On the “log rank”–Conjecture in Communication Complexity *

Ran Raz †
Dept. of Computer Science
Princeton University

Boris Spieker ‡
Dept. of Applied Mathematics
University of Twente

Abstract

We show the existence of a non–constant gap between the communication complexity of a function and the logarithm of the rank of its input matrix. We consider the following problem: each of two players gets a perfect matching between two n–element sets of vertices. Their goal is to decide whether or not the union of the two matchings forms a Hamiltonian cycle. We prove:

1) The rank of the input matrix over the reals for this problem is $2^{O(n)}$.

2) The non–deterministic communication complexity of the problem is $\Omega(n \log\log n)$.

Our result also supplies a superpolynomial gap between the chromatic number of a graph and the rank of its adjacency matrix. Another conclusion from the second result is an $\Omega(n \log\log n)$ lower bound for the graph connectivity problem in the non–deterministic case.

We make use of the theory of group representations for the first result. The second result is proved by an information theoretic argument.

---

*A preliminary version of the paper appeared in FOCS 93

†Current address: Department of Applied Mathematics, Weizmann Institute, Rehovot 76100, Israel. ranraz@wisdom.weizmann.ac.il

‡This research was done while the second author was visiting DIMACS at Princeton University. His work was supported by the National Operations Research Network in the Netherlands (LNMB).

1 Introduction

In the basic model of communication complexity we consider two finite sets X and Y and a function $f : X \times Y \longrightarrow \{0,1\}$. Given such a function, consider the following game between players I and II: for $(x,y) \in X \times Y$ give x to player I and y to player II. Their goal is to compute the value of $f(x,y)$. The deterministic communication complexity of the function f is the minimum number of bits the players have to exchange between them for the worst case input in any deterministic protocol. The model of communication complexity, first introduced by Yao [Y], has been studied in many papers and seems to have close relationships with certain other complexity issues (see [L] for a survey).

For a function f of the type described above, define the 0–1 input matrix D by $D_{xy} = f(x,y)$. The input matrix D determines the communication problem: both players know the matrix D, player I knows the index x of a row, player II knows the index y of a column and they want to determine the entry D_{xy}. It is a simple observation that $\log_2(\mathrm{rank}(D))$ over any field is a lower bound for the communication complexity of f. This was proved by Mehlhorn and Schmidt [MS], and became one of the most important tools in analyzing the communication complexity of a function. Since then it was a standing open problem whether this lower bound is tight when the rank is computed over $\mathbf{R}$ (the rank over $\mathbf{R}$ is greater or equal to the rank over any finite field). A positive answer for this question means that in order to understand the communication complexity of a 0-1 function it is sufficient to analyze the rank of the corresponding input matrix, and therefore reduces the question: what is the communication complexity ? to the question: what is the rank of the input matrix ? We prove that this is not the case.

The non–deterministic communication complexity of a matrix D (i.e. the non–deterministic communication complexity of the underlying function) is the least natural number t such that the 1's in D can be covered by at most 2^t all–1 submatrices. It is easy to prove that the deterministic communication complexity is greater or equal than the non–deterministic communication complexity.

In this paper we consider the following problem: Let M be the set of all bipartite perfect matchings on $2n$ vertices with two color classes of size n. Each of the players gets a matching in M as an input. The goal is now to determine whether the union of the two matchings forms a Hamiltonian cycle. (The problem was first suggested by Alok Aggarwal in 1986). Define D to be the input matrix for this problem. We present two results:
The first one is an exact computation of the rank of the matrix D over $\mathbf{R}$. We look at the regular representation of the symmetric group S_n. We write D as the sum of matrices corresponding to elements of S_n in this representation, and we employ the theory of group representations and some known results about the characters of S_n. It turns out that the

rank is

$$\binom{2n-2}{n-1} = 2^{O(n)}$$

The second result is an $\Omega(n \log \log n)$ lower bound for the non–deterministic communication complexity of D. We achieve this by proving the existence of a subset of 1–entries in the matrix D which has a small intersection with every all–1 submatrix of D. The proof is based on a combinatorial lemma given in section 3. We believe that the non-deterministic communication complexity of the problem is actually $\Omega(n \log n)$. Also, we believe that this better lower bound can be proved by improving the main lemma.

Combining the two results, we get a non–constant gap between the logarithm of the rank of the input matrix and its non–deterministic communication complexity. [1] Of course, this also is a gap between the logarithm of the rank and the deterministic communication complexity of D. So the "log rank" lower bound is not tight ! The best previous gap was a constant of 4/3 due to Razborov [R]. As pointed out by Razborov, our result also supplies a superpolynomial gap between the chromatic number of a graph and the rank of its adjacency matrix (see [LS]). Earlier results gave a constant gap [AS], and a superlinear gap [R]. The exact power of the log-rank lower bound is still not known. In particular it is very important to find out whether the gap between the logarithm of the rank and the communication complexity can be super polynomial [LS].

One of the best known examples, in which the log–rank lower bound is tight, is the graph connectivity problem: each of the two players is given a graph on n vertices, their goal is to determine whether the union of the two edge sets connects all the n vertices. It was proved by Dowling and Wilson [DW] that the logarithm of the rank of the input matrix for this problem is $\Omega(n \log n)$. Therefore the communication complexity of this problem is $\Omega(n \log n)$ (cf. [HMT]). Since our problem is trivially a subproblem of the problem of graph connectivity, we also achieved a lower bound for the non–deterministic communication complexity of the graph connectivity problem.

2 The Computation of the Rank

In this section we compute the rank of the matrix D. This computation was achieved thanks to the cruical help of Laci Lovasz and John H. Conway.

The rows and the columns of the matrix D are indexed by the $n!$ bipartite perfect matchings. The matchings will be denoted here by the appropriate permutations $\sigma \in S_n$.

[1] Since the appearance of a preliminary version of this paper in FOCS 93, a larger gap of n^ϵ was achieved by Nisan and Wigderson [NW].

For our method to compute the rank of D the following observation is crucial:

$$D_{\pi\tau} = \begin{cases} 1 & \text{if } \pi^{-1} \circ \tau \text{ is an } n\text{-cycle} \\ 0 & \text{otherwise} \end{cases}$$

Define

$$K_n = \{\pi \in S_n \mid \pi \text{ is an } n\text{-cycle}\}.$$

Now consider the regular representation[2] of S_n in $GL(\mathbf{R}^{n!})$ by $S_n \ni \pi \longrightarrow P_\pi$

$$(P_\pi)_{\sigma\tau} = \begin{cases} 1 & \text{if } \pi \circ \tau = \sigma \\ 0 & \text{otherwise} \end{cases}$$

The matrix P_π describes the action of π on S_n, i.e. how π permutes the elements of S_n, regarded as 0–1 incidence vectors of size $n!$. The matrix P_π is a permutation matrix (every row and column contains exactly one 1 and the remaining entries equal 0). The main diagonal of P_π contains no 1 unless $\pi = id$.

The following observations can be easily verified and give the connection between our matrix D and the regular representation of S_n :

1. $\forall\, \pi \in S_n \;\; DP_\pi = P_\pi D$
2. $D = \sum_{\pi \in K_n} P_\pi$

Every group representation decomposes into its irreducible components. Let U be a unitary matrix, of size $n! \times n!$, describing a linear transformation decomposing the regular representation of S_n. For every P_π, look at the conjugate matrix $UP_\pi U^{-1}$ (U transforms all the P_π simultaneously). For this unitary matrix U, every $UP_\pi U^{-1}$ is in a block diagonal form (i.e. a matrix of blocks sitting on the diagonal). The blocks structure is the same for all the P_π, and each block describes an irreducible representation of S_n (this is the meaninig of "decomposing into irreducible components").

The size (i.e. the number of rows or columns) of the block B, say, is also called its *dimension* d_B. Let $d_1,\ldots,d_k$ be the dimensions of the different irreducible representations (i.e. the dimensions of the corresponding blocks) of S_n. In the regular representation the multiplicity of an irreducible component is given by its dimension. This means that the multiplicity of a block in our matrices $UP_\pi U^{-1}$ (i.e. the number of copies of the same block or equivalent ones) equal to its size, and we have the following formula

$$n! \;=\; d_1^2 + \;\ldots\; d_k^2,$$

[2] For notations, the basic theory, and representation theory of the symmetric group cf. [S],[B]. Everything we use can also be found in the first four lectures of [FH].

Denote by B_i a block corresponding to the irreducible representation i of S_n. Then up to a change of blocks order, the matrices $UP_\pi U^{-1}$ are of the form described in Figure 1. w.l.o.g we can assume that for our matrix U all copies of the block B_i, in the matrices $UP_\pi U^{-1}$, are the same.

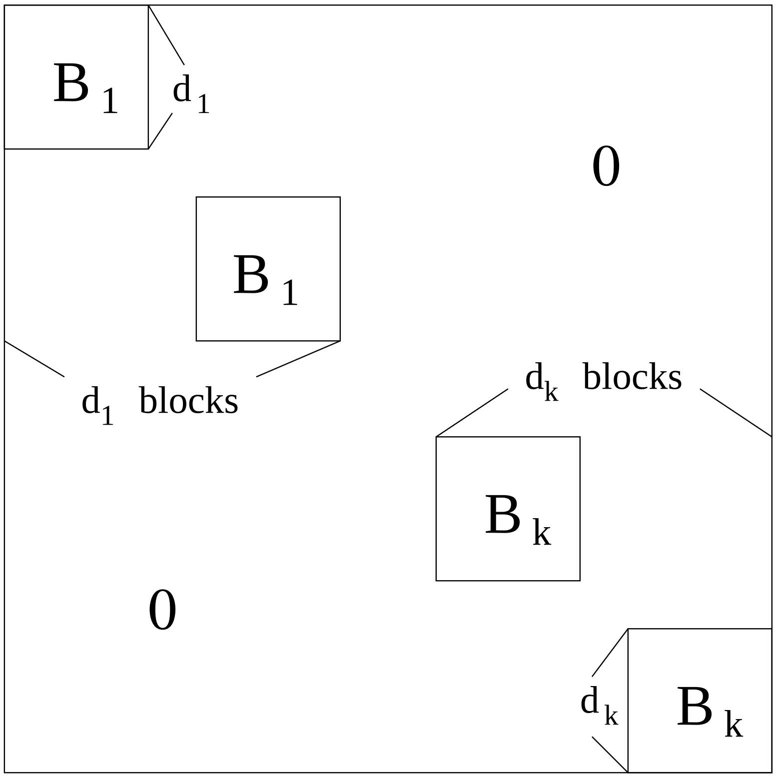

Figure 1: The Matrices $UP_\pi U^{-1}$

Given $\pi \in S_n$ where π is an element of a conjugacy class K, let B_i be one of the blocks of the matrix $UP_\pi U^{-1}$, corresponding to the irreducible representation i of S_n. The trace of B_i is the same for all $\pi \in K$. We write $\chi^K(i)$ for the trace of B_i on K. $\chi^K(i)$ is called the *character* of the (irreducible) representation i on the class K.

The following lemma is used to diagonalize our matrix D.

Lemma : *Let Y be an $n! \times n!$–matrix vanishing outside the blocks of the matrices $UP_\pi U^{-1}$. Assume that in the matrix Y all copies of the block i, corresponding to B_i, are the same. If Y commutes with every $UP_\pi U^{-1}$, $\pi \in S_n$, then Y is a diagonal matrix of the form described in Figure 2.*

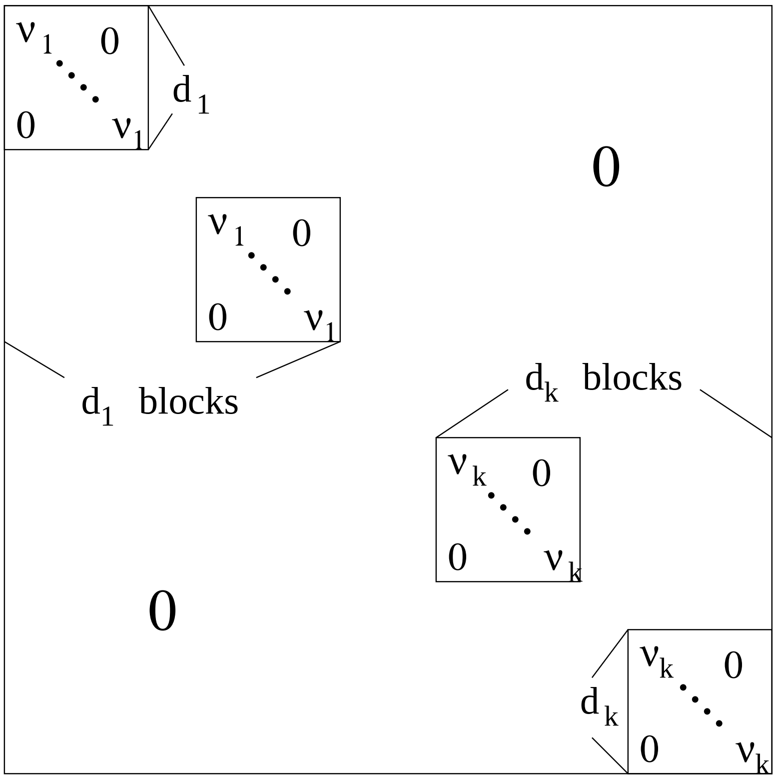

Figure 2: The Matrix UDU^{-1}

Proof: cf. [S], p. 23ff. □

Since $D = \sum_{\pi \in K_n} P_\pi$, UDU^{-1} trivially satisfies the lemma's conditions. Applying the lemma, the matrix UDU^{-1} of our problem is a diagonal matrix with d_j^2 diagonal entries ν_j, (for $j = 1,\ldots,k$). As $D = \sum_{\pi \in K_n} P_\pi$ the important fact is that

$$\chi^{K_n}(i) = 0 \Longleftrightarrow \text{all copies of the block } i \text{ of } UDU^{-1} \text{ are all–zero matrices}$$

Hence our attention is drawn to the character values of the different characters of S_n on the conjugacy class K_n .

Every conjugacy class of S_n is completely determined by the cycles structure

$$\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)\ ,\ \lambda_1 \geq \ldots \geq \lambda_l,$$

of its elements, where λ_j denotes a cycle of length λ_j. Of course, it holds that $n = \lambda_1 + \ldots + \lambda_l$. Hence, there are exactly as many different conjugacy classes in S_n as there are partitions of the integer n.

Definition: Suppose $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$. The *Ferrers diagram*, or *shape* t_λ, of λ is an array of n squares into l left–justified rows, with row j containing λ_j squares for $j \in \{1, \ldots, l\}$. Let δ_λ denote the number of ways in which one can fill the squares of the shape t_λ with the integers $1, \ldots, n$, in strictly increasing order in each of its rows (left to right) and in each of its columns (top down). δ_λ is called the *number of standard tableaux of shape* t_λ.

Example: Figure 3 describes the Ferrers diagram of $\lambda = (3, 2, 2)$, and one legal way of filling its squares with the integers $1, \ldots, 7$.

1	3	7
2	5	
4	6	

Figure 3: Ferrers Diagram

δ_λ can be computed by the so–called "hook"–algorithm:

For every square q in t_λ count the number of squares in the row to its right (not counting q). Add the number of squares in the column below it (not counting q). To this sum add 1. Do this for all the squares in t_λ. δ_λ then equals the ratio of $n!$ and the product of all the numbers just calculated.

Example: For $\lambda = (3, 2, 2)$,

$$\delta_\lambda = 7!/(5 \cdot 4 \cdot 1 \cdot 3 \cdot 2 \cdot 2 \cdot 1) = 21$$

There is a natural 1–1 correspondence between the set of conjugacy classes (partitions of n) of S_n and the set of irreducible representations of S_n . The dimension of the irreducible representation corresponding to the partition λ is given by δ_λ (cf. [S], Chapter 2, Chapter 3). So, for any irreducible representation i, we have $d_i = \delta_i$. It is well known (cf. e.g. [B], p.208ff) that the only characters not vanishing on K_n, are those where the shape of the corresponding partition is a single "turn", i.e. it has at most one row with more than one square and at most one column with more than one square. Given n there are n of these turns. Applying the hook–algorithm we see that the corresponding dimensions are

$$d_j = \binom{n-1}{j} \quad \text{for} \quad j = 0, \ldots, n-1.$$

Hence

$$d_0^2 + \ldots + d_{n-1}^2 = \sum_{j=0}^{n-1} \binom{n-1}{j}^2 = \binom{2n-2}{n-1}$$

The last expression in the previous formula is the rank of the matrix D.

3 The Lower Bound

Theorem 1 *:*
The non-deterministic communication complexity of D is $\Omega(n \log\log n)$.

Proof
Here we think of a matching $f \in M$ as a function $f : [n] \to [n]$. For each $f \in M$ define $g_f \in M$ by $g_f(i+1) = f(i)$, where the sum is taken mod$n$. Give f to player I and g_f to player II, then (f, g_f) form a Hamiltonian cycle and hence $D_{f,g_f} = 1$. Look at the set $E = \{(f, g_f) \mid f \in M\}$. E is a set of 1–entries of the matrix D. Let $M_1 \times M_2 \subset M \times M$ be an all–1 submatrix of D. Let Y be the projection of $E \cap (M_1 \times M_2)$ on the first index:

$$Y = \{f \in M_1 \mid (f, g_f) \in (M_1 \times M_2)\}.$$

Our goal will be to prove that Y is small. We need the following properties of Y:

1. $\forall f, f' \in Y\ ,\quad \forall i : f(i) \neq f'(i+1)$

   Proof

   If this is not the case then by the definition of Y we have $g_f \in M_2$ such that $g_f(i+1) = f(i) = f'(i+1)$. Therefore, the matchings g_f and f' have an edge in common, and do not form a Hamiltonian cycle. So $D_{f'g_f} = 0$, which contradicts the assumption that $M_1 \times M_2$ is an all–1 submatrix.

2. $\forall f, f' \in Y\ ,\quad \forall i,j :$ If $f(i) = f'(i), f(i+1) = f'(i+1), \ldots, f(j-1) = f'(j-1), f(j) = f'(j)$ then $f(i-1) \neq f'(j+1)$ (unless $j+1 = i-1$)

   Proof

   As before. If this is not the case then the image $f'[i, \ldots, j+1]$ is the same as the image $g_f[i, \ldots, j+1]$, so they do not form a Hamiltonian cycle.

By the anti-symmetry Lemma stated below, if n is big enough then for some constant ϵ

$$|Y| \leq \left(\frac{n}{\log^{\epsilon} n}\right)^n$$

Since the projection of E on the first index is M, the number of all–1 submatrices needed to cover the 1–entries of D is at least

$$|M| / \left(\frac{n}{\log^{\epsilon} n}\right)^n = 2^{\Omega(n \log\log n)}$$

□

Lemma 1 *(The anti-symmetry lemma):*
Let $Y \subset [n]^m$, and let $x = (x_1, \ldots, x_m)$ be a random variable uniformly distributed on Y. Suppose that x has the following two $anti-symmetry$ properties (property 2 is a generalization of property 1):

1. $\forall i, k :\quad \Pr(x_i = k) \cdot \Pr(x_{i+1} = k) = 0$

2. $\forall i, j, k, a_i, a_{i+1}, \ldots, a_{j-1}, a_j$ *if* $i - 1 \neq j + 1$ *then*

$$\Pr(x_{i-1} = k \mid x_i = a_i, \ldots, x_j = a_j) \cdot \Pr(x_{j+1} = k \mid x_i = a_i, \ldots, x_j = a_j) = 0$$

(when the sum is taken $\bmod m$*)*

Then

$$|Y| \leq \left(\frac{n}{\sqrt{\alpha}}\right)^m$$

for $\alpha = (\log n)^{1/1000}$ *,if* n *is big enough, and* $m = n$. [3]

Proof
Fix $\alpha = (\log n)^{1/1000}$.
In this proof we use asymptotic notations and assume that n is big enough. By H we denote the entropy function. The log in the proof is always taken base 2. We think of $x_1, \ldots, x_m$ as placed on the circle, so all the sums of indices are taken mod$m$.

Look at the random variable x_i conditioned on its j predecessors $(x_i \mid x_{i-1}, \ldots, x_{i-j})$, and at $(x_i \mid x_{i+1}, \ldots, x_{i+j})$. For i, j, k define A^k_{ij} to be the set of events $(x_{i-1} = a_{i-1}, \ldots, x_{i-j} = a_{i-j})$ that allow $x_i = k$, and B^k_{ij} to be the set of events $(x_{i+1} = a_{i+1}, \ldots, x_{i+j} = a_{i+j})$ that allow $x_i = k$. i.e.

$$(x = \bar{a}) \in A^k_{ij} \iff \Pr(x_i = k \mid x_{i-1} = a_{i-1}, \ldots, x_{i-j} = a_{i-j}) > 0$$

$$(x = \bar{a}) \in B^k_{ij} \iff \Pr(x_i = k \mid x_{i+1} = a_{i+1}, \ldots, x_{i+j} = a_{i+j}) > 0$$

By the definitions

$$j > j' \Rightarrow A^k_{ij} \subseteq A^k_{ij'} \ , \ B^k_{ij} \subseteq B^k_{ij'} \quad .$$

By the second condition of the lemma

$$A^k_{ij} \cap B^k_{i-j-1,j} = \emptyset \quad .$$

Combine these two formulas to get

$$(i - i' - 1)_{\mathrm{mod}\, m} \leq j, j' \quad \Rightarrow \quad A^k_{ij} \cap B^k_{i'j'} = \emptyset \tag{1}$$

In particular we look at A^k_{ij}, B^k_{ij} for $j = l \equiv n/100$.

If $|Y|$ is small then the entropy $H(x)$ is also small. Assume for a contradiction that the entropy $H(x)$ satisfies

$$H(x) > m(\log n - \log \alpha) \tag{2}$$

The proof is divided into three parts: In the first part we show that for a large number of pairs of indices (i, k), A^k_{il} and B^k_{il} are both large. This is finally achieved in Claim 4. More explicitly; we prove (for some C_1, C_2) the existence of a set I of $C_1 nm$ pairs of indices (i, k), with $\Pr(A^k_{il}), \Pr(B^k_{il}) \geq C_2$. The important thing to keep in mind about C_1, C_2 is that each is a polynomial in α^{-1}.

In the second part we work with the set I, and show some properties of this large set of pairs of indices. The goal of this part (claim 6) is to prove (for some C_3, C_4, C_5) the existence of an index $1 \leq i_0 \leq m$, an integer s_0, and $s = s_0/C_5 < n/100$, such that for at least a C_3 fraction of indices $k \in [n]$, the following three conditions hold:

[3] The reason for using different parameters for m and n is explained latter on

- $(i_0, k) \in I$
- For at least C_4 fraction of the indices i in the range $[i_0 - s, i_0 - 1]$; $(i, k) \in I$.
- $\Pr(A^k_{i_0 s}) \geq (1 - C_5) \Pr(A^k_{i_0 s_0})$

The important things to keep in mind about C_3, C_4, C_5 are that each is a polynomial in α^{-1}, and that C_5 is much smaller than all the others.

The heart of the proof is the third part. In this part we make a "random restriction", and fix a small amount of the random variables x_i. More explicitly; an index i_0, and an integer s_0 are carefully chosen as explained before (by the first two parts of the proof). Then all the variables indexed by i in the range $[i_0 - s_0, i_0 - 1]$ (i.e. $x_{i_0-1}, \ldots, x_{i_0-s_0}$) will be fixed. After this "fixing process", we will have a new scheme of $\tilde{m}$ random variables $\tilde{x}_i$ (of all the variables that were not fixed). These new random variables satisfy all the lemma's conditions (as the old ones), except the "unimportant" condition ; $m = n$. The fixing will be chosen to satisfy two conditions: first that the number of elements of Y that are consistent with the fixing is large (at least $|Y|/n^{s_0+3}$), and secondly, that the value of a potential function L goes up substantially. We then repeat the "fixing process" again and again until assumption (2) doesn't hold any more.

We look at the potential function $L = \sum_{i,k} \ln[2 - \ln(\Pr(B^k_{il}) + 1/n)]$. Using the first two parts of the proof, we show that after each step of "fixing process", L is much larger (at least by $\frac{1}{2} n s_0 \alpha^{10}$) for the new scheme of random variables (than for the old one). Since there is a trivial upper bound $L < nm(\ln\ln n + 1)$, it follows that we cannot repeat the "fixing process" more than a small number of times, and that the total number of variables that are fixed is small (less than n/α). Therefore, after a small number of times (and after fixing a small number of random variables) we must end up with a situation in which assumption (2) (our only assumption) is incorrect, i.e. $H(\tilde{x}) \leq \tilde{m}(\log n - \log \alpha)$. Therefore the appropriate set $\tilde{Y}$ is small. By the fact that we repeated the "fixing process" only a small number of times (and that we fixed only a small number of random variables), it will follow that the original set Y is also small.

So actually the proof is by induction. For simplicity, we do not state the exact induction step. We would like to say a few words here on the "unimportant" condition ; $m = n$: one of the conditions of the lemma is that $m = n$, but we do not use this condition until the last argument of the proof (after Claim 10). At that point (after Claim 10), we claim that all the arguments so far could be done for, say, $n \geq m \geq \frac{3}{4}n$. This is important because we want to repeat the "fixing process" more than one time. The reader can "be prepared" for that point, by convincing himself that all the following arguments are really valid for $n \geq m \geq \frac{3}{4}n$ (and not only for $m = n$). This is the reason for using different notations for m and n. It may be thought that it is unnecessary to introduce the parameter m: the parameter m is originally equal to n, but is reduced after fixing some variables. It may be thought that one can keep the same variable set, but replace the set Y by the subset that is consistent with the chosen fixing. The reason that we don't do that is that the density of this subset is much smaller than those of $\hat{Y}$, and Y. Therefore, by the fact that this set is small it would not be possible any more to conclude that Y is small. This could be solved by changing the stopping rule, but we do need assumption (2) to hold.

In the whole proof of the lemma, most of the work is done in the claims. The reader may choose to defer the claims' proofs for a later reading.

Our first goal is to prove that for a large set of pairs of indices (i,k) both A^k_{il}, B^k_{il} are large. This is finally achieved in Claim 4. For the proofs in this part we only use assumption (2) (that $H(x)$ is large), and the definitions. We do not use the properties of the set Y.

Claim 1 *:*

$$\sum_i H(x_i \mid x_{i-1},\ldots,x_{i-l}) \geq m(\log n - \frac{m}{m-l}\log\alpha) \geq m(\log n - 2\log\alpha)$$

$$\sum_i H(x_i \mid x_{i+1},\ldots,x_{i+l}) \geq m(\log n - \frac{m}{m-l}\log\alpha) \geq m(\log n - 2\log\alpha)$$

Proof

We prove the first inequality. The second is the same.

$$H(x) = H(x_1) + H(x_2 \mid x_1) + \ldots + H(x_m \mid x_1,\ldots,x_{m-1})$$

For $i > l$ substitute $H(x_i \mid x_1,\ldots,x_{i-1}) \leq H(x_i \mid x_{i-1},\ldots,x_{i-l})$, and for $i \leq l$ substitute $H(x_i \mid x_1,\ldots,x_{i-1}) \leq H(x_i) \leq \log n$ to get by inequality (2)

$$l\log n + \sum_{i=l+1}^{m} H(x_i \mid x_{i-1},\ldots,x_{i-l}) \geq m(\log n - \log\alpha)$$

This is true for every rotation of the indices, and the claim follows by summing over all possible rotations. $\square$

Define α_i by

$$H(x_i \mid x_{i-1},\ldots,x_{i-l}) = \log n - \log\alpha_i$$

Then by Claim 1

$$\frac{1}{m}\sum_i \log\alpha_i \leq \frac{m}{m-l}\log\alpha \leq 2\log\alpha \tag{3}$$

Claim 2 :

$$\sum_k \Pr(A_{il}^k) \geq n\alpha_i^{-1}$$

Proof

Denote by a the event $(x_{i-1} = a_{i-1}, \ldots, x_{i-l} = a_{i-l})$. Denote by K_a the number of indices k with $\Pr(x_i = k \mid a) > 0$. Then $\sum_k \Pr(A_{il}^k) = E_a(K_a)$ (by changing order of summations) and $H(x_i \mid a) \leq \log K_a$ (by the definition of the entropy function). Hence

$$\log n - \log \alpha_i = H(x_i \mid x_{i-1}, \ldots, x_{i-l}) = E_a[H(x_i \mid a)] \leq$$
$$E_a(\log K_a) \leq \log E_a(K_a) = \log \sum_k \Pr(A_{il}^k)$$

(the last inequality is by the convexity of the log function) □

The next claim should be compared to the previous one. It is not used in the proof, and is given here just as a comment.

Claim 3 :

$$\sum_{i,k} \Pr(A_{il}^k) \geq nm\alpha^{-2} \quad , \quad \sum_{i,k} \Pr(B_{il}^k) \geq nm\alpha^{-2}$$

Proof

By a convexity argument, Claim 2 and inequality (3) we have

$$\log\left[\frac{1}{nm}\sum_{ik} \Pr(A_{il}^k)\right] \geq \frac{1}{m}\sum_i \log[\frac{1}{n}\sum_k \Pr(A_{il}^k)] \geq \frac{1}{m}\sum_i(-\log\alpha_i) \geq -2\log\alpha$$

The second inequality follows in the same way. □

Claim 4 : *There exist a set I of $\frac{1}{2}nm\alpha^{-4}$ pairs of indices (i,k) with*

$$\Pr(A_{il}^k), \Pr(B_{il}^k) \geq \alpha^{-10}$$

Proof

Denote by a bar the average over the index i . Fix i and define:

$$K' = \{k \mid \Pr(A_{il}^k) < \alpha^{-10}\}$$
$$K'' = \{k \mid \Pr(A_{il}^k) \geq \alpha^{-10}\}$$
$$Y' = \{x \in Y \mid x_i \in K'\}$$
$$Y'' = \{x \in Y \mid x_i \in K''\}$$

Let x' be a random variable uniformly distributed on Y', and x'' a random variable uniformly distributed on Y''. Let $\gamma = |Y'|/|Y|$. Our first goal, here, will be to prove an upper bound for the average value of γ (the average is over i).

Define the random variable $\hat{x}$ by $\hat{x} = 0$ if $x_i \in K'$,and $\hat{x} = 1$ if $x_i \in K''$. We partitioned Y into Y', Y'' according to the i'th index. Therefore (notice that $\hat{x}$ is a function of x_i)

$$\begin{aligned}
&H(x_i \mid x_{i-1}, \ldots, x_{i-l}) = H(x_i, \hat{x} \mid x_{i-1}, \ldots, x_{i-l}) = \\
&H(x_i \mid x_{i-1}, \ldots, x_{i-l}, \hat{x}) + H(\hat{x} \mid x_{i-1}, \ldots, x_{i-l}) = \\
&\gamma H(x'_i \mid x'_{i-1}, \ldots, x'_{i-l}) + (1-\gamma) H(x''_i \mid x''_{i-1}, \ldots, x''_{i-l}) + H(\hat{x} \mid x_{i-1}, \ldots, x_{i-l})
\end{aligned}$$

And we get

$$H(x_i \mid x_{i-1}, \ldots, x_{i-l}) \leq \gamma H(x'_i \mid x'_{i-1}, \ldots, x'_{i-l}) + (1-\gamma) H(x''_i \mid x''_{i-1}, \ldots, x''_{i-l}) + 1$$

By the same argument as in Claim 2: $H(x'_i \mid x'_{i-1}, \ldots, x'_{i-l}) \leq \log n - \log \gamma\alpha^{10}$
[Formally the argument is the following: Define ${A'}^k_{il}$ to be the same as A^k_{il}, but for the variable x'. Then the argument in Claim 2 gives $H(x'_i \mid x'_{i-1}, \ldots, x'_{i-l}) \leq \log \sum_k \Pr({A'}^k_{il})$. The last expression is at most $\log n - \log \gamma\alpha^{10}$ because for $k \in K'$, $\Pr(A^k_{il}) < \alpha^{-10}$ (by the definition of K'), and $\Pr({A'}^k_{il}) = \gamma^{-1} \Pr(A^k_{il})$. This last equation is correct because, by the definitions, the two sets ${A'}^k_{il}$ and A^k_{il} are exactly the same thing. We just measure them differently. When measuring according to x', the measure is higher by a factor of γ^{-1} (because x' is distributed on a smaller set, and by the definition of γ the probability for this set is γ). For $k \in K''$, $\Pr({A'}^k_{il}) = 0$ by the definition of Y'. Thus we always have $\Pr({A'}^k_{il}) < \gamma^{-1}\alpha^{-10}$.]
Substitute also: $H(x''_i \mid x''_{i-1}, \ldots, x''_{i-l}) \leq \log n$ to get

$$\begin{aligned}
&H(x_i \mid x_{i-1}, \ldots, x_{i-l}) \leq 1 + \log n - \gamma \log \gamma\alpha^{10} = \\
&\log n - 10\gamma \log \alpha - \gamma \log \gamma + 1 \leq \log n - 10\gamma \log \alpha + 2
\end{aligned}$$

When averaging over i we get by Claim 1 :

$$\log n - 2 \log \alpha \leq \log n - 10\bar{\gamma} \log \alpha + 2 \ ,$$

hence if α is big enough (i.e. n is big enough) then $\bar{\gamma} \leq \frac{1}{4}$. All this is true, in the same way for the sets B^k_{il}.

Now, that we have upper bounds for $\bar{\gamma}$ for the sets A^k_{il}, B^k_{il}, we proceed and define:

$$\begin{aligned}
&K^0 = \{k \mid \Pr(A^k_{il}), \Pr(B^k_{il}) \geq \alpha^{-10}\} \\
&K^{00} = [n] - K^0 \\
&Y^0 = \{x \in Y \mid x_i \in K^0\} \\
&Y^{00} = \{x \in Y \mid x_i \in K^{00}\}
\end{aligned}$$

Our next goal is to give a lower bound for the average size of the set K^0, (again the average is over i).

Let x^0 be a random variable uniformly distributed on Y^0, and x^{00} a random variable uniformly distributed on Y^{00}. Let $\delta = |Y^0|/|Y|$. Then since $\bar{\gamma} \leq \frac{1}{4}$ for the sets A_{il}^k, B_{il}^k we have: $\bar{\delta} \geq \frac{1}{2}$, and as before

$$\log n - \log \alpha_i = H(x_i \mid x_{i-1}, \ldots, x_{i-l}) \leq$$
$$\delta H(x_i^0 \mid x_{i-1}^0, \ldots, x_{i-l}^0) + (1-\delta) H(x_i^{00} \mid x_{i-1}^{00}, \ldots, x_{i-l}^{00}) + 1$$

Substitute $H(x_i^0 \mid x_{i-1}^0, \ldots, x_{i-l}^0) \leq \log |K^0|$,and $H(x_i^{00} \mid x_{i-1}^{00}, \ldots, x_{i-l}^{00}) \leq \log n$ to get

$$\log n - \log \alpha_i \leq \delta \log |K^0| + (1-\delta) \log n + 1$$

So

$$\delta(\log n - \log |K^0|) \leq \log \alpha_i + 1$$

Recall that by the definitions, in all these equations, δ and $|K^0|$ are actually dependent on the fixed index i. Denote by Z the $\frac{1}{2}m$ indices i where $\log n - \log |K^0|$ is smallest. $\delta \leq 1$ and $\bar{\delta} \geq \frac{1}{2}$, therefore when summing the left hand side over i the minimal value we can get is when the weights δ are distributed on the indices in Z. Therefore

$$\sum_{i \in Z} (\log n - \log |K^0|) \leq \sum_i \delta(\log n - \log |K^0|) \leq \sum_i (\log \alpha_i + 1) \leq 2m \log \alpha$$

(the last inequality follows from inequality (3). The $+1$ here is negligible, because we can take in inequality (3) $m/(m-l) < 1.5$ rather than $m/(m-l) < 2$, as stated there. We assume here that n, m, α are large)
Hence (by the convexity of the log function)

$$\frac{1}{2} m \log\left(\frac{1}{\frac{1}{2}m} \sum_{i \in Z} |K^0|\right) \geq \sum_{i \in Z} \log |K^0| \geq \frac{1}{2} m \log n - 2m \log \alpha = \frac{1}{2} m \log n\alpha^{-4}$$

and we get

$$\sum_i |K^0| \geq \sum_{i \in Z} |K^0| \geq \frac{1}{2} mn\alpha^{-4}$$

So, taking $I = \{(i,k) \mid k \in K^0 \ \ for \ \ this \ \ i\}$ we get

$$|I| \geq \frac{1}{2} nm\alpha^{-4}$$

$\square$

We proceed now by working with the set I from the last claim. For every k define

$$I_k = \{i \mid (i,k) \in I\} \ .$$

Define

$$Z = \{k \mid |I_k| \geq \frac{1}{4} m\alpha^{-4}\} \ .$$

Then it is easy to see that

$$|Z| \geq \frac{1}{4} n\alpha^{-4} \ .$$

Redefine I to be

$$I = \{(i,k) \mid k \in Z, i \in I_k\} \ .$$

Now

$$|I| \geq \frac{1}{16} nm\alpha^{-8}.$$

Fix $k \in Z$ and look at the set I_k. For each $i \in I_k$ and for every distance s look at the set of all indices in $I_k \cap [i-1,\ldots,i-s]$. Define

$$J_{is}^k = \{j \mid j \in I_k, j \in [i-1,\ldots,i-s]\}$$

Claim 5 *: For $k \in Z$ and any distance $s \leq m$ there are at least*

$$\frac{1}{2}|I_k| \geq \frac{1}{8} m\alpha^{-4}$$

indices $i \in I_k$ with

$$|J_{is}^k| \geq \frac{s}{2m}|I_k| - 1 \geq \frac{1}{8} s\alpha^{-4} - 1 \ .$$

Proof

The proof follows from the following simple combinatorial claim: [4] Let X be a subset of Z_m (the integers mod m), and let $s \leq m$ be a positive integer. For $i \in X$, let J_i be the intersection of X with the set of the s elements preceding i. Let X' be the subset of X consisting of the indices i, with $|J_i| \geq |X|s/(2m) - 1$. Then $|X'| \geq |X|/2$.

Proof: Define $X'' = X - X'$ (i.e. the complement of X' in X). By averaging, there is an interval of length s that contains at least $|X''|s/m$ elements of X''. The last element i of X'' in this interval has (by the definition of J_i) $|J_i| \geq |X''|s/m - 1$. Since $i \in X''$, this is less than $|X|s/(2m) - 1$. Therefore $|X''|s/m - 1 < |X|s/(2m) - 1$. Thus $|X''| < |X|/2$. $\square$

Fix

$$\alpha_1 = 10^6 \alpha^{100}$$

[4] The following (simpler and more elegant) proof was given by the anonymous referee of this paper

(a big monomial of α) and look at the distances

$$s_r = \alpha_1^r \log n$$

for $r = 1, \ldots, d$ where $d = \alpha_1^4$.
Notice that

$$s_d = \alpha_1^{\alpha_1^4} \log n < l\ .$$

Call (i, k, r) a good triple if $k \in Z, i \in I_k$ and

$$|J_{is_r}^k| \geq \frac{1}{8} s_r \alpha^{-4} - 1$$

(as in the last claim). By the bounds we have for $|Z|$ and by the last claim, the number of good triples is at least

$$|Z| d \frac{1}{8} m \alpha^{-4} \geq \frac{1}{32} nmd\alpha^{-8}$$

For each triple (i, k, r) such that $r > 1$ define:

$$\gamma_{ir}^k = \Pr(A_{is_r}^k \mid A_{is_{r-1}}^k)$$

Call (i, k, r) a nice triple if $r > 1$ and $\gamma_{ir}^k > 1 - \alpha_1^{-1}$.

Claim 6 : *There exist two indices i_0, r_0 and a set W of indices k, such that $|W| \geq 10^{-2} n\alpha^{-8}$ and $\forall k \in W$: (i_0, k, r_0) is good and nice.*

Proof
Fix $k \in Z$ and $i \in I_k$. By the definition of I and the fact that if $j > j'$ we have $A_{ij}^k \subset A_{ij'}^k$ we know

$$\alpha^{-10} \leq \Pr(A_{il}^k) \leq \Pr(A_{is_d}^k) = \Pr(A_{is_1}^k) \Pr(A_{is_2}^k \mid A_{is_1}^k) \cdots \Pr(A_{is_d}^k \mid A_{is_{d-1}}^k) \leq \prod_{r=2}^{d} \gamma_{ir}^k$$

$\alpha^{-10} >> (1 - \alpha_1^{-1})^{\alpha_1^2}$ and therefore $\gamma_{ir}^k \leq 1 - \alpha_1^{-1}$ for at most α_1^2 indices r. For other indices $\gamma_{ir}^k > 1 - \alpha_1^{-1}$.
Summing on all the possibilities for $k \in Z$ and $i \in I_k$, we get that the number of "un-nice" triples (i, k, r) with $k \in Z$ and $i \in I_k$ is at most $\alpha_1^2 nm$. This number is negligible compared to $(\alpha^{-8}/32)nmd$, (because α_1^2/d is negligible compared to $\frac{1}{32}\alpha^{-8}$), which is a lower bound for the number of good triples. Therefore most of the good triples are also nice, and there are more than, say, $10^{-2} nmd\alpha^{-8}$ triples that are good and nice. The claim follows by a pigeon-hole principle. $\square$

We are ready now for the last part of the proof. In this part we are working with the indices i_0, r_0, and the set of indices W, from the last claim.

Define for $0 \leq x \leq 1$

$$f(x) = \ln[2 - \ln(x + 1/n)]$$

Basically this function behaves like $\ln(-\ln(x))$. The constant $1/n$ is added to ensure an upper bound of, say, $f(x) < \ln\ln n + 1$. The constant 2 is added to ensure concavity.
The reader can verify that $f(x)$ is monotone (decreasing) and concave for $0 \leq x \leq 1$.
Look at

$$L = \sum_{i,k} f(\Pr(B_{il}^k))$$

This measures in some sense the average size of B_{il}^k.

Fix i_0, r_0 from the last claim. Fix $s_0 = s_{r_0 - 1}$ and fix W to be the set W from the last claim. Now we are going to fix the values of $x_{i_0-1}, \ldots, x_{i_0-s_0}$. This is done randomly according to the distribution we have from x. When fixing the values of these s_0 variables we get a new scheme of $\tilde{m} = m - s_0$ random variables, for which the same conditions of the lemma trivially still hold. Except the condition $m = n$. For this new scheme define the sets $\tilde{A_{ij}^k}, \tilde{B_{ij}^k}$ exactly as before, and define $\tilde{L}$ exactly as before. $\tilde{L}$ is a random variable now (dependent on the way we fixed $x_{i_0-1}, \ldots, x_{i_0-s_0}$).

We use the following notations for the new scheme: by a we denote the event $x_{i_0-1} = a_{i_0-1}, \ldots, x_{i_0-s_0} = a_{i_0-s_0}$ (i.e. one way of fixing $x_{i_0-1}, \ldots, x_{i_0-s_0}$). The names of the indices in the new scheme will be their names in the old one. Of course some indices are missing (the indices of variables that were fixed). We simply ignore them. So in the new scheme the events $\tilde{A_{il}^k}$ and $\tilde{B_{il}^k}$ are sometimes dependent on a little less than l variables (because some of them were already fixed). This is done for simplifying the notations.

Look at the value of $\tilde{L}$ as a function of the event a. The goal of the next four claims is to show that there exist events a, occurring with a large probability, and for which $\tilde{L}$ is much larger than L. This goal is finally achieved in Claim 10.

Claim 7 *: If x_i was not fixed then*

$$E_a(\Pr(\tilde{B_{il}^k})) \leq \Pr(B_{il}^k)$$

Proof
The left hand side is the probability that $x_i = k$ is allowed after fixing $x_{i_0-1}, \ldots, x_{i_0-s_0}$ and $x_{i+1}, \ldots, x_{i+l}$, and this is trivially smaller than the probability that $x_i = k$ is allowed after fixing only $x_{i+1}, \ldots, x_{i+l}$. □
So, if x_i was not fixed, by the concavity of $f(x)$

$$E_a(f(\Pr(\tilde{B_{il}^k}))) \geq f(E_a(\Pr(\tilde{B_{il}^k}))) \geq f(\Pr(B_{il}^k)) \tag{4}$$

(f is monotonically decreasing)

Claim 8 *:*
If $k \in W$ *,* $i \in J^k_{i_0 s_{r_0}}$ *, and* x_i *was not fixed (i.e.* $i \in J^k_{i_0 s_{r_0}} \cap [i_0 - s_{r_0 - 1} - 1, i_0 - s_{r_0-1} - 2, \ldots, i_0 - s_{r_0}]$ *then*

$$E_a f(\Pr(\tilde{B^k_{il}})) \geq f(\Pr(B^k_{il})) + \frac{1}{10}\alpha^{-10}$$

Proof
Look at the value of $f(\Pr(\tilde{B^k_{il}})) - f(\Pr(B^k_{il}))$ as a function of the event a.
(i_0, k, r_0) is nice so

$$\Pr(A^k_{i_0 s_{r_0}} \mid A^k_{i_0 s_{r_0 - 1}}) > 1 - \alpha_1^{-1} \ .$$

Therefore, if the event a is in $A^k_{i_0 s_{r_0-1}}$ then

$$\Pr(\tilde{A}^k_{i_0 s_{r_0}}) > 1 - \alpha_1^{-1}$$

By formula (1) (from the beginning of the lemma), we know that $\tilde{A}^k_{i_0 s_{r_0}}$ and $\tilde{B^k_{il}}$ are disjoint. Therefore in this case

$$\Pr(\tilde{B^k_{il}}) < \alpha_1^{-1} < (\alpha^{-10})^5$$

But $i \in J^k_{i_0 s_{r_0}} \subset I_k$, hence by the definition of I

$$\Pr(B^k_{il}) \geq \alpha^{-10}$$

and therefore in this case, by the definition of f

$$f(\Pr(\tilde{B^k_{il}})) - f(\Pr(B^k_{il})) > \ln 4$$

(we assume that α is large and therefore $2, 1/n$ are negligible).
The event a is in $A^k_{i_0 s_{r_0-1}}$ with probability of at least α^{-10}, because (i_0, k, r_0) is good. So, the contribution after averaging over the event a is at least $\alpha^{-10} \ln 4$. This last fact is probably the heart of the whole proof, and is the reason for many of the previous definitions.

If a is not in $A^k_{i_0 s_{r_0-1}}$ then $f(\Pr(\tilde{B^k_{il}})) - f(\Pr(B^k_{il}))$ may be negative. By Claim 7 and by the fact that in f we take the log twice, we will show that this negative number is negligible on the average:

Define δ by $\delta = \Pr(B^k_{il})$ (then $\delta \geq \alpha^{-10}$). Look at events a with small $\Pr(\tilde{B^k_{il}})$. Define G to be the set of probability α^{-10} of events a , with the smallest $\Pr(\tilde{B^k_{il}})$. Define G' to be the set of probability $1 - \alpha^{-10}$ of all the other events a . Since for all $a \in A^k_{i_0 s_{r_0-1}}$ we showed that $f(\Pr(\tilde{B^k_{il}})) > f(\delta) + \ln 4$, we have the same for all $a \in G$.

Assume w.l.o.g that $g = \Pr(\tilde{B}_{il}^k)$ is the same for all $a \in G$, and that $g' = \Pr(\tilde{B}_{il}^k)$ is the same for all $a \in G'$ (we can assume this by the concavity of the function f, because we are trying to prove a lower bound for $E_a f(\Pr(\tilde{B}_{il}^k))$, and the value of $E_a f(\Pr(\tilde{B}_{il}^k))$ can only decrease when we rearrange uniformly the mass on G and on G').
Then, since g is a convex combination of the values $f(\Pr(\tilde{B}_{il}^k))$ of the events $a \in G$, we have

$$f(g) > f(\delta) + \ln 4.$$

Also, by Claim 7 we have

$$g' \leq \delta/(1 - \alpha^{-10}).$$

Thus we have (remember that f is monotonically decreasing)

$$\begin{aligned}
&E_a f(\Pr(\tilde{B}_{il}^k)) = \alpha^{-10} f(g) + (1 - \alpha^{-10}) f(g') \geq \\
&\alpha^{-10}(f(\delta) + \ln 4) + (1 - \alpha^{-10}) f(\delta/(1 - \alpha^{-10})) \geq \\
&\alpha^{-10}(f(\delta) + \ln 4) + (1 - \alpha^{-10}) f(\delta(1 + 2\alpha^{-10})) \approx \\
&\alpha^{-10}[f(\delta) + \ln 4] + (1 - \alpha^{-10})[f(\delta) + 2\delta\alpha^{-10} f'_\delta] > \\
&f(\delta) + \alpha^{-10}(\ln 4 + 2\delta f'_\delta)
\end{aligned}$$

(where f'_δ denotes the derivative of the function f at the point δ. f'_δ is negative)

$$f'_\delta = -[[2 - \ln(\delta + (1/n))][\delta + (1/n)]]^{-1} > -[[2 - \ln(\delta + (1/n))][\delta]]^{-1}.$$

Hence

$$2\delta f'_\delta > -2[2 - \ln(\delta + (1/n))]^{-1} > -2[2 - \ln(1 + (1/n))]^{-1} > -1.1$$

and we have

$$E_a f(\Pr(\tilde{B}_{il}^k)) \geq f(\delta) + \alpha^{-10}(\ln 4 - 1.1) \geq f(\delta) + \frac{1}{10}\alpha^{-10}$$

□

Claim 9 *:*

$$E_a(\tilde{L}) > L + n s_0 \alpha^{10}$$

Proof

$$\tilde{L} = \sum_{i,k} f(\Pr(\tilde{B}_{il}^k))$$

(the sum is taken only on indices i that were not fixed)

$$E_a \tilde{L} = \sum_{i,k} E_a f(\Pr(\tilde{B}_{il}^k))$$

(again, the sum is taken only on indices i that were not fixed)
Look at the value of $E_a f(\Pr(\tilde{B_{il}^k})) - f(\Pr(B_{il}^k))$ as a function of (i,k). The first thing to do is to lower bound the number of pairs of indices (i,k) that satisfy the hypotheses of Claim 8.

$$|W| \geq 10^{-2} n\alpha^{-8}$$

(by Claim 6).

$$\forall k \in W: \ |J^k_{i_0 s_{r_0}}| \geq \frac{1}{8} s_{r_0}\alpha^{-4} - 1$$

(by the definition of a good triple). Hence, Claim 8 holds for at least

$$10^{-2} n\alpha^{-8}[\frac{1}{8}s_{r_0}\alpha^{-4} - 1 - s_0] = 10^{-2}n\alpha^{-8}[\frac{1}{8}s_0\alpha_1\alpha^{-4} - 1 - s_0] >$$
$$ns_0[\frac{10^{-2}}{8}\alpha_1\alpha^{-12} - 2] > 10ns_0\alpha^{25}$$

pairs of indices (i,k).
So, for at least $10ns_0\alpha^{25}$ pair of indices (i,k)

$$E_a f(\Pr(\tilde{B_{il}^k})) \geq f(\Pr(B_{il}^k)) + \frac{1}{10}\alpha^{-10}$$

By inequality (4), for all the other pairs of indices

$$E_a f(\Pr(\tilde{B_{il}^k})) \geq f(\Pr(B_{il}^k))$$

We have to remember that in the sum $\tilde{L}$ all the fixed variables are not counted, but their contribution to L is at most $ns_0 f(0) < ns_0(\ln\ln n + 1)$. This is negligible because $\alpha >> \ln\ln n$. $\square$

For each event a denote by P_a the probability for a and by L_a the value of $\tilde{L}$ if a occurs.

Claim 10 *: There exists an event a such that*

$$P_a > n^{-3}n^{-s_0} \ , \ L_a > L + \frac{1}{2}ns_0\alpha^{10} .$$

Proof
$\forall a: \ L_a < nm(\ln\ln n + 1) < n^3$. Denote by γ the probability that $L_a > L + \frac{1}{2}ns_0\alpha^{10}$, then

$$E_a(\tilde{L}) = \sum_a P_a L_a \leq \gamma n^3 + (1-\gamma)(L + \frac{1}{2}ns_0\alpha^{10}) \leq \gamma n^3 + L + \frac{1}{2}ns_0\alpha^{10}$$

Hence, by the previous claim $\gamma n^3 > 1$. But γ is the sum of P_a for at most n^{s_0} possible events a. It follows that the probability of at least one of them is $P_a > n^{-3}n^{-s_0}$. $\square$

In the definition of L_a , in our notations, the sets $\tilde{B}_{il}^k$ are sometimes dependent on a little less than l variables (because some of the variables were already fixed). The "correct" definition of L_a should be similar to the definition of L. Whenever $\tilde{B}_{il}^k$ is dependent on less than l variables, we should redefine it (by "adding variables in the tail"). But by adding more variables we can only decrease the size of $\tilde{B}_{il}^k$, and therefore only increase the value of L_a. Therefore the claim is true also for the "correct" definition of L_a in the new scheme.

Fix a to be the event from Claim 10. We have now a scheme of $\tilde{m} = m - s_0$ random variables for which the conditions of the lemma still hold. The condition $\tilde{m} = n$ is not true but notice that so far we did not use it. All the arguments are correct for, say, $n \geq m \geq \frac{3}{4}n$. So we can repeat this process of fixing some of the random variables. We stop when the assumption $H(\tilde{x}) \geq \tilde{m}(\log n - \log \alpha)$ is not satisfied any more.

Denote by t the total number of times we repeat the "fixing process". Denote by $\tilde{s}$ the total number of variables that we fixed. Then $\tilde{m} = m - \tilde{s}$. Denote by a the event we choose in all the fixed variables. Then by Claim 10, $L_a > \frac{1}{2}n\tilde{s}\alpha^{10}$. But $L_a < nm(\ln\ln n + 1)$, so

$$\tilde{s} < \alpha^{-1}n \ , \ \tilde{m} > n(1 - \alpha^{-1}) \ .$$

By Claim 10

$$P_a > n^{-3t}n^{-\tilde{s}} \geq 2^{-3\tilde{s}}n^{-\tilde{s}}$$

(because $\tilde{s} \geq t \log n$, as we fix at least $\log n$ variables in each step).

By $H(\tilde{x}) \leq \tilde{m}(\log n - \log \alpha)$ we know that when fixing $\tilde{s}$ variables to the event a , we still have a set $\tilde{Y}$, for the variables, such that

$$|\tilde{Y}| \leq (n/\alpha)^{\tilde{m}} \ .$$

(The variable is uniformly distributed on this set).
By $P_a \geq 2^{-3\tilde{s}}n^{-\tilde{s}}$ we get

$$|Y| = |\tilde{Y}|P_a^{-1} \leq \left(\frac{n}{\alpha}\right)^{\tilde{m}} n^{\tilde{s}}2^{3\tilde{s}} \leq \frac{n^n}{\alpha^{\tilde{m}}}2^{3\tilde{s}} \leq \frac{n^n}{\alpha^{\frac{1}{2}n}} = \left(\frac{n}{\sqrt{\alpha}}\right)^m$$

$\square$

Acknowledgments

We would like to thank John H. Conway and Laci Lovasz for a crucial contribution to the proof given in Section 2. We want to express our gratitude to Andy Yao for suggesting us the work on the problem. Our thanks to all of them for helpful discussions and encouraging comments. We would like to thank A.A. Razborov for a very important remark, given in the introduction, and Paul Beame, and the anonymous referee for many important comments that helped to improve the readability of the proof given in section 3.

References

[AS] N. Alon, P.D. Seymour: "A counterexample to the rank-coloring conjecture", J. Graph Theory 13 (1989), pp. 523-525.

[B] Hermann Boerner: "Representation of Groups", North Holland, 1970.

[DW] T.A. Dowling, R.M. Wilson: "Whitney Number Inequalities for Geometric Lattices", Proc. Amer. Math. Soc. 47 (1975), pp. 504-512.

[FH] W. Fulton, J. Harris: "Representation Theory", Springer-Verlag, 1991.

[HMT] A. Hajnal, W. Maass, G. Turan: "On the Communication Complexity of graph Properties", Proc. 20th ACM STOC 1988, pp. 186–191.

[L] L. Lovász: "Communication Complexity: A Survey", in "Paths, Flows and VLSI–layout", Springer 1990, pp. 235–266.

[LS] L. Lovász, M.Saks: "Communication Complexity and Combinatorial Lattice Theory", Journal of Computer and System Sciences 47 (1993), pp. 322-349.

[MS] K. Mehlhorn, E.M. Schmidt: "Las Vegas is Better than Determinism in VLSI and Distributed Computing", Proc. 14th ACM STOC 1982, pp. 330–337.

[NW] N. Nisan, A. Wigderson: "On Rank vs. Communication Complexity", Proc. 35th IEEE FOCS 1994, pp. 831–836.

[R] A. Razborov: "The Gap between the Chromatic Number of a Graph and the Rank of its Adjacency Matrix is Superlinear", Discrete Mathematics 108 (1992), pp. 393–396.

[S] Bruce E. Sagan: "The Symmetric Group", Brooks & Cole, 1991.

[Y] Andrew C. Yao: "Some Complexity Questions Related to Distributive Computing", Proc. 11th ACM STOC 1979, pp. 209-213.