Expansion in the Johnson Graph

Clay Thomas
claytont@cs.princeton.edu

Uma Girish
ugirish@cs.princeton.edu

May 2019

Abstract

The (generalized) Johnson graph is given by slices of the hypercube, and is important for understanding probabilistically checkable proof systems and hardness of approximation. Characterizing the expansion of the Johnson Graph recently served as an important conceptual stepping stone to proving the 2-to-1 games conjecture. Here, we summarize the proof technique used, with a focus on simplicity and clarity of presentation.

The 2-to-1 games conjecture (now theorem) is an immensely important fact in computer science, especially in hardness of approximation. This note mostly ignores this broader context, but here we point out the relevant papers. In [DKK+16] (backed up by a substantial technical report in [DKK+18]), a plan-of-attack on the 2-to-1 games conjecture was formulated, and a combinatorial hypothesis was put forward which would imply the 2-to-1 conjecture. This combinatorial hypothesis seemed related to the small set expansion of the so-called Grassmann graph, and [BKS18] made a simple observation which showed that answering the combinatorial hypothesis was actually equivalent to understanding the small set expansion of the Grassmann graph (the "degree 2 shortcode graph" is another equivalent viewpoint). Classifying the expansion in the Grassmann graph seemed very hard, so first a conceptually related graph called the Johnson graph was studied in [KMMS18]. Finally, [KMS18] proved the necessary result on expansion of the Grassmann graph, completing the proof of the 2-to-1 conjecture.

This note summarizes [KMMS18], giving results which are technically disjoint from those needed to prove the 2-to-1 games conjecture, but which motivated and clarified how the proof of the 2-to-1 games conjecture should go. We also rely heavily on results from [DKK+18], adapted to statements about the Johnson graph instead of the Grassmann.

1 Introduction

Definition 1.1. *For $t < \ell < k$, the (generalized) Johnson graph $J(k, \ell, t)$ has nodes $\binom{[k]}{\ell}$ (i.e. subsets of $\{1, \dots, k\}$ of size exactly ℓ) and edges between nodes A and B if and only if $|A \cap B| = t$.*

Typically, we think of the case when $k \gg \ell$ and $t = \ell/2$.

The central question we are concerned with is this: What are the (small-density) vertex sets of $J(k, \ell, t)$ with small edge-expansion? There's a fairly immediate class of easy examples: for any $R \subseteq [k]$ with $|R| = r = O(1)$, define

$$S_R := \left\{ A \in \binom{[k]}{\ell} \middle| R \subseteq A \right\}$$

Observe that the density of this set $S_R/\binom{k}{\ell}$ is approximately $(\ell/k)^r$, a quantity that goes to zero as $k \to \infty$. However, the expansion of S_R is approximately $1 - e^{-(t/\ell)r}$, which is bounded away from 1. Thus S_R is a small, poorly-expanding set of vertices in $J(k, \ell, t)$. It will turn out that this is in some sense the *only* example of a small non-expanding set of vertices.

We measure correlation with S_R via the following definition:

Definition 1.2. *For any $f : \binom{[k]}{i} \to \mathbb{R}$ and $J \subseteq [k]$, $|J| < i$, define:*

$$\mu(f) = \mathop{\mathbb{E}}_{|I|=i}[f(I)] \qquad \mu_J(f) = \mathop{\mathbb{E}}_{I \supseteq J, |I|=i}[f(I)]$$

If F is the indicator function of a set of vertices S, then $\mu(F)$ is called the density *of S.*

A set of vertices is called pseudorandom when it does not correlate highly with any S_R, i.e. if it does not become significantly more dense when we restrict our attention to supersets of R:

Definition 1.3. *For any $r \leq \ell, \epsilon > 0$, a set of vertices S of $J(k, \ell, t)$ is called (r, ϵ)-pseudorandom when for every $R \subseteq [k]$ with $|R| \leq r$, we have*

$$|\mu_R(S) - \mu(S)| \leq \epsilon$$

The main theorem states that pseudorandom sets are good (edge) expanders in the Johnson graph. Conversely to this theorem, we see that the only way to be a bad expander in $J(k, \ell, t)$ is to non-pseudorandom, i.e. to be similar to some set S_R.

Theorem 1.4 (Main Theorem (Qualitative Version))**.** *For all $\eta, \delta \in (0, 1/2)$ and $\ell, t \in \mathbb{N}$, there exists $r \in \mathbb{N}$, $\epsilon > 0$, and $k \gg \ell$ such that if S is a set of vertices in $J(k, \ell, t)$ which has $\mu(S) = \delta$ and is (r, ϵ)-pseudorandom, then $\Phi(S) \geq 1 - \eta$.*

Theorem 1.5 (Main Theorem (Quantitative Version))**.** *Fix $\ell, t \in \mathbb{N}$ and $\delta \in (0, 1/2)$. If $k > \exp(\mathrm{poly}(\ell)) \,\mathrm{poly}(1/\delta)$, then if S with $\mu(S) = \delta$ is (r, ϵ) pseudorandom in $J(k, \ell, t)$, then*

$$\Phi(S) \geq 1 - (t/\ell)^{r+1} - \exp(r)\epsilon^{1/4} - \frac{\mathrm{poly}(2^{\mathrm{poly}(\ell)}, 1/\delta)}{k^{\Omega(1)}}$$

Note that the qualitative version immediately implies the quantitative version. The three terms subtracted from 1 above are controlled as follows: the first corresponds to r, the second correspond to ϵ, and the third term is $o(1)$ as k grows.

2 Rephrasing in terms of eigenspaces

In this section we outline the proof of theorem 1.5. Define $\langle F, G \rangle = \mathbb{E}_A [F[A]G[A]]$. If the graph $J(k, \ell, t)$ is clear from the context, we will denote the normalized adjacency matrix of $J(k, \ell, t)$ by J, and we will let $A \sim B$ denote that A, B are adjacent vertices of $J(k, \ell, t)$. We conceptualize J as the linear operator on functions $\binom{[k]}{\ell} \to \mathbb{R}$ such that $(JF)[A] = \mathbb{E}_{B \sim A} [F[B]]$. We use the notation $\mathrm{negl}(k)$ for any quantity that goes to zero as k goes to infinity and other parameters are held constant (usually, it will converge at least as fast as $1/k^{\Omega(1)}$).

Let S be an (r, ϵ) pseudorandom set in $J(k, \ell, t)$. As standard first step to proving bounds on expansion is the following observation: Let J be the adjacent matrix of $J(k, \ell, t)$, and let F be the indicator function of some set S of vertices. Then

$$\delta(1 - \Phi(S)) = \mathop{\mathbb{P}}_A [A \in S] \mathop{\mathbb{P}}_{A \in S, B \sim A} [B \in S] = \mathop{\mathbb{P}}_{A \sim B} [A, B \in S] = \langle F, JF \rangle$$

To show that $\Phi(S)$ is large, we'll want to show that $\langle F, JF \rangle$ is small. In the next section, we'll describe the (approximate) eigendecomposition $J_{=0} \oplus J_{=1} \oplus \ldots \oplus J_{=\ell}$ of J which has (approximate) eigenvalues λ_i. Let $F = F_{=0} + F_{=1} + \ldots + F_{=\ell}$ be the projection of F onto these spaces. Then

$$\langle F, JF \rangle = \sum_{i=1}^{\ell} \lambda_i \langle F_{=i}, F_{=i} \rangle \qquad (*)$$

Now, we won't actually be able to prove that $J_{=i}$ give the eigenspaces of $J(k, \ell, t)$. However, we'll be able to show that the above equation holds up to an additive $o(1)$ factor (theorem 3.5). Moreover, we'll see that λ_i decays exponentially, approximately like $\lambda_i \leq (t/\ell)^i$. Thus, for $i > r$, we'll simply rely on the eigenvalues being small to bound (*) (plus an application of Parseval to get $\sum_i \langle F_{=i}, F_{=i} \rangle \leq \delta$). So far we have

$$\langle F, JF \rangle \leq \sum_{i=1}^{r} (t/\ell)^r \langle F_{=i}, F_{=i} \rangle + (t/\ell)^{r+1}\delta + \mathrm{negl}(k)$$

For $i \leq r$, pseudorandomness will be needed. Our goal will be to prove that (r, ϵ) pseudorandomness implies an upper bound on the weight of F on the ith level (for $i \leq r$) of the form

$$\langle F_{=i}, F_{=i} \rangle \leq \delta \epsilon^{\Omega(1)} f(i) + \mathrm{negl}(k)$$

The δ will cancel out and we'll be left with an ϵ term to force $1 - \Phi(S)$ to be small. The exact bound we will get is

Theorem 2.1 (Theorem 2.15 from [KMMS18]). *For any fixed ℓ, t, r, ϵ, there exists k large enough and δ small enough such that for any (r, ϵ)-pseudorandom S with $\mu(S) = \delta$ in $J(k, \ell, t)$, we have*

$$\langle F_{=i}, F_{=i} \rangle \leq \exp(i) \delta \epsilon^{1/4} + \mathrm{negl}(k)$$

for $i = 0, 1, \ldots, r$.

This is the main technical results, which we discuss in section 5. We are left with:

$$\begin{aligned} 1 - \Phi(S) = \frac{1}{\delta} \langle F, JF \rangle &\leq \sum_{i=1}^{r} \epsilon^{1/4} \exp(i) + (t/\ell)^{r+1} + \mathrm{negl}(k) \\ &\leq \epsilon^{1/4} \exp(r) + (t/\ell)^{r+1} + \mathrm{negl}(k) \end{aligned}$$

The next section describes the eigenspaces $J_{=i}$. Unfortunately, we also won't be able to work with the exact projections $F_{=i}$ onto the eigenspaces. Instead, we'll consider approximate projections $F_{\approx i}$ which you can prove are within a $\mathrm{negl}(k)$ factor of $F_{=i}$. The section after next describes these approximate decomposition, which are one of the main tools to proving Theorem 2.15 from [KMMS18].

3 The eigendecomposition

Definition 3.1. *Define the "space spanned by the first i levels" $J_{\leq i}$ as follows: for $F : \binom{[k]}{\ell} \to \mathbb{R}$, let $F \in J_{\leq i}$ if and only if there exists some $f : \binom{[k]}{i} \to \mathbb{R}$ such that, for all $A \in \binom{[k]}{\ell}$, we have*

$$F[A] = \sum_{I \subseteq A, |I| = i} f(I)$$

Moreover, define the "space of level i functions" $J_{=i} = J_{\leq i} \cap J_{\leq i-1}^{\perp}$.

Note the following: $J_{\leq i}$ is a linear subspace, $J_{\leq i} \subseteq J_{\leq i+1}$ (you can prove this by an averaging argument), and $J_{\leq \ell}$ consists of all functions on $J(k, \ell, t)$. Thus, we see that the space of real valued functions on $J(k, \ell, t)$ decomposes as $J_{=0} \oplus J_{=1} \oplus \ldots \oplus J_{=\ell}$.

Lets look at the first few levels as examples:

- $J_{=0}$ is the set of all constant functions
- $J_{=1}$ can be thought of as all additive functions on $[k]$ with average zero. That is, $F[A] = \sum_{a \in A} f(a)$ where we also have $\mathbb{E}_{i \in [k]}[f(a)] = 0$.
- $J_{=2}$ is a bit less natural, but is defined by a set of weights on all pairs over $[k]$, and has zero correlation with with any additive function.

Here's an important alternative characterization of $J_{=i}$: it's those functions in $J_{\leq i}$ whose f function is "orthogonal to the μ_R operator" for every $|R| < i$.

Lemma 3.2. *$F \in J_{=i}$ if and only if $F \in J_{\leq i}$ and $\mu_R(F) = 0$ for every $R \subseteq [k]$ with $|R| < i$.*

Proof. Consider an $F' \in J_{\leq j}$ for some $j < i$, say $F'[A] = \sum_{J \subseteq A, |J|=j} f'(J)$. We have

$$\begin{aligned}
\langle F, F' \rangle &= \mathop{\mathbb{E}}_{A}\left[F[A] \sum_{J \subseteq A, |J|=j} f'(J) \right] \\
&= \frac{1}{\binom{k}{\ell}} \sum_{A, |A|=\ell} \sum_{J \subseteq A, |J|=j} F[A] f'(J) \\
&= \frac{1}{\binom{k}{\ell}} \sum_{|J|=j} \sum_{A \supseteq J, |A|=\ell} F[A] f'(J) \\
&= \frac{1}{\binom{k}{\ell}} \sum_{|J|=j} f'(J) \binom{k}{\ell - j} \mu_J(F)
\end{aligned}$$

Now, for different F' in different $J_{\leq j}$, we can set any value we want for each $f'(J)$. Thus, $\langle F, F' \rangle$ will be zero for all such F' if and only if $\mu_J(F) = 0$ for all J with $|J| < i$, as desired. □

Moreover, the above lemma is true if we replace $F \in J_{=i}$ with any f such that $F[A] = \sum_{I \subseteq A, |I|=i} f(I)$. This generalizes the "additive with average zero" characterization that we provided for $J_{=1}$. (This is actually the more important property – the above lemma is provided as a much more simple special case).

Lemma 3.3 (Lemma 2.6 in [KMMS18], analogue of Lemma 2.19 from [DKK⁺18])**.** *Let $k > \exp(\ell)$, $\ell \geq 2$, and suppose $F \in J_{\leq i}$ is given by $F[A] = \sum_{I \subseteq A, |I|=i} f(I)$. Then $F \in J_{=i}$ if and only if for all $R \subseteq [k]$, $|R| < i$, we have $\mu_R(f) = 0$.*

The proof is actually only given for the Grassman Graph in [DKK⁺18].

It turns out that $J_{=i}$ are the eigenspaces of J, the adjacency matrix of $J(k, \ell, t)$. However, this is (apparently) difficult to prove, and it suffices to show that these are approximate eigenspaces up to an additive $o(1)$ error.

Definition 3.4. *For $i = 0, \ldots, t$, define $\lambda_i = \lambda_i(k, \ell, t) = \binom{t}{i} / \binom{\ell}{i}$. For $i > t$, define $\lambda_i = \lambda_i(k, \ell, t) = 0$.*

Theorem 3.5 (Theorem 2.7 in [KMMS18])**.** *For any $F \in J_{=i}$ such that $F[A] = \sum_{I \subseteq A, |I|=i} f(I)$, we have*

$$\|JF - \lambda_i F\|_\infty \leq \frac{2^{2\ell}}{k} \|f\|_\infty$$

Proof. The proof procedes like this: $JF[A]$ is the average over neighbors B of A, and $F[B]$ is (essentially) the average of f over subsets I of B with $|I| = i$. You can look separately at terms with different sizes of $I \cap A$, and see that in the end there is a big contribution only when $I \subseteq A$. When $I \subseteq A$ on the other hand, the resultant term is exactly $\lambda_i F[A]$. Details can be found in [KMMS18]. □

Note that

$$\lambda_i = \frac{t(t-1)\ldots(t-i+1)}{\ell(\ell-1)\ldots(\ell-i-1)} \leq (t/\ell)^i$$

as needed.

4 Approximate Decompositions

The projection of $F : \binom{[k]}{\ell} \to \mathbb{R}$ onto the ith eigenspace $J_{=i}$ is denoted by $F_{=i}$. The corresponding function $f_{=i} : \binom{[k]}{i} \to \mathbb{R}$ is such that $F_{=i}[A] = \sum_{I \subseteq A, |I|=i} f_{=i}(I)$. (This must exist by definition. It's not clear that $f_{=i}$ should be unique, but any such function will work for all the claims we make.) We call the formula $F = F_0 + F_1 + \ldots + F_\ell$ the decomposition of F.

Unfortunately, there is no formula for $F_{=i}$ in general. However, the following formulas will turn out to give "approximate decompositions" which will be close enough for what we need. Define $F_{\approx 0}[A] = \mu(F) = f_{\approx 0}$, then inductively define

$$f_{\approx i}(I) := \mu_I(F) - \sum_{J \subsetneq I} f_{\approx |J|}(J)$$
$$F_{\approx i}[A] := \sum_{I \subseteq A, |I|=i} f_{\approx i}(I)$$

Let's discuss some intuition for why this should work. As for any graph, the projection onto the space of constant functions is given by $F_{=0}[A] = \mu(F)$ for every A. Unfortunately, this formula already fails to be exact when $i = 1$. Let's take a look at this case in detail:

Claim 4.1 (Adapted from Lemma 2.22 of [DKK+18])**.** *For approximate decomposition $i = 1$, we have $f_{\approx i}(x) = f_{=i}(x) + \text{negl}(k)$.*

More precicely, let $f_1(x) = c(\mu_{\{x\}}(F) - \mu(F))$ for scaling factor $c = (k-\ell)/(k-1) \approx 1$. Then $f_1 = f_{=1}$.

Proof. Let $G[A] = F[A] - \sum_{x \in A} f_1(x) - \mu(F)$. By lemma 3.2, it suffices to show that $\mu_{\{x\}}(G) = 0$ for every $x \in [k]$ (as then G is orthogonal to $J_{\leq i}$). Evaluating gets us:

$$\begin{aligned}
\mu_{\{x\}}(G) &= \mu_{\{x\}}(F) - \mu(F) - \mathop{\mathbb{E}}_{A \ni x}\left[\sum_{y \in A} f_1(y)\right] \\
&= \mu_{\{x\}}(F) - \mu(F) - f_1(x) - \mathop{\mathbb{E}}_{A \ni x}\left[\sum_{y \in A, y \neq x} f_1(y)\right] \\
&= (1-c)[\mu_{\{x\}}(F) - \mu(F)] - (\ell - 1) \mathop{\mathbb{E}}_{\substack{A \ni x \\ y \in A, y \neq x}} [f_1(y)] \\
&= (1-c) f_1(x) - (\ell - 1) \mathop{\mathbb{E}}_{y \in [k] \setminus \{x\}} [f_1(y)]
\end{aligned}$$

Now, to get intuition for our less precise claim, it suffices to see whey the last term above is negligible. Indeed, the expectation of $f_1(y)$, uniform over all y, is zero by construction. The above distribution over y above is *almost* uniform over all y, and the statistical difference between y uniform over $[k]$ and y uniform over $[k] \setminus x$ goes to zero as $k \to \infty$.

For the more precise claim we are trying to show, we evaluate the last expectation above via:

$$\begin{aligned}
\mathop{\mathbb{E}}_{y \in [k]} [f_1(y)] &= c\left[\mathop{\mathbb{E}}_{y \in [k], A \ni y} [F[A]] - \mu(F)\right] = 0 \\
\mathop{\mathbb{E}}_{y \in [k]} [f_1(y)] &= \frac{k-1}{k} \mathop{\mathbb{E}}_{y \in [k] \setminus x} [f_1(y)] + \frac{1}{k} f_1(x) \\
\implies \mathop{\mathbb{E}}_{y \in [k] \setminus x} [f_1(y)] &= -\frac{f_1(x)}{k-1}
\end{aligned}$$

(if we had any weak bound on $f_1(x)$ this would also make the claim that this term is negligible in k precise). Putting this together gets us exactly what we want:

$$\mu_{\{x\}}(G) = \left(1 - c - \frac{\ell - 1}{k - 1}\right) f_1(x) = 0$$

□

We can provide one more piece of intuition for how these approximate decompositions work, in term of an analogy to the boolean hypercube. The hypercube has an extremely nice eigenbasis, namely the familiar Fourier functions $\{\chi_S\}_{S \subseteq [n]}$. The analogous operator to $\mu_R(f)$ would be $\mathbb{E}_{y \geq r}[f(y)]$, which uses the natural identification of the boolean hypercube with subsets of $[n]$. If you're given the "low-level" Fourier coefficients $\widehat{f}(J)$ for $|J| < i$, then you can use this expectation operator to calculate the coefficient for some I with $|I| = i$ as follows:

$$\begin{aligned}
&\mathop{\mathbb{E}}_{y \geq \mathbb{1}[I]}[\chi_J(y)] = \begin{cases} 1 & J \subseteq I \\ 0 & \text{otherwise} \end{cases} \\
\implies &\mathop{\mathbb{E}}_{y \geq \mathbb{1}[I]}[f(y)] = \sum_{J \subseteq [n]} \widehat{f}(J) \mathop{\mathbb{E}}_{y \geq \mathbb{1}[I]}[\chi_J(y)] = \sum_{J \subseteq I} \widehat{f}(J) \\
\implies &\widehat{f}(I) = \mathop{\mathbb{E}}_{y \geq \mathbb{1}[I]}[f(y)] - \sum_{J \subsetneq I} \widehat{f}(J)
\end{aligned}$$

When you move to the Johnson graph, the μ_R operator still "zeros out the higher levels" (i.e. lemma 3.2) as desired. However, you no longer get exact equality on the lower levels, making $f_{approxi}$ imprecise.

In the next section we will require the following claims.

Fact 2.12 from [KMMS18], the analogue of Lemma A.2 from [DKK$^+$18]:

$$|f_{\approx i}(I)| \leq \|F\|_\infty 2^{i^2}$$

This result simply states that $f_{\approx i}$ cannot blow up too much as a function of the values of F. The proof is by induction.

Fact 2.11 from [KMMS18], the analogue of Claim A.1 from [DKK$^+$18]: For all $|J| < i$ we have:

$$|\mu_J(f_{\approx i})| \leq \|F\|_\infty \frac{2^{10i^2}}{k} = \|F\|_\infty \operatorname{negl}(k)$$

This result states that approximate decompositions satisfy an approximate version of lemma 3.3.

5 Analysis Of Higher Moments

Recall the main theorem we wish to prove, that a pseudorandom set cannot put too much mass on the lower levels. Let η be the mass $\langle F_{=i}, F_{=i} \rangle$ of $F_{=i}$ on the i-th level.

Theorem (Theorem 2.1, aka Theorem 2.15 from [KMMS18])**.** *For any fixed ℓ, t, r, ϵ, there exists k large enough and δ small enough such that for any (r, ϵ)-pseudorandom S with $\mu(S) = \delta$ in $J(k, \ell, t)$, we have*

$$\langle F_{=i}, F_{=i} \rangle \leq \exp(i)\delta\epsilon^{1/4} + \operatorname{negl}(k)$$

for $i = 0, 1, \ldots, r$.

The proof of this is to analyze the fourth moment of $F_{\approx i}$.

Lemma 5.1 (Lower Bound)**.**

$$\mathbb{E}_A[F_{\approx i}^4(A)] \geq \frac{\eta^5}{\delta^4}$$

Proof Idea. Since F puts a lot of weight on the i-th level, we have $F_i \approx_{\ell_2} F$. Since $F_{\approx i} \approx_{\ell_2} F_{=i}$, we have $F_{\approx i} \approx_{\ell_2} F$. Since F is 1 with probability at least δ, this will tell us that $F_{\approx i}$ is close to 1 with at least $\delta/2$ probability for appropriate choice of parameters. This means that the fourth moment is large. Since this is simple to execute, we don't present the exact details.

Lemma 5.2 (Upper Bound)**.**

$$\mathbb{E}_A[F_{\approx i}^4[A]] \leq \exp(i)\eta\epsilon + \frac{\exp(i)}{\sqrt{k}}$$

It is not too hard to see that putting these two bounds together gives us the required bound. We now upper bound the third moment $\mathbb{E}_A F^3_{\approx i}(A)$, instead of the fourth bound as in the previous theorem. We remark that the analysis is similar, but more involved and not any more illuminating.

Proof. Recall that $F_{\approx i}(A) = \sum_{I\subseteq} f_{\approx i}(A)$ because of the way we defined the approximate projections. We have to analyze:

$$\mathop{\mathbb{E}}_A F^3_{\approx i}[A] = \mathop{\mathbb{E}}_A \sum_{\substack{I_1,I_2,I_3\subseteq A\\ |I_1|=|I_2|=|I_3|=i}} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)$$

The expression on the right-hand-side involves picking three uniformly random sets I_1, I_2, I_3 and then analyzing the expected value of $f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)$. To do this, it will help to divide this expectation into terms where each term corresponds to a fixed pattern of intersection of I_1, I_2, I_3. More precisely, an intersection pattern σ is a 4-tuple of numbers. A pattern corresponding to three sets A, B, C is the tuple $(|A\cap B|, |A\cap C|, |B\cap C|, |A\cap B\cap C|)$. For every possible intersection pattern, we will bound the expectation. We now describe the details of this proof.

$$\mathop{\mathbb{E}}_A F^3_{\approx i}[A]$$

...By definition, $F_{\approx i}(A) = \sum_{I\subseteq A, |I|=i} f_{\approx i}(I)$

$$= \mathop{\mathbb{E}}_A \sum_{\substack{I_1,I_2,I_3\subseteq A\\ |I_1|=|I_2|=|I_3|=i}} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)$$

...Instead of summing over uniformly random $I_1, I_2, I_3 \subseteq A$ of size i,
we sum over a union size $d \in [i, 3i]$
and over uniformly random $I_1, I_2, I_3 \subseteq A$ whose union is of size d.

$$= \sum_{d=i}^{3i} \mathbb{E}_A \sum_{\substack{|I_1\cup I_2\cup I_3|=d\\ |I_1|=|I_2|=|I_3|=i\\ I_1,I_2,I_3\subseteq A}} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)$$

...We loosely upper bound the number of ways to choose a such I_1, I_2, I_3
by choosing their union in $\binom{\ell}{d}$ ways,
and then choosing I_1, I_2, I_3 in $\binom{d}{i}^3$ ways

$$\leq \sum_{d=i}^{3i} \binom{\ell}{d}\binom{d}{i}^3 \left| \mathop{\mathbb{E}}_{\substack{A,I_1,I_2,I_3\\ I_1,I_2,I_3\subseteq A\\ |I_1\cup I_2\cup I_3|=d\\ |I_1|=|I_2|=|I_3|=i}} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)\right|$$

$$= \sum_{d=i}^{3i} \binom{\ell}{d}\binom{d}{i}^3 \left| \mathop{\mathbb{E}}_{\substack{|I_1\cup I_2\cup I_3|=d\\ |I_1|=|I_2|=|I_3|=i}} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)\right|$$

...We sample an intersection pattern σ from the distribution $\gamma(d)$ of patterns
induced by uniformly random triples of sets I_1, I_2, I_3 whose union is of size d,
and then pick a uniformly random I_1, I_2, I_3 satisfying this pattern.

$$\leq \sum_{d=i}^{3i} \binom{\ell}{d}\binom{d}{i}^3 \sum_{\sigma\sim\gamma(d)} \mathbb{P}[\sigma] \left| \mathop{\mathbb{E}}_{(I_1,I_2,I_3)\in\sigma} f_{\approx i}(I_1)f_{\approx i}(I_2)f_{\approx i}(I_3)\right|$$

We first estimate the coefficient in this expression.

$$\begin{aligned}
&\binom{\ell}{d}\binom{d}{i}^3 \\
&\leq \frac{(e\ell)^d}{d^d}\frac{(ed)^{3i}}{i^{3i}} \\
&\leq \ell^d e^{d+3i}\frac{d^{3i-d}}{i^{3i}} \\
&\leq \ell^d e^{d+3i}\frac{(3i)^{3i-d}}{i^{3i}} \\
&\leq \ell^d exp(i)
\end{aligned}$$

Lemma 5.3. *For all intersection patterns $\sigma \in \gamma(d)$, the quantity in the expectation is bounded, that is:*

$$\left| \mathop{\mathbb{E}}_{(I_1,I_2,I_3)\in\sigma} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \right| \leq \frac{\eta\epsilon \exp(i)}{\ell^d} + neg\ell(k)$$

Suppose we had this lemma, the previous calculation shows us that $\mathbb{E}_A F^3_{\approx i}[A]$ would be bounded by

$$\Big(\frac{\eta\epsilon \exp(i)}{\ell^d} + neg\ell(k)\Big) \times \ell^d exp(i) \lesssim \eta\epsilon \exp(i) + neg\ell(k)$$

This gives us the desired upper bound on the fourth moment. It now suffices to prove Lemma 4.3 □

5.1 Proof Of Lemma 5.3

We now fix d and a pattern $\sigma \sim \gamma(d)$ and estimate

$$\left| \mathop{\mathbb{E}}_{I_1,I_2,I_3\in\sigma} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \right|$$

As we saw before, this distribution on I_1, I_2, I_3 can alternatively be thought of as first sampling $D = \{x_1, \ldots, x_d\}$ and then $I_1, I_2, I_3 \subseteq D$ satisfying σ. We wish to estimate:

$$\left| \mathop{\mathbb{E}}_{|D|=d} \mathop{\mathbb{E}}_{\substack{I_1\cup I_2\cup I_3=D \\ (I_1,I_2,I_3)\in\sigma}} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \right|$$

Case 1: The intersection pattern σ is such that there is some element x_1 that appears in exactly one of the sets, say I_1.

Let $J = I_1 \cap (I_2 \cup I_3) \subseteq \{x_2, \ldots, x_d\}$ be all those elements in I_1 that are not exclusive to I_1. Consider:

$$\mathop{\mathbb{E}}_{|D|=d,} \mathop{\mathbb{E}}_{\substack{I_1\cup I_2\cup I_3=D \\ (I_1,I_2,I_3)\in\sigma}} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) = \mathop{\mathbb{E}}_{J} f_{\approx i}(I_2) f_{\approx i}(I_3) \Big[\mathop{\mathbb{E}}_{I_1 - J} f_{\approx i}(I_1) \Big]$$

The distribution of I_1 in the right hand side is basically a uniform random set of size i conditioned on containing J, but not intersecting with $I_2 \cup I_3 - J$. That is,

$$\mathop{\mathbb{E}}_{I_1 - J} f_{\approx i}(I_1) = \mathop{\mathbb{E}}_{\substack{I_1 : J \subseteq I_1 \\ I_1 \subseteq \{x_2,\ldots,x_d\}^C \cup J}} f_{\approx i}(I_1)$$

We would like to replace $\mathop{\mathbb{E}}_{\substack{I_1 : J \subseteq I_1 \\ I_1 \subseteq \{x_2,\ldots,x_d\}^C \cup J}} f_{\approx i}(I_1)$ by $\mathop{\mathbb{E}}_{J\subseteq I_1} f_{\approx i}(I_1)$. To analyze the error incurred by this substitution, we try to understand the difference between distribution of I_1 conditioned on containing J but not any of $\{x_2, \ldots, x_d\} - J$ as opposed to the distribution of I_1 conditioned just on containing J. We claim that these distributions are $\frac{d^2}{k}$ apart in total variational distance. This is because the probability that a random set B containing J, also contains an element of $\{x_2,\ldots,x_d\}-J$ is at most $\frac{d}{k}$ (the probability that we pick an element of $\{x_2,\ldots,x_d\}-J$) times d (the maximum number of elements of the set B). Therefore, the distribution of I such that $J \subseteq I_2 \subseteq \{x_2,\ldots,x_d\}^C \cup J$ is obtained by conditioning the uniform distribution of I such that $J \subseteq I$ on an event that has negligible ($\leq \frac{d^2}{k}$) probability, thus their total variational distance is at most $\frac{d^2}{k}$. This tells us:

$$\left| \mathop{\mathbb{E}}_{\substack{I_1 : J \subseteq I_1 \\ I_1 \subseteq \{x_2,\ldots,x_d\}^C \cup J}} f_{\approx i}(I_1) \right| \leq \left| \mathop{\mathbb{E}}_{J \subseteq I_1} f_{\approx i}(I_1) \right| + \frac{d^2}{k} \|f_{\approx i}\|_\infty$$

The latter quantity $\|f_{\approx i}\|_\infty$ is upper bounded by $2^{i^2}\|F\|_\infty \leq 2^{i^2}$ due to fact 2.12. The former $|\mathbb{E}_{J\subseteq I} f_{\approx i}(I)|$ is upper bounded by $\frac{2^{10i^2}}{k}\|F\|_\infty$ because $f_{\approx i}$ is 'approximately in $J_{\leq i}$' due to fact 2.11. Fact 2.12 also tells us that

$$|f_{\approx i}(I_2)|, |f_{\approx i}(I_3)| \leq \|f_{\approx i}\|_\infty \leq 2^{i^2}$$

Using $i \leq d \leq 3i$ gives us a total bound of:

$$\frac{9i^2}{k} 2^{3i^2} + \frac{2^{10i^2}}{k} = neg\ell(k)$$

Case 2. The intersection pattern σ is such that every x belongs to at least two sets in $\{I_1, I_2, I_3\}$.

Let $H_3 = I_1 \cap I_2 \cap I_3$ be those elements that belong to all the sets and $H_2 = (I_1 \cap I_2) \cup (I_2 \cap I_3) \cup (I_3 \cap I_1)$ be those elements that belong to exactly two of them and $H = I_1 \cup I_2 \cup I_3$ be those that belong to any of them. Note that $I_1 \cup I_2 \cup I_3 = H_2 \cup H_3 = H$.

$$\begin{aligned} &\mathbb{E} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \\ &= \mathbb{E}_{H_3} \mathbb{E}_{H_2} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \end{aligned}$$

Claim.

$$\mathbb{E}_{H_3} \mathbb{E}_{H_2} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \leq \mathbb{E}_{H_3} \sqrt{\mathbb{E}_{H_2} f_{\approx i}(I_1)^2} \sqrt{\mathbb{E}_{H_2} f_{\approx i}(I_2)^2} \sqrt{\mathbb{E}_{H_2} f_{\approx i}(I_3)^2}$$

Proof. The reason is that every term in H_2 appears in exactly two of the sets so we can apply Cauchy-Schwartz on those two sets and continue. We show the proof below. Let $P_{i,j} = I_i \cap I_j - H_3$ be those elements that appear in I_i and I_j but not in all three sets. We first look at $P_{i,j}$ and apply Cauchy Schwartz on those elements. We remark that all these inequalities must hold with $|f_{\approx i}(I)|$ instead of $f_{\approx i}(I)$ but for ease of notation, we just use $f_{\approx i}(I)$.

$$\begin{aligned} &\mathbb{E}_{H_3} \mathbb{E}_{H_2} f_{\approx i}(I_1) f_{\approx i}(I_2) f_{\approx i}(I_3) \\ &= \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2}} f_{\approx i}(I_3) \Big[\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1) f_{\approx i}(I_2) \Big] \qquad \ldots \text{ Because } I_3 \text{ does not depend on } I_1 \cap I_2 - H_3 = P_{1,2} \\ &\leq \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2}} f_{\approx i}(I_3) \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_2)^2} \qquad \ldots \text{Cauchy-Schwarz} \\ &= \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} f_{\approx i}(I_3) \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_2)^2} \end{aligned}$$

We now apply the same with $I_2 \cap I_3$ to obtain:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} f_{\approx i}(I_3) \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_2)^2} \\
&= \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2} - P_{2,3}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \Big[\mathop{\mathbb{E}}_{P_{2,3}} f_{\approx i}(I_3) \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_2)^2} \Big] \\
&\leq \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2} - P_{2,3}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{2,3}} f_{\approx i}(I_3)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,2} \cup P2,3} f_{\approx i}(I_2)^2} \\
&= \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2} - P_{2,3}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2} \cup P_{2,3}} f_{\approx i}(I_2)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{2,3}} f_{\approx i}(I_3)^2} \\
&= \mathop{\mathbb{E}}_{H_3, H_2 - P_{1,2} - P_{2,3}} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{2,3}} f_{\approx i}(I_3)^2} \\
&= \mathop{\mathbb{E}}_{H_3} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2} \mathop{\mathbb{E}}_{P_{1,3}} \sqrt{\mathop{\mathbb{E}}_{P_{1,2}} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{2,3}} f_{\approx i}(I_3)^2} \\
&\leq \mathop{\mathbb{E}}_{H_3} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,2} \cup P_{1,3}} \mathbb{E} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{P_{1,3} \cup P_{2,3}} f_{\approx i}(I_3)^2} \\
&= \mathop{\mathbb{E}}_{H_3} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_3)^2}
\end{aligned}
$$

□

We now apply Cauchy-Schwartz for the last time.

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{H_3} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_1)^2} \sqrt{\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_3)^2} \\
&\leq \sqrt{\mathop{\mathbb{E}}_{H_3} \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2 \Big) \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_1)^2 \Big)} \sqrt{\mathop{\mathbb{E}}_{H_2 \cup H_3} f_{\approx i}(I_3)^2}
\end{aligned}
$$

We first bound $\mathop{\mathbb{E}}_{H_3} \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2 \Big) \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_1)^2 \Big)$. We let $I_1 = H_3 \cup B$, that is $I_1 - (I_1 \cap I_2 \cap I_3) = B$.

$$
\begin{aligned}
&= \mathop{\mathbb{E}}_{H_3} \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2 \Big) \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_1)^2 \Big) \\
&\leq \Big(\mathop{\mathbb{E}}_{H_3} \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(I_2)^2 \Big) \Big) \Big(\max_{H_3} \Big(\mathop{\mathbb{E}}_{H_2} f_{\approx i}(B \cup H_3)^2 \Big) \\
&\leq \Big(\mathop{\mathbb{E}}_{H} f_{\approx i}(I_2)^2 \Big) \Big(\max_{H_3} \mathop{\mathbb{E}}_{B} f_{\approx i}(B \cup H_3)^2 \Big)
\end{aligned}
$$

Thus, the final bound is:

$$
\sqrt{\Big(\mathbb{E}_H f_{\approx i}(I_2)^2 \Big) \Big(\mathbb{E}_H f_{\approx i}(I_3)^2 \Big) \max_{H_3} \mathbb{E}_B f_{\approx i}(B \cup H_3)^2}
$$

We will show that the first two terms are bounded by $\frac{\eta \exp(i)}{\ell^i}$ each and the last term is bounded by $\frac{\epsilon \exp(i)}{\ell^{i-a}}$. We will use the following lemma whose proof appears in the next section.

Lemma 5.4. *Let $|A| = a \leq i$. Then:*

$$
\mathop{\mathbb{E}}_{\substack{B \subseteq \bar{A} \\ |B| = i-a}} f_{\approx i}(B \cup A)^2 \leq 2^a \sum_{Y \subseteq A} \frac{W^{i-a}[F|_{A-Y}]}{\binom{\ell}{i-a}} + neg\ell(k)
$$

Corollary 5.5. *If F is pseudorandom with respect to sets of size $i \geq a$, we can bound the R.H.S of the previous expression (up to $neg\ell(k)$ factors) by:*

$$
2^a \sum_{Y \subseteq A} \frac{\mathbb{E}[F|_{A-Y}]}{\binom{\ell}{i-a}} \quad \leq \frac{2\epsilon \exp(i)}{\ell^{i-a}}
$$

To apply this lemma to the first two terms, we simply set $A = \emptyset$. For the last term, we let B be of size $b = |I_1| - |H_3| = i - a$, and A be H_3 of size $a = |H_3|$. Putting all this together gives us a bound of:

$$\sqrt{\frac{\exp(i)\eta^2\epsilon}{\ell^{3i-a}} \pm negl(k)}$$

Since every element appears in at least 2 sets, we have:

$$3i - a = |I_1| + |I_2| + |I_3| - |I_1 \cap I_2 \cap I_3| = 2|I_1 \cup I_2 \cup I_3| \geq 2d$$

Plugging this back into the previous bound, we get:

$$\sqrt{\frac{\eta^2\epsilon\exp(i)}{\ell^{2d}}} \pm negl(k)$$

$$= \frac{\eta\sqrt{\epsilon}\exp(i)}{\ell^d} \pm negl(k)$$

This completes the proof of Lemma 5.3.

5.2 Restrictions

In order to prove Lemma 5.4, we will have to understand how restricted functions interact with level i weights. Let us define a restriction of a function $F|_X : \binom{k}{\ell - |X|} \to \mathbb{R}$, which captures conditioning $F : \binom{k}{\ell} \to \mathbb{R}$ on containing a given set X.

$$F|_X(A) := F(X \cup A)$$

Lemma 5.6.

$$f_{\approx i+1,F}(I \cup \{x\}) = f_{\approx i,F|_{\{x\}}}(I) - f_{\approx i,F}(I)$$

Proof:

$$\begin{aligned}
& f_{\approx i+1,F}(I \cup \{x\}) \\
&= \sum_{J \subseteq I \cup \{x\}} (-1)^{|I \cup \{x\}| - |J|} \mathop{\mathbb{E}}_{K \supseteq J} F(K) \\
&= \sum_{J \subseteq I} (-1)^{|I|+1-|J|} \mathop{\mathbb{E}}_{K \supseteq J} F(K) + \sum_{J \subseteq I} (-1)^{|I|+1-|J \cup \{x\}|} \mathop{\mathbb{E}}_{K \supseteq J \cup \{x\}} F(K) \\
&= -\sum_{J \subseteq I} (-1)^{|I|-|J|} \mathop{\mathbb{E}}_{K \supseteq J} F(K) + \sum_{J \subseteq I} (-1)^{|I|-|J|} \mathop{\mathbb{E}}_{K \supseteq J} F|_{\{x\}}(K) \\
&= f_{\approx i,F|_{\{x\}}}(I) - f_{\approx i,F}(I)
\end{aligned}$$

Lemma 5.7.

$$f_{\approx i+|X|,F}(I \cup X) = \sum_{Y \subseteq X} (-1)^{|X-Y|} f_{\approx i,F|_Y}(I)$$

The proof follows by doing a similar calculation as before on the expression

$$\sum_{J \subseteq I \cup X} (-1)^{|I|+|X|-|J|} \mathop{\mathbb{E}}_{K \supseteq J} F(K) = \sum_{\substack{Y \subseteq X \\ J \subseteq I}} (-1)^{|I|+|X|-|J|-|Y|} \mathbb{E}_{K \supseteq J \cup Y} F(K)$$

5.3 Proof Of Lemma 5.4

Lemma 5.8.

$$\mathbb{E}_{|I|=i} f_{\approx i}(I)^2 = \frac{W^{=i}[F]}{\binom{\ell}{i}} \pm \text{negl}(k)$$

Proof.

$$
\begin{aligned}
&W^{=i}[F] \\
&\approx \mathbb{E}_A[F^2_{\approx i}(A)] \\
&= \mathbb{E}_A\Big(\sum_{I\subseteq A, |I|=i} f_{\approx i}(A)\Big)^2 \\
&= \mathbb{E}_A \sum_{I\subseteq A, |I|=i} f_{\approx i}(A)^2 + \binom{\ell}{i} \text{ cross terms} \\
&= \binom{\ell}{i}\mathbb{E}_I f_{\approx i}(I)^2 + \text{negl}(k)
\end{aligned}
$$

The reason behind the cross terms begin negligible is very similar to the analysis of Case 1 intersections. We wish to analyze the expression $\mathbb{E}_A f_{\approx i}(I_1) f_{\approx i}(I_2)$ where $I_1 \neq I_2$. Fix the intersection pattern of I_1, I_2 to be σ. The previous expression is the same as $\mathbb{E}_{(I_1,I_2)\in\sigma} f_{\approx i}(I_1) f_{\approx i}(I_2)$. As before, we fix the elements in $I_1 - I_2$ and then take an expectation of $f_{\approx i}(I_2)$ over the elements of $I_2 - I_1$. The distribution of I_2 here is a random set of size i containing $I_1 \cap I_2$ and avoiding $I_1 - I_2$. We replace this distribution by that of a random set of size i containing $I_1 \cap I_2$ by incurring $neg\ell(k)$ error. By claim 2.11, $\mathbb{E}_{I_2 - I_1} f_{\approx i}(I_2)$ is $neg\ell(k)$. □

This tells us that analyzing the expected weight of $f_{\approx i}(I)^2$ over random sets I of size i is equivalent to understanding the level i weight of F. We will now show that analyzing the expected weight of $f_{\approx i}(I)^2$ over random sets I of size i containing A is equivalent to understanding the level i weight of $F|_{A-Y}$ over subsets $Y \subseteq A$.

Lemma 5.4. *Let $|A| = a \leq i$. Then:*

$$
\mathop{\mathbb{E}}_{\substack{B\subseteq \bar{A} \\ |B|=i-a}} f_{\approx i}(B \cup A)^2 \leq 2^a \sum_{Y\subseteq A} \frac{W^{i-a}[F|_{A-Y}]}{\binom{\ell}{i-a}} + neg\ell(k)
$$

Corollary 5.4. *If F is pseudorandom with respect to sets of size $i \geq a$, we can bound the R.H.S of the previous expression (up to $neg\ell(k)$ factors) by:*

$$
2^a \sum_{Y\subseteq A} \frac{\mathbb{E}[F|_{A-Y}]}{\binom{\ell}{i-a}} \quad \leq 2^a \sum_{Y\subseteq A} \frac{\mathbb{E}[F]+\epsilon}{\binom{\ell}{i-a}} \leq \frac{2\epsilon 2^a}{\binom{\ell}{i-a}} \leq \frac{2\epsilon \exp(i)}{\ell^{i-a}}
$$

Proof Of Lemma 5.7 Consider:

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{\substack{B\subseteq \bar{A} \\ |B|=i-a}} f_{\approx i}(B \cup A)^2 \\
&= \mathop{\mathbb{E}}_{\substack{B\subseteq \bar{A} \\ |B|=i-a}} \Big(\sum_{Y\subseteq A} (-1)^{|Y|} f_{\approx i, F|_{A-Y}}(B)\Big)^2 \qquad \ldots \text{from equation (1)} \\
&\leq \mathop{\mathbb{E}}_{\substack{B\subseteq \bar{A} \\ |B|=i-a}} 2^a \Big(\sum_{Y\subseteq A} f_{\approx i, F|_{A-Y}}(B)^2\Big) \qquad \ldots \text{by Cauchy Schwarz} \\
&= 2^a \sum_{Y\subseteq A} \mathop{\mathbb{E}}_{\substack{B\subseteq \bar{A} \\ |B|=i-a}} f_{\approx i, F|_{A-Y}}(B)^2 \\
&\leq 2^a \sum_{Y\subseteq A} \frac{W^{i-a}[F|_{A-Y}]}{\binom{\ell}{i-a}} + neg\ell(k)
\end{aligned}
$$

References

[BKS18] Boaz Barak, Pravesh K. Kothari, and David Steurer. Small-set expansion in shortcode graph and the 2-to-2 conjecture. *CoRR*, abs/1804.08662, 2018.

[DKK+16] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. Towards a proof of the 2-to-1 games conjecture? *Electronic Colloquium on Computational Complexity (ECCC)*, 23:198, 2016.

[DKK+18] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. On non-optimally expanding sets in grassmann graphs. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018, pages 940–951, New York, NY, USA, 2018. ACM.

[KMMS18] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Small set expansion in the johnson graph. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:78, 2018.

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 592–601, 2018.