Agreement tests on graphs and hypergraphs

Irit Dinur* Yuval Filmus† Prahladh Harsha‡

November 16, 2017

Abstract

Agreement tests are a generalization of low degree tests that capture a local-to-global phenomenon, which forms the combinatorial backbone of most PCP constructions. In an agreement test, a function is given by an ensemble of local restrictions. The agreement test checks that the restrictions agree when they overlap, and the main question is whether average agreement of the local pieces implies that there exists a global function that agrees with most local restrictions.

There are very few structures that support agreement tests, essentially either coming from algebraic low degree tests or from direct product tests (and recently also from high dimensional expanders). In this work, we prove a new agreement theorem which extends direct product tests to higher dimensions, analogous to how low degree tests extend linearity testing. As a corollary of our main theorem, we show that an ensemble of small graphs on overlapping sets of vertices can be glued together to one global graph assuming they agree with each other on average.

Our agreement theorem is proven by induction on the dimension (with the dimension 1 case being the direct product test, and dimension 2 being the graph case). A key technical step in our proof is a new hypergraph pruning lemma which allows us to treat dependent events as if they are disjoint, and may be of independent interest.

Beyond the motivation to understand fundamental local-to-global structures, our main theorem is used in a completely new way in a recent paper by the authors [DFH17] for proving a structure theorem for Boolean functions on the p-biased hypercube. The idea is to approximate restrictions of the Boolean function on simpler sub-domains, and then use the agreement theorem to glue them together to get a single global approximation.

*Weizmann Institute of Science, ISRAEL. email: `irit.dinur@weizmann.ac.il`.

†Technion Israel Institute of Technology, ISRAEL. email: `yuvalfi@cs.technion.ac.il`

‡Tata Institute of Fundamental Research, INDIA. email: `prahladh@tifr.res.in`

1 Introduction

Agreement tests are a type of PCP tests and capture a fundamental local-to-global phenomenon. In this paper, we study an agreement testing question that is a new extension of direct product testing to higher dimensions.

It is a basic fact of computation that any global computation can be broken down into a sequence of local steps. The PCP theorem [AS98, ALM+98] says that moreover, this can be done in a robust fashion, so that as long as *most* steps are correct, the entire computation checks out. At the heart of this is a *local-to-global* argument that allows deducing a global property from local pieces that fit together only approximately.

A key example is the line vs. line [GLR+91, RS96] low degree test in the proof of the PCP theorem. In the PCP construction, a function on a large vector space is replaced by an ensemble of (supposed) restrictions to all possible affine lines. These restrictions are supplied by a prover and are not a priori guaranteed to agree with any single global function. This is taken care of by the "low degree test", which checks that restrictions on intersecting lines agree with each other, i.e. they give the same value to the point of intersection. The crux of the argument is the fact that the local agreement checks *imply* agreement with a single global function. Thus, the low degree test captures a local-to-global phenomenon.

In what other scenarios does such a local-to-global theorem hold? This question was first asked by Goldreich and Safra [GS00] who studied a combinatorial analog of the low degree test. Let us describe the basic framework of agreement testing in which we will study this question. In agreement testing, a global function is given by an ensemble of local functions. There are two key aspects of agreement testing scenarios:

- Combinatorial structure: for a given ground set V of size n, the combinatorial structure is a collection H of subsets $S \subset V$ such that for each $S \in H$ we get a local function. For example, if V is the points of a vector space then H can be the collection of affine lines.
- Allowed functions: for each subset $S \in H$, we can specify a space $\mathcal{F}_S$ of functions on S that are allowed. The input to the agreement test is an ensemble of functions $\{f_S\}$ such that for every $S \in H$, $f_S \in \mathcal{F}_S$. For example, in the line vs. line low degree test we only allow local functions on each line that have low degree.

Given the ensemble $\{f_S\}$, the intention is that f_S is the restriction to S of a global function $F\colon V \to \Sigma$. Indeed, a local ensemble is called *global* if there is a global function $F\colon V \to \Sigma$ such that

$$\forall S \in H, \qquad f_S = F|_S.$$

An *agreement check* for a pair of subsets S_1, S_2 checks whether their local functions agree, denoted $f_{S_1} \sim f_{S_2}$. Formally,

$$f_{S_1} \sim f_{S_2} \qquad \Longleftrightarrow \qquad \forall x \in S_1 \cap S_2, \quad f_{S_1}(x) = f_{S_2}(x).$$

A local ensemble which is global passes all agreement checks. The converse is also true: a local ensemble that passes *all* agreement checks must be global.

An *agreement test* is specified by giving a distribution $\mathcal{D}$ over pairs (or triples, etc.) of subsets S_1, S_2. We define the agreement of a local ensemble to be the probability of agreement:

$$\mathrm{agree}_{\mathcal{D}}(\{f_S\}) := \Pr_{S_1,S_2\sim D}\left[f_{S_1} \sim f_{S_2}\right].$$

An agreement theorem shows that if $\{f_S\}_S$ is a local ensemble with $agree_{\mathcal{D}}(\{f_S\}) > 1-\varepsilon$ then it is close to being global.

Example: direct product tests Perhaps the simplest agreement test to describe is the direct product test, in which H contains all possible k-element subsets of V. For each S, we let $\mathcal{F}_S$ be all possible functions on S, that is $\mathcal{F}_S = \{f\colon S \to \Sigma\}$. The input to the test is an ensemble of local functions $\{f_S\}$, and a natural testing distribution is to choose S_1, S_2 so that they intersect on $t = \Theta(k)$ elements. Suppose that $\mathrm{agree}(\{f_S\}) \geq 1-\varepsilon$. Is there a global function $F\colon V \to \Sigma$ such that $F|_S = f_S$ for most subsets S? This is the content of the direct product testing theorem of Dinur and Steurer [DS14]:

Theorem 1.1 (Agreement theorem, dimension 1)**.** *There exists constants $C > 1$ such that for all $\alpha, \beta \in (0, 1)$ satisfying $\alpha + \beta \leq 1$, all positive integers $n \geq k \geq t$ satisfying $n \geq Ck$ and $t \geq \alpha k$ and $k - t \geq \beta k$, and all finite alphabets Σ, the following holds: Let $f = \{f_S : S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $agree_{\nu_{n,k,t}}(f) \geq 1 - \varepsilon$, that is,*

$$\Pr_{S_1, S_2 \sim \nu_{n,k,t}}[f_{S_1}|_{S_1 \cap S_2} = f_{S_2}|_{S_1 \cap S_2}] \geq 1 - \varepsilon,$$

where $\nu_{n,k,t}$ is the uniform distribution over pairs of k-sized subsets of $[n]$ of intersection exactly t. Then there exists a global function $F : [n] \to \Sigma$ satisfying $\Pr_{S \in \binom{[n]}{k}}[f_S = F|_S] = 1 - O_{\alpha,\beta}(\varepsilon)$.

The qualitatively strong aspect of this theorem is that in the conclusion, the global function agrees *perfectly* with $1-O(\varepsilon)$ of the local functions. Achieving a weaker result where perfect agreement $f_S = F|_S$ is replaced by approximate one $f_S \approx F|_S$ would be significantly easier but also less useful. Quantitatively, this is manifested in that the fraction of local functions that end up disagreeing with the global function F is at most $O(\varepsilon)$ and is *independent of n and k*. It would be significantly easier to prove a weaker result where the closeness is $O(k\varepsilon)$ (via a union bound on the event that $F(i) = f_S(i)$). This theorem is proven [DS14] by imitating the proof of the parallel repetition theorem [Raz98]. This theorem is also used as a component in the recent work on agreement testing on high dimensional expanders [DK17].

Our Results

In order to motivate our extension of Theorem 1.1, let us describe it in a slightly different form. The global function F can be viewed as specifying the coefficients of a linear form $\sum_{i=1}^{n} F(i)x_i$ over variables $x_1, \ldots, x_n$. For each S, the local function f_S specifies the partial linear form only over the variables in S. This f_S is supposed to be equal to F on the part of the domain where $x_i = 0$ for all $i \notin S$. Given an ensemble $\{f_S\}$ whose elements are promised to agree with each other on average, the agreement theorem allows us to conclude the existence of a global linear function that agrees with most of the local pieces.

This description naturally leads to the question of extending this to higher degree polynomials. Now, the global function is a degree d polynomial with coefficients in Σ, namely $F = \sum_T F(T)x_T$, where we sum over subsets $T \subset [n]$, $|T| \leq d$. The local functions f_S will be polynomials of degree $\leq d$, supposedly obtained by zeroing out all variables outside S. Two local functions f_{S_1}, f_{S_2} are said to agree, denoted $f_{S_1} \sim f_{S_2}$, if every monomial that is induced by $S_1 \cap S_2$ has the same coefficient in both polynomials. Our new agreement theorem says that in this setting as well, local agreement implies global agreement.

Theorem 1.2 (Main)**.** *For every positive integer d and alphabet Σ, there exists a constant $C > 1$ such that for all $\alpha, \beta \in (0, 1)$ satisfying $\alpha + \beta \leq 1$ and all positive integers $n \geq k \geq t$ satisfying $n \geq Ck$ and $t \geq \max\{\alpha k, 2d\}$ and $k - t \geq \max\{\beta k, d\}$, the following holds: Let $f = \{f_S : \binom{S}{\leq d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $agree_{\nu_{n,k,t}}(f) \geq 1 - \varepsilon$, that is,*

$$\Pr_{S_1, S_2 \sim \nu_{n,k,t}}[f_{S_1}|_{S_1 \cap S_2} = f_{S_2}|_{S_1 \cap S_2}] \geq 1 - \varepsilon,$$

where $\nu_{n,k,t}$ is the uniform distribution over pairs of k-sized subsets of $[n]$ of intersection exactly t. Then there exists a global function $G : \binom{[n]}{\leq d} \to \Sigma$ satisfying $\Pr_{S \in \binom{[n]}{k}}[f_S = G|_S] = 1 - O_{d,\alpha,\beta}(\varepsilon)$.

Here, $F|_S$ refers to the restriction $F|_{\binom{S}{\leq d}}$.

Furthermore, we may assume that the global function G is the one given by "popular vote", namely for each $A \in \binom{[n]}{\leq d}$ set $G(A)$ to be the most frequently occurring value among $\{f_S(A) \mid S \supset A\}$ (breaking ties arbitrarily).

For $d = 1$, this theorem is precisely Theorem 1.1 (but for the "furthermore" clause). The additional "furthermore" clause strengthens our theorem by naming the popular vote function as a candidate global function that explains most of the local functions. This addendum strengthens also Theorem 1.1 and turns out important for an application [DFH17] of our theorem which we describe later in the introduction.

Let us spell out how this theorem fits into the framework described above. The ground set is $V = \binom{[n]}{\leq d}$, and the collection of subsets H is the collection of all induced hypergraphs on k elements. In particular, if we focus on $\Sigma = \{0, 1\}$, we can view the local function of a subset $S \subset [n]$, $|S| = k$, as specifying a hypergraph on the vertices of S with hyperedges of size up to d. The theorem says that if these small hypergraphs agree with each other most of the time, then there is a global hypergraph that they nearly all agree with.

For the special case of $d = 2$ and $\Sigma = \{0, 1\}$, we get an interesting statement about combining small pieces of a graph into a global one.

Corollary 1.3 (Agreement test for graphs)**.** *There exist a constant $C > 1$ such that for all $\alpha, \beta \in (0, 1)$ satisfying $\alpha + \beta \le 1$ and all for all positive integers $n \ge k \ge t \ge 4$ satisfying $n \ge Ck$, $t \ge \alpha k$ and $k - t \ge \max\{\beta k, 2\}$ the following holds:*

Let $\{G_S\}$ be an ensemble of graphs, where S is a k element subset of $[n]$ and G_S is a graph on vertex set S. Suppose that

$$\Pr_{\substack{S_1, S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}} [G_{S_1}|_{S_1 \cap S_2} = G_{S_2}|_{S_1 \cap S_2}] \ge 1 - \varepsilon.$$

Then there exists a single global graph $G = ([n], E)$ satisfying $\Pr_{S \in \binom{[n]}{k}}[G_S = G|_S] = 1 - O(\varepsilon)$.

Here too we emphasize that the strength of the statement is in that the conclusion talks about *exact* agreement between the global graph and the local graphs, i.e. $G_S = G|_S$ and not $G_S \approx G|_S$, for a fraction of $1 - O(\varepsilon)$ of the sets S. It is also important that there is no dependence in the $O(\cdot)$ on either n or k. A similar agreement testing statement can be made for hypergraphs of any uniformity $\le d$.

A technical component in our proof which we wish to highlight is a new hypergraph pruning lemma, which may be of independent interest. The lemma can be interpreted by viewing a hypergraph as specifying the minterms of a monotone DNF (of width at most d). The lemma allows to prune the DNF so that the new sub-DNF still has similar density (the fraction of inputs on which it is 1), but also has a structural property which we call *bounded branching factor* and which implies that for typical inputs, only a single minterm is responsible for the function evaluating to 1.

Lemma 1.4 (hypergraph pruning lemma)**.** *Fix constants $\varepsilon > 0$ and $d \ge 1$. There exists $p_0 > 0$ (depending on d, ε) such that for every $n \ge k \ge 2d$ satisfying $k/n \le p_0$ and every d-uniform hypergraph H on $[n]$ there exists a subhypergraph H' obtained by removing hyperedges such that*

1. $\Pr_{S \sim \nu_{n,k}}[H'|_S \ne \emptyset] = \Omega_{d,\varepsilon}(\Pr_{S \sim \nu_{n,k}}[H|_S \ne \emptyset])$.
2. *For every $e \in H'$,* $\Pr_{S \sim \nu_{n,k}}[H'|_S = \{e\} \mid S \supset e] \ge 1 - \varepsilon$.

Here $H'|_S$ is the hypergraph induced on the vertices of S.

We illustrate an application of this lemma later on.

Context and Motivation

Agreement tests were first studied implicitly in the context of PCP theorems. In fact, every PCP construction that has a composition step invariably relies on an agreement theorem. This is because in a typical PCP construction, the proof is broken into small pieces that are further encoded e.g. by composition or by a gadget. The soundness analysis decodes each gadget separately, thereby obtaining a collection of local views. Then, essentially through an agreement theorem, these are stitched together into one global NP witness. Similar to locally testable codes, agreement tests are a combinatorial question that is related to PCPs. Interestingly, this relation has recently been made formal by Dinur *et al.* [DKK+16], where it is proved that a certain agreement test (whose correctness is hypothesized there) *formally implies* a certain rather strong unique games PCP theorem. Such a formal connection is not known to exist between LTCs and PCPs. For example, even if someone manages to construct the "ultimate" locally testable codes with linear length and distance, and testable with a constant number of queries, this is not known to have any implications for constructing linear size PCPs (although one may hope that such codes will be useful toward that goal).

Beyond their role in PCPs, we believe that agreement tests capture a fundamental local-to-global phenomenon, and merit study on their own. Exploring new structures that support agreement theorems seems to be an important direction.

Application for structure theorems In a very recent work [DFH17], the authors have found a totally different application for agreement tests (in particular, for Theorem 1.2) that is outside the PCP domain. Theorem 1.2 is applied towards proving a certain structure theorem on Boolean functions in the p-biased hypercube. Given a function on the p-biased hypercube, the key is to look at restrictions of the global function to small sub-cubes that are identical to the *uniform* hypercube. On the uniform hypercube, there are previously known structure theorems which give us a local approximation of our function separately on each sub-cube. One ends up with an ensemble $\{f_S\}$ of simple functions (juntas, actually) that locally approximate the function, and then Theorem 1.2 is used to stitch all of the local junta approximations into one nice global function.

The interplay between the global structure of a function and how it behaves on (random) restrictions is a powerful tool that is well studied for proving circuit complexity lower bounds. Although agreement tests have not so far been useful in that arena, this seems like an interesting possibility.

Relation to Property Testing Agreement testing is similar to property testing in that we study the relation between a global object and its local views. In property testing we have access to a single global object, and we restrict ourselves to look only at random local views of it. In agreement tests, we don't get access to a global object but rather to an *ensemble of local functions* that are not apriori guaranteed to come from a single global object. Another difference is that unlike in property testing, in an agreement test the local views are pre-specified and are a part of the problem description, rather than being part of the algorithmic solution.

Still, there is an interesting interplay between Corollary 1.3, which talks about combining an ensemble of local graphs into one global graph, and graph property testing. Suppose we focus on some testable graph property, and suppose further that the test proceeds by choosing a random set of vertices and reading all of the edges in the induced subgraph, and checking that the property is satisfied there (many graph properties are testable this way, for example bipartiteness [GGR98]). Suppose we only allow ensembles $\{G_S\}$ where for each subset S, the local graph G_S satisfies the property (e.g. it is bipartite). This fits into our formalism by specifying the space of allowed functions $\mathcal{F}_S$ to consist only of accepting local views. This is analogous to requiring, in the low degree test, that the local function on each line has low degree as a univariate polynomial. By Corollary 1.3, we know that if these local graphs agree with each other with probability $1-\varepsilon$, there is a global graph G that agrees with $1-O(\varepsilon)$ of them. In particular, this graph *passes the property test*, so must itself be close to having the property! At this point it is absolutely crucial that the agreement theorem provides the stronger guarantee that $G|_S = G_S$ (and not $G|_S \approx G_S$) for $1-O(\varepsilon)$ of the S's. We can thus conclude that not only is there a global graph G, but actually that this global G is close to having the property.

This should be compared to the low degree agreement test, where we only allow local functions with low degree, and the conclusion is that there is a global function that itself has low degree.

Technical Contribution

Our proof of Theorem 1.2 proceeds by induction on the dimension d. For $d = 1$, this is the direct product test theorem of Dinur and Steurer [DS14], which we reprove in a way that more readily generalizes to higher dimension. Given an ensemble $\{f_S\}$, it is easy to define the global function G, by popular vote ("majority decoding"). The main difficulty is to prove that for a typical set S, f_S agrees with $G|_S$ on all elements $i \in S$ (and later on all d-sets).

Our proof doesn't proceed by defining G as majority vote right away. Instead, like in many previous proofs [DG08, IKW12, DS14], we condition on a certain event (focusing say on all subsets that contain a certain set T, and such that $f_S|_T = \alpha$ for a certain value of α), and define a "restricted global" function, for each T, by taking majority just among the sets in the conditioned event. This boosts the probability of agreement inside this event. After this boost, we can afford to take a union bound and safely get agreement with the restricted global function G_T. The proof then needs to perform another agreement step which stitches the restricted global functions $\{G_T\}_T$ into a completely global function. The resulting global function does not necessarily equal the majority vote function G, and a separate argument is then carried out to show that the conclusion is correct also for G.

In higher dimensions $d > 1$, these two steps of agreement (first to restricted global and then to global) become a longer sequence of steps, where at each step we are looking at restricted functions that are defined over larger and larger parts of the domain.

The technical main difficulty is that a single event $f_S = F|_S$ consists of $\binom{k}{d}$ little events, namely $f_S(A) = F(A)$ for all $A \in \binom{S}{d}$, that each have some probability of failure. We thus need an even larger boost, to bound the failure probability by about ε/k^d so that we can afford to take a union bound on the $\binom{k}{d}$ different sub-events. How do we get this large boost? Our strategy is to proceed by induction, where at each stage, we condition on the global function from the previous stage, boosting the probability of success further.

Hypergraph pruning lemma An important component that yields this boosting is the hypergraph pruning lemma (Lemma 1.4) that was described earlier. The lemma allows approximating a given hypergraph H by a subhypergraph $H' \subset H$ that has a *bounded branching factor*.

Definition 1.5 (branching factor)**.** *For any $\rho \geq 1$, a hypergraph H over a vertex set V is said to have* branching factor ρ *if for all subsets $A \subset V$ and integers $r \geq 0$, there are at most ρ^r hyperedges in H of cardinality $|A| + r$ containing A.*

Our proof of the hypergraph pruning lemma produces a sub-hypergraph with branching factor $\rho = O(n/k)$. The branching factor is responsible for the second item in the lemma, which guarantees that usually if a set S contains a hyperedge from H, it contains a *unique* hyperedge from H'.

The importance of this is roughly for "inverting union bound arguments". It essentially allows us to estimate the probability of an event of the form "S contains some hyperedge of H'" as the sum, over all hyperedges, of the probability that S contains a specific hyperedge.

The proof of the lemma is subtle and proceeds by induction on the dimension d. It essentially describes an algorithm for obtaining H' from H and the proof of correctness uses the FKG inequality. We illustrate how Lemma 1.4 is used by its application to majority decoding.

Majority decoding The most natural choice for the global function F in the conclusion of Theorem 1.2 is the majority decoding, where $F(A)$ is the most common value of $f_S(A)$ over all S containing A. This is the content of the "furthermore" clause in the statement of the theorem. Neither the proof strategy of [DS14] nor our generalization promises that the produced global function F is the majority decoding. Our inductive strategy produces a global function which agrees with most local functions, but we cannot guarantee immediately that this global function corresponds to majority decoding. What we are able to show is that *if* there is a global function agreeing with most of the local functions *then* the function obtained via majority decoding also agrees with most of the local functions. We outline the argument below. Suppose that $\{f_S\}$ is an ensemble of local functions that mostly agree with each other, and suppose that they also mostly agree with some global function F. Let G be the function obtained by majority decoding: $G(A)$ is the most common value of $f_S(A)$ over all S containing A. Our goal is to show that G also mostly agrees with the local functions, and we do this by showing that F and G mostly agree.

Suppose that $F(A) \neq G(A)$. We consider two cases. If the distribution of $f_S(A)$ is very skewed toward $G(A)$, then $f_S(A) \neq F(A)$ will happen very often. If the distribution of $f_S(A)$ is very spread out, then $f_{S_1}(A) \neq f_{S_2}(A)$ will happen very often. Since both events $f_S(A) \neq F(A)$ and $f_{S_1}(A) \neq f_{S_2}(A)$ are known to be rare, we would like to conclude that $F(A) \neq G(A)$ happens for very few A's.

Here we face a problem: the bad events (either $f_S(A) \neq F(A)$ or $f_{S_1}(A) \neq f_{S_2}(A)$) corresponding to different A's are not necessarily disjoint. A priori, there might be many different A's such that $F(A) \neq G(A)$, but the bad events implied by them could all coincide.

The hypergraph pruning lemma enables us to overcome this difficulty. Let $H = \{A : F(A) \neq G(A)\}$, and apply the hypergraph pruning lemma to obtain a subhypergraph H'. The lemma states that with constant probability, a random set S sees at most one disagreement between F and G. This implies that the bad events considered above can be associated, with constant probability, with a *unique* A. In this way, we are able to obtain an upper bound on the probability that F, G disagree on an input from H'. The hypergraph pruning lemma then guarantees that the probability that F, G disagree (on *any* input) is also bounded.

Organization

The rest of this paper is organized as follows. We begin by reproving Theorem 1.1 of Dinur and Steurer [DS14] in Section 2 a manner that generalizes to higher dimension. We then generalize the proof of the $d = 1$ theorem to higher dimensions (Theorem 3.1) in Section 3. This almost proves Theorem 1.2 but for the "furthermore" clause. In Section 4, we prove the hypergraph pruning lemma, a crucial ingredient in the generalization to higher dimensions. Finally, in Section 5, we use the hypergraph pruning lemma (again) to prove the "furthermore" clause of Theorem 1.2 thus completing the proof of our main result. We also show how the agreement theorem can be extended to the μ_p biased setting in Section 5.

2 One-dimensional agreement theorem

In this section, we prove the following direct product agreement testing theorem for dimension one in the uniform setting. This theorem is a special case of the more general theorem (Theorem 3.1) proved in the next section and also follows from the work of Dinur and Steurer [DS14]. However, we give the proof for the dimension one case as it serves as a warmup to the general dimension case.

Theorem 1.1 (Restated) (Agreement theorem, dimension 1). *There exists constants $C > 1$ such that for all $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$, all positive integers n, k, t satisfying $n \geq Ck$ and $t \geq \alpha k$ and $k - t \geq \beta k$, and all finite alphabets Σ, the following holds: Let $f = \{f_S \colon S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying $\mathrm{agree}_{\nu_{n,k,t}}(f) \geq 1 - \varepsilon$, that is,*

$$\Pr_{S_1, S_2 \sim \nu_{n,k,t}} [f_{S_1}|_{S_1 \cap S_2} = f_{S_2}|_{S_1 \cap S_2}] \geq 1 - \varepsilon,$$

where $\nu_{n,k,t}$ is the uniform distribution over pairs of k-sized subsets of $[n]$ of intersection exactly t. Then there exists a global function $F \colon [n] \to \Sigma$ satisfying $\Pr_{S \in \binom{[n]}{k}}[f_S = F|_S] = 1 - O_{\alpha,\beta}(\varepsilon)$.

The distribution $\nu_{n,k.t}$ is the distribution induced on the pair of sets $(S_1, S_2) \in \binom{[n]}{k}^2$ by first choosing uniformly at random a set $U \subset [n]$ of size t and then two sets S_1 and S_2 of size k of $[n]$ uniformly at random conditioned on $S_1 \cap S_2 = U$. We can think of picking these two sets as first choosing uniformly at random a set T of size $t-1$, then a random element $i \in [n] \setminus T$, setting $U = T + i$ and then choosing two sets S_1 and S_2 such that $S_1 \cap S_2 = T + i$. Clearly, the probability that the functions f_{S_1} and f_{S_2} disagree is the sum of the probabilities of the following two events: (A) $f_{S_1}|_T \neq f_{S_2}|_T$, (B) $f_{S_1}|_T = f_{S_2}|_T$ but $f_{S_1}(i) \neq f_{S_2}(i)$. This motivates the following definitions for any $T \in \binom{[n]}{t-1}$ and $i \in [n] \setminus T$.

$$\begin{aligned} \varepsilon_T(\emptyset) &= \Pr_{\substack{S_1, S_2 \sim \nu(k,t) \\ S_1 \cap S_2 \supseteq T}} [f_{S_1}|_T \neq f_{S_2}|_T], \\ \varepsilon_T(i) &= \Pr_{\substack{S_1, S_2 \sim \nu(k,t) \\ S_1 \cap S_2 = T + i}} [f_{S_1}|_T = f_{S_2}|_T \text{ and } f_{S_1}(i) \neq f_{S_2}(i)]. \end{aligned}$$

It is easy to see that for a typical T, both $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i \notin T}[\varepsilon_T(i)]$ is $O(\varepsilon)$. This suggests the following strategy to prove Theorem 1.1. For each typical T, construct a "global" function $g_T \colon [n] \to \Sigma$ based on the most popular value of f_S among the f_S's that agree on T (see Section 2.2 for details) and show that most g_T's agree with each other. More precisely, we prove the theorem in 3 steps as follows: In the first step (Section 2.1), we bound $\varepsilon_T(\emptyset)$ and $\varepsilon_T(i)$ for typical T's and i. In the second step (Section 2.2), we construct for a typical T, a "global" function g_T that explains most "local" $\{f_S\}_{S \supset T}$. In the final step (Section 2.3), we show that the global functions corresponding to most pairs of typical T's agree with each other, thus demonstrating the existence of a single global function F (in particular a random global function g_T) that explains most of the "local" functions f_S even corresponding to S's which do not contain T.

2.1 Step 1: Bounding $\varepsilon_T(\emptyset)$ and $\varepsilon_T(i)$

We begin by showing that for a typical T of size $t - 1$, we can upper bound $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i \notin T}[\varepsilon_T(i)]$.

Lemma 2.1. *We have $\mathbb{E}_T[\varepsilon_T(\emptyset)] \leq \varepsilon$ and $\mathbb{E}_{T, i \notin T}[\varepsilon_T(i)] \leq \frac{\varepsilon}{t}$.*

Proof. For a non-negative integer j, let ε_j be the probability that the functions f_{S_1} and f_{S_2} corresponding to a pair of sets (S_1, S_2) picked according to the distribution $\nu_n(k,t)$ disagree on exactly j elements in

$S_1 \cap S_2$. By assumption of Theorem 1.1, we have $\sum_{j=1}^{t} \varepsilon_j \leq \varepsilon$. Furthermore, it is easy to see that $\mathbb{E}_T[\varepsilon_T(\emptyset)] = \left(1 - \frac{1}{t}\right)\varepsilon_1 + \sum_{j>1} \varepsilon_j$ and $\mathbb{E}_{T,i}[\varepsilon_T(i)] = \varepsilon_1/t$. The lemma follows from these observations. $\square$

We will need the following auxiliary lemma in our analysis.

Lemma 2.2. *Let $c \in (0,1)$ and $n \geq 4k/c$. Consider the bipartite inclusion graph between $[n]$ and $\binom{[n]}{k}$ (ie., (i,S) is an edge if $i \in S$). Let $B \subset [n]$ and $T \subset \binom{[n]}{k}$ be such that for each $i \in B$, the set of neighbours of i in T (denoted by $T_i := \{S \in T \mid S \ni i\}$) is of size at least $c\binom{n-1}{k-1}$. Then either*

$$\Pr_{S \sim \nu_{n,k}}[S \in T] \geq \max\left\{\frac{ck}{2} \cdot \Pr_i[i \in B], \frac{c^2}{16}\right\}.$$

Proof. Let S be a random set of size k. To begin with, we can assume that $|B| \leq n/2$ since otherwise $\Pr_S[S \in T] \geq c/2 \geq c^2/16$ and we are done. Let i be any element in B. The probability that $S \cap B = \{i\}$ conditioned on the event that S contains i is given as follows:

$$\Pr[S \cap B = \{i\} \mid i \in S] = \prod_{i=1}^{|B|-1}\left(1 - \frac{k-1}{n-i}\right) \geq \left(1 - \frac{k-1}{n-|B|}\right)^{|B|} \geq 1 - \frac{k}{n/2}|B|.$$

Hence, for any $i \in B$, $\Pr[S \in T_i \text{ and } S \cap B = \{i\} \mid i \in S] \geq c - \frac{2k}{n} \cdot |B|$. It follows that

$$\Pr[S \in T] \geq \sum_{i \in B} \Pr[S \in T_i \text{ and } S \cap B = \{i\}] \geq \frac{k}{n}\sum_{i \in B} \Pr[S \in T_i \text{ and } S \cap B = \{i\} \mid i \in S] \geq \frac{k}{n}|B|\left(c - \frac{2k}{n}|B|\right).$$

If the above is true for B, it is also true for any $B' \subset B$. Now, if $|B| \geq cn/4k$, then consider $B' \subset B$ of size $\lfloor cn/4k \rfloor \geq cn/8k$. Then applying the above inequality for B', we have $\Pr[S \in T] \geq \frac{c}{8} \cdot \frac{c}{2} = \frac{c^2}{16}$. Other wise $|B| < cn/4k$, now again appealing to the above inequality, we have $\Pr[S \in T] \geq \frac{ck}{2} \cdot \Pr[i \in B]$. $\square$

2.2 Step 2: Constructing global functions for typical T's

We prove the following lemma in this section.

Lemma 2.3. *For all $\alpha \in (0,1)$ and positive integers n,k,t satisfying $n \geq 8k$ and $t \geq \alpha k$ and alphabet Σ the following holds: Let $\{f_S \colon S \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying*

$$\Pr_{\substack{S_1,S_2 \in \binom{[n]}{k} \\ |S_1 \cap S_2| = t}}[f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon,$$

then there exists an ensemble $\{g_T \colon [n] \to \Sigma \mid T \in \binom{[n]}{t-1}\}$ of global functions such when a random $T \in \binom{[n]}{t-1}$ and $S \in \binom{[n]}{k}$ are chosen such that $S \supset T$, then $\Pr[g_T|_S \neq f_S] = O_\alpha(\varepsilon)$.

By Lemma 2.1, we know that a typical T of size $t-1$ satisfies $\varepsilon_T(\emptyset) = O(1)$. We prove the above lemma, by constructing for each such typical T a global function g_T that explains most local functions f_S for $S \supset T$. For the rest of this section fix such a T.

Given $X = \binom{[n]}{k}$, let $X_T := \{S \in X \mid S \supset T\}$. Let $n' = n - (t-1)$ and $k' = k - (t-1)$. For $i \notin T$, let $X_{T,i} := X_{T+i} = \{S \in X_T \mid i \in S\}$.

We now define the "global" function $g_T : [n] \to \Sigma$ as follows. We first define the value of g_T (we will drop the subscript T when T is clear from context) for $i \in T$ and then for each $i \notin T$. Define $g|_T : T \to \Sigma$ to be the most popular restriction of the functions $f_S|_T$ for $S \in X_T$. In other words, $g|_T$ is the function that maximizes $\Pr_{S \in X_T}[g|_T = f_S|_T]$. Let $X^{(0)} := \{S \in X_T \mid f_S|_T = g|_T\}$ be the set of S's that agree with this most popular value. For each $i \notin T$, let $X^{(0)}_{T,i} := X^{(0)} \cap X_{T,i}$. For each such i, define $g(i)$ to be the most popular value $f_S(i)$ among $S \in X^{(0)}_{T,i}$. This completes the definition of the function g.

We now show that if $\varepsilon_T(\emptyset)$ is small, then the function g_T agrees with most functions $f_S, S \in X_T$.

$$\begin{aligned}
\Pr_{S\in X_T}[f_S \neq g|_S] &\leq \Pr_{S\in X_T}[f_S|_T \neq g|_T] + \sum_{i\notin T} \Pr_{S\in X_T}[i \in S \text{ and } f_S|_T = g|_T \text{ and } f_S(i) \neq g(i)] \\
&= \Pr_{S\in X_T}[f_S|_T \neq g|_T] + \frac{k'}{n'} \sum_{i\notin T} \Pr_{S\in X_{T,i}}[f_S|_T = g|_T \text{ and } f_S(i) \neq g(i)] \\
&= \Pr_{S\in X_T}[f_S|_T \neq g|_T] + \frac{k'}{n'} \sum_{i\notin T} \Pr_{S\in X_{T,i}}[S \in X^{(0)}_{T,i}] \cdot \Pr_{S\in X^{(0)}_{T,i}}[f_S(i) \neq g(i)] \qquad (1)
\end{aligned}$$

This motivates the definition of the following quantities which we need to bound.

$$\gamma(\emptyset) := \Pr_{S\in X_T}[f_S|_T \neq 1g|_T]; \qquad \gamma(i) := \Pr_{S\in X^{(0)}_{T,i}}[f_S(i) \neq g(i)]; \qquad \rho(i) := \Pr_{S\in X_{T,i}}[S \in X^{(0)}_{T,i}].$$

We now bound $\gamma(\emptyset)$ and $\gamma(i)$ in terms $\varepsilon_T(\emptyset)$ and $\mathbb{E}_{i\notin T}[\varepsilon_T(i)]$ via the following (disagreement) probabilities.

$$\kappa(\emptyset) := \Pr_{S_1,S_2\in X_T}[f_{S_1}|_T \neq f_{S_2}|_T]; \qquad \kappa(i) := \Pr_{S_1,S_2\in X^{(0)}_{T,i}}[f_{S_1}(i) \neq f_{S_2}(i)].$$

Claim 2.4 (Bounding $\gamma(\emptyset)$)**.** $\gamma(\emptyset) \leq \kappa(\emptyset) \leq 2\varepsilon_T(\emptyset)$.

Proof. By definition, we have $\kappa(\emptyset) = \mathbb{E}_{S_1\in X_T}\left[\Pr_{S_2\in X_T}[f_{S_T}|_T \neq f_{S_2}|_T]\right] \geq \gamma(\emptyset)$ since $g|_T$ is the most popular value among $f_S|_T$ for $S \in X_T$. The only difference between $\kappa(\emptyset)$ and $\varepsilon_T(\emptyset)$ is the distribution from which the pairs (S_1, S_2) are drawn; for $\kappa(\emptyset)$, (S_1, S_2) is drawn uniformly from all pairs $X_T \times X_T$ while for $\varepsilon_T(\emptyset)$, (S_1, S_2) is drawn from $\nu_n(k,t)$. To complete the argument, we choose $S_1, S_2, S \in X_T$ in the following coupled fashion such that $(S_1, S_2) \sim X_T^2$ while $(S_1, S), (S_2, S) \sim \nu_n(k,t)$. First choose $S_1, S_2 \in X_T$ at random, then choose $i_1 \in S_1 \setminus T$ and $i_2 \in S_2 \setminus T$ at random, and choose $S \in X_T$ at random such that $S_1 \cap S = T + i_1$ and $S_2 \cap S = T + i_2$. We now have $(S_1, S), (S_2, S) \sim \nu_n(k,t)$. Clearly, if $f_{S_1}|_T \neq f_{S_2}|_T$, then either $f_{S_1}|_T \neq f_S|_T$ or $f_{S_2}|_T \neq f_S|_T$. Hence, $\kappa(\emptyset) \leq 2\varepsilon_T(\emptyset)$. ☐

Claim 2.5 (Bounding $\gamma(i)$)**.** *If $3k - 2t \leq n$, then $\gamma(i) \leq \kappa(i) \leq 2\varepsilon_T(i)/\rho(i)^3$.*

Proof. The proof of this claim proceeds similar to the proof of the previous claim. By definition, we have $\kappa(i) = \mathbb{E}_{S_1\in X^{(0)}_{T,i}}\left[\Pr_{S_2\in X^{(0)}_{T,i}}[f_{S_T}(i) \neq f_{S_2}(i)]\right] \geq \gamma(i)$ since $g(i)$ is the most popular value among $f_S(i)$ for $S \in X^{(0)}_{T,i}$. We then observe that

$$\kappa(i) = \Pr_{S_1,S_2\in X_{T,i}}[f_{S_1}(i) \neq f_{S_2}(i) \mid S_1, S_2 \in X^{(0)}_{T,i}] = \frac{\Pr_{S_1,S_2\in X_{T,i}}[S_1, S_2 \in X^{(0)}_{T,i} \text{ and } f_{S_1}(i) \neq f_{S_2}(i)]}{\rho(i)^2}$$

We now choose S_1, S_2, S in a coupled fashion as follows. Let $\mathbf{B}$ be the distribution of $|S_1 \cap S_2|$ when S_1, S_2 are chosen at random from $X_{T,i}$. First choose $S \in X^{(0)}_{T,i}$ at random. Then choose $B \sim \mathbf{B}$, so $B \geq t$. Choose disjoint sets I, I_1, I_2 disjoint from S of sizes $B - t, k - B, k - B$ respectively, and let $S_j = I_j \cup I \cup\cup T\{i\}$ for $j \in \{1, 2\}$. Here, we have used the fact that $3k - B - t \leq n$. The joint distribution (S_1, S_2, S) satisfy that $(S_1, S_2) \sim X_{T,i} \times X_{T,i}$ and $(S_j, S) \sim \nu_n(k,t)$ conditioned on $S_j \in X_{T,i}$ and $S \in X^{(0)}_{T,i}$. Furthermore, if $S_1, S_2 \in X^{(0)}_{T,i}$ (i.e., $f_{S_1}|_T = f_{S_2}|_T = g|_T$) and $f_{S_1}(i) \neq f_{S_2}(i)$ then one of the following must hold:

1. $f_{S_1}|_T = f_S|_T$ and $f_{S_1}(i) \neq f_S(i)$, or
2. $f_{S_2}|_T = f_S|_T$ and $f_{S_2}(i) \neq f_S(i)$.

(The first parts always hold, and the second parts cannot both not hold.) This shows that $\kappa(i)$ is bounded above by

$$\begin{aligned}
\kappa(i) &\leq \frac{2}{\rho(i)^2} \cdot \Pr_{\substack{S_1\in X_{T,i} \\ S\in X^{(0)}_{T,i} \\ S_1\cap S=\{i\}}}[f_{S_1}|_T = f_S|_T \text{ and } f_{S_1}(i) \neq f_S(i)] \\
&\leq \frac{2}{\rho(i)^3} \cdot \Pr_{\substack{S_1,S\in X_{T,i} \\ S_1\cap S=\{i\}}}[f_{S_1}|_T = f_S|_T \text{ and } f_{S_1}(i) \neq f_S(i)] = \frac{2\varepsilon_T(i)}{\rho(i)^3}.
\end{aligned}$$

☐

Claim 2.6. *If $8k \le n$ and $\varepsilon_T(\emptyset) \le \frac{1}{128}$, then $\Pr_{i \notin T}\left[\rho(i) \le \frac{1}{2}\right] \le O(\varepsilon_T(\emptyset)/k')$.*

Proof. This follows from an application of Lemma 2.2 by setting $c = \frac{1}{2}$ and $B := \{i \notin T \mid \rho(i) \le \frac{1}{2}\}$. Then, either $\gamma(\emptyset) \ge 1/64$ or $\Pr[i \in B] \le 4\gamma(\emptyset)/k' \le 8\varepsilon_T(\emptyset)/k'$. $\square$

We now return to bounding $\Pr[f_S \ne g|_{S \cup T}]$ from (1) as follows:

Claim 2.7. *If $n \ge 8k$ and $\varepsilon_T(\emptyset) \le \frac{1}{128}$, then $\Pr_{S,T\colon S \supset T}[f_S \ne g_T|_S] = O(\varepsilon_T(\emptyset + k' \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)])$.*

Proof.

$$\begin{aligned}
\Pr[f_S \ne g_T|_S] &\le \Pr_{S \in X_T}[f_S|_T \ne g|_T] + \frac{k'}{n'} \sum_{i \notin T} \Pr_{S \in X_{T,i}}[S \in X_{T,i}^{(0)}] \cdot \Pr_{S \in X_{T,i}^{(0)}}[f_S(i) \ne g(i)] \\
&= \gamma(\emptyset) + \frac{k'}{n'} \sum_{i \notin T, \rho(i) \le 1/2} 1 + \frac{k'}{n'} \sum_{i \notin T, \rho(i) > 1/2} \rho(i) \cdot \gamma(i) \\
&\le 2\varepsilon_T(\emptyset) + 8\varepsilon_T(\emptyset) + \frac{k'}{n'} \sum_{i \notin T, \rho(i) > 1/2} \frac{2\varepsilon_T(i)}{\rho(i)^2} = O\left(\varepsilon_T(\emptyset) + k' \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right).
\end{aligned}$$ $\square$

We now complete the proof of the main lemma of this section.

Proof of Lemma 2.3. By Lemma 2.1, we have $\mathbb{E}_T[\varepsilon_T(\emptyset)] \le \varepsilon$. Hence, $\Pr_T[\varepsilon_T(\emptyset) \le \frac{1}{128}] = 1 - O(\varepsilon)$. We call such a T typical. For non-typical T, we define g_T arbitrarily (this happens with probability at most $O(\varepsilon)$). For every typical T, we have from the global function g_T satisfies

$$\Pr_{S \in X_T}[f_S \ne g_T|_S] = O\left(\varepsilon_T(\emptyset) + (k - (t-1)) \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right).$$

If $t \ge k\alpha$, the right hand side of the above inequality can be further bounded (using Lemma 2.1) as $O\left(\varepsilon_T(\emptyset) + (k - (t-1)) \cdot \mathbb{E}_{i \notin T}[\varepsilon_T(i)]\right) = O(\varepsilon + k \cdot \varepsilon/t) = O_\alpha(\varepsilon)$. This completes the proof of Lemma 2.3. $\square$

2.3 Step 3: Obtaining a single global function

In the final step, we show that the global function g_T corresponding to a random typical T explains most local functions f_S corresponding to S's not necessarily containing T. We will first prove this under the assumption that $k - 2(t-1) = \Omega(k)$. For concreteness, let us assume $t \le k/3$. We will then show to extend it to any t satisfying $k - t \ge \beta k$.

Suppose we choose two $(t-1)$-sets T_1, T_2 at random, and a k-set S containing $T_1 \cup T_2$ at random (here we use $2(t-1) \le k$). Then,

$$\Pr[g_{T_1}|_S \ne g_{T_2}|_S] = O(\varepsilon).$$

This prompts defining

$$\delta_{T_1,T_2} := \Pr_{S \supseteq T_1 \cup T_2}[g_{T_1}|_S \ne g_{T_2}|_S],$$

so that $\mathbb{E}[\delta_{T_1,T_2}] = O(\varepsilon)$.

If g_{T_1}, g_{T_2} disagree on $T_1 \cup T_2$ then $\delta_{T_1,T_2} = 1$, which happens with probability at most $O(\varepsilon)$. Assume this is not the case. Denote by B the set of points of $\overline{T_1 \cup T_2}$ on which g_{T_1}, g_{T_2} disagree, and let $n' = n - |T_1 \cup T_2| = \Theta(n)$, $k' = k - |T_1 \cup T_2| = \Theta(k)$. Applying Lemma 2.2 (with $c = 1$) shows that unless $\delta_{T_1,T_2} > 1/8$ (which happens with probability at most $O(\varepsilon)$), we have $|B|/n' = O(\delta_{T_1,T_2}/k')$, and so $|B|/n = O(\delta_{T_1,T_2}/k)$. This shows that if $\delta_{T_1,T_2} \le 1/8$ then

$$\Pr_{i \in [n]}[g_{T_1}(i) \ne g_{T_2}(i) \mid \delta_{T_1,T_2} \le 1/8] \le O(\delta_{T_1,T_2}/k).$$

Choose a random $S \in \binom{[n]}{k}$ containing a random T_2 (but not necessarily T_1). Then

$$\begin{aligned}
\mathbb{E}_{T_1}\left[\Pr_{T_2,S\colon S\supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S]\right] &= \Pr_{T_1,T_2,S\colon S\supset T_2}[g_{T_1}|_S \neq g_{T_2}|_S] \\
&\leq \Pr[\delta_{T_1,T_2} > 1/8] + \Pr[\exists i, i \in S \text{ and } g_{T_1}(i) \neq g_{T_2}(i) \mid \delta_{T_1,T_2} \leq 1/8] \\
&= O(\varepsilon) + n \cdot \frac{(k-(t-1))}{(n-(t-1))} \cdot \frac{O(\mathbb{E}[\delta_{T_1,T_2} \mid \delta_{T_1,T_2} \leq 1/8])}{k} \\
&= O\left(\varepsilon + \frac{\mathbb{E}[\delta_{T_1,T_2}]}{\Pr[\delta_{T_1,T_2} \leq 1/8]}\right) = O(\varepsilon).
\end{aligned}$$

Choose a set T_1 such that the above probability holds with respect to T_2, S, and define $F = g_{T_1}$. Then

$$\Pr[f_S \neq F|_S] \leq \Pr[f_S \neq g_{T_2}|_S] + \Pr[g_{T_1}|_S \neq g_{T_2}|_S] = O(\varepsilon).$$

We have proved the following lemma.

Lemma 2.8. *For all $\alpha \in (0, 1/3)$ if $n \geq 4k$ and $\alpha k \leq t \leq k/3$, there exists a function $F\colon [n] \to \Sigma$ such that $\Pr[f_S \neq F|_S] = O_\alpha(\varepsilon)$.*

Proof of Theorem 1.1. Consider the following coupling argument. Let $S_1, S_2 \sim \nu_n(k, t')$. Let S be a random set of size k containing $S_1 \cap S_2$ as well as $t - t'$ random elements from S_1, S_2 each and the rest of the elements chosen from $\overline{S_1 \cup S_2}$. This can be done as long as $k \geq 2(t-t') + t' = 2t - t'$. Clearly, $(S, S_j) \sim \nu_n(k, t)$ for $j = 1, 2$. Furthermore,

$$\Pr[f_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] \leq \Pr[f_{S_1}|_{S_1\cap S} \neq f_S|_{S_1\cap S}] + \Pr[f_{S_2}|_{S_2\cap S} \neq f_S|_{S_2\cap S}] \leq 2\varepsilon.$$

This demonstrates that if the hypothesis for the agreement theorem is true for a particular choice of n, k, t, then the hypothesis is also true for n, k, t' by increasing ε to 2ε provided $k - t \geq (k-t')/2$. Thus, given the hypothesis is true for some t satisfying $k - t \geq \beta k$, we can perform the above coupling argument a constant number of times to to reduce t to less than $k/3$ and then conclude using Lemma 2.8. □

3 Agreement theorem for high dimensions

Theorem 3.1 (Agreement theorem)**.** *There exists constants $C > 1$ such that for all positive integers d and $\alpha, \beta \in (0,1)$ satisfying $\alpha + \beta \leq 1$ and all positive integers n, k, t satisfying $n \geq Ck$ and $t \geq \max\{\alpha k, d\}$ and $k - t \geq \max\{\beta k, d\}$ and alphabet Σ the following holds: Let $\{f_S\colon \binom{S}{\leq d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of functions satisfying*

$$\Pr_{\substack{S_1,S_2\in\binom{[n]}{k}\\ |S_1\cap S_2|=t}}[f_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] \leq \varepsilon,$$

then there exists a function $F\colon \binom{[n]}{\leq d} \to \Sigma$ satisfying $\Pr_{S\in\binom{[n]}{k}}[f_S \neq F|_S] = O_{\alpha,\beta,d}(\varepsilon)$. Here, $F|_S$ refers to the restriction $F|_{\binom{S}{\leq d}}$.

As before, we let $\nu_n(k,t)$ denote the distribution induced on the pair of sets $(S_1, S_2) \in \binom{[n]}{k}^2$ by first choosing uniformly at random a set $U \subset [n]$ of size t and then two sets S_1 and S_2 of size k of $[n]$ uniformly at random conditioned on $S_1 \cap S_2 = U$. The proof of this theorem proceeds similar to the dimension one setting in three steps. In the first step (Section 3.1), we prove some preliminary lemmas which help in bounding the error of a "typical" subset T of $[n]$ of size $t - d$. In the second step (Section 3.2), we define for each $T \subset [n]$ of size $t - d$, a "global" function $g_T\colon \binom{[n]}{\leq d} \to \Sigma$ such that when we pick a random pair $T \subset S$ where $|T| = t - d$ and $|S| = k$, then $\Pr_{T,S\colon T\subset S}[g_T|_S = f_S] = O(\varepsilon)$. In other words, for a random $T \subset S$, the global function explains the local function. Finally, in step (Section 3.3), we argue that a random "global" function g_T explains most "local" functions f_S corresponding to S (not necessarily ones that contain T).

First for some notation. Let $n' := n - (t-d)$ and $k' := k - (t-d)$. For any set $T \subset [n]$ of size $t - d$, we let $\overline{T} := [n] \setminus T$. Let $X_T := \{S \in \binom{[n]}{k} \mid S \supset T\}$. For $A \subset \overline{T}$, $|A| = i \leq d$, we define $X_{T,A} := X_{T\cup A} = \{S \in \binom{[n]}{k} \mid S \supset T \cup A\}$.

For $i = -1, 0, \dots, d$, Define $T^{(i)} := \{U \in \binom{[n]}{\le d} \mid |U \setminus T| \le i\}$. Clearly, $\emptyset = T^{(-1)} \subset \binom{T}{\le d} = T^{(0)} \subset T^{(1)} \subset \dots \subset T^{(d-1)} \subset T^{(d)} = \binom{[n]}{\le d}$. For $A \subset \overline{T}$ and $|A| = i$, define $T^{(A)} := \{U \in \binom{[n]}{\le d} \mid U \setminus T \subset A\} = \binom{T \cup A}{\le d}$. Clearly, $T^{(i)} = \bigcup_{A \in \binom{\overline{T}}{i}} T^{(A)}$. For $S \in X_{(A)}$, let $f_S|_{T,A}$ denote the restriction $f_S|_{T^{(A)} \cap \binom{S}{\le d}}$. Similarly, $f_S|_{T,i} := f_S|_{T^{(i)} \cap \binom{S}{\le d}}$. Note that $f_S|_{T,i}$ refers to the restriction of f_S to the set of all subsets of size at most d which have at most i elements outside T. Given two local functions f_{S_1} and f_{S_2}, we say that they agree (denoted by $f_{S_1} \sim f_{S_2}$) if they agree on the intersection of their domains (ie., $f_{S_1}(a) = f_{S_2}(a)$ for all $a \in \binom{S_1 \cap S_2}{\le d}$). Similarly, we say that two restrictions $f_{S_1}|_{T,i}$ and $f_{S_2}|_{T,i}$ agree (denoted by $f_{S_1}|_{T,i} \sim f_{S_2}|_{T,i}$) if $f_{S_1}(a) = f_{S_2}(a)$ for all $a \in \binom{S_1 \cap S_2}{\le d} \cap T^{(i)}$.

3.1 Step 1: some preliminary lemmas

Lemma 3.2. *For all $0 \le i \le d$,*

$$\Pr_{\substack{S_1, S_2 \sim \nu_n(k,t) \\ T \subseteq S_1 \cap S_2, |T| = t-d}} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1} \not\sim f_{S_2}] = O_{d,\alpha}(k^{-i}\varepsilon).$$

Proof. We can rewrite the above probability as

$$\Pr_{S_1, S_2 \sim \nu_n(k,t)} [f_{S_1} \not\sim f_{S_2}] \cdot \mathbb{E}_{\substack{S_1, S_2 \sim \nu_n(k,t) \\ f_{S_1} \not\sim f_{S_2}}} \left[\Pr_{T \subseteq S_1 \cap S_2, |T| = t-d} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1}] \right].$$

The first factor is clearly at most ε. Now consider any S_1, S_2 of size k intersecting at a set of size t such that $f_{S_1} \not\sim f_{S_2}$, say $f_{S_1}(A) \ne f_{S_2}(A)$ for some $A \subseteq S_1 \cap S_2$. Hence, if f_{S_1} and f_{S_2} agree on all sets in $T^{(i-1)} \cap \binom{S_1 \cap S_2}{\le d}$, it must be the case that $|A \setminus T| \ge i$. Hence,

$$\Pr_{T \subseteq S_1 \cap S_2, |T| = t-d} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1}] \le \Pr_{T \subseteq S_1 \cap S_2, |T| = t-d} [|A \setminus T| \ge i].$$

Let $U = S_1 \cap S_2$. We can estimate the probability on the right by

$$\Pr_{T \subseteq U, |T| = t-d} [|A \setminus T| \ge i] \le \sum_{B \subseteq A, |B| = i} \Pr_{T \subseteq U, |T| = t-d} [U \setminus T \supseteq B] = \binom{d}{i} \frac{d(d-1)\cdots(d-i+1)}{t(t-1)\cdots(t-i+1)} = O_d(t^{-i}) = O_{d,\alpha}(k^{-i}),$$

wherein the last step we have used the fact $t \ge \alpha k$. □

We deduce the following corollaries.

Corollary 3.3. *Let $|T| = t - d$ and $|A| = i \le d$ be disjoint sets. Define*

$$\varepsilon_{T,A} := \Pr_{\substack{S_1, S_2 \sim \nu(k,t) \\ S_1 \cap S_2 \supseteq T \cup A}} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}].$$

Then $\mathbb{E}_{T,A}[\varepsilon_{T,A}] = O(k^{-i}\varepsilon)$ where the expectation it taken over T and A such that $|T| = t - d, |A| = i$ and $T \cap A = \emptyset$.

Proof. This follows from the simple observation that

$$\begin{aligned}
\mathbb{E}_{T,A}[\varepsilon_{T,A}] &= \mathbb{E}_{T,A} \left[\Pr_{\substack{S_1, S_2 \sim \nu(k,t) \\ S_1 \cap S_2 \supseteq T \cup A}} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}] \right] \\
&\le \mathbb{E}_{T,A} \left[\Pr_{\substack{S_1, S_2 \sim \nu(k,t) \\ S_1 \cap S_2 \supseteq T \cup A}} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1} \not\sim f_{S_2}] \right] \\
&= \Pr_{\substack{S_1, S_2 \sim \nu_n(k,t) \\ T \subseteq S_1 \cap S_2, |T| = t-d}} [f_{S_1}|_{T,i-1} \sim f_{S_2}|_{T,i-1} \text{ and } f_{S_1} \not\sim f_{S_2}] \\
&= O(k^{-i}\varepsilon).
\end{aligned}$$

□

Corollary 3.4. *Let $|T| = k - d$ and let $0 \le i \le d$. Define $\varepsilon_{T,i} := \mathbb{E}_{A \subset \overline{T}, |A|=i}[\varepsilon_{T,A}]$. Then $\mathbb{E}_T[\varepsilon_{T,i}] = O(k^{-i}\varepsilon)$.*

We also need the following lemma (which in some sense is the generalization of Lemma 2.2 to general d). However the proof of this lemma is far more elaborate and requires the hypergraph pruning lemma (Lemma 1.4 proved in Section 4).

Lemma 3.5. *Fix $d \ge 1$ and $c > 0$. There exists $p_0 > 0$ (depending on c, d) such that the following holds for every $n \ge k \ge 2d$ satisfying $k/n \le p_0$.*

Let F be a d-uniform hypergraph, and for each $A \in F$, let $Y_A \subseteq X_A = \{S \in \binom{[n]}{k} \mid S \supseteq A\}$ have density at least c in X_A. Then

$$\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A \in F} X_A\right] = O_{c,d}\left(\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A\in F} Y_A\right]\right).$$

Proof. Let $\varepsilon = c/2$, and apply the uniform hypergraph pruning lemma (Lemma 1.4) setting $H := F$ to get a subhypergraph F' of F. For every $A \in F'$,

$$\Pr_{S\colon |S|=k}[S \in Y_A \text{ and } F'|_S = \{A\} \mid S \in X_A] \ge c - \Pr_{S\colon |S|=k}[F'|_S \ne \{A\} \mid S \in X_A] \ge c - \varepsilon = c/2.$$

Summing over all $A \in F'$, we get

$$\Pr_{S\colon |S|=k}\left[S \in \bigcup_{A\in F} Y_A\right] \ge \sum_{A \in F'} \Pr_{S\colon |S|=k}[S \in Y_A \text{ and } F'|_S = \{A\}] \ge$$
$$\frac{c}{2}\sum_{A\in F'} \Pr_{S\colon |S|=k}[S \in X_A] \ge \frac{c}{2}\Pr_{S\colon |S|=k}[F'|_S \ne \emptyset] = \Omega_{c,d}\left(\Pr_{S\colon |S|=k}[F|_S \ne \emptyset]\right).$$

This completes the proof since the right-hand side is exactly the left-hand side of the statement of the lemma. ☐

3.2 Step 2: Constructing a global function for a typical T

We prove the following lemma in this section.

Lemma 3.6. *For all $\alpha, \beta \in (0,1)$ and positive integers d, there exists a constant $p \in (0,1)$ such that for all positive integers n, k, t satisfying $k \le pn$, $t \ge \max\{\alpha k, d\}$, $k - t \ge \max\{\beta k, d\}$ and alphabet Σ the following holds: Let $\{f_S\colon \binom{S}{\le d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ be an ensemble of local functions satisfying*

$$\Pr_{\substack{S_1,S_2 \in \binom{[n]}{k}\\ |S_1\cap S_2| = t}}[f_{S_1}|_{S_1\cap S_2} \ne f_{S_2}|_{S_1\cap S_2}] \le \varepsilon,$$

then there exists an ensemble $\{g_T\colon \binom{[n]}{\le d} \to \Sigma \mid T \in \binom{[n]}{t-d}\}$ of global functions such when a random $T \in \binom{[n]}{t-d}$ and $S \in \binom{[n]}{k}$ are chosen such that $S \supset T$, then $\Pr[g_T|_S \ne f_S] = O_{\alpha,\beta,d}(\varepsilon)$.

We now define the "global" function $g_T\colon \binom{[n]}{\le d} \to \Sigma$. We will drop the subscript T for ease of notation. We will define g incrementally by first defining $g|_{T^{(-1)}}$ (the empty function) and then inductively extending the definition of g from the domain $T^{(i-1)}$ to $T^{(i)}$ (recall that $T^{(-1)} \subset T^{(0)} \subset \cdots \subset T^{(d)} = \binom{[n]}{\le d}$). To begin with set $X^{(-1)} := X_T$ and $\delta_{-1} := 1 - \frac{|X^{(-1)}|}{|X_T|} = 0$. Let $g\colon T^{(-1)} \to \Sigma$ be the empty function. For $i := 0 \ldots d$ do, we inductively extend the definition of g from $T^{(i-1)}$ to $T^{(i)}$ as follows. If $\delta_{i-1} > \frac{1}{2}$, set $g := \perp$ and exit. For each $A \in \overline{T}, |A| = i$, let

$$X^{(i-1)}_{(A)} := \{S \in X^{(i-1)} \mid S \supset A\},$$

and g_A be the most popular $f_S|_{T,A}$ among $S \in X_{(A)}^{(i-1)}$ (breaking ties arbitrarily). Let $\gamma(A)$ denote the probability that a random value in $X_{(A)}^{(i-1)}$ is not the popular value, more precisely

$$\gamma(A) := \Pr_{S \in X_{(A)}^{(i-1)}} [f_S|_{T,A} \neq g_A],$$

and $\rho(A) := \frac{|X_{(A)}^{(i-1)}|}{|X_{(A)}|}$. Note that $g_A \colon \binom{T \cup A}{\leq d} \to \Sigma$ and g_A agrees with g on the domain $T^{(i-1)}$ (ie., the domain where it has been defined so far). We now extend g from $T^{(i-1)}$ to $T^{(i)}$ as follows: for each $B \in T^{(i)} \setminus T^{(i-1)}$, let A be the unique subset in $\binom{\overline{T}}{i}$ such that $B = B' \cup A$ for some $B' \in T$. Set $g(B) := g_A(B)$. Set

$$X^{(i)} := \left\{ S \in X^{(i-1)} \mid \forall A \subset S \setminus T, |A| = i, f_S|_{T,A} = g|_{T^{(A)} \cap \binom{S}{\leq d}} \right\},$$

and $\delta_i := 1 - \frac{|X^{(i)}|}{|X_T|}$ before proceeding to the next i. Thus, $X^{(i)}$ refers to the set of S's where the global function $g \colon T^{(i)} \to \Sigma$ agrees with local functions f_S and δ_i is the density of those S's that disagree with the global function.

We would like to bound the probability that the global function g defined above agrees with local functions, namely $\Pr_{S \colon S \supset T} [g_T|_S \neq f_S]$. Note that this probability is upper bounded by the probability δ_d. We now inductively bound $\delta_i, i = 0, \dots, d$. First we need the following claims on $\gamma(A)$ and $\rho(A)$.

Claim 3.7 (Estimating $\gamma(A)$)**.** *If $t + d \leq k$ and $3k \leq n$, then $\gamma(A) \leq 2\varepsilon_{T,A}/\rho(A)^3$.*

Proof. By definition, we have $\gamma(A) = \min_\alpha \Pr_{S \in X_{(A)}^{(i-1)}} [f_S|_{T,A} \neq \alpha]$. Hence, we have

$$\gamma(A) \leq \Pr_{S_1, S_2 \in X_{(A)}^{(i-1)}} [f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}] \leq \frac{1}{\rho(A)^2} \cdot \Pr_{S_1, S_2 \in X_{(A)}} \left[S_1, S_2 \in X_{(A)}^{(i-1)} \text{ and } f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A} \right].$$

Let $\mathbf{M}$ be the distribution of $|S_1 \cap S_2|$ when S_1, S_2 are chosen at random from $X_{(A)}$. Choose $S \in X_{(A)}^{(i-1)}$ at random, and draw $m \sim \mathbf{M}$ (so $m \geq t - d + i$). Choose two disjoint subsets R_1, R_2 of $S \setminus (T \cup A)$ of size $d - i$, two disjoint subsets I_1, I_2 of $\overline{S}$ of size $k - m - d + i$, and a subset I disjoint from I_1, I_2, S of size $m - i - t + d$; this is possible since $t + d \leq k$ and $3k \leq n$. Let $S_j = A \cup R_j \cup I_j \cup I \cup T$ (which have size $i + (d - i) + (k - m - d + i) + (m - i - t + d) + (t - d) = k$, so that $S_1 \cap S_2 = A \cup I \cup T$ has size $i + (m - i - t + d) + (t - d) = m$ and $S_j \cap S = A \cup R_j \cup T$ have size $i + (d - i) + (t - d) = t$. The joint distribution (S_1, S_2, S) satisfy that $(S_1, S_2) \sim X_{(A)} \times X_{(A)}$ and $(S_j, S) \sim \nu_n(k, t)$ conditioned on $S_j \in X_{(A)}$ and $S \in X_{(A)}^{(i-1)}$. Furthermore, if $f_{S_1}|_{T,A} \not\sim f_{S_2}|_{T,A}$ and $S_1, S_2 \in X_{(A)}^{(i-1)}$ (i.e., for all $A_1 \in S_1 \setminus T$ of size i, $f_{S_1}|_{T,A_1} = g|_{T^{(A_1)} \cap \binom{S}{\leq d}}$ and for all $A_2 \in S_2 \setminus T$ of size i, $f_{S_2}|_{T,A} = g|_{T^{(A)} \cap \binom{S}{\leq d}}$), then one of the following must hold:

1. $f_{S_1}|_{T,i} \sim f_S|_{T,i}$ and $f_{S_1}|_{T,A} \not\sim f_S|_{T,A}$, or
2. $f_{S_2}|_{T,i} \sim f_S|_{T,i}$ and $f_{S_2}|_{T,A} \not\sim f_S|_{T,A}$.

Hence,

$$\gamma(A) \leq \frac{2}{\rho(A)^2} \cdot \Pr_{\substack{S_1 \in X_{(A)} \\ S \in X_{(A)}^{(i-1)} \\ |S_1 \cap S| = t}} [f_{S_1}|_{T,i} \sim f_S|_{T,i} \text{ and } f_{S_1}|_{T,A} \not\sim f_S|_{T,A}] \leq$$

$$\frac{2}{\rho(A)^3} \cdot \Pr_{\substack{S_1, S \in X_{(A)} \\ |S_1 \cap S| = t}} [f_{S_1}|_{T,i} \sim f_S|_{T,i} \text{ and } f_{S_1}|_{T,A} \not\sim f_S|_{T,A}] \leq \frac{2\varepsilon_{T,A}}{\rho(A)^3}. \qquad \square$$

Claim 3.8 (Estimating $\rho(A)$)**.** *If $k \geq t+d$ and $k \leq p_0 n$, then* $\Pr_{S \in X_T} \left[\exists A \subset \overline{T}, |A| = i, S \supset A, \rho(A) < \frac{1}{2} \right] = O(\delta_{i-1})$.

Proof. Let $F = \{|A| = i \mid \rho(A) \leq 1/2\}$. Define $Y_{(A)} = \{S \in X_{(A)} \mid S \notin X^{(i-1)}_{(A)}\}$. If $A \in F$ then $|Y_{(A)}|/|X_{(A)}| = 1 - \rho(A) \geq 1/2$. Then applying Lemma 3.5 (setting $d = d, c = 1/2, n = n - (t-d), k = k - (t-d)$, we have

$$\Pr_{S \in X_T}[S \supseteq A \text{ for some } A \in F] = O(\Pr_{S \in X_T}[S \in Y_{(A)} \text{ for some } A \in F]).$$

The conditions for Lemma 3.5 require $k - (t-d) \geq 2d$ and $k - (t-d) \leq p_0(n-(t-d))$ which are satisfied if $k \geq t + d$ and $k \leq p_o n$. If $S \in Y_{(A)}$ for *any* A then $S \notin X^{(i-1)}$, and so the probability on the right is at most $\Pr_{S \in X_T}[S \notin X^{(i-1)}] = \delta_{i-1}$. Therefore

$$\Pr_{S \in X_T}\left[\rho(A) < 1/2 \text{ for some } A \in \binom{S \setminus T}{i}\right] = O(\delta_{i-1}).$$

□

Claim 3.9. *If $k - t \geq \beta k$ and $\delta_{i-1} \leq \frac{1}{2}$, then $\delta_i = O_\beta(\delta_{i-1} + k^i \varepsilon_{T,i})$.*

Proof.

$$\begin{aligned}
\delta_i &= \Pr_{S \in X_T}\left[S \notin X^{(i)}\right] = \Pr_{S \in X_T}\left[S \notin X^{(i-1)}\right] + \Pr_{S \in X_T}\left[S \in X^{(i-1)} \text{ and } S \notin X^{(i)}\right] \\
&= \delta_{i-1} + \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A \text{ and } S \in X^{(i-1)} \text{ and } f_S|_{T,A} \neq g|_{T^{(A)} \cap \binom{S}{\leq d}}\right] \\
&= \delta_{i-1} + \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A, \rho(A) < \frac{1}{2}\right] \\
&+ \Pr_{S \in X_T}\left[\exists A \in \overline{T}, |A| = i, S \supset A, \rho(A) \geq \frac{1}{2}, S \in X^{(i-1)} \text{ and } f_S|_{T,A} \neq g|_{T^{(A)} \cap \binom{S}{\leq d}}\right] \\
&= O\left(\delta_{i-1}\right) + \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \Pr_{S \in X_T}\left[S \supset A, S \in X^{(i-1)} \text{ and } f_S|_{T,A} \neq g|_{T^{(A)} \cap \binom{S}{\leq d}}\right] \quad \text{[By Claim 3.8]} \\
&= O\left(\delta_{i-1}\right) + \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \Pr_{S \in X_T}\left[S \in X_{(A)}\right] \cdot \Pr_{S \in X_{(A)}}\left[S \in X^{(i-1)}_{(A)}\right] \cdot \Pr_{S \in X^{(i-1)}_{(A)}}\left[f_S|_{T,A} \neq g|_{T^{(A)} \cap \binom{S}{\leq d}}\right] \\
&\leq O\left(\delta_{i-1}\right) + \frac{\binom{n'-i}{k'-i}}{\binom{n'}{k'}} \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \rho(A) \cdot \gamma(A) \\
&\leq O\left(\delta_{i-1}\right) + \left(\frac{k'}{n'}\right)^i \sum_{A\colon A \in \binom{\overline{T}}{i}, \rho(A) > \frac{1}{2}} \frac{2\varepsilon_{T,A}}{\rho(A)^2} \quad \text{[By Claim 3.7]} \\
&\leq O\left(\delta_{i-1}\right) + 8\left(\frac{k'}{n'}\right)^i \sum_{A\colon A \in \binom{\overline{T}}{i}} \varepsilon_{T,A} = O_\beta\left(\delta_{i-1} + k^i \varepsilon_{T,i}\right) \quad \text{[Since } k' = k - (t-d) = \Theta(k)\text{]} \\
&= O_\beta\left(\delta_{i-1} + k^i \varepsilon_{T,i}\right) \quad \text{[By Corollary 3.4].}
\end{aligned}$$

□

We are now ready to complete the proof of Lemma 3.6

Proof of Lemma 3.6. Given T, we have shown above how to construct a function g_T, given that $\delta_i \leq c_\delta$ for all i. If the latter condition fails, define g_T arbitrarily.

We have defined above a sequence $\delta_{-1} = 0, \delta_0, \ldots, \delta_d$. We have defined δ_i only given $\delta_{i-1} \leq \frac{1}{2}$. If $\delta_{i-1} > \frac{1}{2}$, we define $\delta_i = 1$. Note that $\Pr[f_S \neq g_T|_S] \leq \delta_d$.

We have shown above that if $\delta_{i-1} \leq \frac{1}{2}$ then $\delta_i = O(\delta_{i-1} + k^i \varepsilon_{T,i})$. It is always the case that $\delta_i = O(\delta_{i-1} + k^i \varepsilon_{T,i}) + 1_{\delta_{i-1} > \frac{1}{2}}$. We now prove by induction on i that $\mathbb{E}_T[\delta_i] = O(\varepsilon)$. This clearly holds when $i = -1$. Assuming that it holds for $i-1$, for i we get

$$\mathbb{E}[\delta_i] = O(\mathbb{E}[\delta_{i-1}] + k^i \mathbb{E}[\varepsilon_{T,i}]) + \Pr[\delta_{i-1} > \frac{1}{2}] = O(\varepsilon).$$

We conclude that $\Pr_{T,S}[g_T|S \neq f_S] \leq \mathbb{E}[\delta_d] = O(\varepsilon)$. □

3.3 Step 3: Obtaining a single global function

Given the set of local functions $\{f_S\}_{S\in\binom{[n]}{k}}$, we constructed a set of global functions $\{g_T\}_{T\in\binom{[n]}{t-d}}$ such that for most pairs $S\supset T$, the global function g_T agrees with the local function f_S (Lemma 3.6). In this step, we conclude that a random global function g_T agrees with most local functions f_S (not necessarily S's that contain T).

We will first prove this under the assumption that $k-2(t-1)=\Omega(k)$. For concreteness, let us assume $t\le k/3$. We will then show to extend it to any t satisfying $k-t\ge\beta k$. To begin with, we observe that Lemma 3.6 immediately implies the following claim.

Claim 3.10. *For T_1,T_2 of size $t-d$, define $\delta_{T_1,T_2}:=\Pr_{S\supseteq T_1\cup T_2}[g_{T_1}|_S\neq g_{T_2}|_S]$. Then $\mathbb{E}_{T_1,T_2}[\delta_{T_1,T_2}]=O(\varepsilon)$.*

We now move to more general S in the following sense: S contains T_2 but not necessarily T_1.

Claim 3.11. *For all T_1,T_2, $\Pr_{|S|=k,S\supseteq T_2}[g_{T_1}|_S\neq g_{T_2}|_S]=O(\delta_{T_1,T_2})$.*

Proof. We will prove this be choosing $L=O_d(1)$ collection of k-sets $(S,S_1,\ldots,S_L)$ in a coupled fashion such that each S is a random k-set containing T_1 and for each $j\ge 1$, S_j is a random k-set containing $T_1\cup T2$ with the additional property that $\binom{S}{\le d}\subseteq\bigcup_{j\ge1}\binom{S}{\le d}$. Given such a distribution, the lemma follows by a union bound.

The coupled distribution is obtained in the following fashion. Let $k-|T_1\cup T_2|\ge k/3$. We proceed to find a collection of $O(1)$ subsets $R_i\subseteq[k]$ of size at most $k/3$ such that $\binom{[k]}{d}=\bigcup_i\binom{R_i}{d}$. The idea is to split $[k]$ into $O(d)$ parts of size at most $k/(3d)$, and to take as R_i the union of any d of these. Given a random k-set $S\in\binom{[n]}{d}$ containing T_2, choose a random permutation mapping $[k]$ to S, apply it to the R_i, remove from the resulting sets any elements of $T_1\cup T_2$, and complete them to sets $\tilde{R}_i$ of size $k-|T_1\cup T_2|$ randomly and set $S_j=\tilde{R}_j\cup T_1\cup T_2$. Clearly, if S is a random k-set containing T_2, the sets S_j are individually random sets of size k containing $T_1\cup T_2$. ☐

We can now complete the proof of Theorem 3.1

Proof of Theorem 3.1. As in the dimension one setting, we first prove Theorem 3.1 if $\alpha k\le t\le k/3$ and then extend it to any t satisfying $k-t\ge\beta k$. From Claim 3.10 and Claim 3.11, we have that

$$\mathbb{E}_{T_1}\left[\Pr_{T_2,S\colon S\supset T_2}[g_{T_1}|_S\neq g_{T_2}|_S]\right]=O(\delta_{T_1,T_2})=O(\varepsilon).$$

Choose a T_1 such that the inner probability is $O(\varepsilon)$ and set $F=g_{T_1}$. We now have,

$$\begin{aligned}\Pr_S[f_S\neq F|_S]&=\Pr_{S,T_2\colon S\supset T_2}[f_S\neq F|_S]\\&\le\mathbb{E}_{T_2}\left[\Pr_{S\colon S\supset T_2}[F|_S\neq g_{T_2}|_S]\right]+\mathbb{E}_{T_2}\left[\Pr_{S\colon S\supset T_2}[f_S\neq g_{T_2}|_S]\right]=O(\varepsilon).\end{aligned}$$

This completes the proof for $t\le k/3$ (in particular to any t satisfying $k-2t=\Omega(k)$.

To extend the proof to all t satisfying $k-t=\Omega(k)$, we employ the following coupling argument as in the dimension one setting. Let $S_1,S_2\sim\nu_n(k,t')$. Let S be a random set of size k containing $S_1\cap S_2$ as well as $t-t'$ random elements from S_1,S_2 each and the rest of the elements chosen from $\overline{S_1\cup S_2}$. This can be done as long as $k\ge 2(t-t')+t'=2t-t'$. Clearly, $(S,S_j)\sim\nu_n(k,t)$ for $j=1,2$. Furthermore,

$$\Pr[f_{S_1}|_{S_1\cap S_2}\neq f_{S_2}|_{S_1\cap S_2}]\le\Pr[f_{S_1}|_{S_1\cap S}\neq f_S|_{S_1\cap S}]+\Pr[f_{S_2}|_{S_2\cap S}\neq f_S|_{S_2\cap S}]\le 2\varepsilon.$$

This demonstrates that if the hypothesis for the agreement theorem is true for a particular choice of n,k,t, then the hypothesis is also true for n,k,t' by increasing ε to 2ε provided $k-t\ge(k-t')/2$. Thus, given the hypothesis is true for some t satisfying $k-t\ge\beta k$, we can perform the above coupling argument a constant number of times to to reduce t to less than $k/3$ and then conclude using the above argument for $t\le k/3$. ☐

4 Hypergraph Pruning Lemma

We begin with a a few definitions. The number of hyperedges in a hypergraph H is denoted $|H|$. For a vertex set V, μ_p refers to the biased distribution over subsets S of V defined by choosing each $v \in V$ to be in S independently with probability p while $\nu_{n,k}$ refers to the uniform distribution over subsets S of V of size k. For a hypergraph H and a subset S of the vertices, $H|_S$ is the subhypergraph induced by the vertices in S while $H|_{S=\emptyset}$ is obtained by removing all vertices in S from all hyperedges of H. For a hypergraph H, $\iota_p(H) := \Pr_{S\sim\mu_p}[H|_S \neq \emptyset]$. And finally, we recall the definition of branching factor from the introduction. For any $\rho \geq 1$, a hypergraph H over a vertex set V is said to have *branching factor* ρ if for all subsets $A \subset V$ and integers $k \geq 0$, there are at most ρ^k hyperedges in H of cardinality $|A| + k$ containing A.

The main goal of this section is to prove the following two hypergraph pruning lemmas; one under the biased μ_p distribution and the other under the uniform $\nu_{n,k}$ distribution, which was stated in the introduction. These pruning lemma show that any hypergraph H has a subgraph H' with bounded branching factor with almost the same $\iota_p(H)$.

Lemma 4.1 (hypergraph pruning lemma (biased setting))**.** *Fix constants $c > 0$ and $d \geq 0$. There exists $p_0 > 0$ (depending on c, d) such that for every $p \in (0, p_0)$ and every d-uniform hypergraph H there exists a subhypergraph H' obtained by removing hyperedges such that*

1. *H' has branching factor c/p.*
2. *$\iota_p(H') = \Omega_{c,d}(\iota_p(H))$.*

Lemma 1.4 (Restated) (hypergraph pruning lemma (uniform setting)) *Fix constants $\varepsilon > 0$ and $d \geq 1$. There exists $p_0 > 0$ (depending on d, ε) such that for every $n \geq k \geq 2d$ satisfying $k/n \leq p_0$ and every d-uniform hypergraph H on $[n]$ there exists a subhypergraph H' obtained by removing hyperedges such that*

1. *$\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset])$.*
2. *For every $e \in H'$, $\Pr_{S\sim\nu_{n,k}}[H'|_S = \{e\} \mid S \supset e] \geq 1 - \varepsilon$.*

Here $H'|_S$ is the hypergraph induced on the vertices of S.

4.1 Proof in the μ_p biased setting

The hypergraph pruning lemma (Lemma 4.1) is proved by induction on d. The proof is divided into several steps, expressed in the following lemmata. We begin with an easy claim.

Claim 4.2. *If H has branching factor ρ then $H|_{A=\emptyset}$ has branching factor $2^{|A|}\rho$.*

Proof. It's enough to prove the theorem when $A = \{i\}$. Let B, k be given. We will show that the number of hyperedges in $H|_{i=\emptyset}$ extending B by k elements is at most $(2\rho)^k$. If $k = 0$ then this is clear. Otherwise, for each such hyperedge e, either e or $e + i$ belongs in H. The former case includes all hyperedges of H extending B by k elements, and the latter all hyperedges of H extending $B + i$ by k elements. Since H has branching factor ρ, we can upper bound the number of hyperedges by $2\rho^k \leq (2\rho)^k$. □

The first lemma identifies a "critical depth" for H.

Lemma 4.3. *For every integer d, $c > 0$ and $p \in (0, 1)$ the following holds. Let H be a d-uniform hypergraph. Then, either H has a subhypergraph H' with branching factor c/p such that $\iota_p(H') \geq \iota_p(H)/(d+1)$, or for some there $1 \leq r \leq d$, there exists a $(d - r)$-uniform hypergraph I, and a subhypergraph H' of H such that*

1. *Each hyperedge in I has at least $(c/p)^r$ extensions in H'.*
2. *For every $e \in I$ and every $A \neq \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H'.*
3. *$\iota_p(I) \geq \iota_p(H)/(d + 1)$.*

Proof. We define a sequence of graphs H_r, B_r for $0 \le r \le d$ as follows:

- $H_0 = H$ and B_0 is the empty d-uniform hypergraph.
- B_r contains all sets $|A| = d - r$ which have at least $(c/p)^r$ extensions in H_{r-1}.
- H_r contains all hyperedges in H_{r-1} which are not extensions of a set in B_r.

It's not hard to check that $\iota_p(H_r) \le \iota_p(H_{r+1}) + \iota_p(B_{r+1})$, and so

$$\iota_p(H) \le \iota_p(B_1) + \cdots + \iota_p(B_d) + \iota_p(H_d).$$

Hence one of the values on the right-hand side is at least $\iota_p(H)/(d+1)$.

The construction guarantees that for every r, every set A of size at least $d-r$ has at most $(c/p)^{d-|A|}$ extensions in H_r. In particular, H_d has branching factor c/p. This completes the proof when $\iota_p(H_d) \ge \iota_p(H)/(d+1)$. If $\iota_p(B_r) \ge \iota_p(H)/(d+1)$ for some $r \ge 1$ then we take $I = B_r$ and $H' = H_{r-1}$. The first property in the statement of the lemma follows directly from the construction of B_r, and the second follows from the guarantee stated earlier for H_{r-1} applied to $e \cup A$, which has size $d - r + |A|$ which is at least $d - (r-1)$. □

The strategy now is to apply induction on I to reduce its branching factor, and then to "complete" it to a d-uniform hypergraph. The completion is accomplished in two steps. The first step adds all hyperedges which can be associated with more than one hyperedge of the pruned I.

Lemma 4.4. *For every integer d, $c > 0$ and $p \in (0,1)$ the following holds. Let H be a d-uniform hypergraph and I a $(d-r)$-uniform hypergraph for some $1 \le r \le d$ such that*

1. *For every $e \in I$ and every $A \neq \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H.*
2. *I has branching factor c/p.*

Then the subhypergraph K of H consisting of all hyperedges of H which extend at least two hyperedges of I has branching factor $O_d(c/p)$.

Proof. Fix a set B of size $d - s$, where $s \ge 1$. We have to bound the number of extensions of B in K. Each of these extensions belongs to one of the following types:

- Type 1: Extends $e_1 \neq e_2 \in I$, where $B \not\subseteq e_1$.
- Type 2: Extends $e_1 \neq e_2 \in I$, where $B \subseteq e_1 \cap e_2$.

We consider each of these types separately.

Type 1. Let $B' = B \cap e_1$. There are at most $2^{|B|} \le 2^d$ choices for B'. Since I has branching factor c/p and $e_1 \supseteq B'$, given $B' \subseteq B$ there are at most $(c/p)^{d-r-|B'|}$ choices for e_1. By assumption, $A := B \setminus e_1$ is non-empty, and moreover $|A| = |B| - |B \cap e_1| = d - s - |B'|$. Hence the first property of I implies that $e_1 \cup B = e_1 \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{r+s-d+|B'|}$ extensions in H. In total, we have counted at most $2^d \cdot (c/p)^{d-r-|B|'} \cdot (c/p)^{r+s-d+|B'|} = 2^d (c/p)^s$ extensions.

Type 2. Since $e_1 \supseteq B$ and I has branching factor c/p, there are at most $(c/p)^{(d-r)-(d-s)} = (c/p)^{s-r}$ choices for e_1. Let $e_\cap = e_1 \cap e_2$, and note that given e_1, there are at most $2^{|e_1|} \le 2^d$ choices for $e_\cap$. Given $e_\cap$, since I has branching factor c/p, there are at most $(c/p)^{d-r-|e_\cap|}$ choices for e_2. By assumption, $A := e_2 \setminus e_1$ is non-empty, and moreover $|A| = |e_2| - |e_\cap| = d - r - |e_\cap|$. Hence the first property of I implies that $e_1 \cup e_2 = e_1 \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{2r-d+|e_\cap|}$ extensions in H. In total, we have counted at most $(c/p)^{s-r} \cdot 2^d \cdot (c/p)^{d-r-|e_\cap|} \cdot (c/p)^{2r-d+|e_\cap|} = 2^d (c/p)^s$ extensions.

Summing over both types, there are at most $2^{d+1}(c/p)^s \le (2^{d+1}c/p)^s$ extensions, completing the proof. □

The second completion step guarantees that the completion contains enough hyperedges.

Lemma 4.5. *For every integer d, $c > 0$, there exists $p_0 = p_0(c,d) \in (0,1)$ such that the following holds for all $p \in (0, p_0)$. Let H be a d-uniform hypergraph and I a $(d-r)$-uniform hypergraph for some $1 \le r \le d$ such that*

1. Each hyperedge in I has at least $(c/p)^r$ extensions in H.

2. For every $e \in I$ and every $A \neq \emptyset$ disjoint from e, $e \cup A$ has at most $(c/p)^{r-|A|}$ extensions in H.

3. I has branching factor c/p.

Then there exists a subhypergraph K of H such that

1. K contains $\Omega_d(|I|(c/p)^r)$ hyperedges.

2. K has branching factor $O_d(c/p)$.

Proof. We choose p_0 so that $\lfloor (c/p)^r \rfloor \geq (c/p)^r/2$.[1]

Let K' be the subhypergraph constructed in Lemma 4.4. Every hyperedge in $H \setminus K'$ extends at most one hyperedge of I. For every hyperedge $e \in I$, let n_e be the number of extensions of e in K', let $m_e = \max(\lfloor (c/p)^r \rfloor - n_e, 0)$, and let H_e be a set of m_e extensions of e in $H \setminus K'$. We let $K = K' \cup \bigcup_{e \in I} H_e$.

By construction, every $e \in I$ has at least $(c/p)^r/2$ extensions in K. A given hyperedge can extend at most 2^d many hyperedges of I, so K contains at least $|I|(c/p)^r/2^{d+1}$ hyperedges.

It remains to bound the branching factor of K. Fix a set B of size $d-s$, where $s \geq 1$. We will bound the number of extensions of B in $K \setminus K'$.

Let $B' = B \cap e$. There are at most $2^{|B|} \leq 2^d$ choices for B'. Since I has branching factor c/p, given B' there are at most $(c/p)^{d-r-|B'|}$ choices for e. Let $A := B \setminus e$, so that $|A| = |B| - |B \cap e| = d - s - |B'|$. If $A \neq \emptyset$ then the second property of I implies that $e \cup B = e \cup A$ has at most $(c/p)^{r-|A|} = (c/p)^{r+s-d+|B'|}$ extensions in H and so in $K \setminus K'$. If $A = \emptyset$ then we get the same conclusion by construction since $e \cup B = e$. In total, we have counted at most $2^d \cdot (c/p)^{d-r-|B'|} \cdot (c/p)^{r+s-d+|B'|} = 2^d (c/p)^s \leq (2^d c/p)^s$ extensions, completing the proof. $\square$

We will argue about the completion using the following fundamental lemma, which is also important for applications.

Lemma 4.6. *For every integer d, $c > 0$ and $\varepsilon \in (0,1)$, there exists $f(c,d,\varepsilon) \in (0,1)$ satisfying $\lim_{c \to 0} f(c,d,\varepsilon) = 1$ for every d, ε such that the following holds. Let H be a d-uniform hypergraph, and let $p \in (0, 1-\varepsilon)$. If H has branching factor c/p then for every hyperedge $e \in H$, $\Pr_{S \sim \mu_p}[H|_S = \{e\}] \geq f(c,d,\varepsilon) p^d$.*

Before proceeding to the proof of the lemma, we first recall the statement of FKG inequality.

Lemma 4.7 (FKG inequality)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotonically increasing (or decreasing) family of subsets. Then*

$$\mu_p(\mathcal{A} \cap \mathcal{B}) \geq \mu_p(\mathcal{A}) \cdot \mu_p(\mathcal{B}).$$

Proof of Lemma 4.6. Let $K := H|_{e=\emptyset} \setminus \emptyset = (H-e)|_{e=\emptyset}$. Note that $\Pr_{S \sim \mu_p}[H|_S = \{e\}] = p^d \Pr_{S \sim \mu_p}[K|_S = \emptyset]$. Claim 4.2 shows that $H|_{e=\emptyset}$ has branching factor $O_d(c/p)$. In particular, for every s it has at most $O_d(c/p)^s$ hyperedges of cardinality s. For every hyperedge $e' \in K$, let $E_{e'}$ denote the event $e' \notin K|_S$ (i.e., $S \not\supseteq e'$), where $S \sim \mu_p$. Note that

$$\Pr[E_{e'}] = 1 - p^s = \exp\left(\frac{\log(1-p^s)}{p^s} p^s\right).$$

Now $\log(1-x)/x = -1 - x/2 - \cdots$ is decreasing (its derivative is $-1/2 - 2x/3 - \cdots$), and so $p^s \leq p \leq 1 - \varepsilon$ implies that $\log(1-p^s)/p^s \geq \log \varepsilon/(1-\varepsilon)$. In other words, $\Pr[E_{e'}] \geq e^{-O_\varepsilon(p^s)}$.

Since the events $E_{e'}$ are monotone decreasing, the FKG lemma shows that they positively correlate, hence

$$\Pr_{S \sim \mu_p}[K|_S = \emptyset] \geq \prod_{s=1}^{d} (1-p^s)^{O_d(c/p)^s} \geq \prod_{s=1}^{d} e^{-O_{d,\varepsilon}(c^s)} =: f(c,d,\varepsilon).$$

The lemma follows since clearly $\lim_{c \to 0} f(c,d,\varepsilon) = 1$. $\square$

We can now complete the inductive proof of Lemma 4.1.

[1] Another possibility, which slightly affects the proof, is to choose p_0 so that $\lceil (c/p)^r \rceil \leq 2(c/p)^r$.

Proof of Lemma 4.1. The proof is by induction on d. When $d = 0$ we can take $H' = H$, so we can assume that $d \geq 1$. Let $\gamma = c/M_d$, where $M_d \geq 1$ will be chosen later. We apply Lemma 4.3 to H with $c := \gamma$. If H has a subhypergraph H' with branching factor γ/p such that $\iota_p(H') \geq \iota_p(H)/(d+1)$ then we are done, so suppose that there exists some $d-r$ uniform hypergraph I and a subhypergraph H' satisfying the properties of the lemma. Apply the induction hypothesis to construct a subhypergraph I' of I that has branching factor γ/p and satisfies $\iota_p(I') = \Omega_{\gamma,d}(\iota_p(I)) = \Omega_{\gamma,d}(\iota_p(H))$ (this requires $p \leq p_0'(\gamma, d)$). Next, apply Lemma 4.5 with $c := \gamma$, $H := H'$, and $I := I'$ (this requires $p \leq p_0''(\gamma, d)$) to obtain a subhypergraph K of H' (and so of H) satisfying

- K contains $\Omega_d(|I'|(\gamma/p)^r)$ hyperedges.
- K has branching factor $O_d(\gamma/p)$.

We choose M_d so that K has branching factor c/p, and let $p_0 = \min(p_0'(\gamma, d), p_0''(\gamma, d))$, which depends only on c, d.

We will take $H' := K$, so it remains to show that $\iota_p(K) = \Omega_{c,d}(\iota_p(H))$. Since $p \leq p_0$, Lemma 4.6 shows that for every hyperedge $e \in K$, $\Pr_{S\sim\mu_p}[K|_S = \{e\}] = \Omega_{c,d}(p^d)$. For different hyperedges these events are disjoint, hence $\iota_p(K) = \Omega_{c,d}(|K|p^d) = \Omega_{c,d}(|I'|p^{d-r})$. On the other hand, the union bound shows that $\iota_p(I') \leq |I'|p^{d-r}$, and so $\iota_p(K) = \Omega_{c,d}(\iota_p(I')) = \Omega_{c,d}(\iota_p(H))$, completing the proof. □

As a corollary, we obtain the following useful result.

Corollary 4.8. *Fix constants $\varepsilon > 0$ and $d \geq 0$. There exists $p_0 > 0$ (depending on d, ε) such that for every $p \in (0, p_0)$ and every d-uniform hypergraph H there exists a subhypergraph H' obtained by removing hyperedges such that*

1. $\iota_p(H') = \Omega_{d,\varepsilon}(\iota_p(H))$.
2. *For every $e \in H'$,* $\Pr_{S\sim\mu_p}[H'|_S = \{e\}] \geq (1-\varepsilon)p^d$.

Proof. Let $c > 0$ be a constant to be chosen later, and define $p_0 \leq 1/2$ so that the theorem applies. The theorem gives us a subhypergraph satisfying the first property. Moreover, for every $e \in H'$, Lemma 4.6 (applied with $\varepsilon := 1/2$) shows that $\Pr_{S\sim\mu_p}[H|_S = \{e\}] \geq f(c, d)p^d$, where $\lim_{c\to 0} f(c, d) = 1$. Take c so that $f(c, d) > 1 - \varepsilon$ to complete the proof. □

4.2 Proof in the uniform setting

We now use Corollary 4.8 to transfer the hypergraph pruning lemma to the uniform setting (Lemma 1.4). Recall that distribution $\nu_{n,k}$ refers to the uniform distribution over $\binom{[n]}{k}$.

Proof of Lemma 1.4. Let $p = k/n$. Notice that

$$\Pr_{S\sim\mu_p}[H|_S \neq \emptyset] \geq \sum_{\ell=k}^{n} \Pr[\mathrm{Bin}(n,p) = \ell] \Pr_{S\sim\nu_{n,\ell}}[H|_S \neq \emptyset] \geq \Pr[\mathrm{Bin}(n,p) \geq k] \Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset].$$

It is well-known that the median[2] of $\mathrm{Bin}(n, p)$ is one of $\lfloor np \rfloor, \lceil np \rceil$. Since $np = k$, we deduce that the median is k and $\Pr[\mathrm{Bin}(n, p) \geq k] \geq 1/2$. Therefore $\iota_p(H) \geq \Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/2$. Applying Corollary 4.8 with $\varepsilon := \min(\varepsilon/2, 1/2)$, we thus get a subhypergraph H' such that

$$\iota_p(H') = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]),$$

which implies that

$$|H'| = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/p^d).$$

Let now $e \in H'$ be an arbitrary hyperedge. We are given that $\Pr_{S\sim\mu_p}[H'|_S = \{e\} \mid e \in S] \geq 1 - \varepsilon/2$. For $K = H'|_{e=\emptyset} \setminus \{\emptyset\}$, the left-hand side is $\Pr_{S\sim\mu_p}[K|_S = \emptyset]$. As before, we have

$$\Pr_{S\sim\nu_{n,k}}[K|_S \neq \emptyset] \leq 2 \Pr_{S\sim\mu_p}[K|_S \neq \emptyset] \leq \varepsilon,$$

[2] The median of a distribution X on the integers is the integer m such that $\Pr[X \geq m], \Pr[X \leq m] \geq 1/2$.

and so we get the second property. For the first property, we have

$$\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] \geq \sum_{e\in H'} \Pr_{S\sim\nu_{n,k}}[H'|_S = \{e\}] \geq (1-\varepsilon)|H'|\frac{k^{\underline{d}}}{n^{\underline{d}}}.$$

By assumption $k^{\underline{d}}/n^{\underline{d}} \geq (p/2)^d$, and so

$$\Pr_{S\sim\nu_{n,k}}[H'|_S \neq \emptyset] \geq (1-\varepsilon)\cdot\Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]/p^d)\cdot(p/2)^d = \Omega_{d,\varepsilon}(\Pr_{S\sim\nu_{n,k}}[H|_S \neq \emptyset]). \qquad \square$$

5 Agreement theorem via majority decoding

A nice application of the hypergraph pruning lemma is to show that majority decoding always works for agreement testing. In particular, if the agreement theorem (Theorem 3.1) holds, then one might without loss of generality assume that the global function is the one obtained by majority/plurality decoding.

Lemma 5.1. *For every positive integer d and alphabet Σ, there exists a $p \in (0,1)$ such that for $\alpha \in (0,1)$ and all positive integers n, k, t satisfying $n \geq k \geq t \geq \max\{2d, \alpha k\}$ and $k \leq pn$ the following holds.*

Suppose an ensemble of local functions $\{f_S\colon \binom{S}{d} \to \Sigma \mid S \in \binom{[n]}{k}\}$ and a global function $F\colon \binom{[n]}{d} \to \Sigma$ satisfy

$$\Pr_{S_1,S_2\sim\nu_{n,k,t}}[f_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] = \varepsilon, \qquad \Pr_{S\sim\nu_{n,k}}[f_S \neq F|_S] = \delta.$$

Then, the global function $G\colon \binom{[n]}{d} \to \Sigma$ defined by plurality decoding (ie., $G(T)$ is the most popular value of $f_S(T)$ over all S containing T, breaking ties arbitrarily) satisfies

$$\Pr_{S\sim\nu_{n,k}}[f_S \neq G|_S] = O_{d,\alpha}(\varepsilon+\delta).$$

Proof. All probabilities below, unless specified otherwise, are over $S \sim \nu_{n,k}$.

Since $\Pr[f_S \neq G|_S] \leq \Pr[f_S \neq F|_S] + \Pr[F|_S \neq G|_S] = \delta + \Pr[F|_S \neq G|_S]$, it suffices to bound $\Pr[F|_S \neq G|_S]$. Let $H := \{T : G(T) \neq F(T)\}$, so that $\Pr[F|_S \neq G|_S] = \Pr[H|_S \neq \emptyset]$. Note that F and G are functions, while H is a hypergraph. Apply Lemma 1.4 on the hypergraph H, for a constant $\varepsilon = \eta := 1/(2|\Sigma|) > 0$, to get a subhypergraph H' ($p = p_0(d,\varepsilon)$ is chosen such that $k \leq pn$).

For any edge $T \in H'$ and $\sigma \in \Sigma$, define the following quantities

$$p(T,\sigma) := \Pr[H'|_S = \{T\} \text{ and } f_S(T) = \sigma \mid S \supseteq T], \qquad p(T) := \max_\sigma p(T,\sigma)$$
$$q(T,\sigma) := \Pr[f_S(T) = \sigma \mid S \supseteq T], \qquad q(T) := \max_\sigma q(T,\sigma)$$

Note that $G(T)$ by definition satisfies $q(T) = q(T, G(T))$. Since by the hypergraph pruning lemma, we have $\Pr[H'|_S = \{T\}|S \supset T] \geq 1-\eta$, we have $q(T,\sigma) \geq (1-\eta)\cdot p(T,\sigma)$ for all σ. Hence, $q(T,G(T)) = q(T) \geq (1-\eta)\cdot p(T)$. On the other hand for any σ, $p(T,\sigma) \geq q(T,\sigma) - \eta$. In particular, $p(T,G(T)) \geq q(T,G(T)) - \eta \geq q(T,G(T))/2$ (since $q(T,G(T)) \geq 1/|\Sigma|$ and $\eta \leq 1/(2|\Sigma|)$). Combining these, we have that for all $T \in H'$,

$$p(T,G(T)) \geq (1-\eta)\cdot p(T)/2. \tag{2}$$

We now relate the probabilities $p(T)$ and $p(T,G(T))$ to δ and ε in the lemma statement.

By the hypergraph pruning lemma, we have $\Pr[H'|_S = \{T\}|S \supset T] \geq 1-\eta$ or equivalently $\sum_\sigma p(T,\sigma) \geq 1-\eta$. For each hyperedge $T \in H'$, we have

$$\Pr_{S_1,S_2\sim\nu_{n,k}}[f_{S_1}(T) \neq f_{S_2}(T) \text{ and } H'|_{S_1} = H'|_{S_2} = \{T\} \mid S_1\cap S_2 \supseteq T] = \sum_{\sigma_1\neq\sigma_2} p(T,\sigma_1)p(T,\sigma_2)$$
$$\geq \sum_{\sigma_1} p(T,\sigma_1)(1-\eta-p(T,\sigma_1)) \geq \sum_{\sigma_1} p(T,\sigma_1)(1-\eta-p(T)) \geq (1-\eta)(1-\eta-p(T)).$$

Consider now the following coupling. Choose $S_1, S_2 \sim \nu_{n,k}$ containing T, and choose a set S intersecting each of S_1, S_2 in exactly t elements including T (this is possible since k/n is small enough). If $f_{S_1}(T) \neq f_{S_2}(T)$ then either $f_{S_1}(T) \neq f_S(T)$ or $f_{S_2}(T) \neq f_S(T)$, and so

$$(1-\eta)(1-\eta-p(T)) \leq 2\Pr_{S_1,S\sim\nu_{n,k,t}}[f_{S_1}(T) \neq f_S(T) \text{ and } H'|_{S_1} = \{T\} \mid S_1\cap S \supseteq T].$$

Summing over all edges in H', we deduce that

$$\varepsilon \geq \sum_{T \in H'} \frac{(1-\eta)(1-\eta-p(T))}{2} \Pr_{S_1,S_2 \sim \nu_{n,k,t}}[S_1 \cap S_2 \supseteq T] = \sum_{T \in H'} \frac{(1-\eta)(1-\eta-p(T))}{2} \Omega_\alpha(\Pr[S \supseteq T]), \tag{3}$$

since $t \geq \alpha k$.

We now relate δ to $p(T, H(T))$. We clearly have

$$\Pr_{S \sim \nu_{n,k}}[f_S(T) \neq F(T) \text{ and } H'|_S = \{T\} \mid S \supseteq T] \geq \Pr_{S \sim \nu_{n,k}}[f_S(T) = G(T) \text{ and } H'|_S = \{T\} \mid S \supseteq T] = p(T, G(T)).$$

Summing over all edges in H', we deduce that

$$\delta \geq \sum_{T \in F'} p(T, G(T)) \cdot \Pr[S \supseteq T]. \tag{4}$$

Either $p(T) \leq 1/3$ in which case $(1-\eta)(1-\eta-p(T))/2 = \Omega(1)$ or $p(T) \geq 1/3$ and hence $p(T, G(T)) \geq 1/6 = \Omega(1)$ (from (2)). Thus, in either case, adding (4) and (3), we have

$$\varepsilon + \delta \geq \sum_{T \in H'} \Omega_\alpha(\Pr[S \supseteq T]) = \Omega_\alpha(\Pr[H'|_S \neq \emptyset]) = \Omega_{d,\alpha}(\Pr[H|_S \neq \emptyset]).$$

We conclude that $\Pr[H|_S \neq \emptyset] = O_{d,\alpha}(\varepsilon + \delta)$, completing the proof. □

We can now combine the above lemma with the agreement theorem (Theorem 3.1) proved earlier to obtain the agreement theorem (Theorem 1.2) as stated in the introduction, with the "furthermore" clause.

Proof of Theorem 1.2. By Theorem 3.1, we have a global function $F\colon \binom{[n]}{\leq d} \to \Sigma$ (not necessarily G) satisfying

$$\Pr_{S \in \binom{[n]}{k}}[f_S \neq F|_S] = O(\varepsilon).$$

For each $i \in \{0, 1, \ldots, d\}$, let $f^{(i)}|_S := f_S|_{\binom{S}{i}}, F^{(i)} := F|_{\binom{[n]}{i}}$ and $G^{(i)} := G|_{\binom{[n]}{i}}$. Clearly, we have for each i,

$$\Pr_{S_1,S_2 \sim \nu_{n,k,t}}[f^{(i)}_{S_1}|_{S_1 \cap S_2} \neq f^{(i)}_{S_2}|_{S_1 \cap S_2}] = \varepsilon, \qquad \Pr_{S \sim \nu_{n,k}}[f^{(i)}_S \neq F^{(i)}|_S] = O(\varepsilon).$$

Hence, by Lemma 5.1, we have

$$\Pr_{S \sim \nu_{n,k}}[f^{(i)}_S \neq G^{(i)}|_S] = O(\varepsilon).$$

This implies $\Pr_{S \sim \nu_{n,k}}[f_S \neq G|_S] = d \cdot O(\varepsilon) = O_d(\varepsilon)$. □

The entire discussion in this paper so far has been with respect to the distribution $\nu_{n,k}$, the uniform distribution over k-sized subsets of $[n]$. We can extend these results to the biased setting μ_p using a trick. In this setting, the distribution $\nu_{n,k,t}$ is replaced by the distribution $\mu_{p,q}$, which is a distribution over pairs S_1, S_2 of subsets of $[n]$ defined as follows. For each element x independently, we put x only in S_1 or only in S_2 with probability $p(1-q)$ (each), and we put x in both with probability pq. This is possible if $p(2-q) \leq 1$ (we assume below $p \leq 1/2$ and hence $p(2-q) \leq 1$). Note that if sets S_1, S_2 are picked according to the distribution $\mu_{p,q}$ then the marginal distribution of each of S_1 and S_2 is μ_p.

Theorem 5.2 (Agreement theorem via majority decoding in the biased setting)**.** *For every positive integer d and alphabet Σ, there exists constants $p_0 \in (0, 1/2)$ such that for all $p \in (0, p_0)$ and $q \in (0, 1)$ and sufficiently large n the following holds: Let $\{f_S\colon \binom{S}{\leq d} \to \Sigma \mid S \in \{0,1\}^n\}$ be an ensemble of functions satisfying*

$$\Pr_{S_1,S_2 \sim \mu_{p,q}}[f_{S_1}|_{S_1 \cap S_2} \neq f_{S_2}|_{S_1 \cap S_2}] \leq \varepsilon,$$

then the global function $G\colon \binom{[n]}{\le d} \to \Sigma$ *defined by plurality decoding (ie.,* $G(T)$ *is the most popular value of* $f_S(T)$ *over all* S *containing* T*, chosen according to the distribution* $\mu_p([n])$ *ie.,* $\Pr_{S\sim\mu_p}[f_S(T) = G(T)] = \max_\sigma \Pr_{S\sim\mu}[f_S(T) = \sigma]$[3]*) satisfies*

$$\Pr_{S\sim\mu_p}[f_S \neq G|_S] = O_{d,q}(\varepsilon).$$

Proof. Let N be a large integer and define $K = \lfloor pN \rfloor$, $T = \lfloor pqN \rfloor$. For every $S \in \binom{[N]}{K}$, define $\tilde{f}_S = f_{S\cap[n]}$. In other words, for all $A \subset S \in \binom{[N]}{K}, |A| \le d$. let $\tilde{f}_S(A) = f_{S\cap[n]}(A\cap[n])$. If $S_1, S_2 \sim \nu_{N,K,T}$ then the distribution of $S_1 \cap [n], S_2 \cap [n]$ is close to $\mu_{p,q}$, and so for large enough N we have

$$\Pr_{S_1,S_2\sim\nu_{N,K,T}}[\tilde{f}_{S_1}|_{S_1\cap S_2} \neq f_{S_2}|_{S_1\cap S_2}] \le \varepsilon/2.$$

Hence, the ensemble of functions $\{\tilde{f}_S\}_{S\in\binom{[N]}{K}}$ satisfies the hypothesis of the ageement theorem (Theorem 1.2) with ε replaced by $3\varepsilon/2$. Hence, by Theorem 1.2, if we define $\tilde{G}\colon \binom{[N]}{\le d} \to \Sigma$ by plurality decoding then $\Pr_{S\sim\nu_{N,K}}[\tilde{f}_S \neq \tilde{G}|_S] = O_d(\varepsilon)$. Since $\tilde{f}_S$ depends only on $S \cap [n]$, there exists a function $\hat{G}\colon \binom{[n]}{d} \to \Sigma$ such that $\tilde{G}(T) = \hat{G}(T \cap [n])$. Moreover, for large enough N the distribution of $S \cap [n]$ approaches μ_p, and so $\hat{G} = G$.[4] This completes the proof. □

References

[ALM⁺98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. *Proof verification and the hardness of approximation problems.* J. ACM, 45(3):501–555, May 1998. (Preliminary version in *33rd FOCS*, 1992). eccc:1998/TR98-008, doi:10.1145/278298.278306. 1

[AS98] Sanjeev Arora and Shmuel Safra. *Probabilistic checking of proofs: A new characterization of NP.* J. ACM, 45(1):70–122, January 1998. (Preliminary version in *33rd FOCS*, 1992). doi:10.1145/273865.273901. 1

[DFH17] Irit Dinur, Yuval Filmus, and Prahladh Harsha. *Low degree functions that are μ_p-nearly Boolean are sparse juntas*, 2017. 1, 2, 4

[DG08] Irit Dinur and Elazar Goldenberg. *Locally testing direct product in the low error range.* In *Proc. 49th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 613–622. 2008. doi:10.1109/FOCS.2008.26. 4

[DK17] Irit Dinur and Tali Kaufman. *High dimensional expanders imply agreement expanders.* In *Proc. 58th IEEE Symp. on Foundations of Comp. Science (FOCS)*. 2017. (To appear). eccc:2017/TR17-089. 2

[DKK⁺16] Irit Dinur, Subhash Khot, Guy Kindler, Dor Minzer, and Muli Safra. *Towards a proof of the 2-to-1 games conjecture?* Technical Report 2016/TR16-198, Elect. Colloq. on Comput. Complexity (ECCC), 2016. eccc:2016/TR16-198. 3

[DS14] Irit Dinur and David Steurer. *Direct product testing.* In *Proc. 29th IEEE Conf. on Comput. Complexity*, pages 188–196. 2014. eccc:2013/TR13-179, doi:10.1109/CCC.2014.27. 1, 2, 4, 5, 6

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. *Property testing and its connection to learning and approximation.* J. ACM, 45(4):653–750, July 1998. (Preliminary version in *37th FOCS*, 1996). eccc:1996/TR96-057, doi:10.1145/285055.285060. 4

[GLR⁺91] Peter Gemmell, Richard J. Lipton, Ronitt Rubinfeld, Madhu Sudan, and Avi Wigderson. *Self-testing/correcting for polynomials and for approximate functions.* In *Proc. 23rd ACM Symp. on Theory of Computing (STOC)*, pages 32–42. 1991. doi:10.1145/103418.103429. 1

[GS00] Oded Goldreich and Shmuel Safra. *A combinatorial consistency lemma with application to proving the PCP theorem.* SIAM J. Comput., 29(4):1132–1154, 2000. (Preliminary version in *RANDOM*, 1997). eccc:1996/TR96-047, doi:10.1137/S0097539797315744. 1

[3] More formally, what we actually prove is that *some* choice of the most common value works, though we conjecture that the result holds for an arbitrary choice of the common value.

[4] There's a fine point here: there could be several most common values. Fortunately, this doesn't invalidate the proof — just choose the correct G.

[IKW12] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. *New direct-product testers and 2-query PCPs*. SIAM J. Comput., 41(6):1722–1768, 2012. (Preliminary version in *41st STOC*, 2009). eccc:2009/TR09-090, doi:10.1137/09077299X. 4

[Raz98] Ran Raz. *A parallel repetition theorem*. SIAM J. Comput., 27(3):763–803, June 1998. (Preliminary version in *27th STOC*, 1995). doi:10.1137/S0097539795280895. 2

[RS96] Ronitt Rubinfeld and Madhu Sudan. *Robust characterizations of polynomials with applications to program testing*. SIAM J. Comput., 25(2):252–271, April 1996. (Preliminary version in *23rd STOC*, 1991 and *3rd SODA*, 1992). doi:10.1137/S0097539793255151. 1