol for which all functions are monotone.

To prove the claim we condition on the value of $x_2, \ldots, x_n$ at all the nodes v_i and let α_i be the remaining bit at v_i. For simplicity we will denote the functions of this bit by f_i and g_i.

Let

$$S_1 = \{i : f_i(-1) = f_i(1) = 1\}, \quad S_2 = \{i : f_i(-1) = f_i(1) = -1\},$$

$$S_3 = \{i : f_i(-1) = -1, f_i(1) = 1\}, \quad S_4 = \{i : f_i(-1) = 1, f_i(1) = -1\}.$$

If S_1 and S_2 are both nonempty, then the agreement probability for both f and g is 0. Now without loss of generality, assume that S_2 is empty. Assume first that S_1 is nonempty. Then the agreement probability for g is $\mathbf{P}[\alpha_i = 1 \quad \forall i \in S_3 \cup S_4)$ while the agreement probability for f is $\mathbf{P}[\alpha_i = 1 \quad \forall i \in S_3, \alpha_i = -1 \ \ \forall i \in S_4]$. By FKG, the first probability is at least $\mathbf{P}[\alpha_i = 1 \ \ \forall i \in S_3]\mathbf{P}[\alpha_i = 1 \ \ \forall i \in S_4]$ while the second probability is at most $\mathbf{P}[\alpha_i = 1 \quad \forall i \in S_3]\mathbf{P}[\alpha_i = -1 \ \ \forall i \in S_4]$. By symmetry, the two second factors are the same, completing the proof when S_1 is nonempty. An easy modification, left to the reader, takes care of the case when S_1 is also empty. ■

Remark: The last step in the proof above may be replaced by a more direct calculation showing that in fact we have strict inequality unless the sets U', U'' are empty. This is similar to the monotonicity proof in [31]. This implies that every optimal protocol must consist of monotone functions (in general, it may be monotone increasing in some coordinates and monotone decreasing in the other coordinates).

Remark: The above proof works in a much more general setup than just our tree-indexed Markov chain case. One can take any measure on $\{-1,1\}^m$ satisfying the FKG lattice condition with all marginals having mean 0, take n independent copies of this and define everything analogously in this more general framework. The proof of Theorem 6.3 extends to this context.

6.3 Monotonicity in the number of parties

Our last theorem yields a certain monotonicity when comparing the simple dictator protocol $\mathcal{D}$ and the simple protocol MAJ_r, which is majority on the first r bits. The result is not very strong – it is interesting mainly because it allows to compare protocols behavior for different number of parties. It shows that if MAJ_r is a better protocol than dictatorship for k_1 parties on the star, then it is also better than dictatorship for k_2 parties if $k_2 > k_1$.

Theorem 6.6 *Fix ρ and n and suppose k_1 and r are such that*

$$\mathcal{P}(\mathrm{Star}_{k_1}, \rho, n, \mathrm{Star}_{k_1}, \mathsf{MAJ}_r) \geq (>)\ \mathcal{P}(\mathrm{Star}_{k_1}, \rho, n, \mathrm{Star}_{k_1}, \mathcal{D}).$$

Then for all $k_2 > k_1$,

$$\mathcal{P}(\mathrm{Star}_{k_2}, \rho, n, \mathrm{Star}_{k_2}, \mathsf{MAJ}_r) \geq (>)\ \mathcal{P}(\mathrm{Star}_{k_2}, \rho, n, \mathrm{Star}_{k_2}, \mathcal{D}).$$

Note that it suffices to prove the theorem assuming $r = n$. In order to prove the theorem, we first introduce or recall some necessary definitions including the notion of stochastic domination.

Definitions and set-up: We define an ordering on $\{0, 1, \ldots, n\}^I$, writing $\eta \preceq \delta$ if $\eta_i \leq \delta_i$ for all $i \in I$. If ν and μ are two probability measures on $\{0, 1, \ldots, n\}^I$, we say μ *stochastically dominates* ν, written $\nu \preceq \mu$, if there exists a probability measure m on $\{0, 1, \ldots, n\}^I \times \{0, 1, \ldots, n\}^I$ whose first and second marginals are respectively ν and μ and such that m is supported on $\{(\eta, \delta) : \eta \preceq \delta\}$. Fix ρ, $n \geq 3$, and any tree T. Let our tree-indexed Markov chain be $\{x_v\}_{v \in T}$, where $x_v \in \{-1, 1\}^n$ for each $v \in T$. Let $A \subseteq \{-1, 1\}^n$ be the strings which have a majority of 1's. Let X_v denote the number of 1's in x_v. Given $S \subseteq T$, let μ_S be the conditional distribution of $\{X_v\}_{v \in T}$ given $\cap_{v \in S}\{x_v \in A\}$ $(= \cap_{v \in S}\{X_v \geq n/2\})$.

The following lemma is key and might be of interest in itself. It can be used to prove (perhaps less natural) results analogous to Theorem 6.6 for general trees. Its proof will be given later.

Lemma 6.7 *In the above setup, if $S_1 \subseteq S_2 \subseteq T$, we have*

$$\mu_{S_1} \preceq \mu_{S_2}.$$

Before proving the lemma or showing how it implies Theorem 6.6, a few remarks are in order.

- Note that if $\{x_k\}$ is a Markov chain on $\{-1, 1\}^n$ with transition matrix T_ρ, then if we let X_k be the number of 1's in x_k, then $\{X_k\}$ is also a Markov chain on the state space $\{0, 1, \ldots, n\}$ (although it is certainly not true in general that a function of a Markov chain is a Markov chain.) In this way, with a slight abuse of notation, we can think of T_ρ as a transition matrix for $\{X_k\}$ as well as for $\{x_k\}$. In particular, given a probability distribution μ on $\{0, 1, \ldots, n\}$ we will write μT_ρ for the probability measure on $\{0, 1, \ldots, n\}$ given by one step of the Markov chain.

- We next recall the easy fact that the Markov chain T_ρ on $\{-1, 1\}^n$ is *attractive* meaning that if ν and μ are probability measures on $\{-1, 1\}^n$ with $\nu \preceq \mu$, then it follows that $\nu T_\rho \preceq \mu T_\rho$. (This is easily verified for one coordinate and the one coordinate case easily implies the n-dimensional case.) The same is true for the Markov chain $\{X_k\}$ on $\{0, 1, \ldots, n\}$.

Along with these observations, Lemma 6.7 is enough to prove Theorem 6.6:

Proof: Let $v_0, v_1, \ldots, v_k$ be the vertices of Star_k, where v_0 is the center. Clearly, $\mathcal{P}(\mathrm{Star}_k, \rho, \mathrm{Star}_k, \mathcal{D}) = (\frac{1}{2} + \frac{1}{2}\rho)^k$. On the other hand, a little thought reveals that

$$\mathcal{P}(\mathrm{Star}_k, \rho, n, \mathrm{Star}_k, \mathsf{MAJ}_n) = \prod_{\ell=0}^{k-1} (\mu_{\{v_0,\ldots,v_\ell\}} \mid_{v_0} T_\rho)(A),$$

where by $\mu \mid_v$ we mean the x_v marginal of a distribution μ (recall that $A \subseteq \{-1,1\}^n$ is the strings which have a majority of 1's). By Lemma 6.7 and the attractivity of the process, the terms $(\mu_{\{v_0,\ldots,v_\ell\}} \mid_{v_0} T_\rho)(A)$ (which do not depend on k as long as $\ell \leq k$) are nondecreasing in ℓ. Therefore if

$$\mathcal{P}(\mathrm{Star}_k, \rho, n, \mathrm{Star}_k, \mathsf{MAJ}_n) \geq (>)(\tfrac{1}{2} + \tfrac{1}{2}\rho)^k,$$

then $(\mu_{\{v_0,\ldots,v_{k-1}\}} \mid_{v_0} T_\rho)(A) \geq (>)\frac{1}{2} + \frac{1}{2}\rho$ which implies in turn that for every $k' \geq k$, $(\mu_{\{v_0,\ldots,v_{k'-1}\}} \mid_{v_0} T_\rho)(A) \geq (>)\frac{1}{2} + \frac{1}{2}\rho$ and thus for all $k' > k$

$$\mathcal{P}(\mathrm{Star}_{k'}, \rho, n, \mathrm{Star}_{k'}, \mathsf{MAJ}_n) \geq (>)(\tfrac{1}{2} + \tfrac{1}{2}\rho)^{k'}.$$

■

Before proving Lemma 6.7, we recall the definition of *positive associativity*. If μ is a probability measure on $\{0, 1, \ldots, n\}^I$, μ is said to be *positively associated* if any two monotone functions on $\{0, 1, \ldots, n\}^I$ are positively correlated. This is equivalent to the fact that if $B \subseteq \{0, 1, \ldots, n\}^I$ is an upset, then μ conditioned on B is stochastically larger than μ. (It is immediate to check that this last condition is equivalent to monotone *events* being positively correlated. However, it is well known that monotone events being positively correlated implies that monotone functions are positively correlated; this is done by writing out a monotone function as a positive linear combination of indicator functions.)

Proof of Lemma 6.7: It suffices to prove this when S_2 is S_1 plus an extra vertex z. We claim that for any set S, μ_S is positively associated. Given this claim, we form μ_{S_2} by first conditioning on $\cap_{v \in S_1}\{x_v \in A\}$, giving us the measure μ_{S_1}, and then further conditioning on $x_z \in A$. By the claim, μ_{S_1} is positively associated and hence the last further conditioning on $X_z \in A$ stochastically increases the measure, giving $\mu_{S_1} \preceq \mu_{S_2}$.

To prove the claim that μ_S is positively associated, we first claim that the distribution of $\{X_v\}_{v \in T}$, which is just a probability measure on $\{0, 1, \ldots, n\}^T$, satisfies the FKG lattice condition (18).

Assuming the FKG lattice condition holds for $\{X_v\}_{v \in T}$, it is easy to see that the same inequality holds when we condition on the sublattice $\cap_{v \in S}\{X_v \geq n/2\}$ (it is crucial here that the set $\cap_{v \in S}\{X_v \geq n/2\}$ is a sublattice meaning that η, δ being in this set implies that $\eta \vee \delta$ and $\eta \wedge \delta$ are also in this set).

The FKG theorem, which says that the FKG lattice condition (for any distributive lattice) implies positive association, can now be applied to this conditioned measure to conclude that the conditioned measure has positive association, as desired.

Finally, by Lemma 6.5, in order to prove that the distribution of $\{X_v\}_{v \in T}$ satisfies the FKG lattice condition, it is enough to check this for "smallest boxes" in the lattice, i.e., for η and δ that agree at all but two locations. If these two locations are not neighbors, it is easy to check that we have equality. If they are neighbors, it easily comes down to checking that if $a > b$ and $c > d$, then

$$\mathbf{P}[X_1 = c | X_0 = a]\mathbf{P}[X_1 = d | X_0 = b]$$

is greater or equal to

$$\mathbf{P}[X_1 = d | X_0 = a]\mathbf{P}[X_1 = c | X_0 = b]$$

where $\{X_0, X_1\}$ is the distribution of our Markov chain on $\{0, 1, \ldots, n\}$ restricted to two consecutive times. It is straightforward to check that for $\rho \in (0, 1)$, the above Markov chain can be embedded into a continuous time Markov chain on $\{0, 1, \ldots, n\}$ which only takes steps of size 1. The last claim now follows from Lemma 6.8 stated and proved below. ■

Lemma 6.8 *If $\{X_t\}$ is a continuous time Markov chain on $\{0, 1, \ldots, n\}$ which only takes steps of size 1, then if $a > b$ and $c > d$, it follows that*

$$\mathbf{P}[X_1 = c | X_0 = a]\mathbf{P}[X_1 = d | X_0 = b]$$

is greater or equal to

$$\mathbf{P}[X_1 = d | X_0 = a]\mathbf{P}[X_1 = c | X_0 = b]$$

(Of course, by time scaling, X_1 can be replaced by any time X_t.)

Proof: Let $R_{a,c}$ be the set of all possible realizations of our Markov chain during $[0, 1]$ starting from a and ending in c. Define $R_{a,d}$, $R_{b,c}$ and $R_{b,d}$ analogously. Letting $\mathbf{P}_x$ denote the measure on paths starting from x, we need to show that

$$\mathbf{P}_a(R_{a,c})\mathbf{P}_b(R_{b,d}) \geq \mathbf{P}_a(R_{a,d})\mathbf{P}_b(R_{b,c})$$

or equivalently that

$$\mathbf{P}_a \times \mathbf{P}_b[R_{a,c} \times R_{b,d}] \geq \mathbf{P}_a \times \mathbf{P}_b[R_{a,d} \times R_{b,c}]$$

We do this by giving a measure preserving injection from $R_{a,d} \times R_{b,c}$ to $R_{a,c} \times R_{b,d}$. We can ignore pairs of paths where there is a jump in both paths at the same time since these have $\mathbf{P}_a \times \mathbf{P}_b$ measure 0. Given a pair of paths in $R_{a,d} \times R_{b,c}$, we can switch the paths after their first meeting time. It is clear that this gives an injection from $R_{a,d} \times R_{b,c}$ to $R_{a,c} \times R_{b,d}$ and the Markov property guarantees that this injection is measure preserving, completing the proof. ■

7 Conclusions and open questions

In this paper we have exactly analyzed the NICD problem on the path and asymptotically analyzed the NICD problem on the star. However, we have seen that results on more complicated trees may be hard to come by. Many problems are still open. We list a few:

- Is it true that for every tree NICD instance, there is an optimal protocol in which each player uses some majority rule? This question was already raised in [31] for the special case of the star.

- Our analysis for the star is quite tight. However, one can ask for more. In particular, what is the best bound that can be obtained on

$$r_k = \frac{\mathcal{M}(\mathrm{Star}_k, \rho, S_k)}{\lim_{\substack{n\to\infty \\ n \text{ odd}}} \mathcal{P}(\mathrm{Star}_k, \rho, n, S_k, \mathsf{MAJ}_n)}$$

  for fixed value of ρ. Our results show that $r_k = k^{o(1)}$. Is it true that $\lim_{k\to\infty} r_k = 1$?

- Finally, we would like to find more applications of the reverse Bonami-Beckner inequality in computer science and combinatorics.

8 Acknowledgments

We thank David Aldous, Christer Borell, Svante Janson, Yuval Peres, and Oded Schramm for helpful discussions. We also thank the referee for a careful reading and a number of suggestions.

References

[1] M. Abramowitz and I. Stegun. *Handbook of mathematical functions*. Dover, 1972.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. Deterministic simulation in LOGSPACE. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing*, pages 132–140, 1987.

[3] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Computational Complexity*, pages 60–75, 1995.

[4] N. Alon, G. Kalai, M. Ricklin, and L. Stockmeyer. Lower bounds on the competitive ratio for mobile user tracking and distributed job scheduling. *Theoretical Computer Science*, 130:175–201, 1994.

[5] N. Alon, U. Maurer, and A. Wigderson. Unpublished results, 1991.

[6] N. Alon and J. Spencer. *The Probabilistic Method*. 2nd ed., Wiley, 2000.

[7] K. Amano and A. Maruoka. On learning monotone Boolean functions under the uniform distribution. *Lecture Notes in Computer Science*, 2533:57–68, 2002.

[8] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math.*, pages 159–182, 1975.

[9] M. Ben-Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation*. Academic Press, New York, 1990.

[10] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[11] S. Bobkov and F. Götze. Discrete isoperimetric and Poincaré-type inequalities. *Prob. Theory and Related Fields*, 114:245–277, 1999.

[12] A. Bonami. Études des coefficients Fourier des fonctiones de $L^p(G)$. *Ann. Inst. Fourier*, 20(2):335–402, 1970.

[13] C. Borell. Positivity improving operators and hypercontractivity. *Math. Zeitschrift*, 180(2):225–234, 1982.

[14] J. Bourgain. An appendix to *Sharp thresholds of graph properties, and the k-sat problem*, by E. Friedgut. *J. American Math. Soc.*, 12(4):1017–1054, 1999.

[15] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson, and N. Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77:55–64, 1992.

[16] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geom. and Func. Analysis*, 7:438–461, 1997.

[17] N. Bshouty, J. Jackson, and C. Tamon. Uniform-distribution attribute noise learnability. In *Proc. 12th Ann. Workshop on Comp. Learning Theory*, pages 75–80, 1999.

[18] I. Dinur, V. Guruswami, and S. Khot. Vertex Cover on k-uniform hypergraphs is hard to approximate within factor $(k - 3 - \varepsilon)$. ECCC Technical Report TR02-027, 2002.

[19] I. Dinur and S. Safra. The importance of being biased. In *Proc. 34th Ann. ACM Symp. on the Theory of Computing*, pages 33–42, 2002.

[20] W. Feller. *An introduction to probability theory and its applications*. 3rd ed., Wiley, 1968.

[21] C. Fortuin, P. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[22] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):474–483, 1998.

[23] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993–3002, 1996.

[24] H. O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.

[25] G. Hardy, J. Littlewood, and G. Pólya. *Inequalities*. 2nd ed. Cambridge University Press, 1952.

[26] J. Håstad. Some optimal inapproximability results. *J. ACM*, 48:798–869, 2001.

[27] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proc. 29th Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 68–80, 1988.

[28] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34th Ann. ACM Symp. on the Theory of Computing*, pages 767–775, 2002.

[29] D. Kleitman. Families of non-disjoint subsets. *J. Combin. Theory*, 1:153–155, 1966.

[30] A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. In *Proc. 43rd Ann. IEEE Symp. on Foundations of Comp. Sci.*, pages 177–186, 2002.

[31] E. Mossel and R. O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. To appear, 2003.

[32] A. Naor, E. Friedgut, and G. Kalai. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. Appl. Math.*, 29(3):427–437, 2002.

[33] R. O'Donnell. Hardness amplification within NP. In *Proc. 34th Ann. ACM Symp. on the Theory of Computing*, pages 751–760, 2002.

[34] R. O'Donnell. *Computational applications of noise sensitivity*. PhD thesis, Massachusetts Institute of Technology, 2003.

[35] R. O'Donnell and R. Servedio. Learning monotone decision trees. Manuscript, 2004.

[36] R. Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3-4):205–221, 1995.

[37] V. Sazonov. *Normal approximation — some recent advances*. Springer-Verlag, 1981.

[38] M. Talagrand. On Russo's approximate 0-1 law. *Annals of Probability*, 22:1476–1387, 1994.

[39] K. Yang. On the (im)possibility of non-interactive correlation distillation. In *Proc. of LATIN 2004*.

A Proof of the reverse Bonami-Beckner inequality

Borell's proof of the reverse Bonami-Beckner inequality [13] follows the same lines as the traditional proofs of the usual Bonami-Beckner inequality [12, 8]. Namely, he proves the result in the case $n = 1$ (i.e., the "two-point inequality") and then shows that this can be tensored to produce the full theorem. The usual proof of the tensoring is easily modified by replacing Minkowski's inequality with the reverse Minkowski inequality [25, Theorem 24]. Hence, it is enough to consider functions $f : \{-1, 1\} \to \mathbb{R}^{\geq 0}$ (i.e., $n = 1$). By monotonicity of norms, it suffices to prove the inequality in the case that $\rho = (1-p)^{1/2}/(1-q)^{1/2}$; i.e., $\rho^2 = (1-p)/(1-q)$. Finally, it turns out that it suffices to consider the case where $0 < q < p < 1$ (see Lemma A.3).

Lemma A.1 *Let $f : \{-1, 1\} \to \mathbb{R}^{\geq 0}$ be a nonnegative function, $0 < q < p < 1$, and $\rho^2 = (1-p)/(1-q)$. Then $\|T_\rho f\|_q \geq \|f\|_p$.*

Proof (Borell): If f is identically zero the lemma is trivial. Otherwise, using homogeneity we may assume that $f(x) = 1 + ax$ for some $a \in [-1, 1]$. We shall consider only the case $a \in (-1, 1)$; the result at the endpoints follows by continuity. Note that $T_\rho f(x) = 1 + \rho a x$.

Using the Taylor series expansion for $(1+a)^q$ around 1, we get

$$\begin{aligned} \|T_\rho f\|_q^q &= \frac{1}{2}\left((1+a\rho)^q + (1-a\rho)^q\right) \qquad (19)\\ &= \frac{1}{2}\left((1+\sum_{n=1}^{\infty}\binom{q}{n}a^n\rho^n) + (1+\sum_{n=1}^{\infty}\binom{q}{n}(-a)^n\rho^n)\right)\\ &= 1+\sum_{n=1}^{\infty}\binom{q}{2n}a^{2n}\rho^{2n}. \qquad (20)\end{aligned}$$

(Absolute convergence for $|a| < 1$ lets us rearrange the series.) Since $p > q$, it holds for all $x > -1$ that $(1+x)^{p/q} \geq 1 + px/q$. In particular, from (20) we obtain that

$$\|T_\rho f\|_q^p = \left(1+\sum_{n=1}^{\infty}\binom{q}{2n}a^{2n}\rho^{2n}\right)^{p/q} \geq 1+\sum_{n=1}^{\infty}\frac{p}{q}\binom{q}{2n}a^{2n}\rho^{2n}. \qquad (21)$$

Similarly to (20) we can write

$$\|f\|_p^p = 1+\sum_{n=1}^{\infty}\binom{p}{2n}a^{2n}. \qquad (22)$$

From (21) and (22) we see that in order to prove the theorem it suffices to show that for all $n \geq 1$

$$\frac{p}{q}\binom{q}{2n}\rho^{2n} \geq \binom{p}{2n}. \tag{23}$$

Simplifying (23) we see the inequality

$$(q-1)\cdots(q-2n+1)\rho^{2n} \geq (p-1)\cdots(p-2n+1),$$

which is equivalent in turn to

$$(1-q)\cdots(2n-1-q)\rho^{2n} \leq (1-p)\cdots(2n-1-p). \tag{24}$$

Note that we have $(1-p) = (1-q)\rho^2$. Inequality (23) would follow if we could show that for all $m \geq 2$ it holds that $\rho(m-q) \leq (m-p)$. Taking the square and recalling that $\rho^2 = (1-p)/(1-q)$ we obtain the inequality

$$(1-p)(m-q)^2 \leq (m-p)^2(1-q),$$

which is equivalent to

$$m^2 - 2m + p + q - pq \geq 0.$$

The last inequality holds for all $m \geq 2$ thus completing the proof. ∎

We also prove the two-function version promised in Section 3.1. Recall first the reverse Hölder inequality [25, Theorem 13] for discrete measure spaces:

Lemma A.2 *Let f and g be nonnegative functions and suppose $1/p + 1/p' = 1$, where $p < 1$ ($p' = 0$ if $p = 0$). Then*

$$\mathbf{E}[fg] = \|fg\|_1 \geq \|f\|_p \|g\|_{p'},$$

where equality holds if $g = f^{p/p'}$.

Proof of Corollary 3.3: By definition, the left-hand side of (2) is $\mathbf{E}[fT_\rho g]$. We claim it suffices to prove (2) for $\rho = (1-p)^{1/2}(1-q)^{1/2}$. Indeed, otherwise, let r satisfy $\rho = (1-p)^{1/2}(1-r)^{1/2}$ and note that $r \geq q$. Then, assuming (2) holds for p, r and ρ we obtain:

$$\mathbf{E}[fT_\rho g] \geq \|f\|_p \|g\|_r \geq \|f\|_p \|g\|_q,$$

as needed.

We now assume $\rho = (1-p)^{1/2}(1-q)^{1/2}$. Let p' satisfy $1/p + 1/p' = 1$. Applying the reverse Hölder inequality we get that $\mathbf{E}[fT_\rho g] \geq \|f\|_p \|T_\rho g\|_{p'}$. Note that, since $1/(1-p') = 1-p$, the fact that $\rho = (1-p)^{1/2}(1-q)^{1/2}$ implies $\rho = (1-q)^{1/2}(1-p')^{-1/2}$. Therefore, using the reverse Bonami-Beckner inequality with $p' \leq q \leq 1$, we conclude that

$$\mathbf{E}[f(x)g(y)] \geq \|f\|_p \|T_\rho g\|_{p'} \geq \|f\|_p \|g\|_q.$$

■

Lemma A.3 *It suffices to prove (1) for* $0 < q < p < 1$.

Proof: Note first that the case $p = 1$ follows from the case $p < 1$ by continuity. Recall that $1 - p = \rho^2(1-q)$. Thus, $p > q$. Suppose (1) holds for $0 < q < p < 1$. Then by continuity we obtain (1) for $0 \leq q < p < 1$. From $1 - p = \rho^2(1-q)$, it follows that $1 - q' = 1/(1-q) = \rho^2/(1-p) = \rho^2(1-p')$. Therefore if $p \leq 0$, then $p' = 1 - 1/(1-p) \geq 0$ and $q' = 1 - \rho^2/(1-p) > p' \geq 0$. We now conclude that if f is non-negative, then

$$\begin{aligned}
\|T_\rho f\|_q &= \inf\{\|gT_\rho f\|_1 : \|g\|_{q'} = 1, g \geq 0\} \quad \text{(by reverse Hölder)} \\
&= \inf\{\|fT_\rho g\|_1 : \|g\|_{q'} = 1, g \geq 0\} \quad \text{(by reversibility)} \\
&\geq \inf\{\|f\|_p \|T_\rho g\|_{p'} : \|g\|_{q'} = 1, g \geq 0\} \quad \text{(by reverse Hölder)} \\
&\geq \|f\|_p \inf\{\|g\|_{q'} : \|g\|_{q'} = 1, g \geq 0\} = \|f\|_p \\
&\quad \text{(by (1) for } 0 \leq p' < q' < 1\text{)}.
\end{aligned}$$

We have thus obtained that (1) holds for $p \leq 0$. The remaining case is $p > 0 > q$. Let $r = 0$ and choose ρ_1, ρ_2 such that $(1-p) = \rho_2^2(1-r)$ and $(1-r) = \rho_1^2(1-q)$. Note that $0 < \rho_1, \rho_2 < 1$ and that $\rho = \rho_1\rho_2$. The latter equality implies that $T_\rho = T_{\rho_1}T_{\rho_2}$ (this is known as the "semi-group property"). Now

$$\|T_\rho f\|_q = \|T_{\rho_1}T_{\rho_2} f\|_q \geq \|T_{\rho_2} f\|_r \geq \|f\|_p,$$

where the first inequality follows since $q < r \leq 0$ and the second since $p > r \geq 0$.

We have thus completed the proof. ■