When is a sum of double cosets of S_n almost Boolean?

Yuval Filmus

18 Oct 2021

Abstract

A function $f\colon S_n \to \mathbb{R}$ of the form $f(\pi) = \sum_{ij} c_{ij} 1_{\pi(i)=j}$ is Boolean if it is a *dictator* (depends on a single $\pi(i)$ or on a single $\pi^{-1}(j)$). What can we say about f if it is *almost* Boolean? We answer this question for several different notions of *almost*: L_2, L_0, L_∞.

1 Motivation: Erdős–Ko–Rado theorem for permutations

We say that two permutations α, β *intersect* if they agree on the image of some point: $\alpha(i) = \beta(i)$ for some i. A family $\mathcal{F}$ of permutations is *intersecting* if every two permutations in $\mathcal{F}$ intersect.

> How large can an intersecting family of permutations on n elements be?

A simple argument along the lines of Katona's circle argument [Kat72], due to Frankl and Deza [FD77], shows that any such family contains at most $(n-1)!$ permutations. The idea is to decompose the group S_n of all permutations on n elements into $(n-1)!$ many subsets of length n, each consisting of a single permutation and its cyclic shifts. For example, if $n = 5$, then one such subset is

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 1 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 1 & 2 & 3 \\ 5 & 1 & 2 & 3 & 4 \end{matrix}$$

These permutations are pairwise non-intersecting, and so every intersecting family contains at most one of them. Since S_n can be decomposed into $(n-1)!$ such subsets, it follows that every intersecting family contains at most $(n-1)!$ many permutations. This is the *Erdős–Ko–Rado theorem for permutations*, generalizing a classical result on intersecting families of sets [EKR61].

Conversely, the following intersecting family contains $(n-1)!$ permutations: all permutations mapping i to j. We call such families *double cosets*,[1] and they also go by other names, such as *stars*, *links*, and *canonical intersecting families*. This suggests a natural follow-up question:

> What are all intersecting families of the maximum size $(n-1)!$? Are they all canonical?

The answer to this question is positive, as shown by Cameron and Ku [CK03], Larose and Malvenuto [LM04], Godsil and Meagher [GM09], and Ellis, Friedgut and Pilpel [EFP11], using different techniques. Some of these proofs are more ad hoc than others. The proof that spurred our work is the one due to Ellis, Friedgut and Pilpel, which uses spectral techniques, and is arguably more conceptual than the others.

We present their proof below. First, however, let us mention that in some settings, not all intersecting families of maximum size are canonical, and so the question is not redundant. A case in point is the

[1] They are double coset of the point stabilizer of 1.

alternating group A_n, which is the group of all even permutations. Ku and Wong [KW07] gave an example of an intersecting family of maximum size on 4 points which is not canonical:

$$\begin{matrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \\ 2 & 3 & 1 & 4 \end{matrix}$$

They also showed that non-canonical examples occur only for $n = 4$.

Spectral proof Ku [KW07] proposed a proof of the Erdős–Ko–Rado theorem for permutations using the Hoffman bound, which was realized by Renteln [Ren07]. The idea of the proof (replacing the Hoffman bound with the slightly simpler Lovász bound) is to construct a symmetric $S_n \times S_n$ matrix A with the following properties:

1. If α, β are intersecting then $A_{\alpha\beta} = 1$.
2. The maximal eigenvalue of A is $(n-1)!$.

Suppose we have constructed such a matrix (we explain below how to do it). Let f be the indicator vector of an intersecting family $\mathcal{F}$. Then

$$|\mathcal{F}|^2 = f^T A f \leq (n-1)! f^T f = (n-1)!|\mathcal{F}|,$$

and so $|\mathcal{F}| \leq (n-1)!$. Moreover, if $|\mathcal{F}| = (n-1)!$, then f must lie in the eigenspace of $(n-1)!$. Ellis, Friedgut and Pilpel used this additional property to classify all intersecting families of maximum size.

But first, let us explain how to construct A.[2] Our starting point is the all-ones matrix J, and the adjacency matrix B of the *derangement graph*, which is the graph on S_n in which two permutations are connected if they *do not* intersect. The maximum eigenvalue of B is clearly D_n, the number of derangements (permutations without any fixed points), and the corresponding eigenspace $V_{\max}$ consists of all constant vectors (this is because the derangement graph is connected). Renteln showed that the minimum eigenvalue is $-D_n/(n-1)$, as conjectured by Ku [KW07], and described its eigenspace $V_{\min}$ explicitly when $n \geq 5$.

Given this data, we take

$$A = J - \frac{(n-1)(n-1)!}{D_n} B,$$

which clearly satisfies the first property above. The all-ones matrix J has two eigenspace: $V_{\max}$ corresponds to the eigenvalue $n!$, and $V_{\max}^{\perp}$ corresponds to the eigenvalue 0. This means that J and B have common eigenspaces, allowing us to determine the maximal eigenvalue of A. There are two competing options for this eigenvalue: the one corresponding to $V_{\max}$, and the one corresponding to $V_{\min}$. The eigenvalue corresponding to $V_{\max}$ is $n! - \frac{(n-1)(n-1)!}{D_n} D_n = (n-1)!$, and the one corresponding to $V_{\min}$ is $-\frac{(n-1)(n-1)!}{D_n} \cdot -\frac{D_n}{n-1} = (n-1)!$. The two eigenvalues coincide, showing that the maximum eigenvalue of A is indeed $(n-1)!$. Furthermore, the corresponding eigenspace is $U_1 := V_{\max} \oplus V_{\min}$.

Summarizing, if $\mathcal{F}$ is an intersecting family of size $(n-1)!$, then its characteristic function f belongs to the eigenspace U_1 (for $n \geq 5$). We can describe this eigenspace explicitly: it is the linear span of the characteristic functions of double cosets! We can therefore express f in the following form:

$$f = \sum_{i,j} c_{ij} x_{ij},$$

where x_{ij} is the indicator of "i goes to j", that is, $x_{ij} = 1$ if the input permutation sends i to j. In other words, x_{ij} is the characteristic function of the double coset consisting of all permutations mapping i to j. Another way of thinking of x_{ij} is as the (i,j)'th entry of the input permutation, encoded as a *permutation matrix*.

[2] Grayed out paragraphs can be skipped on first reading.

Every function on S_n can be written as *some* polynomial in the variables x_{ij}. The *degree* of a function is the minimum degree of such a polynomial. A function belongs to U_1 if it has degree at most 1; we also say that it is linear. You might wonder why we don't allow a constant coefficient. Such a coefficient is not needed, since it can be expressed via the identity $\sum_j x_{1j} = 1$; however, constant coefficients will be useful at some point below.

Reiterating, if $\mathcal{F}$ is an intersecting family of size $(n-1)!$, then its characteristic function f is a Boolean function (that is, a function attaining the values 0 and 1) of degree at most 1.

What are the Boolean degree 1 functions on S_n?

In the case of the Boolean cube $\{0,1\}^n$, this is a simple exercise. A function on the Boolean cube has degree at most 1 if it can be written in the form

$$f(x_1,\ldots,x_n) = c + \sum_{i=1}^n c_i x_i.$$

(This time we cannot get rid of the constant coefficient.) If f is Boolean, then at most one of the c_i's can be non-zero, since otherwise the function will attain at least three different values. Therefore f must depend on at most one coordinate x_i, and consequently, it is one of the following functions: $0, 1, x_i, 1-x_i$; we call such a function a *dictator* (this includes constants).

The case of the symmetric group is more complicated: it is no longer true that at most one of the c_{ij}'s can be non-zero, since there are pairs of variables $x_{i_1j_1}, x_{i_2j_2}$ which cannot equal 1 at the same time; we say that they are *disjoint*, since the corresponding double cosets are disjoint. A pair of variables is disjoint if the two variables are on the same row ($i_1 = i_2$) or column ($j_1 = j_2$).

A sum of pairwise disjoint x_{ij}'s is always Boolean. This gives rise to two more examples of Boolean degree 1 functions:

$$\sum_{j \in J} x_{ij} \text{ for some } 1 \leq i \leq n \text{ and } J \subseteq \{1,\ldots,n\},$$
$$\sum_{i \in I} x_{ij} \text{ for some } 1 \leq j \leq n \text{ and } I \subseteq \{1,\ldots,n\}.$$

These classes of functions are closed under negation ($f \mapsto 1-f$), so in contrast to the case of the Boolean cube, there is no need to explicitly list their negations. We call such functions *dictators*, since in the first case, the value of the function on a permutation π depends on $\pi(i)$, and in the second case, it depends on $\pi^{-1}(j)$.

Ellis, Friedgut and Pilpel [EFP11] showed that all Boolean degree 1 functions are of this form, using the Birkhoff–von Neumann theorem. This theorem states that every doubly stochastic matrix is a linear combination of permutation matrices. Equivalently, the polytope $P \subset \mathbb{R}^{n^2}$ whose vertices are the permutation matrices is represented by the inequalities $x_{ij} \geq 0$ for all i,j and by the equalities $\sum_j x_{ij} = 1$ for all i and $\sum_i x_{ij} = 1$ for all j.

Now suppose that $f = \sum_{i,j} c_{ij} x_{ij}$ is Boolean and non-zero. Let F be the linear span of $f^{-1}(0)$, viewed as a set of permutation matrices. Since f is linear, $f(F) = 0$, and so F is a face of the polytope (it cannot be all of $\mathbb{R}^{n^2}$ since f is non-zero), and so it is contained in some facet $x_{ij} = 0$. Consequently, if $x_{ij} = 1$ then $f(x) \neq 0$; since f is Boolean, in fact $f(x) = 1$ for such x. Therefore $f - x_{ij}$ is another Boolean degree 1 function, with fewer 1's. Continuing in this way, we get that f is a sum of x_{ij}'s, which must be disjoint since f is Boolean.

The only sum of disjoint x_{ij}'s which is the characteristic function of a family of size $(n-1)!$ is a sum of length 1, and so an intersecting family of size $(n-1)!$ must be a double coset.

Stability The next question that comes to mind is:

How do intersecting families of size $(1-\epsilon)(n-1)!$ look like? Are they all "essentially" canonical?

We can attack this question using the spectral technique: if f is the characteristic function of an intersecting family of size $(1-\epsilon)(n-1)!$, then f must be *close* to a degree 1 function, in the sense that $\mathbb{E}[(f-g)^2] = O(\epsilon)$ for some degree 1 function g;[3] we call this *closeness in* L_2, and it is the default notion of closeness that we use below.

What can we say about Boolean functions which are close to degree 1? This is the question that will occupy us in the rest of the talk, rephrased from the point of view of g. Namely, instead of studying Boolean functions which are close to degree 1, we will study degree 1 functions which are close to Boolean; the two questions are equivalent.

First, however, we should mention that Ellis [Ell12a] gave a very strong answer to the question about intersecting families of permutations. Verifying a conjecture of Cameron and Ku [CK03], Ellis showed that every intersecting family of size at least $(1-1/e+o(1))(n-1)!$ is in fact *contained* in a double coset; this is optimal by considering the family obtained by starting with the double coset x_{11}, adding a permutation $\pi \notin x_{11}$, and throwing out all permutations in x_{11} which do not intersect π.

Ellis's proof combines an isoperimetric inequality on the transposition graph (the graph on S_n in which two permutations are connected if they differ by a transposition) with the *cross-intersecting* version of the Erdős–Ko–Rado theorem, which states that if $\mathcal{F}, \mathcal{G}$ are such that every permutation in $\mathcal{F}$ intersects every permutation in $\mathcal{G}$, then $|\mathcal{F}| \cdot |\mathcal{G}| \leq (n-1)!^2$; this cross-intersecting version can be proved using the spectral argument outlined above, together with the additional information that the *minimal* eigenvalue of A is at least $-(n-1)!$.

2 Linear functions close to Boolean

Here is the research question which we will answer in this talk:

Suppose that f is a degree 1 function which is ϵ-close to Boolean, that is, $\mathbb{E}[(f-F)^2] \leq \epsilon$ for some Boolean function F. What can we say about f?

A few years ago, together with David Ellis and Ehud Friedgut we came up with two different answers to this question:

- **Sparse regime [EFF15a]:** If f has degree 1, $\mathbb{E}[f] = c/n$, and f is $\epsilon c/n$-close to Boolean, then f is $O(c^2\sqrt{\epsilon}/n + (c/n)^2)$-close to a sum $x_{i_1j_1} + \cdots + x_{i_mj_m}$, where $m \approx c$ and the summands are *not necessarily disjoint.*
- **Dense regime [EFF15b]:** If f has degree 1, $\eta \leq \mathbb{E}[f] \leq 1-\eta$, and f is ϵ-close to Boolean, then f is $O(\epsilon^{1/7}/\eta + n^{-1/3}/\eta)$-close to a dictator.

The two results are most effective in two different regimes. For the first result, we think of c as small, even constant. The function f is very sparse, and so we have to compare its distance from Boolean to its sparsity (otherwise it is trivially close to the zero function). For constant c, we get that f is $O(\sqrt{\epsilon}/n + 1/n^2)$-close to a sum of possibly non-disjoint x_{ij}'s. We cannot guarantee that the summands are disjoint, since a function of the form $x_{11} + x_{22}$ is quite close to Boolean: the distance is only $O(1/n^2)$. If c is linear in n then the first result is completely useless.

For the second result, we think of η as constant, and so the function f is balanced. The result then states that f is $O(\epsilon^{1/7} + 1/n^{1/3})$-close to a dictator. In contrast to the first result, here the approximation is a bona fide dictator, intuitively since a sum of many non-disjoint x_{ij}'s is *not* close to Boolean; this will come up again below. If we take $\eta = O(1/n)$, in order to accommodate the kind of sparse functions handled by the first result, then the second result becomes completely useless.

[3]In order to deduce this we also need A to have an *eigenvalue gap*: the second largest eigenvalue must be bounded away from $(n-1)!$. This can be shown using the method of Renteln.

There are many problems with these two results. First of all, they do not cover the entire range of sparsity: what happens if f is neither very sparse nor balanced? Second, when ϵ gets very small, the approximation error gets "stuck", since there is a dependence on n. Third, the dependence on ϵ is not optimal: ideally, we would like the approximation error to be proportional to the distance from Boolean.

In short, our dream result would look as follows:

> There is an explicit class $\mathcal{F}_\epsilon$ of functions (depending on n) such that:
>
> - If f is a degree 1 function which is ϵ-close to Boolean, then f is $O(\epsilon)$-close to some function in $\mathcal{F}_\epsilon$.
> - Every function in $\mathcal{F}_\epsilon$ is $O(\epsilon)$-close to Boolean.

Such a result exists for the Boolean cube: the Friedgut–Kalai–Naor (FKN) theorem [FKN02].

> If $f\colon \{0,1\}^n \to \mathbb{R}$ is ϵ-close to Boolean, then it is $O(\epsilon)$-close to a dictator.

Our goal is to obtain an analogous result for the symmetric group. Taking a look at the two results mentioned above, it is not so clear what $\mathcal{F}_\epsilon$ should be. The first result suggests taking as $\mathcal{F}_\epsilon$ an arbitrary sum of x_{ij}'s, while the second result suggests taking as $\mathcal{F}_\epsilon$ the set of dictators, matching the FKN theorem. How do we bridge this gap?

We will take as inspiration another extension of the FKN theorem, this time to the (n,k)-slice, the domain consisting of all vectors in $\{0,1\}^n$ with exactly k many 1s. This domain, also known as the *Johnson association scheme*, is often denoted by $\binom{[n]}{k}$ or $J(n,k)$. We define $p = k/n$, and usually assume that $p \leq 1/2$. A function on the slice has degree 1 if it can be written in the form

$$f(x_1,\ldots,x_n) = \sum_{i=1}^n c_i x_i.$$

(Once again, we don't need a constant coefficient, since $x_1 + \cdots + x_n = k$.) Here is the FKN theorem for the slice [Fil16]:

> If $f\colon \binom{[n]}{pn} \to \mathbb{R}$ is a degree 1 function which is ϵ-close to Boolean, where $p \leq 1/2$, then f is $O(\epsilon)$-close to a function of the form
> $$x_{i_1} + \cdots + x_{i_m} \text{ or } 1 - x_{i_1} - \cdots - x_{i_m},$$
> where $m \leq 1$ or $m = O(\sqrt{\epsilon}/p)$.

Where does the mysterious bound $O(\sqrt{\epsilon}/p)$ come from? Consider the function

$$f = x_1 + \cdots + x_m,$$

where we think of m as "not too large". Since every triplet of coordinates is almost independent (unless pn is extremely small),

$$\binom{m}{2}p^2 \gtrsim \Pr[f \geq 2] \gtrsim \binom{m}{2}p^2 - \binom{m}{3}p^3.$$

If mp is large then f will not be close to Boolean (this will follow from the discussion below), and otherwise $\Pr[f \geq 2] \approx \binom{m}{2}p^2$.

If $m \leq 1$ then f is Boolean. Otherwise, $\Pr[f \geq 2] \approx (mp)^2/2$, and so for f to be ϵ-close to Boolean, we need $(mp)^2 \leq 2\epsilon$, or $m \leq \sqrt{2\epsilon}/p$; conversely, for such values of m, calculation shows that f is indeed ϵ-close to Boolean (essentially because the distribution of f is roughly Poisson).

An analogous calculation shows that $\Pr[f \geq 1] \approx mp = O(\sqrt{\epsilon})$, and so we obtain the following curious corollary:

If $f\colon \binom{[n]}{pn} \to \mathbb{R}$ is ϵ-close to Boolean then f is $O(\sqrt{\epsilon})$-close to a dictator.

(We cannot say that f is close to a constant, since we also have to consider the case $m = 1$.)

What this corollary shows is that when looked at from afar (at a scale of $O(\sqrt{\epsilon})$), the only functions close to Boolean are dictators, while up close, a more interesting picture emerges. This, it turns out, explains the discrepancy between the different results for the symmetric group in the sparse and dense regimes.

At this point we can hazard a guess at what $\mathcal{F}_\epsilon$ should look like: it should consist of all sums of x_{ij}'s which "look Boolean", together with their negation. In contrast to the situation on the slice, in which any two x_i's are non-disjoint, on the symmetric group the notion of non-disjointness is non-trivial. If $x_{i_1j_1}, x_{i_2j_2}$ are two non-disjoint variables then they contribute $1/n^2$ to $\Pr[f \geq 2]$, and so the number of such pairs should be limited to $O(\epsilon n^2)$. Moreover, for technical reasons, we will need to limit the number of summands to $O(n)$. We arrive at the following definition:

The class $\mathcal{F}_\epsilon(K)$ consists of all sums $x_{i_1j_1} + \cdots + x_{i_mj_m}$ and $1 - x_{i_1j_1} - \cdots - x_{i_mj_m}$ satisfying

1. $m \leq Kn$.
2. The number of non-disjoint pairs $x_{i_sj_s}, x_{i_tj_t}$ is at most $K\epsilon n^2$.

Our main result states that an analog of the FKN theorem holds for the symmetric group with respect to this class, for some constant $K > 0$:

- If f is a degree 1 function which is ϵ-close to Boolean, then f is $O(\epsilon)$-close to a function in $\mathcal{F}_\epsilon(K)$.
- Every function in $\mathcal{F}_\epsilon(K)$ is $O(\epsilon)$-close to Boolean.

The second part is an easy calculation, in which we need to use the bound $m \leq Kn$. We describe the proof of the first part, which is the hard direction, in the rest of the talk.

But first, when does a collection $\mathcal{C}$ of $m \leq Kn$ double cosets contain only $K\epsilon n^2$ many non-disjoint pairs? We consider two cases: either there is a "prominent line" $\mathcal{L}$, which is a row or column containing more than $\sqrt{\epsilon}n$ many double cosets, or no line is prominent.

In the first case, any double coset outside of $\mathcal{L}$ intersect all double cosets in $\mathcal{L}$, save possibly one, and so there are at least $(m-|\mathcal{L}|)(|\mathcal{L}|-1) \geq (m-|\mathcal{L}|)\sqrt{\epsilon}n$ many non-disjoint pairs. Consequently, $m-|\mathcal{L}| \leq K\sqrt{\epsilon}n$, that is, almost all of $\mathcal{C}$ lies inside the prominent line $\mathcal{L}$.

In the second case, every double coset in $\mathcal{C}$ intersects at most $2\sqrt{\epsilon}n$ other double cosets in $\mathcal{C}$, and so there are at least $|\mathcal{C}|(|\mathcal{C}| - 2\sqrt{\epsilon}n)/2$ many non-disjoint pairs. Hence either $|\mathcal{C}| \leq 4\sqrt{\epsilon}n$, or the number of non-disjoint pairs is at least $|\mathcal{C}|^2/4$, and consequently $|\mathcal{C}| \leq 2\sqrt{\epsilon}n$.

Suppose that g is the sum of the x_{ij}'s corresponding to the double cosets in $\mathcal{C}$. In the first case, by throwing out all double cosets outside of $\mathcal{L}$, we can approximate g up to an error of $O(\sqrt{\epsilon})$ by the sum of double cosets in $\mathcal{L}$, which is a dictator. In the second case, g is sparse: it is $O(\sqrt{\epsilon})$-close to the zero function. Taking g to be the function in $\mathcal{F}_\epsilon(K)$ promised by our FKN theorem, we conclude the following corollary:

If a degree 1 function on the symmetric group is ϵ-close to Boolean, then it is $O(\sqrt{\epsilon})$-close to a dictator.

This corollary is tight, as demonstrated by the function

$$f = \sum_{i=1}^{\sqrt{\epsilon}n} x_{ii}.$$

Once again, at an error scale of $O(\sqrt{\epsilon})$, the only degree 1 functions which are almost Boolean are dictators, while at a scale of $O(\epsilon)$, a more nuanced picture emerges.

3 On the proof

The main idea behind the proof, borrowed from [Fil16], is to reduce the FKN theorem on the symmetric group to the FKN theorem on the Boolean cube, by finding an "equitable" covering of the symmetric group by high-dimensional Boolean cubes.

Suppose for simplicity that n is even. We can construct a copy of the $(n/2)$-dimensional Boolean cube $\{0,1\}^{n/2}$ as follows. Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be two arbitrary permutations of $1, \ldots, n$. Given a point $y \in \{0,1\}^{n/2}$, we construct a permutation $\pi \in S_n$ using the following rules:

1. If $y_1 = 1$ then $\pi(a_1) = b_1$ and $\pi(a_2) = b_2$, and otherwise $\pi(a_1) = b_2$ and $\pi(a_2) = b_1$.
2. If $y_2 = 1$ then $\pi(a_3) = b_3$ and $\pi(a_4) = b_4$, and otherwise $\pi(a_3) = b_4$ and $\pi(a_4) = b_3$.
3. ...
4. If $y_{n/2} = 1$ then $\pi(a_{n-1}) = b_{n-1}$ and $\pi(a_n) = b_n$, and otherwise $\pi(a_{n-1}) = b_n$ and $\pi(a_n) = b_{n-1}$.

In other words, π always sends $\{a_1, a_2\}$ to $\{b_1, b_2\}$, $\{a_3, a_4\}$ to $\{b_3, b_4\}$, and so on; y_1 determines whether a_1 gets sent to b_1 and a_2 to b_2, or vice versa; and so on. We denote the resulting set of $2^{n/2}$ permutations by $\mathcal{D}_{a,b}$.

Suppose that f is a degree 1 function given by the formula

$$f = \sum_{i,j} c_{ij} x_{ij}.$$

What happens when we restrict f to $\mathcal{D}_{a,b}$? We get

$$\begin{aligned} f|_{\mathcal{D}_{a,b}} &= y_1(c_{a_1b_1} + c_{a_2b_2}) + (1-y_1)(c_{a_1b_2} + c_{a_2b_1}) + \cdots + y_{n/2}(c_{a_{n-1}b_{n-1}} + c_{a_nb_n}) + (1-y_{n/2})(c_{a_{n-1}b_n} + c_{a_nb_{n-1}}) \\ &= y_1(c_{a_1b_1} + c_{a_2b_2} - c_{a_1b_2} - c_{a_2b_1}) + \cdots + y_{n/2}(c_{a_{n-1}b_{n-1}} + c_{a_nb_n} - c_{a_{n-1}b_n} - c_{a_nb_{n-1}}) + C_{f,a,b}, \end{aligned}$$

where $C_{f,a,b}$ is some constant independent of y.

The crucial observation is that since f is close to Boolean, for random a, b, the function $f|_{\mathcal{D}_{a,b}}$ is also close to Boolean, and so the FKN theorem for the Boolean cube applies to it. The reason is that if we choose a, b at random and a random point in $\mathcal{D}_{a,b}$, then we just get a random permutation (this is what we meant by an equitable covering). The FKN theorem tells us something about the coefficients of the function $f|_{\mathcal{D}_{a,b}}$, which will allow us to understand the coefficients c_{ij} of f.

The FKN theorem implies that $f|_{\mathcal{D}_{a,b}}$ is close to a function of one of the following forms: $0, 1, x_i, 1 - x_i$. In particular, the coefficients in front of the y_i are either all close to zero, or one of them is close to ± 1 and the rest are close to zero. Since there are $n/2$ such coefficients, we get that for random i_1, i_2, j_1, j_2, the quantity

$$c_{i_1j_1} + c_{i_2j_2} - c_{i_1j_2} - c_{i_2j_1}$$

is $O(\epsilon/n)$-close to 0 or to ± 1 on average, and moreover, it is close to ± 1 only for a $O(1/n)$ fraction of these expressions.

How do we turn information about such expressions to information about individual coefficients? The idea is to fix a "good" choice $i_2 = I$, $j_2 = J$, namely one for which the two properties listed above still hold when we fix i_2, j_2 and vary only over i_1, j_1. We can thus "decode" $c_{i,j}$, for every $i \neq I$ and $j \neq J$, into a value $d_{i,j} \in \{0, \pm 1\}$ such that

$$c_{ij} \approx d_{ij} + c_{iJ} + c_{Ij} - c_{IJ},$$

and so

$$\begin{aligned} f &\approx \sum_{\substack{i \neq I \\ j \neq J}} (d_{ij} + c_{iJ} + c_{Ij} - c_{IJ}) x_{ij} \\ &= \sum_{\substack{i \neq I \\ j \neq J}} d_{ij} x_{ij} + \sum_i c_{iJ} + \sum_j c_{Ji} - n c_{IJ}. \end{aligned}$$

What happened here? Every permutation contributes exactly c_{iJ} to the sum $\sum_{ij} c_{iJ}x_{ij}$, since exactly one of $x_{i1},\dots,x_{in}$ is equal to 1. Similarly, every permutation contributes exactly c_{Ij} to the sum $\sum_{ij} c_{Ij}x_{ij}$, and exactly nc_{IJ} to the sum $\sum_{ij} c_{IJ}$.

A simple argument shows that the constant coefficient must be close to an integer, and so defining $d_{iJ} = d_{Ij} = 0$, we can approximate f as follows:

$$f \approx d + \sum_{ij} d_{ij}x_{ij},$$

where d is an integer, $d_{ij} \in \{0,\pm 1\}$, and only $O(n)$ many of the d_{ij}'s are non-zero. The latter property is a reflection of the original fraction of non-zero $c_{i_1j_1} + c_{i_2j_2} - c_{i_1j_2} - c_{i_2j_1}$ being $O(1/n)$ (there are $(n-1)^2$ many choices for i_1, j_1). Furthermore, careful calculation shows that the approximation error is $O(\epsilon)$.

At this point, it looks like we are almost done: up to the constant coefficient d, we have managed to approximate f by a difference of two sums of x_{ij}'s. However, our stakes are higher: we would like to approximate f either by a single sum, or by one minus a single sum.

Before we continue, let us pause to consider what happens in the case of the (n,k)-slice. Assuming $k \leq n/2$, there is a very similar reduction to the k-dimensional cube: choose a random permutation $a_1,\dots,a_n$, and consider all vectors x satisfying $x_{a_1} + x_{a_2} = x_{a_3} + x_{a_4} = \cdots = x_{a_{2k-1}} + x_{a_{2k}} = 1$ and $x_{a_{2k+1}} = \cdots = x_{a_n} = 0$; the coordinate y_1 determines whether $(x_{a_1}, x_{a_2}) = (1,0)$ or $(x_{a_1}, x_{a_2}) = (0,1)$, and so on.

Repeating essentially the same argument, we deduce that a degree 1 function on the slice which is close to Boolean must be close to a function of the form

$$d + \sum_i d_i x_i,$$

where d is an integer, $d_i \in \{0,\pm 1\}$, and $O(n/k)$ of the d_i's are non-zero. We can analyze the distribution of such a sum by "brute force", showing that since the result is close to Boolean, the linear combination must be either a sum of x_i's or one minus such a sum.

The case of the symmetric group is more complicated, since not all x_{ij}'s are related to one another in the same way: some pairs are disjoint, and some intersect. Therefore we will need a more subtle argument.

Let us start an easy case: suppose that there are only $n/2$ many non-zero d_{ij}'s (we can actually guarantee that this happens when f is somewhat unbalanced). In this case, a union bound shows that a random permutation "misses" all non-zero d_{ij}'s with probability at least $1/2$, and consequently $d \in \{0,1\}$.

Given that a random permutation passes through a particular non-zero d_{ij}, with probability $1/2 - o(1)$ it will miss all other non-zero d_{ij}, and so the function will evaluate to $d + d_{ij}$. Consequently, if $d = 0$ then most of the non-zero d_{ij}'s must be 1's, and if $d = 1$ then most of the non-zero d_{ij}'s must be -1's. Throwing out the errant d_{ij}'s, we obtain an approximation of f as either a sum of $O(n)$ many x_{ij}'s, or one minus such a sum. Since f is close to Boolean, the sum contains at most $O(\epsilon n^2)$ many non-disjoint pairs, completing the proof.

It remains to reduce the general case to this special case, in which a random permutation misses all non-zero d_{ij}'s with some constant probability. Can this actually fail to happen? Indeed it can, as the following example demonstrates:

$$1 - \sum_j x_{1j}.$$

This kind of function is not ruled out by the data we have on d and the d_{ij}. However, it is easy to fix such an example by finding a better representation of the same function, namely the zero function.

More generally, we can think of the d_{ij} as populating an $n \times n$ matrix. An obstruction to the hitting property we are aiming at is a line (row or column) with too many non-zero entries. Since each line contains at most three different values $(0,1,-1)$, there is some value which appears at most $n/3$ times. By subtracting this value from the entire line and compensating for it by updating the constant coefficient (using the fact that the x_{ij}'s in a line always sum to 1), we can obtain an alternative representation of the same function without this obstruction.

Performing this process a constant number of times, we obtain a different representation

$$d + \sum_{ij} d_{ij} x_{ij} = e + \sum_{ij} e_{ij} x_{ij},$$

in which e is an integer, the e_{ij} are bounded integers, at most $O(n)$ of them are non-zero, and every line in the matrix formed by the e_{ij}'s contains at most $(2/3)n + O(1)$ many non-zero entries. It turns out that this is enough for the hitting property to hold: a random permutation misses all non-zero e_{ij}'s with constant probability. We have thus reduced the general case to the special case which we have already showed how to handle.

To show that the hitting property holds, we sample a uniform permutation π in a specific way which demonstrates that all e_{ij} are missed with some constant probability. The sampling proceeds in two major stages. The first stage is comprised of several steps, each of which determines one value of π. Thus after t steps, we know t values of π, leaving unknown the value of $\pi(i)$ for all i in some set I of size $n-t$, as well as the value of $\pi^{-1}(j)$ for all j in some set J of size $n-t$. At odd steps, we choose $i \in I$ which maximizes $\#\{j \in J : d_{ij} \neq 0\}$ ("the number of non-zeroes on row i"), and choose $\pi(i) \in J$ at random. We act dually at even steps, choosing $\pi^{-1}(j) \in I$ at random for $j \in J$ which maximizes $\#\{i \in I : d_{ij} \neq 0\}$ ("the number of non-zeroes on column j"). This stage proceeds for $n/4$ steps, and then in the second stage we sample the remaining unknown part of the permutation all at once.

Why does this work? Let us start by analyzing the first stage. Suppose that the line chosen at step t has m_t many non-zeroes. Then the sample at that step misses all non-zeroes on the line with probability

$$1 - m_t/(n-t) \geq e^{-O(m_t/(n-t))} \geq e^{-O(m_t/n)},$$

the approximation holding since $m_t/(n-t) \lesssim (2/3)n/(3/4)n = 8/9$ and $n-t \geq (3/4)n$. Since $\sum_t m_t = O(n)$, all steps are successful with constant probability.

After the first stage, all remaining lines must be very sparse, in the sense that $\#\{j \in J : d_{ij} \neq 0\}$ is small for every $i \in I$ and $\#\{i \in I : d_{ij} \neq 0\}$ is small for every $j \in J$, where I, J are defined as in the first stage. Indeed, suppose that the heaviest remaining line is a row containing B many non-zeroes. This means that every odd step during the first stage, we chose a row with at least B many non-zeroes, and so there are at least $(n/8)B$ many non-zero d_{ij}'s. This shows that $B = O(1)$, that is, every line contains $O(1)$ many non-zeroes. This sparsity allows us to show that the number of non-zero d_{ij}'s hit at the second stage has a roughly Poisson distribution (we do this by estimating moments). Since the expected number of d_{ij}'s hit is $O(n)/n = O(1)$, with probability $e^{-O(1)} = \Omega(1)$, no non-zero d_{ij} is hit at the second stage, completing the proof of the hitting property.

4 Further results

The main result that we have been discussing so far interprets closeness in the L_2 sense: two functions f, g are close if $\mathbb{E}[(f-g)^2]$ is small. This is natural given the application that we started with, to intersecting families of permutations. But there are many other notions of closeness that one can consider. For example:

> What can we say about degree 1 functions f which satisfy $\Pr[f \notin \{0,1\}] \leq \epsilon$?
> What can we say about degree 1 functions f which satisfy $f(\pi) \in [-\epsilon, \epsilon] \cup [1-\epsilon, 1+\epsilon]$ for all $\pi \in S_n$?

In both cases, the answer should state that f is close to a function of a specific form *using the same notion of closeness*. The notion of closeness used in the first question is known as L_0, and the one used in the second question is L_∞.

It turns out that our main theorem holds as stated if we replace closeness in L_2 by closeness in L_0. That is, for some constant $K' > 0$,

- If f is a degree 1 function satisfying $\Pr[f \notin \{0,1\}] \leq \epsilon$ then $\Pr[f \neq g] = O(\epsilon)$ for some $g \in \mathcal{F}_\epsilon(K')$.

- All functions $g \in \mathcal{F}_\epsilon(K')$ satisfy $\Pr[g \notin \{0,1\}] = O(\epsilon)$.

We cannot just deduce this statement from the L_2 statement, since a function close to Boolean in L_0 need not be close to Boolean in L_2. For example, nx_{11} is $1/n$-close to Boolean in L_0, but $\Omega(n)$-far from Boolean in L_2. Therefore we need to reprove the result in this setting.

The proof is very similar to the proof in the L_2 case, the main difference being in the application of the FKN theorem. Recall that we used the FKN theorem to show that for random a, b, the coefficients in front of $y_1, \ldots, y_{n/2}$ in the expression

$$f|_{\mathcal{D}_{a,b}} = y_1(c_{a_1b_1} + c_{a_2b_2} - c_{a_1b_2} - c_{a_2b_1}) + \cdots + y_{n/2}(c_{a_{n-1}b_{n-1}} + c_{a_nb_n} - c_{a_{n-1}b_n} - c_{a_nb_{n-1}}) + C_{f,a,b}$$

are close to $0, \pm 1$. If many of the coefficients are "bad" (far from $0, \pm 1$) then $f|_{\mathcal{D}_{a,b}}$ would be far from Boolean; since on average $f|_{\mathcal{D}_{a,b}}$ is close to Boolean, we know that the average $c_{i_1j_1} + c_{i_2j_2} - c_{i_1j_2} - c_{i_2j_1}$ is good.

In contrast, in the L_0 setting, no matter how many of the coefficients are bad (different from $0, \pm 1$), the probability $\Pr[f|_{\mathcal{D}_{a,b}} \notin \{0,1\}]$ cannot exceed 1. However, it is still true that there cannot be many bad coefficients, since if there were too many, then it would be likely that we see at least two of them when we choose a, b at random (this requires an argument), which we can rule out since on average, $\Pr[f|_{\mathcal{D}_{a,b}} \notin \{0,1\}]$ is small.

What happens if we consider closeness in L_∞? Examples such as $x_{11} + x_{22}$, which are close to Boolean in L_2 and in L_0, are no longer close to Boolean in L_∞. The difference is that closeness in L_∞ is *worst-case*, rather than average-case as in the other two settings. Therefore one could hope for more structure in this case. Indeed, we can show that if ϵ is smaller than some constant, then we can guarantee that f is close to a *dictator* in L_∞. In other words, the following holds for some constant ϵ_0:

If f is a degree 1 function satisfying $f \in [-\epsilon_0, \epsilon_0] \cup [1-\epsilon_0, 1+\epsilon_0]$, then $\mathrm{round}(f, \{0,1\})$ is a dictator.

The proof follows the steps of the L_2 statement, but the details are simpler.

Before moving on to future research, one question which might have crossed your mind is: were the L_0 and L_∞ variants studied in other settings, such as that of the Boolean cube? It turns out that these questions are much less exciting for the Boolean cube, since the L_0 and L_∞ analogs of the classical FKN theorem are much easier to prove than the original L_2 version. The L_0 and L_∞ settings only become interesting in more complicated domains such as the symmetric group.

5 Future research

This work leaves many directions for future research. The two most obvious are:

- What happens when we replace degree 1 functions with degree d functions for larger d?
- Can we generalize this proof technique to other domains?

We deal with these questions in order.

5.1 Higher degrees

The results discussed so far are about degree 1 functions which are close to Boolean. What can we say about degree d functions which are close to Boolean?

This question comes up naturally in the study of *d-intersecting* families of permutations, in which any two permutations agree on the images of d points.[4] One example of such a family is

$$\{\pi \in S_n : \pi(i_1) = j_1, \ldots, \pi(i_d) = j_d\}.$$

[4] A related notion is that of *setwise-d-intersecting families*, in which any two permutations agree on the image of a d-set. These families were studied by Ellis [Ell12b], who proved analogs of all the results mentioned in the sequel in this setting.

In analogy to the case $d = 1$, we call families of this type *canonical d-intersecting families*. Frankl and Deza [FD77] conjectured that for every d, if n is large enough then the maximum size of a d-intersecting family is $(n-d)!$, matching the size of a canonical d-intersecting family. For $d = 2$ this holds for all $n \geq d$, as shown by Meagher and Razafimahatratra [MR20], and a similar behavior is expected when $d = 3$. When $d = 4$, however, canonical d-intersecting families do not have the maximum size for small n: the subset of S_7 consisting of all permutations in which at least 5 of $1, \ldots, 6$ are fixed points is 4-intersecting and contains $7 > (7-4)!$ permutations. Similar examples show that for all $d \geq 4$, we do need n to be large enough in order for the canonical d-intersecting families to be extremal.

Ellis, Friedgut and Pilpel [EFP11] proved the Frankl–Deza conjecture for all values of d, using the same spectral approach that we outlined above; in contrast to the case $d = 1$, we cannot just take $A = I - cB$, where B is the adjacency matrix of the "d-derangement graph", in which any two permutations are connected if they agree on fewer than d points. Instead, we need to weight the edges of the d-derangement graph appropriately.

As in the case $d = 1$, the proof of Ellis, Friedgut and Pilpel implies that for every d, if n is large enough, then the characteristic function of a d-intersecting family of maximum size $(n-d)!$ has degree d, that is, it can be expressed as a polynomial of degree d in the variables x_{ij}. This suggests that we can prove that the canonical d-intersecting families are the unique d-intersecting families of maximum size ("uniqueness") by understanding the structure of Boolean degree d functions. Similarly, the characteristic function of a d-intersecting family of almost maximum size $(1-\epsilon)(n-d)!$ is close to degree d, motivating the study of degree d functions which are close to Boolean.

Ellis, Friedgut and Pilpel claimed that Boolean degree d functions can always be written as a non-negative linear combination of degree d monomials $x_{i_1 j_1} \cdots x_{i_d j_d}$, these monomials are just the characteristic functions of canonical d-intersecting families. If true, this would directly imply the uniqueness claim. However, there is an error in their proof, and in fact the claim is false [Fil17]. Nevertheless, uniqueness follows from an argument of Ellis [Ell11], who actually showed a much stronger stability statement: every d-intersecting family of size at least $(1 - 1/e + o(1))(n-d)!$ must be a subset of some canonical d-intersecting family.

This suggests that before studying degree d functions which are close to Boolean, we study degree d functions which are *exactly* Boolean. For inspiration, let us first consider Boolean degree d functions on the Boolean cube. Nisan and Szegedy [NS94] showed that such functions depend on $O(d2^d)$ coordinates, and this was later tighthened to $O(2^d)$ by Chiarelli, Hatami and Saks [CHS20] and by Wellens [Wel19]. However, we cannot expect such a description in the case of the symmetric group. As an example, the function

$$x_{11} + \sum_{i=2}^{n} x_{1i} x_{i1}$$

is a Boolean degree 2 function, but it does not look like a junta (it depends on $\pi(i)$ for all i and on $\pi^{-1}(j)$ for all j).

A different kind of structure was suggested by Beals, Buhrman, Cleve, Mosca and de Wolf [BBC+01], who showed that a degree d function on the Boolean cube can be expressed as a decision tree of depth $d^{O(1)}$; see also Buhrman and de Wolf's survey [BdW02]. It turns out that this kind of structure does extend to the symmetric group: together with Dafni, Lifshitz, Lindzey, Saurabh, and Vinyals [DFL+21], we showed that a degree d function on the symmetric group can be represented as a decision tree of depth $d^{O(1)}$ whose nodes correspond to queries of the form "$\pi(i) =?$" and "$\pi^{-1}(j) =?$". For example, the function considered above can be expressed by a decision tree which first queries $\pi(1)$, and then either immediately returns 1 (if $\pi(1) = 1$) or queries $\pi(\pi(1))$ and returns 1 if $\pi(\pi(1)) = 1$. Using similar ideas, Dafni et al. gave a different proof of uniqueness.

Now let us get back to our original question: what can we say about degree d functions which are close to Boolean? In the case of the Boolean cube, Kindler and Safra [KS04] showed that if a degree d function is ϵ-close to Boolean, then it is $O(\epsilon)$-close to a Boolean degree d function.[5] This suggests the following conjecture:

[5]What they actually prove is that for some constants $C_d, \epsilon_d > 0$, if $\epsilon \leq \epsilon_d$ then f is $O(\epsilon)$-close to a Boolean function depending on C_d variables. This directly implies our statement, since there are only finitely many Boolean functions depending on C_d variables whose degree is larger than d.

For each d there exists a constant c_d such that for large enough n, if a degree d function f is ϵ-close to Boolean, then f is $O(\epsilon^{c_d})$-close to a Boolean degree d function.

The reason that we introduce the constant c_d is that already when $d = 1$, the conjecture only holds for $c_d = 1/2$; it is not clear what kind of structure emerges when we want an $O(\epsilon)$-approximation to f, though we were able to answer this question in the case of the slice, together with Dinur and Harsha [DFH19, DFH20].

A less bold version of the conjecture would state that f can be approximated by a decision tree of depth $d^{O(1)}$, or equivalently, a Boolean function of degree $d^{O(1)}$.

When f is sparse, say $\mathbb{E}[f] = c/n^d$, we were able to prove this conjecture, together with Ellis and Friedgut [EFF17], in a strong form: if a degree d function is $O(\epsilon c/n)$-close to Boolean, then it is $O_d(c\sqrt{\epsilon}/n + c^2/n^{3/2})$-close to a sum of m many degree-d monomial, where $m \approx c$. The conjecture is wide open in the non-sparse case.

5.2 Other domains

The basic scheme of the proof is quite generic: given a domain, all you have to do is find an equitable covering of the domain by high-dimensional Boolean cubes. Indeed, essentially the same proof works for the slice [Fil16], as well as for the Boolean cube with respect to an arbitrary product measure (by a slight extension of the proof in [Fil21]).

Another domain for which the scheme is expected to work is the “non-bipartite analog of the symmetric group”. We usually think of the symmetric group as the set of all permutations of $\{1, \ldots, n\}$, but we can also think of it as the set of all bijections between two abstract sets of size n, or equivalently, the set of all perfect matchings in $K_{n,n}$. The non-bipartite analog is the set of all perfect matchings in K_{2n}, also known as the *perfect matching scheme*. (Intersecting families of perfect matchings were studied by Meagher and Moura [MM05] and by Lindzey [Lin17, Lin20, Lin18].)

The situation is less clear in the case of *q-analogs*, some of which appear in the following table:

$q = 1$	$q > 1$
Hamming scheme $H(n, m) = \{1, \ldots, m\}^n$	Bilinear scheme $H_q(n, m)$
Johnson scheme $J(n, k)$ (aka the slice)	Grassmann scheme $J_q(n, k)$
Symmetric group S_n	Linear group $(P)GL_n(q), (P)SL_n(q)$

(In the second column, q is a prime power.)

The **bilinear scheme** is the set of all $n \times m$ matrices with entries in the finite field $\mathbb{F}_q$, and a degree 1 function is one of the form

$$f(A) = C + \sum_{x,y} \phi_{x,y}(x^T A y),$$

where $\phi_{x,y} \colon \mathbb{F}_q \to \mathbb{R}$ is arbitrary; when $q = 2$, we can take $\phi_{x,y} = c_{x,y}(-1)^{x^T A y}$, and when q is prime, we replace -1 with a primitive q'th root of unity.

The **Grassmann scheme** is the set of all k-dimensional subspaces of an n-dimensional vector space over $\mathbb{F}_q$. A function has degree 1 if it is of the form

$$f(V) = \sum_v c_v 1_{v \in V}.$$

The Grassmann scheme exhibits an important symmetry: $J_q(n, k)$ is isomorphic to $J_q(n, n-k)$, the correspondence mapping a subspace to its orthogonal complement. This means that degree 1 functions can equivalently be defined as ones of the form

$$f(V) = \sum_v c_v 1_{v \perp V}.$$

Why are these called q-analogs? If we look at certain parameters of these domains (for example, the dimension of the space of functions of degree at most d) and take the limit $q = 1$, then we get the corresponding parameters for the domains listed in the $q = 1$ columns (this is the famous "field of one element").

What do we know about these domains from our perspective? The first question to ask is what are all Boolean degree 1 functions, and this is already hard to answer: in the case of the Grassmann scheme, the answer should be

$$0, 1_{x \in V}, 1_{y \perp V}, 1_{x \in V} + 1_{y \perp V}$$

and their negations, where in the latter case, $x \not\perp y$; however, we only know this for $q = 2, 3, 4, 5$, since the existing proof (work together with Ihringer [FI19]) goes by induction, and the basis is only known for these values of q.

We do not know what all Boolean degree 1 functions look like for the bilinear scheme, nor do we know any FKN theorems in these settings. The main reason for this lack of knowledge is that it is not clear how to reduce these domains to the Boolean cube, and no other proof method is known, although global hypercontractivity [KLLM21] (originally due to Noam Lifshitz) is a promising avenue.

References

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoret. Comput. Sci.*, 288(1):21–43, 2002. Complexity and logic (Vienna, 1998).

[CHS20] John Chiarelli, Pooya Hatami, and Michael Saks. An asymptotically tight bound on the number of relevant variables in a bounded degree Boolean function. *Combinatorica*, 40(2):237–244, 2020.

[CK03] Peter J. Cameron and C. Y. Ku. Intersecting families of permutations. *European J. Combin.*, 24(7):881–890, 2003.

[DFH19] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Analyzing boolean functions on the biased hypercube via higher-dimensional agreement tests. In *ACM-SIAM Symposium on Discrete Algorithms (SODA19)*, 2019.

[DFH20] Irit Dinur, Yuval Filmus, and Prahladh Harsha. Low degree almost Boolean functions are sparse juntas, 2020+.

[DFL+21] Neta Dafni, Yuval Filmus, Noam Lifshitz, Nathan Lindzey, and Marc Vinyals. Complexity measures on symmetric group and beyond. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, 2021.

[EFF15a] David Ellis, Yuval Filmus, and Ehud Friedgut. A quasi-stability result for dictatorships in S_n. *Combinatorica*, 35(5):573–618, 2015.

[EFF15b] David Ellis, Yuval Filmus, and Ehud Friedgut. A stability result for balanced dictatorships in $\mathrm{S_n}$. *Random Structures Algorithms*, 46(3):494–530, 2015.

[EFF17] David Ellis, Yuval Filmus, and Ehud Friedgut. Low-degree Boolean functions on S_n, with an application to isoperimetry. *Forum Math. Sigma*, 5:Paper No. e23, 46, 2017.

[EFP11] David Ellis, Ehud Friedgut, and Haran Pilpel. Intersecting families of permutations. *J. Amer. Math. Soc.*, 24(3):649–682, 2011.

[EKR61] P. Erdős, Chao Ko, and R. Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford Ser. (2)*, 12:313–320, 1961.

[Ell11] David Ellis. Stability for t-intersecting families of permutations. *J. Combin. Theory Ser. A*, 118(1):208–227, 2011.

[Ell12a] David Ellis. A proof of the Cameron-Ku conjecture. *J. Lond. Math. Soc. (2)*, 85(1):165–190, 2012.

[Ell12b] David Ellis. Setwise intersecting families of permutations. *J. Combin. Theory Ser. A*, 119(4):825–849, 2012.

[FD77] Péter Frankl and Mikhail Deza. On the maximum number of permutations with given maximal or minimal distance. *J. Combinatorial Theory Ser. A*, 22(3):352–360, 1977.

[FI19] Yuval Filmus and Ferdinand Ihringer. Boolean degree 1 functions on some classical association schemes. *J. Combin. Theory Ser. A*, 162:241–270, 2019.

[Fil16] Yuval Filmus. Friedgut-Kalai-Naor theorem for slices of the Boolean cube. *Chic. J. Theoret. Comput. Sci.*, pages Art. 14, 17, 2016.

[Fil17] Yuval Filmus. A comment on intersecting families of permutations. *arXiv*, 1706.10146, 2017.

[Fil21] Yuval Filmus. Lecture notes on boolean function analysis. `https://yuvalfilmus.cs.technion.ac.il/Courses/BooleanFunctionAnalysis/2021/LectureNotes.pdf`, 2021.

[FKN02] Ehud Friedgut, Gil Kalai, and Assaf Naor. Boolean functions whose Fourier transform is concentrated on the first two levels. *Adv. in Appl. Math.*, 29(3):427–437, 2002.

[GM09] Chris Godsil and Karen Meagher. A new proof of the Erdős-Ko-Rado theorem for intersecting families of permutations. *European J. Combin.*, 30(2):404–414, 2009.

[Kat72] G. O. H. Katona. A simple proof of the Erdős-Chao Ko-Rado theorem. *J. Combinatorial Theory Ser. B*, 13:183–184, 1972.

[KLLM21] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications, 2021.

[KS04] Guy Kindler and Shmuel Safra. Noise-resistant Boolean functions are juntas, 2004. Unpublished manuscript.

[KW07] Cheng Yeaw Ku and Tony W. H. Wong. Intersecting families in the alternating group and direct product of symmetric groups. *Electron. J. Combin.*, 14(1):Research Paper 25, 15, 2007.

[Lin17] Nathan Lindzey. Erdős-Ko-Rado for perfect matchings. *European J. Combin.*, 65:130–142, 2017.

[Lin18] Nathan Lindzey. *Matchings and Representation Theory.* PhD thesis, University of Waterloo, 2018.

[Lin20] Nathan Lindzey. Stability for 1-intersecting families of perfect matchings. *European J. Combin.*, 86:103091, 12, 2020.

[LM04] Benoit Larose and Claudia Malvenuto. Stable sets of maximal size in Kneser-type graphs. *European J. Combin.*, 25(5):657–673, 2004.

[MM05] Karen Meagher and Lucia Moura. Erdős-Ko-Rado theorems for uniform set-partition systems. *Electron. J. Combin.*, 12:Research Paper 40, 12, 2005.

[MR20] Karen Meagher and A. S. Razafimahatratra. 2-intersecting permutations. *arXiv*, 2005.00139, 2020.

[NS94] Noam Nisan and Márió Szegedy. On the degree of Boolean functions as real polynomials. *Comput. Complexity*, 4(4):301–313, 1994. Special issue on circuit complexity (Barbados, 1992).

[Ren07] Paul Renteln. On the spectrum of the derangement graph. *Electron. J. Combin.*, 14(1):Research Paper 82, 17, 2007.

[Wel19] Jake Wellens. A tighter bound on the number of relevant variables in a bounded degree Boolean function. *ArXiv*, 1903.08214, 2019.