The Phase Transition for Recovering a Random Hypergraph from its Edge Data

by

Andrew Yao

S.B. Mathematics and Computer Science and Engineering, MIT, 2024

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2025

Authored by: Andrew Yao
Department of Electrical Engineering and Computer Science
April 8, 2025

Certified by: Yury Polyanskiy
Professor of Electrical Engineering and Computer Science, Thesis Supervisor

Accepted by: Katrina LaCurts
Chair, Master of Engineering Thesis Committee

The Phase Transition for Recovering a Random Hypergraph from its Edge Data

by

Andrew Yao

Submitted to the Department of Electrical Engineering and Computer Science on April 8, 2025 in partial fulfillment of the requirements for the degree of

MASTER OF ENGINEERING IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

ABSTRACT

The *weighted projection* of a hypergraph is the weighted undirected graph with the same vertex set and edge weight equal to the number of hyperedges that contain the edge; the *projection* is the unweighted graph with the same vertex set and edge set consisting of edges with weight at least one. For $d \geq 3$, after observing the unweighted and weighted projection of a random d-uniform hypergraph that is sampled using a generalization of the Erdős–Rényi random model, we study the recovery of a fraction of the hyperedges and the entire hypergraph. For both cases, we show that there is a sharp phase transition in the feasibility of recovery based on the density of the hypergraph, with recovery possible only when the hypergraph is sufficiently sparse. Particularly, we resolve numerous conjectures from [5]. Furthermore, we display an efficient algorithm that is optimal for both exact and partial recovery. We also analyze the phase transition for exact recovery by exhibiting a regime of probabilities that is below the exact recovery threshold by a polylogarithmic factor for which exact recovery is possible.

Thesis supervisor: Yury Polyanskiy
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

I would like to thank my advisor Yury Polyanskiy for his guidance while I worked towards this thesis. During discussions with Yury, I often learned about how to connect my ideas with broader themes and directions for further research. I would also like to thank my collaborators Guy Bresler and Chenghao Guo. Discussions with Guy and Chenghao significantly improved my understanding of various research topics and the writing process.

Finally, I would like to thank my family for their constant support throughout my academic journey. Their encouragement has been a great source of motivation while I have pursued various research endeavors.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A hypergraph consists of a set of vertices and a set of hyperedges, each of which is a subset of the vertex set. Storing the data of hypergraphs can be very challenging due to the potential high-dimensionality of the hyperedges. For example, if we are given that a hypergraph has $n \geq 1$ vertices, the number of potential hypergraphs is 2^{2^n-1}, assuming that the hyperedges must be nonempty. Even if we assume that each hyperedge has size d, the number of potential hypergraphs is $2^{\binom{n}{d}}$. In order to decrease the dimensionality of the dataset, we can consider applying lower-dimensional processing. Then, the question that we pose is, what fraction of the hypergraphs can we recover after applying this processing?

A method to represent the data of a hypergraph is to store its edge data. We refer to the set of edges between any two vertices both present in some hyperedge of the hypergraph as its unweighted edge data and the counts of how many hyperedges contain each edge as its weighted edge data. We also refer to the unweighted edge data as the *(unweighted) projected graph* and the weighted edge data as the *weighted projected graph*. See Section 1.1 for rigorous definitions of these terms.

Observe that we can uniquely determine the edge data of a hypergraph but we cannot always determine a hypergraph given its edge data. In this paper, we study the problem of recovery for the hyperedges of random uniform hypergraphs after storing its unweighted and weighted edge data.

The paper [5] studies the recovery of the hypergraph exactly after storing unweighted edge data and we extend upon its results. Furthermore, [15] studies recovering the communities exactly in the hypergraph stochastic block model (HSBM) after observing the hypergraph and [8, 14, 17] study recovery in the HSBM after observing the weighted edge data of the hypergraph. We study recovering the hypergraph exactly after observing the weighted edge data as well.

The paper [22] considers reconstructing a hypergraph given its projected graph, but the random hypergraph model that it considers differs significantly from the model that we consider. In the random hypergraph model studied by the paper, the hypergraph can have hyperedges of any size and there can be multiple hyperedges with the same vertex set. Furthermore, the number of hyperedges that have a given vertex set is determined by a Poisson distribution with mean determined by the size of the vertex set.

A statistical and machine learning approach is used in [20] to reconstruct a hypergraph given its projected graph. The approach considered involves first selecting cliques from the projected graph and then deciding which of these cliques are hyperedges using a machine learning approach. The paper does not restrict to uniform hypergraphs and does not consider recovering the hypergraph completely but rather optimizing the Jaccard score between the original and predicted hyperedges.

Furthermore, the paper [7] views data as a weighted hypergraph with the goal of analyzing the pretraining of foundational models. The hypergraph model that the paper studies first samples hyperedges from a weighted hypergraph with probabilities proportional to their weights. Afterwards, the goal is to use this data to construct a hypergraph that is isomorphic or approximately isomorphic to the original hypergraph, where the distance between two weighted hypergraphs is the L^1 distance between the vectors of weights assigned to their hyperedges.

A hypergraph with vertex set V has an edge set E such that each hyperedge $e \in E$ is a nonempty subset of V. For $d \geq 2$, each hyperedge of a d-uniform hypergraph contains d vertices. Then, a 2-uniform hypergraph corresponds to a graph. We consider d-uniform hypergraphs for $d \geq 3$.

The random hypergraph model that we consider is introduced in [5] and generalizes the Erdős-Rényi model $G(n, p)$. Suppose $d \geq 3$ and n is a positive integer. In this paper we often denote the set of d-uniform hypergraphs with vertex set $[n]$ as $\{0,1\}^{\binom{[n]}{d}}$. Then, for some $\delta \in (-\infty, d-1)$ and $p \propto (1 + o_n(1))n^{-d+1+\delta}$, the random d-uniform hypergraph $\mathcal{H}$ is sampled from $\mathrm{Ber}(p)^{\binom{[n]}{d}}$. We specify the model in more detail in Section 1.1. After observing the unweighted or weighted edge data of $\mathcal{H}$, the goal is to recover information of $\mathcal{H}$.

We extend upon the *exact recovery* results of the paper [5] and study the problem of *partial recovery* introduced in Appendix C of the paper. Exact recovery refers to recovering each hyperedge of $\mathcal{H}$ and the main quantifier of it that we use is the *probability of exact recovery*. For the definition of this metric, see Subsection 1.2.2. The partial recovery problem refers to recovering a fraction of the hyperedges of $\mathcal{H}$ on average and the main quantifier of partial recovery we use is the *partial recovery loss*. The partial recovery loss is the ratio of the size of the symmetric difference between the hyperedges of $\mathcal{H}$ and the hyperedges we predict to be in $\mathcal{H}$ to the expected number of hyperedges of $\mathcal{H}$, see Subsection 1.2.1.

As mentioned earlier, the papers [8, 14, 15, 17] consider exact recovery for the hypergraph SBM. On the other hand, [21] considers exact and partial recovery for graph matching and discusses "all-or-nothing" thresholds. We discuss these thresholds later in Section 1.3.

We can describe the exact recovery problem for the unweighted projection as $\min(AH, 1) = \mathrm{Proj}(\mathcal{H})$, where $A \in \{0,1\}^{\binom{[n]}{2} \times \binom{[n]}{d}}$ is fixed, $H \in \{0,1\}^{\binom{[n]}{d}}$ is the prediction of $\mathcal{H}$, and $\mathrm{Proj} \in \{0,1\}^{\binom{[n]}{2}}$. In this case, for $h \in \binom{[n]}{d}$ and $e \in \binom{[n]}{2}$, $A_{eh} = \mathbf{1}\{e \subset h\}$. The exact recovery problem for the weighted projection is more interesting interesting, since we can describe it as the linear equation $AH = \mathrm{Proj}_W(\mathcal{H})$, where $\mathrm{Proj}_W \in \mathbb{Z}_{\geq 0}^{\binom{[n]}{2}}$.

Furthermore, observe that this recovery problem is an example of a planted constraint satisfaction problem (CSP). First, we randomly select $\mathcal{H}$, which then deterministically defines the set of constraints. For more works on planted CSPs, see [1–3, 11–13, 16].

In this paper we study thresholds for δ for when the probability of exact recovery and partial recovery rate are asymptotically equal to 0 or 1. We find the threshold values for δ for both types of recovery for all $d \geq 3$. Moreover, we observe that in most contexts, there is an "all-or-nothing" transition at $\frac{d-1}{d+1}$, meaning that the probability of exact recovery is $1 - o_n(1)$ for $\delta < \frac{d-1}{d+1}$ and the partial recovery loss is $1 - o_n(1)$ for $\delta > \frac{d-1}{d+1}$, see Section 1.3. In general, we first focus on recovery for unweighted edge data and afterwards we discuss how to extend the results to weighted edge data.

We also note that this problem has been solved algorithmically. See Section 5.4 for discussion regarding this direction.

1.1 Random model

We define the projected graph and random hypergraph model that are introduced in [5].

For a d-uniform hypergraph $H = (V, E)$ let the *projection* of H be the 2-uniform hypergraph with vertex set V and edge set equal to the set of $\{i, j\}$ such that $i, j \in V$, $i \neq j$, and there exists $h \in E$ such that $\{i, j\} \subset h$; we denote the projection of H by $\mathrm{Proj}(H)$. Let the *weighted projection* of H be the weighted graph with vertex set V and edge weight of $\{i, j\}$ equal to the number of $h \in H$ that contain $\{i, j\}$ for all $i, j \in V$, $i \neq j$; we denote the weighted projection of H by $\mathrm{Proj}_W(H)$.

Suppose $\delta \in (-\infty, d-1)$ and $c \in (0, \infty)$. Unless otherwise stated assume that $p = (c+o_n(1))n^{-d+1+\delta}$. Suppose $\mathcal{H} \subset \{0,1\}^{\binom{[n]}{d}}$ is a random variable such that each element of $\binom{[n]}{d}$ is a hyperedge of $\mathcal{H}$ independently with probability p. Furthermore suppose $\mathcal{H}_c \in \{0,1\}^{\binom{[n]}{d}}$ is the random variable such that each d-clique in $\mathrm{Proj}(\mathcal{H})$ is a hyperedge of $\mathcal{H}_c$. We observe $\mathrm{Proj}(\mathcal{H})$ and use this projected graph to make statistical inferences about $\mathcal{H}$.

Remark 1.1. The regime of p we consider is slightly different from that of the paper [5], which considers when $p = n^{-d+1+\delta}$. Despite this, many of the results from the paper are still true.

1.2 Notation

Next we describe some additional notation that is used throughout this paper. Suppose $d \geq 2$ and H is a d-uniform hypergraph. Let $E(H)$ and $V(H)$ denote the sets of edges and vertices of H, respectively, and let $e(H) = |E(H)|$ and $v(H) = |V(H)|$. Furthermore let $\alpha(H) = \frac{e(H)}{v(H)}$ and $m(H) = \max_{K \leq H} \alpha(H)$, where the maximum is over subgraphs K of H.

Let $\mathcal{G}$ (resp. $\mathcal{G}_W$) be the set of projections (resp. weighted projections) of some d-uniform hypergraph. Also let $q = \Pr[[d] \in E(\mathcal{H}_c)]$. Note that q is the probability that any element of $\binom{[n]}{d}$ is an edge of $\mathcal{H}_c$ by symmetry.

1.2.1 Partial recovery

Suppose $\mathcal{B} : \mathcal{G} \to \{0,1\}^{\binom{[n]}{d}}$ is a partial recovery algorithm for the unweighted projection. We define the loss of $\mathcal{B}$ to be

$$\ell(\mathcal{B}) = \frac{\mathbb{E}[|\mathcal{B}(\Psi)\Delta E(\mathcal{H})|]}{p\binom{n}{d}},$$

where $p\binom{n}{d} = \mathbb{E}[|E(\mathcal{H})|]$.

Observe that

$$\mathbb{E}[|\mathcal{B}(\Psi)\Delta\mathcal{H}|] = \sum_{h\in\binom{[n]}{d}} \mathbf{1}_{h\in\mathcal{B}(\Psi)} \Pr[h \notin \mathcal{H}|\Psi] + (1-\mathbf{1}_{h\in\mathcal{B}(\Psi)}) \Pr[h \in \mathcal{H}|\Psi].$$

Hence the optimal unweighted partial recovery algorithm is

$$\mathcal{B}^* : \mathcal{G} \to \{0,1\}^{\binom{[n]}{d}}, G \mapsto \{h \in \binom{[n]}{d} : \Pr[h \in \mathcal{H}|\mathrm{Proj}(\mathcal{H}) = G] \geq \frac{1}{2}\}.$$

For a partial recovery algorithm $\mathcal{B}_W : \mathcal{G}_W \to \{0,1\}^{\binom{[n]}{d}}$ for the weighted projection, the loss of $\mathcal{B}_W$ is

$$\ell_W(\mathcal{B}_W) = \frac{\mathbb{E}[|\mathcal{B}_W(\mathrm{Proj}_W(\mathcal{H}))\Delta E(\mathcal{H})|]}{p\binom{n}{d}}.$$

The optimal weighted partial recovery algorithm $\mathcal{B}_W^*$ is defined analogously to $\mathcal{B}^*$. That is,

$$\mathcal{B}_W^* : \mathcal{G} \to \{0,1\}^{\binom{[n]}{d}}, G \mapsto \{h \in \binom{[n]}{d} : \Pr[h \in \mathcal{H}|\mathrm{Proj}_W(\mathcal{H}) = G] \geq \frac{1}{2}\}.$$

Definition 1.2. *The partial recovery loss is $\ell(\mathcal{B}^*)$. If $\ell(\mathcal{B}^*) = o_n(1)$, then almost exact recovery is possible.*

The weighted partial recovery loss is $\ell_W(\mathcal{B}_W^*)$. If $\ell_W(\mathcal{B}_W^*) = o_n(1)$, then almost exact weighted recovery is possible.

Remark 1.3. Observe that this definition of almost exact recovery is equivalent to the definition of almost exact recovery in [21].

1.2.2 Exact recovery

Suppose $\mathcal{A} : \mathcal{G} \to \{0,1\}^{\binom{[n]}{d}}$ is an exact recovery algorithm for the unweighted projection. Its probability of error is $\Pr[\mathcal{A}(\mathrm{Proj}(\mathcal{H})) \neq \mathcal{H}]$. As discussed in [5, Section 2.3], the algorithm minimizing the probability of error is the MAP algorithm $\mathcal{A}^* : \mathcal{G} \to \{0,1\}^{\binom{[n]}{d}}$ where

$$\mathcal{A}^*(G) \in \operatorname*{argmax}_{H:\,\mathrm{Proj}(H)=G} \Pr[\mathcal{H} = H] = \operatorname*{argmin}_{H:\,\mathrm{Proj}(H)=G} e(H).$$

Similarly, for an exact recovery algorithm $\mathcal{A}_W : \mathcal{G}_W \to \{0,1\}^{\binom{[n]}{d}}$ for the unweighted projection, its probability of error is $\Pr[\mathcal{A}_W(\mathrm{Proj}_W(\mathcal{H})) \neq \mathcal{H}]$. The algorithim minimizing the probability error is the MAP algorithm $\mathcal{A}_W^* : \mathcal{G}_W \to \{0,1\}^{\binom{[n]}{d}}$ which is analogous to $\mathcal{A}^*$.

Definition 1.4. *The probability of exact recovery is* $\Pr[\mathcal{A}^*(\mathrm{Proj}(\mathcal{H})) = \mathcal{H}]$*. The weighted probability of exact recovery is* $\Pr[\mathcal{A}_W^*(\mathit{Proj}_W(\mathcal{H})) = \mathcal{H}]$.

1.3 Main results

Theorem 1.5 (Partial Recovery)**.** *If $\delta < \frac{d-1}{d+1}$ then the partial recovery loss is $o_n(1)$ and if $\delta > \frac{d-1}{d+1}$ then the partial recovery loss is $1 - o_n(1)$.*

Proof. This follows from Theorem 3.1, Theorem 3.7, and Theorem 3.8. ∎

Theorem 1.6 (Partial Recovery for Weighted Projection)**.** *If $\delta < \frac{d-1}{d+1}$ then the partial recovery loss is $o_n(1)$ and if $\delta > \frac{d-1}{d+1}$ then the weighted partial recovery loss is $1 - o_n(1)$.*

Proof. Since the partial recovery loss is $o_n(1)$ for $\delta < \frac{d-1}{d+1}$ by Theorem 1.5, the same is true for the weighted partial recovery loss, see Lemma 1.10. Afterwards, using Theorem 3.13 finishes the proof. ∎

Furthermore, we prove the following two theorems in Chapter 5.

Theorem 1.7 (Exact Recovery)**.** *Suppose $3 \leq d \leq 5$. If $\delta < \frac{2d-4}{2d-1}$ then the probability of exact recovery is $1 - o_n(1)$ and if $\delta > \frac{2d-4}{2d-1}$ then the probability of exact recovery is $o_n(1)$.*

Suppose $d \geq 5$. If $\delta < \frac{d-1}{d+1}$ then the probability of exact recovery is $1 - o_n(1)$ and if $\delta > \frac{d-1}{d+1}$, the probability of exact recovery is $o_n(1)$.

Remark 1.8. Observe that Theorem 1.7 verifies the conjecture from [5, Appendix C] that the threshold for exact recovery is $\frac{2d-4}{2d-1}$ for $d = 4, 5$. Furthermore, the two cases of the statement overlap for $d = 5$. This is intentional since $\frac{2d-4}{2d-1} = \frac{d-1}{d+1}$ when $d = 5$.

Theorem 1.9 (Exact Recovery for Weighted Projection)**.** *If $\delta < \frac{d-1}{d+1}$ then the probability of weighted exact recovery is $1 - o_n(1)$ and if $\delta > \frac{d-1}{d+1}$ the probability of weighted exact recovery is $o_n(1)$.*

The contributions of this paper as compared to previously established exact recovery thresholds are summarized in Table 1.1.

d	Previous work (exact recovery)	New (exact recovery)	Partial recovery Weighted exact recovery
3	$2/5$	$2/5$	$1/2$
4	$[1/2, 4/7]$	$4/7$	$3/5$
5	$[1/2, 2/3]$	$2/3$	$2/3$
≥ 6	$[\frac{d-3}{d}, \frac{d^2-d-2}{d^2-d+2}]$	$\frac{d-1}{d+1}$	$\frac{d-1}{d+1}$

Table 1.1: This table compares exact recovery thresholds discovered in [5] with the contributions of this paper. Note that the exact recovery thresholds correspond to the problem of unweighted exact recovery. Furthermore, the partial recovery thresholds are the same for observing the unweighted or weighted projection.

An all-or-nothing phase transition occurs at a threshold if, for example, we can recover all of the input with high probability below the threshold, but we cannot recover any constant fraction of the input with constant probability above. As mentioned earlier, [21] finds that the all-or-nothing threshold occurs for the problem of graph-matching. All-or-nothing thresholds are also discovered in [19] for the problem of recovering a sparse binary vector with linear regression, where the metric is the mean squared error. Furthermore, [4] finds all-or-nothing thresholds for recovery in the spiked Wigner model.

We observe an all-or-nothing threshold at $\frac{d-1}{d+1}$ for unweighted recovery when $d \geq 5$ and weighted recovery when $d \geq 3$. This is the case because we transition from having a probability of exact recovery of $1 - o_n(1)$ below the threshold to a partial recovery loss of $1 - o_n(1)$ above.

Furthermore, observe that the exact recovery results do not consider when $\delta = \min(\frac{d-1}{d+1}, \frac{2d-4}{2d-1})$. We exhibit regimes of probabilities satisfying this condition for which the probability of exact recovery is $1 - o_n(1)$ in Chapter 6.

1.4 Introduction to the partial recovery loss

We discuss some fundamental equalities for the partial recovery loss introduced in Subsection 1.2.1. First, we have that

$$\mathbb{E}_{\mathcal{H}|\Psi}[|\mathcal{B}^*(\Psi)\Delta\mathcal{H}|] = \sum_{h \in \binom{[n]}{d}} \min(\Pr[h \in \mathcal{H}|\Psi], \Pr[h \notin \mathcal{H}|\Psi]). \tag{1.1}$$

Using this gives that

$$\begin{aligned} p\binom{n}{d}\ell(\mathcal{B}^*) &= \mathbb{E}_\Psi[\mathbb{E}_{\mathcal{H}|\Psi}[|\mathcal{B}^*(\Psi)\Delta\mathcal{H}|]] \\ &= \sum_{G \in \mathcal{G}} \sum_{h \in \binom{[n]}{d}} \Pr[\Psi = G] \min(\Pr[h \in \mathcal{H}|\Psi], \Pr[h \notin \mathcal{H}|\Psi]) \\ &= \sum_{h \in \binom{[n]}{d}} \sum_{G \in \mathcal{G}} \Pr[\Psi = G] \min(\Pr[h \in \mathcal{H}|\Psi], \Pr[h \notin \mathcal{H}|\Psi]) \\ &= \binom{n}{d} \sum_{G \in \mathcal{G}} \Pr[\Psi = G] \min(\Pr[[d] \in \mathcal{H}|\Psi], \Pr[[d] \notin \mathcal{H}|\Psi]) \\ &= \binom{n}{d} \sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left|\Pr[[d] \in \mathcal{H}|\Psi] - \frac{1}{2}\right|\right). \end{aligned} \tag{1.2}$$

Furthermore, $\ell(\mathcal{B}^*) \leq 1$ since the trivial algorithm that always outputs the empty set has a loss of 1.

Lemma 1.10. *$\ell(\mathcal{B}^*) \geq \ell_W(\mathcal{B}^*_W)$ and $\Pr[\mathcal{A}^*(Proj(\mathcal{H})) = \mathcal{H}] \leq \Pr[\mathcal{A}^*_W(Proj_W(\mathcal{H})) = \mathcal{H}]$.*

Proof. It is straightforward to deduce that $\Pr[\mathcal{A}^*(\mathrm{Proj}(\mathcal{H})) = \mathcal{H}] \leq \Pr[\mathcal{A}^*_W(\mathrm{Proj}_W(\mathcal{H})) = \mathcal{H}]$. For $\ell(\mathcal{B}^*) \geq \ell_W(\mathcal{B}^*_W)$, using (1.2) and its analogue for weighted projections gives that it suffices to prove that

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left|\Pr[[d] \in \mathcal{H}|\Psi = G] - \frac{1}{2}\right|\right)$$

$$\geq \sum_{G \in \mathcal{G}_W} \Pr[\mathrm{Proj}_W(\mathcal{H}) = G] \left(\frac{1}{2} - \left| \Pr[[d] \in \mathcal{H} | \mathrm{Proj}_W(\mathcal{H}) = G] - \frac{1}{2} \right| \right),$$

which is equivalent to

$$\sum_{G \in \mathcal{G}_W} \left| \Pr[[d] \in \mathcal{H} | \mathrm{Proj}_W(\mathcal{H}) = G] - \frac{1}{2} \right| \geq \sum_{G \in \mathcal{G}} \left| \Pr[[d] \in \mathcal{H} | \Psi = G] - \frac{1}{2} \right|.$$

This can be proved using Jensen's inequality since $x \mapsto \left| x - \frac{1}{2} \right|$ is convex. ■

1.5 Thesis organization

This thesis is based on the contents of the preprint [6], a paper that I am a primary author of. Chapter 6 does not appear in the paper, but all other content of this thesis appears in the paper.

In Chapter 2, we enumerate combinatorial structures in Ψ and prove the concentration result Theorem 2.9; this concentration result implies Corollary 2.10, which is used to justify the upper bounds of the partial recovery thresholds. Afterwards, in Chapter 3, we prove the partial recovery results. In Chapter 4, we prove extremal combinatorial results related to ambiguous graphs which are essential to showing the lower bounds of the exact recovery thresholds. In Chapter 5, we display the proofs of the exact recovery results, and in particular prove the upper bounds of the exact recovery thresholds in Theorem 5.7. Furthermore, we discuss efficient algorithms for partial and exact recovery in Section 5.4. Next, in Chapter 6, we analyze the phase transition for exact recovery by exhibiting a regime of probabilities that are below the exact recovery threshold by a polylogarithmic factor for which the exact recovery loss is $1 - o_n(1)$. In Appendix A, we analyze the conditional entropy $H(\mathcal{H}|\Psi)$.

Chapter 2

Structures of the Projected Graph

2.1 Projection covers and relaxations

First, we give two lemmas that generalize ideas from [5, Proof of Lemma 39]. We do not include the proofs.

Lemma 2.1. *Suppose $k \geq 2$ and $\mathcal{U} \subset 2^{[k]}$. The probability that there exists $h \in E(\mathcal{H})$ such that $h \cap [k] = u$ for all $u \in \mathcal{U}$ is*

$$\prod_{u \in \mathcal{U}} (1 - (1-p)^{\binom{n-k}{d-|u|}}).$$

Furthermore, this probability is at most

$$p^{|\mathcal{U}|} \prod_{u \in \mathcal{U}} \binom{n-k}{d-|u|} = \Theta_n(n^{(1+\delta)|\mathcal{U}| - \sum_{u \in \mathcal{U}} |u|}).$$

Lemma 2.2. *Suppose $k \geq 2$ and $\mathcal{E} \subset \binom{[k]}{2}$. The probability that $\mathcal{E} \subset E(Proj(\mathcal{H}))$ is at most*

$$\sum_{\mathcal{U}} p^{|\mathcal{U}|} \prod_{u \in \mathcal{U}} \binom{n-k}{d-|u|},$$

where the sum is over $\mathcal{U} \subset 2^{[k]}$ satisfying the following conditions:

- *For all $u \in \mathcal{U}$, $|u| \geq 2$ and $\binom{u}{2} \not\subset \bigcup_{u' \in \mathcal{U} \setminus \{u\}} \binom{u'}{2}$.*
- *$\mathcal{E} \subset \bigcup_{u \in \mathcal{U}} \binom{u}{2}$.*

Remark 2.3. $\mathrm{Proj}(\mathcal{H})$ is a 2-uniform hypergraph, so $E(\mathrm{Proj}(\mathcal{H}))$ is a set of edges, which are hyperedges with size 2.

Next we discuss a technique from [5] that involves using relaxation to establish a convex optimization problem after applying Lemmas 2.1 and 2.2. Suppose $k \geq 2$ and $\mathcal{E} \subset \binom{[k]}{2}$. The goal is to upper bound the probability that $\mathcal{E} \subset E(\mathrm{Proj}(\mathcal{H}))$ using Lemma 2.2. Suppose $\mathcal{U}$ satisfies the conditions of Lemma 2.2; the lemma implies that an upper bound on

$p^{|\mathcal{U}|}\prod_{u\in\mathcal{U}}\binom{n-k}{d-|u|}$ is an upper bound on the probability that $\mathcal{E}\subset E(\mathrm{Proj}(\mathcal{H}))$ after scaling by some constant since the number of $\mathcal{U}$ is finite.

First observe that $|\mathcal{E}|\leq\sum_{u\in\mathcal{U},|u|\geq 2}\binom{|u|}{2}$. The relaxation technique is to replace $\binom{x}{2}$ for some real variable $x\geq 2$ with the real variable $y\geq 1$; that is, $y=\binom{x}{2}$ and $x=\frac{1+\sqrt{1+8y}}{2}$. If $y_u=\binom{x_u}{2}$ for $u\in\mathcal{U}$ such that $|u|\geq 2$ then $|\mathcal{E}|\leq\sum_{u\in\mathcal{U},|u|\geq 2}y_u$. From Lemma 2.1,

$$p^{|\mathcal{U}|}\prod_{u\in\mathcal{U}}\binom{n-k}{d-|u|}=\Theta_n(n^{(1+\delta)|\mathcal{U}|-\sum_{u\in\mathcal{U}}|u|})=\Theta_n(n^{(1+\delta)|\mathcal{U}|-\sum_{u\in\mathcal{U}}\frac{1+\sqrt{1+8y_u}}{2}}).$$

Suppose $M=|\mathcal{U}|$; the conditions of Lemma 2.2 imply that $1\leq M\leq|\mathcal{E}|$. Then maximizing the quantity $p^{|\mathcal{U}|}\prod_{u\in\mathcal{U}}\binom{n-k}{d-|u|}$ corresponds to maximizing $(1+\delta)M-\sum_{i=1}^{M}\frac{1+\sqrt{1+8y_i}}{2}$ given that $\sum_{i=1}^{M}y_i\geq|\mathcal{E}|$ and $y_i\geq 1$ for $1\leq i\leq M$, which is a convex optimization problem. Particularly, since the function $(1+\delta)M-\sum_{i=1}^{M}\frac{1+\sqrt{1+8y_i}}{2}$ is convex, it is maximized at a vertex of the set of inputs.

Some results of [5] that are proved using the methods described in this subsection are [5, Lemmas 35, 39, and 40]. In this paper we prove results in Chapter 2, Theorem 3.3, and Lemma 4.5 using these methods.

Remark 2.4. For the proof of Lemma 2.6, we also impose the constraint that $|u|\leq d$ for all $u\in\mathcal{U}$, which corresponds to $y_i\leq\binom{d}{2}$ for $1\leq i\leq M$, because the hyperedges of $\mathcal{H}$ have size d. It is necessary to impose this constraint because the value k in Lemma 2.1 and Lemma 2.2 may be greater than d in the context of Lemma 2.6. We also impose this constraint in the proof of Lemma 4.5.

2.2 Structure results

The main goal of this section is to prove that $e(\mathcal{H}_c)$ is concentrated around its mean if $\delta>\frac{d-1}{d+1}$, see Corollary 2.10. Observe that Lemmas 2.5 and 2.6 have similar statements and proofs as [5, Lemma 35 and Lemma 40], but some differences are that we require different bounds and only consider when $\delta>\frac{d-1}{d+1}$. Furthermore, we use methods discussed in Section 2.1 in this section.

Lemma 2.5. *Suppose $\delta>\frac{d-1}{d+1}$. Suppose m is an integer such that $0\leq m\leq d-1$. Assume that $\{K_i:1\leq i\leq M\}$ is a set of subsets S of $[d]$ such that $2\leq|S|\leq d$. Assume that*

$$\bigcup_{i=1}^{M}\binom{K_i}{2}\supset\binom{[d]}{2}\backslash\binom{[m]}{2}.$$

Then

$$(1+\delta)M-\sum_{i=1}^{M}|K_i|\leq\left(\binom{d}{2}-\binom{m}{2}\right)(\delta-1).$$

Equality occurs if and only if the K_i are distinct and $\{K_i:1\leq i\leq M\}=\binom{[d]}{2}\backslash\binom{[m]}{2}$.

Proof. First we may assume that $\binom{K_i}{2} \cap \left(\binom{[d]}{2} \backslash \binom{[m]}{2}\right) \not\subset \bigcup_{j\in[M]\backslash\{i\}} \binom{K_j}{2}$ for $1 \leq i \leq M$. We can assume this because if the condition is not true we can remove K_i and increase $(1+\delta)M - \sum_{i=1}^{M} |K_i|$. This condition implies that the K_i are distinct and $M \leq \binom{d}{2} - \binom{m}{2}$.

Suppose $y_i = \binom{|K_i|}{2}$ for $1 \leq i \leq M$, then $y_i \geq 1$ for $1 \leq i \leq m$ and $\sum_{i=1}^{M} y_i \geq \binom{d}{2} - \binom{m}{2}$. We have that

$$(1+\delta)M - \sum_{i=1}^{M} |K_i| = (1+\delta)M - \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}.$$

Let

$$f(y_1,\ldots,y_M) = (1+\delta)M - \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}.$$

Since f is convex, its maximal value occurs at a vertex. Suppose $(y_i)_{1\leq i\leq M}$ is the vertex such that $y_i = 1$ for $1 \leq i \leq M-1$ and $y_M = \binom{d}{2} - \binom{m}{2} - M + 1$. Then

$$f(y_1,\ldots,y_M) = (\delta-1)M + 2 - \frac{1+\sqrt{1+8(\binom{d}{2} - \binom{m}{2} - M + 1)}}{2}.$$

Since $f(y_1,\ldots,y_M)$ is convex with respect to M, the maximum value of $f(y_1,\ldots,y_M)$ for $1 \leq M \leq \binom{d}{2} - \binom{m}{2}$ occurs when $M \in \{1, \binom{d}{2} - \binom{m}{2}\}$.

Suppose $M = 1$. Then

$$f(y_1,\ldots,y_M) = 1 + \delta - \frac{1+\sqrt{1+8(\binom{d}{2} - \binom{m}{2})}}{2}.$$

Because the number of edges in $[d]$ but not $[m]$ is greater than M, we also must prove that equality does not hold. Because $\delta > \frac{d-1}{d+1}$ it suffices to prove that

$$1 + \frac{d-1}{d+1} - \frac{1+\sqrt{1+8(\binom{d}{2} - \binom{m}{2})}}{2} \leq -\left(\binom{d}{2} - \binom{m}{2}\right)\frac{2}{d+1}.$$

This can be proved using expansion.

Next suppose $M = \binom{d}{2} - \binom{m}{2}$. Then $y_M = 1$ so

$$f(y_1,\ldots,y_M) = \left(\binom{d}{2} - \binom{m}{2}\right)(\delta - 1).$$

Afterwards it is straightforward to verify the equality case. ■

Lemma 2.6. *Suppose $\delta > \frac{d-1}{d+1}$. Suppose m and k are integers such that $m \in \{0, 1, 2, d-1\}$ and $m \leq k \leq d-1$. Assume that $\{K_i : 1 \leq i \leq M\}$ is a set of subsets S of $[1, 2d-k]$ such that $2 \leq |S| \leq d$. Assume that*

$$\bigcup_{i=1}^{M} \binom{K_i}{2} \supset \left(\binom{[d]}{2} \bigcup \binom{[k] \cup \{i : d+1 \leq i \leq 2d-k\}}{2}\right) \backslash \binom{[m]}{2}$$

Then
$$(1+\delta)M - \sum_{i=1}^{M} |K_i| \le k - m + (d(d-1) - m(m-1))(\delta - 1). \tag{2.1}$$
Equality occurs if and only if $k = m$, the K_i are distinct, and $\{K_i : 1 \le i \le M\} = \left(\binom{[d]}{2} \bigcup \binom{[k]\cup\{i:d+1\le i\le 2d-k\}}{2}\right) \setminus \binom{[m]}{2}$.

Proof. Let
$$\mathcal{E} = \left(\binom{[d]}{2} \cup \binom{[k]\bigcup\{i : d+1 \le i \le 2d-k\}}{2}\right) \setminus \binom{[m]}{2}.$$
Similarly to the proof of Lemma 2.5, assume the condition $(*1)$ that for $1 \le i \le M$, $\binom{K_i}{2} \cap \mathcal{E} \not\subset \bigcup_{j\in[M]\setminus\{i\}} \binom{K_j}{2}$. Note that $(*1)$ implies that the K_i are distinct and that
$$M \le |\mathcal{E}| = 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}.$$
We must prove that the equality case of (2.1) occurs if and only if $M = 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}$. Furthermore, since each element of $[2d-k]$ must be a vertex of one of the K_i and $2d-k > d$, $M \ge 2$.

Case 1: $m \in \{0,1,2\}$**,** $m < d-1$

Suppose $m \in \{0,1,2\}$ and $m < d-1$. (We do not consider when $m = 2$ and $d = 3$.)

Suppose $y_i = \binom{|K_i|}{2}$ for $1 \le i \le M$. Then, $1 \le y_i \le \binom{d}{2}$ for $1 \le i \le M$ and
$$\sum_{i=1}^{M} y_i \ge 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}.$$
Furthermore, the left hand side of (2.1) is
$$(1+\delta)M - \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}.$$
Let $\mathcal{R}_M$ be the set of $(y_i)_{1\le i\le M}$ such that $1 \le y_i \le \binom{d}{2}$ for $1 \le i \le M$ and $\sum_{i=1}^{M} y_i \ge 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}$. Additionally, let
$$f(y_1,\ldots,y_M) = (1+\delta)M - \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}.$$
Observe that f is convex so the maximum value of f over $\mathcal{R}_M$ occurs at a vertex of $\mathcal{R}_M$. Suppose $(y_i)_{1\le i\le M}$ is a vertex of $\mathcal{R}_M$ such that $y_i \in \{1, \binom{d}{2}\}$ for $1 \le i \le M-1$.

Suppose $2 \le M \le \binom{d}{2} - \binom{k}{2} - \binom{m}{2}$. If $y_i = 1$ for $1 \le i \le M-1$ then
$$y_M \ge 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2} - M + 1 > \binom{d}{2},$$
which is a contradiction. Assume that one of the y_i, $1 \le i \le M-1$ equals $\binom{d}{2}$; it is clearly not optimal if two distinct y_i for $1 \le i \le M-1$ equal $\binom{d}{2}$ since we will then have that

$\sum_{i=1}^{M} y_i > 2\binom{d}{2}$, so we can increase the value of f by decreasing some of the y_i. Without loss of generality, assume that $y_i = 1$ for $1 \le i \le M-2$, $y_{M-1} = \binom{d}{2}$, and

$$y_M = \binom{d}{2} - \binom{k}{2} - \binom{m}{2} - M + 2.$$

Then,

$$f(y_1, \dots, y_M) = (\delta - 1)M + 2(\delta + 1) + 4 - d - \frac{1 + \sqrt{1 + 8y_M}}{2}.$$

Since $M < 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}$, we must prove that equality does not occur. Therefore, we must prove that

$$(\delta - 1)M + 4 - d - \frac{1 + \sqrt{1 + 8y_M}}{2} < k - m + (d(d-1) - m(m-1))(\delta - 1)$$

It suffices to prove that this inequality is true for $2 \le M \le \binom{d}{2} - \binom{k}{2} - \binom{m}{2} + 1$ (observe that we add $\binom{d}{2} - \binom{k}{2} - \binom{m}{2} + 1$ as a value for M to simplify calculations). The left hand side is convex with respect to M, so it suffices to prove that the inequality is true for $M \in \{2, \binom{d}{2} - \binom{k}{2} - \binom{m}{2} + 1\}$.

First, suppose $M = 2$. The required inequality is

$$\begin{aligned} &2(\delta - 1) + 4 - d - \frac{1 + \sqrt{1 + 8(\binom{d}{2} - \binom{k}{2} - \binom{m}{2})}}{2} \\ &< k - m + (d(d-1) - m(m-1))(\delta - 1). \end{aligned}$$

Since $\delta > \frac{d-1}{d+1}$ and $d(d-1) - m(m-1) > 2$, it suffices to prove that

$$-\frac{4}{d+1} + 4 - d - \frac{1 + \sqrt{1 + 8(\binom{d}{2} - \binom{k}{2} - \binom{m}{2})}}{2} \le k - m - \frac{2}{d+1}(d(d-1) - m(m-1)).$$

This inequality can be verified by expansion. Note that equality occurs if and only if $k = m = 0$ or $k = m = 1$.

Although Lemma 2.6 is true for $m \in \{0, 1, 2, d-1\}$, we only use the case $m = 0$, see Remark 2.7. Therefore, for completeness, we include the expansion for the case $m = 0$ and equivalently $m = 1$. It suffices to prove that

$$\begin{aligned} &-\frac{4}{d+1} + 4 - d - \frac{1 + \sqrt{1 + 4d^2 - 4d - 4k^2 + 4k}}{2} \le k - \frac{2d(d-1)}{d+1} \\ \Leftrightarrow &\frac{2d^2 - 2d - 4}{d+1} + 4 - d - k - \frac{1}{2} \le \frac{\sqrt{1 + 4d^2 - 4d - 4k^2 + 4k}}{2} \\ \Leftrightarrow &2d - 2k - 1 \le \sqrt{1 + 4d^2 - 4d - 4k^2 + 4k} \Leftrightarrow 8k^2 \le 8dk, \end{aligned}$$

which follows from $0 \le k \le d-1$.

Next, suppose $M = \binom{d}{2} - \binom{k}{2} - \binom{m}{2} + 1$. Then, $y_M = 1$ so the required inequality is

$$(\delta - 1)\left(\binom{d}{2} - \binom{k}{2} - \binom{m}{2} + 1\right) + 2 - d < k - m + (d(d-1) - m(m-1))(\delta - 1).$$

This is equivalent to

$$(1-\delta)\left(\binom{d}{2}-\binom{m}{2}+\binom{k}{2}-1\right)+2<k-m+d.$$

We need to prove this inequality for $\delta > \frac{d-1}{d+1}$, so it suffices to prove that

$$\frac{2}{d+1}\left(\binom{d}{2}-\binom{m}{2}+\binom{k}{2}-1\right)+2\leq k-m+d.$$

This inequality is equivalent to

$$(k-m)(d+2-k-m)\geq 0,$$

which is true since $m \leq k \leq d-1$ and $m \leq 2$.

Suppose $\binom{d}{2}-\binom{k}{2}-\binom{m}{2}+1 \leq M \leq 2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}$. It is clearly not optimal for two of the y_i for $1 \leq i \leq M-1$ to equal $\binom{d}{2}$. Suppose one of the y_i for $1 \leq i \leq M-1$ equals $\binom{d}{2}$. Without loss of generality, suppose $y_i = 1$ for $1 \leq i \leq M-2$ and $y_{M-1} = \binom{d}{2}$. Then,

$$y_M \geq \binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+2.$$

Since $\binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+2 \leq 1$,

$$\begin{aligned} f(y_1,\ldots,y_{M-2},y_{M-1},y_M) &\leq f\left(y_1,\ldots,y_{M-2},\binom{d}{2},1\right)\\ &\leq f\left(y_1,\ldots,y_{M-2},2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+1,1\right)\\ &= f\left(y_1,\ldots,y_{M-2},1,2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+1\right). \end{aligned}$$

Then, we may assume that $y_i = 1$ for $1 \leq i \leq M-1$. We have that

$$y_M = 2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+1.$$

Furthermore,

$$f(y_1,\ldots,y_M) = (\delta-1)M+2-\frac{1+\sqrt{1+8(2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}-M+1)}}{2}.$$

Observe that f convex with respect to M. Therefore, f is maximized over $\binom{d}{2}-\binom{k}{2}-\binom{m}{2}+1 \leq M \leq 2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}$ when $M \in \{\binom{d}{2}-\binom{k}{2}-\binom{m}{2}+1, 2\binom{d}{2}-\binom{k}{2}-\binom{m}{2}\}$.

Suppose $M = \binom{d}{2}-\binom{k}{2}-\binom{m}{2}+1$. Then $y_M = \binom{d}{2}$ so this case is equivalent to the previous case we considered where $(y_1,\ldots,y_M) = (1,\ldots,1,\binom{d}{2},1)$.

Next suppose $M = 2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}$. Then $y_M = 1$ so the required inequality is

$$(\delta - 1)\left(2\binom{d}{2} - \binom{k}{2} - \binom{m}{2}\right) \leq k - m + (d(d-1) - m(m-1)\Big)(\delta - 1).$$

This is equivalent to

$$(1 - \delta)\left(\binom{k}{2} - \binom{m}{2}\right) \leq k - m.$$

If $k = m$ then it is clear that equality occurs. Suppose $k > m$. We must prove that

$$(1 - \delta)\left(\binom{k}{2} - \binom{m}{2}\right) < k - m.$$

Since $\delta > \frac{d-1}{d+1}$, it suffices to prove that

$$\frac{2}{d+1}\left(\binom{k}{2} - \binom{m}{2}\right) \leq k - m.$$

This is equivalent to

$$\frac{1}{d+1}(k-m)(k+m-1) \leq k - m,$$

which is true since $m \leq 2$ and $k \leq d - 1$.
Case 2: $m = d - 1$

Next suppose $m = d - 1$. Since $k = m = d - 1$, we must prove that

$$(1 + \delta)M - \sum_{i=1}^{M} |K_i| \leq 2(d-1)(\delta - 1)$$

and that equality occurs if and only if $\{K_i : 1 \leq i \leq M\} = \mathcal{E}$. Note that

$$\mathcal{E} = \{\{d, j\} : j \in [d-1]\} \cup \{\{d+1, j\} : j \in [d-1]\} \subset \bigcup_{i=1}^{M} \binom{K_i}{2}.$$

Suppose

$$\{K_i : 1 \leq i \leq M\} = S_a \sqcup S_b \sqcup S_{ab},$$

where $K_i \in S_a$ if $K_i \cap \{d, d+1\} = \{d\}$, $K_i \in S_b$ if $K_i \cap \{d, d+1\} = \{d+1\}$, and $K_i \in S_{ab}$ if $K_i \cap \{d, d+1\} = \{d, d+1\}$ for $1 \leq i \leq M$.

Let $Z = \{j : j \in [d-1], \exists k \in S_{ab}$ such that $j \in k\}$. For all $j \in Z$, both edges in $\mathcal{E}$ that contain j are covered by an element of S_{ab}. Suppose $k \in S_a$ contains an element j of Z. If $k = \{d, j\}$ then ($*$1) will be contradicted so $|k| \geq 3$. If we remove j from k then the left hand side of (2.1) will decrease but all of the edges of $\mathcal{E}$ will remain covered. Hence, we can assume that no element of S_a contains an element of Z. We can similarly assume that no element of S_b contains an element of Z.

Furthermore, assume that $j \in [d-1]$ and $k_1, k_2 \in S_a$ satisfy $k_1 \neq k_2$ and $j \in k_1 \cap k_2$. If $k_1 = \{d, j\}$ then ($*$1) will be contradicted so $|k_1| \geq 3$. If we remove j from k_1 then the left hand side of (2.1) will decrease but all of the edges of $\mathcal{E}$ will remain covered. Hence, we can assume that no element of $[d-1]$ is contained in two elements of S_a. We can similarly assume that no element of $[d-1]$ is contained in two elements of S_b and that no element of $[d-1]$ is contained in two elements of S_{ab}.

Suppose $k \in S_a$ and $|k| \geq 3$. Suppose $j \in [d-1] \cap k$. Suppose we remove j from k and add $\{d, j\}$ to $\{K_i : 1 \leq i \leq M\}$. Then, the left hand side of (2.1) will increase by δ. Hence, we can assume that $|k| = 2$ for all $k \in S_a$ and similarly that $|k| = 2$ for all $k \in S_b$.

Since each element of S_a is $\{d, j\}$ for some $j \in [d-1]\backslash Z$, $|S_a| = d - 1 - |Z|$. Similarly, $|S_b| = d - 1 - |Z|$. Furthermore, the left hand side of (2.1) is

$$2(\delta-1)(d-1-|Z|) + \sum_{k \in S_{ab}} 1 + \delta - |k| = (\delta-1)(2d-2-2|Z|+|S_{ab}|) - |Z|.$$

The inequality

$$(\delta-1)(2d-2-2|Z|+|S_{ab}|) - |Z| \leq (\delta-1)(2d-2)$$

is equivalent to

$$-(1-\delta)|S_{ab}| \leq (2\delta-1)|Z|.$$

If $|Z| = |S_{ab}| = 0$, then the inequality holds with equality. Suppose $|Z| > 0$. Then, the inequality is strict because $\delta > \frac{d-1}{d+1} \geq \frac{1}{2}$. Hence, equality holds if and only if $|Z| = |S_{ab}| = 0$. Furthermore, $|Z| = |S_{ab}| = 0$ if and only if $\{K_i : 1 \leq i \leq M\} = \mathcal{E}$. ■

Remark 2.7. It may be possible to generalize the previous result to more values of m. We use the case $m = 0$ and intermediate results to prove Corollary 2.10, which is used in Section 3.3.

Theorem 2.8. *Suppose $\delta > \frac{d-1}{d+1}$. Suppose m is an integer such that $0 \leq m \leq d-1$ and l is an integer such that $l \geq m$. Let X be the set of elements h of $\binom{[n]}{d}$ such that $h \cap [l] = [m]$ and $\binom{h}{2}\backslash\binom{[m]}{2} \subset e(\Psi)$, where $[0]$ is the empty set. Then*

$$\mathbb{E}[|X|] = (1+o_n(1))\binom{n}{d-m}\left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}-\binom{m}{2}}.$$

Proof. The proof of this theorem has the same structure as the proof of Theorem 3.3. Suppose $h \in \binom{[n]}{d}$ and $h \cap [l] = [m]$. The number of h is $\binom{n-l}{d-m}$. Hence, it suffices to prove that

$$\Pr[h \in X] = \Pr[\binom{h}{2}\backslash\binom{[m]}{2} \subset e(\Psi)] = (1+o_n(1))\left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}-\binom{m}{2}}.$$

Using the Harris inequality gives that

$$\begin{aligned}
\Pr[h \in X] &\geq (1-(1-p)^{\binom{n-2}{d-2}})^{\binom{d}{2}-\binom{m}{2}} \\
&\geq \left(\binom{n-2}{d-2}p - p - O_n((n^{d-2}p)^2)\right)^{\binom{d}{2}-\binom{m}{2}} \\
&\geq (1-o_n(1))\left(\binom{n-2}{d-2}p\right)^{\binom{d}{2}-\binom{m}{2}}
\end{aligned}$$

$$= (1 - o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{\binom{d}{2}-\binom{m}{2}}.$$

From Lemma 2.2 with $[k]$ replaced by h and $\mathcal{E}$ replaced by $\binom{h}{2}\backslash\binom{[m]}{2}$,

$$\Pr[\binom{h}{2}\backslash\binom{[m]}{2} \subset e(\mathcal{H}_c)] \le \sum_{\mathcal{U}} p^{|\mathcal{U}|} \prod_{u\in\mathcal{U}} \binom{n-d}{d-|u|},$$

where the sum is over $\mathcal{U} \subset 2^h$ satisfying the conditions of the lemma.

If $\mathcal{U} = \binom{h}{2}\backslash\binom{[m]}{2}$, then

$$p^{|\mathcal{U}|} \prod_{u\in\mathcal{U}} \binom{n-d}{d-|u|} \le (1 + o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{\binom{d}{2}-\binom{m}{2}}.$$

Otherwise

$$p^{|\mathcal{U}|} \prod_{u\in\mathcal{U}} \binom{n-d}{d-|u|} = \Theta_n(n^{(1+\delta)|\mathcal{U}|-\sum_{u\in\mathcal{U}}|u|}) = o_n\left(n^{(\binom{d}{2}-\binom{m}{2})(\delta-1)}\right)$$

by Lemma 2.5. Therefore

$$\Pr[[d] \in X] \le (1 + o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{\binom{d}{2}-\binom{m}{2}},$$

which finishes the proof. ∎

Theorem 2.9. *Suppose $\delta > \frac{d-1}{d+1}$. Suppose m is an integer such that $m \in \{0, 1, 2, d-1\}$ and l is an integer such that $l \ge m$. Let X be the set of elements h of $\binom{[n]}{d}$ such that $h \cap [l] = [m]$ and $\binom{h}{2}\backslash\binom{[m]}{2} \subset e(\Psi)$, where $[0]$ is the empty set. Then Var$[|X|] = o_n(\mathbb{E}[|X|]^2)$.*

Proof. Let $\mathcal{S}$ be the set of elements of $\binom{[n]}{d}$ that have intersection with $[l]$ equal to $[m]$. We have that

$$\mathbb{E}[|X|^2] = \sum_{a,b\in\mathcal{S}} \Pr[a, b \in X].$$

Suppose $m \le k \le d-1$. Suppose $a, b \in \mathcal{S}$ and $k = |a \cap b|$. From Lemma 2.2 with $[k]$ replaced by $a \cup b$ and $\mathcal{E}$ replaced by $(a \cup b)\backslash\binom{[m]}{2}$,

$$\Pr[a, b \in X] = \Pr[(\binom{a}{2} \cup \binom{b}{2})\backslash\binom{[m]}{2} \subset E(\mathcal{H}_c)] \le \sum_{\mathcal{U}} p^{|\mathcal{U}|} \prod_{u\in\mathcal{U}} \binom{n-2d+k}{d-|u|}, \tag{2.2}$$

where the sum is over $\mathcal{U} \subset 2^{a\cup b}$ satisfying the conditions of the lemma. For convenience, denote the set of such $\mathcal{U}$ by $\mathcal{P}$.

Suppose $k > m$. From Lemma 2.6, for $\mathcal{U} \in \mathcal{P}$ we have that

$$(1 + \delta)|\mathcal{U}| - \sum_{u\in\mathcal{U}} |u| < k - m + (d(d-1) - m(m-1))(\delta - 1),$$

where the equality case of the lemma cannot occur. Using (2.2) then gives that

$$\Pr[a, b \in X] = o_n(n^{k-m+(d(d-1)-m(m-1))(\delta-1)}).$$

Suppose $k = m$. From Lemma 2.6, for $\mathcal{U} \in \mathcal{P}$ we have that

$$(1+\delta)|\mathcal{U}| - \sum_{u \in \mathcal{U}} |u| \leq (d(d-1) - m(m-1))(\delta - 1)$$

with equality if and only if $\mathcal{U} = (\binom{a}{2} \cup \binom{b}{2}) \setminus \binom{[m]}{2}$. Using (2.2) then gives that

$$\Pr[a, b \in X] \leq (1 + o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{d(d-1)-m(m-1)}.$$

We therefore have that

$$\begin{aligned}
\mathbb{E}_{\mathcal{H}}[|X|^2] = \sum_{a,b \in \mathcal{S}} \Pr[a, b \in X] &= \sum_{k=m}^{d} \sum_{a,b \in \mathcal{S},\, |a \cap b| = k} \Pr[a, b \in X] \\
\leq &\mathbb{E}[|X|] + \sum_{k=m+1}^{d-1} \binom{n-l}{d-m} \binom{d-m}{k-m} \binom{n-l-d+m}{d-k} o_n(n^{k-m+d(d-1)(\delta-1)}) \\
&+ \binom{n-l}{d-m} \binom{n-l-d+m}{d-m} (1 + o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{d(d-1)-m(m-1)} \\
= &\mathbb{E}[|X|] + (1 + o_n(1)) \binom{n}{d-m}^2 \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{d(d-1)-m(m-1)}.
\end{aligned}$$

Furthermore, from Theorem 2.8,

$$\mathbb{E}[|X|] \geq (1 - o_n(1)) \binom{n}{d-m} \left(\frac{cn^{\delta-1}}{(d-2)!} \right)^{\binom{d}{2} - \binom{m}{2}}.$$

It follows that

$$\mathrm{Var}[|X|] = \mathbb{E}[|X|^2] - \mathbb{E}[|X|]^2 = o_n(\mathbb{E}[|X|]^2),$$

which finishes the proof. ■

Corollary 2.10. *Suppose $\delta > \frac{d-1}{d+1}$. Then $Var[|e(\mathcal{H}_c)|] = o_n(\mathbb{E}[|e(\mathcal{H}_c)|]^2)$.*

Proof. This follows from Theorem 2.9 with $l = m = 0$. ■

Remark 2.11. We expect Corollary 2.10 to be true for $\delta \leq \frac{d-1}{d+1}$ as well although we omit a rigorous proof. For $\delta < \frac{d-1}{d+1}$, the idea is that almost all hyperedges of $\mathcal{H}_c$ are hyperedges of $\mathcal{H}$, the number of which we know is concentrated.

Chapter 3

Partial recovery

3.1 Dense regime

In this section we consider when $\delta \geq 1$, in which case $\mathcal{H}$ is very dense. In any other parts of the paper, it is assumed that $\delta < 1$.

Theorem 3.1. *Suppose $1 \leq \delta < d-1$. The partial recovery loss is $1 - o_n(1)$.*

Proof. This follows from Lemma 1.10 and Theorem 3.2. ■

Theorem 3.2. *Suppose $1 \leq \delta < d-1$. The weighted partial recovery loss is $1 - o_n(1)$.*

Proof. Suppose the function $f : \mathcal{G}_W \to \{0,1\}$ satisfies the condition that if $G \in \mathcal{G}$ then $f(G) = \mathbf{1}_{[d] \in \mathcal{B}^*(G)}$. Using (1.2) gives that

$$\mathbb{E}[|\mathcal{B}^*(\mathrm{Proj}_W(\mathcal{H}))\Delta\mathcal{H}|] = \binom{n}{d} \Pr_{\mathcal{H}}[f(\mathrm{Proj}_W(\mathcal{H})) \neq \mathbf{1}_{[d]\in e(\mathcal{H})}]. \tag{3.1}$$

From Fano's inequality,

$$H(\mathbf{1}_{[d]\in e(\mathcal{H})}|\mathrm{Proj}_W(\mathcal{H})) \leq H_B(\Pr_{\mathcal{H}}[f(\mathrm{Proj}_W(\mathcal{H})) \neq \mathbf{1}_{[d]\in e(\mathcal{H})}]).$$

Using this gives that

$$\begin{aligned}\binom{n}{d} H_B(\Pr_{\mathcal{H}}[f(\mathrm{Proj}_W(\mathcal{H})) \neq \mathbf{1}_{[d]\in e(\mathcal{H})}]) &\geq \sum_{h \in \binom{[n]}{d}} H(\mathbb{1}\{h \in \mathcal{H}\})|\mathrm{Proj}_W(\mathcal{H})) \\ &\geq H(\mathcal{H}|\mathrm{Proj}_W(\mathcal{H})) = H(\mathcal{H}) - H(\mathrm{Proj}_W(\mathcal{H})).\end{aligned} \tag{3.2}$$

First suppose $\delta > 1$. Then, for all $i, j \in [n]$, $i \neq j$, we have that

$$H(|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}|) \leq \log\left(\binom{n-2}{d-2}\right).$$

Thus,

$$H(\mathrm{Proj}_W(\mathcal{H})) \leq \sum_{1 \leq i < j \leq n} H(|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}|) \leq \binom{n}{2} \log\left(\binom{n-2}{d-2}\right).$$

Since $H(\mathcal{H}) = \binom{n}{d} H_B(p) = \Omega_n(n^{1+\delta}\log(n))$, we then have that $H(\mathrm{Proj}_W(\mathcal{H})) = o_n(H(\mathcal{H}))$. Hence, (3.2) gives that $\Pr[f(\mathrm{Proj}_W(\mathcal{H})) \neq \mathbf{1}_{[d]\in e(\mathcal{H})}] = (1-o_n(1))p$ so using (3.1) gives that

$$\ell(\mathcal{B}^*) = \frac{\mathbb{E}_{\mathcal{H}}[|\mathcal{B}^*(\Psi)\Delta\mathcal{H}|]}{p\binom{n}{d}} = \frac{\Pr_{\mathcal{H}}[f(\Psi) \neq \mathbf{1}_{[d]\in e(\mathcal{H})}]}{p} = 1 - o_n(1).$$

Next suppose $\delta = 1$. Suppose $i, j \in [n]$, $i \neq j$. We have that

$$|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}| \sim \text{Binomial}\left(\binom{n-2}{d-2}, p\right)$$

so

$$H(|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}|) = O_n(1)$$

since $|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}|$ has mean $\binom{n-2}{d-2}p = O_n(1)$, see [18, Exercise I.4]. Then,

$$H(\mathrm{Proj}_W(\mathcal{H})) \leq \sum_{1\leq i<j\leq n} H(|\{h \in E(\mathcal{H}) : \{i,j\} \subset h\}|) = O_n(n^2).$$

Since $H(\mathcal{H}) = \Omega_n(n^2\log(n))$, $H(\mathrm{Proj}_W(\mathcal{H})) = o_n(H(\mathcal{H}))$ and we conclude similarly as the case $\delta > 1$. ■

3.2 Preliminary results

Theorem 3.3. *Recall that $q = \Pr[\binom{[d]}{2} \subseteq \Psi]$. Then*

$$q = (1+o_n(1))\left(p + \left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}} + o_n\left(p + n^{\binom{d}{2}(\delta-1)}\right)\right).$$

Proof. First we lower bound q. Note that if $[d] \in e(\mathcal{H})$ then $[d] \in e(\mathcal{H}_c)$ so $q \geq p$. Suppose $[d] \notin e(\mathcal{H})$; this event occurs with probability $1-p$. Then, $[d] \in e(\mathcal{H}_c)$ if and only if for each edge $\{i,j\}$ for $1 \leq i < j \leq d$, there exists $h \in \binom{[n]}{d}\backslash\{[d]\}$ such that $\{i,j\} \subset h$ and $h \in e(\mathcal{H})$.

Suppose $1 \leq i < j \leq d$. The probability there exists $h \in \binom{[n]}{d}\backslash\{[d]\}$ such that $\{i,j\} \subset h$ and $h \in e(\mathcal{H})$ is $1-(1-p)^{\binom{n-2}{d-2}-1}$. Using the Harris inequality gives that

$$\begin{aligned}
\Pr[[d] \in e(\mathcal{H}_c) | [d] \notin e(\mathcal{H})] &\geq \left(1-(1-p)^{\binom{n-2}{d-2}-1}\right)^{\binom{d}{2}} \\
&\geq \left(\binom{n-2}{d-2}p - p - O_n((n^{d-2}p)^2)\right)^{\binom{d}{2}} \\
&\geq (1-o_n(1))\left(\binom{n-2}{d-2}p\right)^{\binom{d}{2}} \\
&= (1-o_n(1))\left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}}.
\end{aligned}$$

Hence,

$$q = \Pr[[d] \in e(\mathcal{H}_c)] \geq p + (1-p)(1-o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}}$$
$$= p + (\frac{cn^{(\delta-1)}}{(d-2)!})^{\binom{d}{2}} + o_n(n^{\binom{d}{2}(\delta-1)}). \tag{3.3}$$

Next we upper bound q using the technique discussed in Section 2.1. From Lemma 2.2 with $[k]$ replaced by $[d]$ and $\mathcal{E}$ replaced by $\binom{[d]}{2}$,

$$q = \Pr[[d] \in e(\mathcal{H}_c)] \leq \sum_{\mathcal{U}} p^{|\mathcal{U}|} \prod_{u \in \mathcal{U}} \binom{n-d}{d-|u|},$$

where the sum is over $\mathcal{U} \subset 2^{[d]}$ satisfying the conditions of the lemma. For convenience, denote the set of such $\mathcal{U}$ by $\mathcal{P}$.

Suppose $\mathcal{U} \in \mathcal{P}$ and $M = |\mathcal{U}|$. It is clear that $1 \leq M \leq \binom{d}{2}$. If $M = 1$ ($\mathcal{U} = \{[d]\}$) the probability is p and if $M = \binom{d}{2}$ ($\mathcal{U} = \binom{[d]}{2}$) then the probability is $(1+o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}}$. Using the union bound gives that

$$q \leq p + (1+o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}} + \sum_{\substack{2 \leq M \leq \binom{d}{2}-1, \\ \mathcal{U} \in \mathcal{P}, |\mathcal{U}|=M}} \Theta_n(n^{(1+\delta)M - \sum_{u \in \mathcal{U}} |u|}). \tag{3.4}$$

Suppose $\mathcal{U} \in \mathcal{P}$ and $M = |\mathcal{U}|$; recall that $1 \leq M \leq \binom{d}{2}$. Furthermore suppose $y_u = \binom{|u|}{2}$ for $u \in \mathcal{U}$. Then $y_u \geq 1$ for $u \in \mathcal{U}$ and $\sum_{u \in \mathcal{U}} y_u \geq \binom{d}{2}$. Furthermore

$$M(1+\delta) - \sum_{u \in \mathcal{U}} |u| = M(1+\delta) - \sum_{u \in \mathcal{U}} \frac{1 + \sqrt{8y_u + 1}}{2}. \tag{3.5}$$

Suppose $1 \leq M \leq \binom{d}{2}$. Let $\mathcal{R}_M$ be the set of $(y_i)_{1 \leq i \leq M}$ such that $y_i \geq 1$ for $1 \leq i \leq M$ and $\sum_{i=1}^{M} y_i \geq \binom{d}{2}$. Let

$$f(y_1, \ldots, y_M) = M(1+\delta) - \sum_{i=1}^{m} \frac{1 + \sqrt{8y_i + 1}}{2}.$$

An upper bound of (3.5) is the maximal value of bound f over $\mathcal{R}_M$. Since f is convex, this maximal value occurs at the vertex $y_i = 1$ for $1 \leq i \leq M-1$ and $y_M = \binom{d}{2} - M + 1$. The value of f at this vertex is

$$g(M) := M(1-\delta) + 2 - \frac{1 + \sqrt{8(\binom{d}{2} - M + 1) + 1}}{2}$$

and $\max_{y \in \mathcal{R}_M} f(y) = g(M)$. Observe that g is convex in M over $[1, \binom{d}{2}]$. Hence, the maximum value of g for $M \in [1, \binom{d}{2}]$ is $g(1) = -d + 1 + \delta$ or $g(\binom{d}{2}) = \binom{d}{2}(\delta - 1)$. Particularly, if $1 < M < \binom{d}{2}$ then $g(M) < \max(-d+1+\delta, \binom{d}{2}(\delta-1))$. Using the fact that an upper bound of (3.5) is $g(M)$ then gives that if $\mathcal{U} \in \mathcal{P}$ and $1 < |\mathcal{U}| < \binom{d}{2}$,

$$|\mathcal{U}|(1+\delta) - \sum_{u \in \mathcal{U}} |u| < \max(-d+1+\delta, \binom{d}{2}(\delta-1)).$$

Using (3.4) then gives that

$$\Pr[[d] \in e(\mathcal{H}_c)] \leq p + (1+o_n(1)) \left(\frac{cn^{\delta-1}}{(d-2)!}\right)^{\binom{d}{2}} + o_n(p + n^{\binom{d}{2}(\delta-1)}).$$

Using this inequality and (3.3) completes the proof. ■

Remark 3.4. Observe that Theorem 3.3 for $\delta > \frac{d-1}{d+1}$ is implied by Theorem 2.8 with $l = m = 0$.

Corollary 3.5. *The asymptotic expressions*

$$q = \begin{cases} (1+o_n(1))p & \text{if } \delta < \frac{d-1}{d+1} \\ p + \Theta_n(p) & \text{if } \delta = \frac{d-1}{d+1} \\ \omega_n(p) & \text{if } \delta > \frac{d-1}{d+1} \end{cases}$$

are true.

Proof. This follows from Theorem 3.3 and the fact that $\binom{d}{2}(\delta-1) > -d+1+\delta$ if and only if $\delta > \frac{d-1}{d+1}$. ■

Lemma 3.6. *Almost exact recovery is possible if and only if*

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H}|\Psi]^2 = p - o_n(p).$$

Furthermore if

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H}|\Psi]^2 = o_n(p)$$

then the partial recovery loss is $1 - o_n(1)$.

Proof. Since

$$1 \leq \frac{\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left|\Pr[[d] \in \mathcal{H}|\Psi = G] - \frac{1}{2}\right|\right)}{\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{4} - \left(\Pr[[d] \in \mathcal{H}|\Psi = G] - \frac{1}{2}\right)^2\right)} \leq 2$$

we have that

$$1 \leq \frac{\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left|\Pr[[d] \in \mathcal{H}|\Psi = G] - \frac{1}{2}\right|\right)}{p - \sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H}|\Psi = G]^2} \leq 2. \tag{3.6}$$

From (1.2), almost exact recovery is possible if and only if

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left| \Pr[[d] \in \mathcal{H} | \Psi = G] - \frac{1}{2} \right| \right) = o_n(p).$$

Then, from (3.6) almost exact recovery is possible if and only if

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H} | \Psi = G]^2 = p - o_n(p).$$

Furthermore from (1.2) the partial recovery loss is $1 - o_n(1)$ if and only if

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left| \Pr[[d] \in \mathcal{H} | \Psi = G] - \frac{1}{2} \right| \right) = p - o_n(p).$$

Then (3.6) gives that

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H} | \Psi = G]^2 = o_n(p)$$

implies that

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \left(\frac{1}{2} - \left| \Pr[[d] \in \mathcal{H} | \Psi = G] - \frac{1}{2} \right| \right) \geq p - o_n(p).$$

This quantity is also at most p, which implies that partial recovery loss is $1 - o_n(1)$. ■

Theorem 3.7. *Almost exact recovery, which means $o_n(1)$ partial recovery loss, is possible if $\delta < \frac{d-1}{d+1}$.*

Proof. Suppose $\delta < \frac{d-1}{d+1}$. We have that

$$\sum_{G \in \mathcal{G}} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H} | \Psi = G]^2 = \sum_{G \in \mathcal{G}, \binom{[d]}{2} \subset e(G)} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H} | \Psi = G]^2.$$

Using Cauchy-Schwarz and Corollary 3.5 gives that

$$\sum_{G \in \mathcal{G}, \binom{[d]}{2} \subset e(G)} \Pr[\Psi = G] \Pr[[d] \in \mathcal{H} | \Psi = G]^2 \geq \frac{p^2}{\Pr[\binom{[d]}{2} \subset e(\Psi)]} = \frac{p^2}{q} = (1 - o_n(1))p.$$

Then, almost exact recovery is possible if $\delta < \frac{d-1}{d+1}$ from Lemma 3.6. ■

3.3 Partial recovery

The goal of this section is to prove the following result.

Theorem 3.8. *Suppose $\delta > \frac{d-1}{d+1}$. Then the partial recovery loss is $1 - o_n(1)$.*

Towards this direction, we analyze when the partial recovery loss is $1 - o_n(1)$ by using Lemma 3.6. First, assume that $\mathcal{H}'$ is sampled from the conditional distribution $p_{\mathcal{H}|\Psi}$ independently of $\mathcal{H}$. Then, we obtain the Markov chain

$$\mathcal{H} \to \mathrm{Proj}(\mathcal{H}) = \mathrm{Proj}(\mathcal{H}') \to \mathcal{H}'.$$

The following lemma is essential.

Lemma 3.9.

$$\begin{aligned}&\binom{n}{d}\sum_{G\in\mathcal{G}}\Pr[\Psi = G]\Pr[[d]\in\mathcal{H}|\Psi = G]^2\\ &= \sum_{\substack{H,H'\in\{0,1\}^{\binom{[n]}{d}},\\ \mathit{Proj}(H)=\mathit{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathit{Proj}(\mathcal{H})=\mathit{Proj}(H)]}|E(H)\cap E(H')|.\end{aligned}$$

Lemma 3.10. *Suppose $\mathcal{U}_n \subset \{0,1\}^{\binom{[n]}{d}}$, $n \geq 1$ satisfy $\Pr[\mathcal{H}\in\mathcal{U}_n] = o_n(1)$. Then*

$$\sum_{\substack{H\in\mathcal{U}_n, H'\in\{0,1\}^{\binom{[n]}{d}}\\ \mathit{Proj}(H)=\mathit{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathit{Proj}(\mathcal{H})=\mathit{Proj}(H)]}|E(H)\cap E(H')| = o_n\left(p\binom{n}{d}\right).$$

Proof. Using the Cauchy-Schwarz inequality gives that

$$\begin{aligned}&\sum_{\substack{H\in\mathcal{U}_n, H'\in\{0,1\}^{\binom{[n]}{d}}\\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})]}|E(H)\cap E(H')| \leq \sum_{H\in\mathcal{U}_n}\Pr[\mathcal{H}=H]e(H)\\ &\leq \left(\sum_{H\in\{0,1\}^{\binom{[n]}{d}}}\Pr[\mathcal{H}=H]\right)^{\frac{1}{2}}\left(\sum_{H\in\mathcal{U}_n}\Pr[\mathcal{H}=H]e(H)^2\right)^{\frac{1}{2}} = o_n\left(p\binom{n}{d}\right).\end{aligned}$$

■

First observe that $e(\mathcal{H})$ is concentrated around its mean $p\binom{n}{d}$ and Corollary 2.10 gives that $e(\mathcal{H}_c)$ is concentrated around its mean $q\binom{n}{d}$. Suppose $\epsilon = o_n(1)$ satisfies $e(\mathcal{H}) \in [(1-\epsilon)p\binom{n}{d}, (1+\epsilon)p\binom{n}{d}]$ and $e(\mathcal{H}_c) \in [(1-\epsilon)q\binom{n}{d}, (1+\epsilon)q\binom{n}{d}]$ with probability $1 - o_n(1)$. Let $\mathcal{Z}$ be the set of $H \in \{0,1\}^{\binom{[n]}{d}}$ such that $e(H) \in [(1-\epsilon)p\binom{n}{d}, (1+\epsilon)p\binom{n}{d}]$ and the number of d-cliques in $\mathrm{Proj}(H)$ is in $[(1-\epsilon)q\binom{n}{d}, (1+\epsilon)q\binom{n}{d}]$. From Lemma 3.10,

$$\begin{aligned}&\sum_{\substack{H,H'\in\{0,1\}^{\binom{[n]}{d}},\\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]}|E(H)\cap E(H')|\\ &= \sum_{\substack{H,H'\in\mathcal{Z},\\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]}|E(H)\cap E(H')| + o_n\left(p\binom{n}{d}\right).\end{aligned} \tag{3.7}$$

Lemma 3.11. *Suppose $\delta > \frac{d-1}{d+1}$. Then,*

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathit{Proj}(H)=\mathit{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathit{Proj}(\mathcal{H})=\mathit{Proj}(H)]^2} \leq \left(\frac{(1+o_n(1))e^{\binom{d}{2}}}{q}\right)^{(1+\epsilon)p\binom{n}{d}}.$$

Proof. Let $\mathcal{G}' = \mathrm{Proj}(\mathcal{Z})$. We have that

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]^2} \leq \sum_{H\in\mathcal{Z}} \frac{\Pr[\mathcal{H}=H]}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]} \leq |\mathcal{G}'|.$$

The maximum number of edges in some $G \in \mathcal{G}'$ is $(1+\epsilon)\binom{d}{2}p\binom{n}{d}$. Therefore,

$$|\mathcal{G}'| \leq (1+\epsilon)\binom{d}{2}p\binom{n}{d}\binom{\binom{n}{2}}{(1+\epsilon)\binom{d}{2}p\binom{n}{d}},$$

if n is sufficiently large so that $(1+\epsilon)\binom{d}{2}p\binom{n}{d} < \frac{1}{2}\binom{n}{2}$. Stirling's approximation gives that $k! \geq \left(\frac{k}{e}\right)^k$. Thus using Theorem 3.3 and the fact that $\delta > \frac{d-1}{d+1}$ gives that

$$\binom{\binom{n}{2}}{(1+\epsilon)\binom{d}{2}p\binom{n}{d}} \leq \left(\frac{e\binom{n}{2}}{(1+\epsilon)\binom{d}{2}p\binom{n}{d}}\right)^{(1+\epsilon)\binom{d}{2}p\binom{n}{d}} = \left(\frac{(1+o_n(1))e^{\binom{d}{2}}}{q}\right)^{(1+\epsilon)p\binom{n}{d}}.$$

■

Lemma 3.12. *Suppose $\delta > \frac{d-1}{d+1}$ and $M \in (0,1)$. Then*

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathit{Proj}(H)=\mathit{Proj}(H'), \\ E(H)\cap E(H') \geq Mp\binom{n}{d}}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathit{Proj}(\mathcal{H})=\mathit{Proj}(H)]} = o_n(1).$$

Proof. If $H' \in \mathcal{Z}$ then

$$\Pr[\mathcal{H}=H'] \leq \left(\frac{p}{1-p}\right)^{(1-\epsilon)p\binom{n}{d}} (1-p)^{\binom{n}{d}}.$$

Suppose $I \in [Mp\binom{n}{d}, (1+\epsilon)p\binom{n}{d}]$ is an integer. Next we upper bound the number of choices for H' given H and $|E(H)\cap E(H')| = I$.

Suppose $H \in \mathcal{Z}$. The number of choices for $E(H)\cap E(H')$ is at most

$$\binom{\lfloor(1+\epsilon)p\binom{n}{d}\rfloor}{I}.$$

Since $E(H') \subset E(H^c)$, the number of choices for $E(H')\backslash E(H)$ is at most

$$\binom{\lfloor(1+\epsilon)q\binom{n}{d}\rfloor}{\lfloor(1+\epsilon)p\binom{n}{d}\rfloor - I}$$

assuming that n is sufficiently large so that $p \leq \frac{q}{2}$. Hence

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H'), \\ E(H)\cap E(H')=I}} \Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']$$

$$\leq \sum_{H\in\mathcal{Z}} \Pr[\mathcal{H}=H]\left(\frac{p}{1-p}\right)^{(1-\epsilon)p\binom{n}{d}}(1-p)^{\binom{n}{d}}\binom{\lfloor(1+\epsilon)p\binom{n}{d}\rfloor}{I}\binom{\lfloor(1+\epsilon)q\binom{n}{d}\rfloor}{\lfloor(1+\epsilon)p\binom{n}{d}\rfloor - I}$$

$$\leq \left(\frac{p}{1-p}\right)^{(1-\epsilon)p\binom{n}{d}} e^{-p\binom{n}{d}}\left(\frac{e(1+\epsilon)}{I/(p\binom{n}{d})}\right)^{I}\left(\frac{e(1+\epsilon)\frac{q}{p}}{1+\epsilon-I/(p\binom{n}{d})}\right)^{(1+\epsilon-I/(p\binom{n}{d}))p\binom{n}{d}}$$

$$\leq (1+o_n(1))p^{(I/(p\binom{n}{d})-2\epsilon)p\binom{n}{d}}q^{(1+\epsilon-I/(p\binom{n}{d}))p\binom{n}{d}}$$

$$\cdot\left(\frac{1+o_n(1)}{(I/(p\binom{n}{d}))^{I/(p\binom{n}{d})}(1+\epsilon-I/(p\binom{n}{d}))^{1+\epsilon-I/(p\binom{n}{d})}}\right)^{p\binom{n}{d}}.$$

Observe that the inequalities are true even for the edge-case $I = \lfloor(1+\epsilon)p\binom{n}{d}\rfloor$. Let $\alpha = \max_{m\in[M,1)}\frac{2}{m^m(1-m)^{1-m}}$. Then if n is sufficiently large,

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H'), \\ E(H)\cap E(H')=I}} \Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H'] \leq (1+o_n(1))p^{(I/(p\binom{n}{d})-2\epsilon)p\binom{n}{d}}q^{(1+\epsilon-I/(p\binom{n}{d}))p\binom{n}{d}}\alpha^{p\binom{n}{d}}. \tag{3.8}$$

Using this inequality, Lemma 3.11, and Cauchy-Schwarz gives that

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H'), \\ Mp\binom{n}{d}\leq|E(H)\cap E(H')|\leq(1+\epsilon)p\binom{n}{d}}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=H]}$$

$$\leq \left(\sum_{Mp\binom{n}{d}\leq I\leq(1+\epsilon)p\binom{n}{d}}\ \sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H'), \\ E(H)\cap E(H')=I}} \Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']\right)^{\frac{1}{2}}$$

$$\cdot\left(\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]^2}\right)^{\frac{1}{2}}$$

$$\leq (\sum_{Mp\binom{n}{d}\leq I\leq(1+\epsilon)p\binom{n}{d}} (1+o_n(1))p^{(I/(p\binom{n}{d})-2\epsilon)p\binom{n}{d}}q^{(1+\epsilon-I/(p\binom{n}{d}))p\binom{n}{d}}\alpha^{p\binom{n}{d}}$$

$$\cdot\left(\frac{(1+o_n(1))e^{\binom{d}{2}}}{q}\right)^{(1+\epsilon)p\binom{n}{d}})^{\frac{1}{2}}$$

$$= \left(\sum_{Mp\binom{n}{d} \le I \le (1+\epsilon)p\binom{n}{d}} \left((1+o_n(1)) \left(\frac{p}{q}\right)^{I/(p\binom{n}{d})} p^{-2\epsilon} \alpha e^{\binom{d}{2}} \right)^{p\binom{n}{d}} \right)^{\frac{1}{2}}.$$

Observe that using Theorem 3.3 gives that

$$\frac{q}{p} \ge (1+o_n(1)) \frac{c^{\binom{d}{2}-1}}{(d-2)!^{\binom{d}{2}}} n^{\binom{d}{2}(\delta-1)+d-1-\delta}.$$

Because $\delta > \frac{d-1}{d+1}$, $\binom{d}{2}(\delta-1)+d-1-\delta > 0$. Suppose $Mp\binom{n}{d} \le I \le (1+\epsilon)p\binom{n}{d}$. Then $I/(p\binom{n}{d}) \ge M$ so

$$(1+o_n(1)) \left(\frac{p}{q}\right)^{I/(p\binom{n}{d})} p^{-2\epsilon} \alpha e^{\binom{d}{2}} = O_n(n^{-M\left(\binom{d}{2}(\delta-1)+d-1-\delta\right)+2\epsilon(d-1-\delta)}).$$

Particualrly, if n is sufficiently large then

$$(1+o_n(1)) \left(\frac{p}{q}\right)^{I/(p\binom{n}{d})} p^{-2\epsilon} \alpha e^{\binom{d}{2}} = O_n(n^{-\frac{M}{2}\left(\binom{d}{2}(\delta-1)+d-1-\delta\right)})$$

since $\epsilon = o_n(1)$. We therefore have that

$$\sum_{\substack{H,H' \in \mathcal{Z}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H'), \\ Mp\binom{n}{d} \le |E(H)\cap E(H')| \le (1+\epsilon)p\binom{n}{d}}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=H]}$$

$$\le \left((1+\epsilon)p\binom{n}{d} O_n(n^{-\frac{M}{2}\left(\binom{d}{2}(\delta-1)+d-1-\delta\right)})^{p\binom{n}{d}} \right)^{\frac{1}{2}} = o_n(1).$$

■

Proof of Theorem 3.8. From Lemma 3.6 and Lemma 3.9 it suffices to prove that

$$\sum_{\substack{H,H' \in \{0,1\}^{\binom{[n]}{d}}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]} |E(H)\cap E(H')| = o_n\left(p\binom{n}{d}\right).$$

From Lemma 3.12, for all $M \in (0,1)$ we have that

$$\sum_{\substack{H,H' \in \{0,1\}^{\binom{[n]}{d}}, \\ \mathrm{Proj}(H)=\mathrm{Proj}(H')}} \frac{\Pr[\mathcal{H}=H]\Pr[\mathcal{H}=H']}{\Pr[\mathrm{Proj}(\mathcal{H})=\mathrm{Proj}(H)]} |E(H)\cap E(H')| \le (M+o_n(1))p\binom{n}{d}.$$

Considering the limit $M \to 0$ completes the proof. ■

3.4 Weighted partial recovery

For weighted partial recovery, we can prove a similar result as Theorem 3.8 using a very similar method. This is the statement of the result.

Theorem 3.13. *Suppose $\delta > \frac{d-1}{d+1}$. Then the weighted partial recovery loss is $1 - o_n(1)$.*

Proof. Suppose $k \geq 1$. For $i, j \in [n]$, $i \neq j$, the probability that $\{i, j\}$ is contained in at least k hyperedges of $\mathcal{H}$ is

$$(1-p)^{\binom{n-2}{d-2}} \sum_{m=k}^{\binom{n-2}{d-2}} \binom{\binom{n-2}{d-2}}{m} \left(\frac{p}{1-p}\right)^m \leq \sum_{m \geq k} \left(\binom{n-2}{d-2} p\right)^m = O_n(n^{k(\delta-1)}).$$

Then, the expected number of $\{i, j\}$ that are contained in at least k hyperedges of $\mathcal{H}$ is $O_n(n^{2+k(\delta-1)})$. By selecting k to be sufficiently large, this expected value will be $o_n(1)$, so the probability that there exists an edge $\{i, j\}$ that is contained in at least k hyperedges is $o_n(1)$.

Suppose k is sufficiently large. Then, set $\mathcal{Z}'$ to be the set of $H \in \mathcal{Z}$ such that each edge $\{i, j\}$ is contained in less than k hyperedges. We have that $\Pr[\mathcal{H} \in \mathcal{Z}'] = 1 - o_n(1)$, so we can use $\mathcal{Z}'$ in place of $\mathcal{Z}$.

We can now essentially use the same proof as Theorem 3.8, of course after replacing Proj with Proj_W and $\mathcal{Z}$ with $\mathcal{Z}'$. The only significant step is to prove the analogue of Lemma 3.12; the remaining steps are straightforward to verify. For this proof, we can follow the framework given in the proof of Lemma 3.12.

The most important step is to justify the analogue of Lemma 3.11. Letting $\mathcal{G}'' = \mathrm{Proj}_W(\mathcal{Z}')$, since each edge of $\mathrm{Proj}(H)$ is contained in less than k hyperedges and the number of edges in $\mathrm{Proj}(H)$ is at most $(1+\epsilon)\binom{d}{2}p\binom{n}{d}$ for $H \in \mathcal{Z}'$, we have that

$$|\mathcal{G}''| \leq (k-1)^{(1+\epsilon)\binom{d}{2}p\binom{n}{d}} |\mathcal{G}'|,$$

so using Lemma 3.11 gives that

$$\sum_{\substack{H, H' \in \mathcal{Z}', \\ \mathrm{Proj}_W(H) = \mathrm{Proj}_W(H')}} \frac{\Pr[\mathcal{H} = H] \Pr[\mathcal{H} = H']}{\Pr[\mathrm{Proj}_W(\mathcal{H}) = \mathrm{Proj}_W(H)]^2} \leq |\mathcal{G}''| \leq \left(\frac{(1 + o_n(1))(ke)^{\binom{d}{2}}}{q}\right)^{(1+\epsilon)p\binom{n}{d}}.$$

Furthermore, it is clear that

$$\sum_{\substack{H, H' \in \mathcal{Z}', \\ \mathrm{Proj}_W(H) = \mathrm{Proj}_W(H'), \\ |E(H) \cap E(H')| = I}} \Pr[\mathcal{H} = H] \Pr[\mathcal{H} = H'] \leq \sum_{\substack{H, H' \in \mathcal{Z}, \\ \mathrm{Proj}(H) = \mathrm{Proj}(H'), \\ |E(H) \cap E(H')| = I}} \Pr[\mathcal{H} = H] \Pr[\mathcal{H} = H']$$

for all $I \in [Mp\binom{n}{d}, (1+\epsilon)p\binom{n}{d}]$, so the analogue

$$\sum_{\substack{H, H' \in \mathcal{Z}, \, \mathrm{Proj}_W(H) \\ = \mathrm{Proj}_W(H'), \\ |E(H) \cap E(H')| = I}} \Pr[\mathcal{H} = H] \Pr[\mathcal{H} = H'] \leq (1 + o_n(1)) p^{(I/(p\binom{n}{d})-2\epsilon)p\binom{n}{d}} q^{(1+\epsilon-I/(p\binom{n}{d}))p\binom{n}{d}} \alpha^{p\binom{n}{d}}$$

of (3.8) is true. Afterwards, we can follow the same steps to prove the analogue of Lemma 3.12. ■

3.5 Correlation inequality

It is in fact possible to extend the results from Theorem 3.8 from one hyperedge to multiple hyperedges. As explained in the proof of the following result, the special case $k = 1$ is equivalent to Theorem 3.8.

Theorem 3.14. *Suppose Suppose $\delta > \frac{d-1}{d+1}$ and $k \geq 1$. Then*

$$\sup_{\substack{h_1,\ldots,h_k \in \binom{[n]}{d} \\ \text{distinct}}} \sum_{G \in \mathcal{G}} \Pr[h_1, \ldots, h_k \in E(\mathcal{H}) | \mathit{Proj}(\mathcal{H}) = G]^2 \Pr[\mathit{Proj}(\mathcal{H}) = G] = o_n(p^k).$$

Proof. First we resolve the case $k = 1$. From Lemma 3.6, it suffices to prove that $\mathbb{E}_{\mathcal{H},\mathcal{H}'}[E(\mathcal{H}) \cap E(\mathcal{H}')] = o_n\left(p\binom{n}{d}\right)$, which we show in the proof of Theorem 3.8.

Next, assume that $k \geq 2$. Suppose $h_1, \ldots, h_k \in \binom{[n]}{d}$. Suppose S_i, $1 \leq i \leq k$ are disjoint sets of $\lfloor \frac{n}{k} \rfloor - d$ vertices that are disjoint from h_i, $1 \leq i \leq k$, assuming $n \geq kd$. Let $T_i = S_i \cup h_i$ for $1 \leq i \leq k$. Let $\mathcal{H}_i$ be $\mathcal{H}$ with vertex set restricted to T_i for $1 \leq i \leq k$. Observe that the $\mathcal{H}_i$ do not have any overlapping hyperedges. Furthermore, let $\mathcal{H}^C$ denote $\mathcal{H} \setminus \left(\bigcup_{i=1}^k \mathcal{H}_i\right)$, that is, $\mathcal{H}^C$ is $\mathcal{H}$ restricted to $\binom{[n]}{d} \setminus \left(\bigcup_{i=1}^k \binom{T_i}{d}\right)$.

Suppose $G \in \mathcal{G}$. Where G_i is some graph with vertex set T_i for $1 \leq i \leq k$ and G^C is some graph with vertex set $[n]$, we have that

$$\Pr[\mathrm{Proj}(\mathcal{H}) = G] = \sum_{\substack{G_i, 1 \leq i \leq k,\, G^C, \\ \text{projection is } G}} \prod_{i=1}^k \Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i] \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C]$$

and

$$\begin{aligned} &\Pr[h_1, \ldots, h_k \in E(\mathcal{H}), \mathrm{Proj}(\mathcal{H}) = G] \\ &= \sum_{\substack{G_i, 1 \leq i \leq k,\, G^C, \\ \text{projection is } G}} \prod_{i=1}^k \Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)] \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C]. \end{aligned}$$

Using the Cauchy-Schwarz inequality gives that

$$\begin{aligned} &\Pr[h_1, \ldots, h_k \in E(\mathcal{H}) | \mathrm{Proj}(\mathcal{H}) = G]^2 \Pr[\mathrm{Proj}(\mathcal{H}) = G] \\ &= \frac{\left(\sum_{\substack{G_i, 1 \leq i \leq k,\, G^C, \\ \text{projection is } G}} \prod_{i=1}^k \Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)] \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C]\right)^2}{\sum_{\substack{G_i, 1 \leq i \leq k,\, G^C, \\ \text{projection is } G}} \prod_{i=1}^k \Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)] \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C]} \\ &\leq \sum_{\substack{G_i, 1 \leq i \leq k, G^C, \\ \text{projection is } G}} \prod_{i=1}^k \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]} \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C]. \end{aligned}$$

Afterwards summing over the G gives that

$$\begin{aligned}
&\sum_{G\in\mathcal{G}} \Pr[h_1,\dots,h_k \in E(\mathcal{H}) | \mathrm{Proj}(\mathcal{H}) = G]^2 \Pr[\mathrm{Proj}(\mathcal{H}) = G] \\
&\le \sum_{G\in\mathcal{G}} \sum_{\substack{G_i, 1\le i\le k, G^C,\\ \text{projection is } G}} \prod_{i=1}^{k} \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]} \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C] \\
&= \sum_{G_i, 1\le i\le k, G^C} \prod_{i=1}^{k} \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]} \Pr[\mathrm{Proj}(\mathcal{H}^C) = G^C] \\
&= \sum_{G_i, 1\le i\le k} \prod_{i=1}^{k} \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]} \\
&= \prod_{i=1}^{k} \sum_{G_i} \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]}.
\end{aligned}$$

Observe that the $\mathcal{H}_i$ follow the same random model as $\mathcal{H}$ but with a different value of p. The $\mathcal{H}_i$ have $\lfloor \frac{n}{k} \rfloor$ vertices and each hyperedge appears with probability $(c + o_n(1))n^{-d+1+\delta}$. Thus, when the $\mathcal{H}_i$ have n vertices, each hyperedge appears with probability $(c+o_n(1))k^{-d+1+\delta}n^{-d+1+\delta}$. From repeating the case $k = 1$ for this random model,

$$\begin{aligned}
&\sum_{G\in\mathcal{G}} \Pr[h_1,\dots,h_k \in E(\mathcal{H}) | \mathrm{Proj}(\mathcal{H}) = G]^2 \Pr[\mathrm{Proj}(\mathcal{H}) = G] \\
&\le \prod_{i=1}^{k} \sum_{G_i} \frac{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i, h_i \in E(\mathcal{H}_i)]^2}{\Pr[\mathrm{Proj}(\mathcal{H}_i) = G_i]} = o_n(p^k).
\end{aligned}$$

■

Remark 3.15. The previous proof exhibits an advantage of considering the regime $p = (c + o_n(1))n^{-d+1+\delta}$.

Chapter 4

Ambiguous graph results

4.1 Introduction to ambiguous graphs

The goal of this chapter is to prove combinatorial results that will eventually justify lower bounds of the exact recovery thresholds. We apply these results in Chapter 5 to prove Theorems 1.7 and 1.9. First, in Definition 4.1, we define an ambiguous graph, which is introduced in the paper [5]. Afterwards, in Definition 4.2, we define the ambiguous graph $G_{a,d}$, which is also introduced in the paper.

Definition 4.1. *A minimal preimage of a graph G is a hypergraph H such that $Proj(H) = G$ and $|e(H)|$ is minimal. A graph G is ambiguous if there exists two distinct minimal preimages of G.*

Definition 4.2. *Define the hypergraph $H = (V, E)$ as follows:*

- $V = \{w_1, w_2\} \sqcup \{v_1, \ldots, v_{d-1}\} \bigsqcup_{i=1}^{d-1} S_i^1 \bigsqcup_{i=1}^{d-1} S_i^2$.
- *E consists of $\{w_1, v_1, \ldots, v_{d-1}\}$, $\{w_1, v_i\} \sqcup S_i^1$ for $1 \leq i \leq d-1$, and $\{w_2, v_i\} \sqcup S_i^2$ for $1 \leq i \leq d-1$.*

Then, define $G_{a,d} := Proj(H)$.

As explained in [5, Lemma 27], H and H' are minimal preimages of $G_{a,d}$, where H' is H with the hyperedge $\{w_1, v_1, \ldots, v_{d-1}\}$ replaced by $\{w_2, v_1, \ldots, v_{d-1}\}$.

The following result is the main result of this chapter. We dedicate the sections of this chapter to proving various cases of the theorem for $d \geq 4$.

Theorem 4.3. *Suppose $d \geq 3$. For any preimage h of an ambiguous graph, $d - 1 - \frac{1}{m(h)} \geq \frac{2d-4}{2d-1}$.*

Proof. The $d = 3$ case is resolved in [5, Appendix D]. The cases $d = 4$ and $d = 5$ follow from Lemmas 4.8 and 4.9, respectively. ∎

Remark 4.4. Observe that $d - 1 - \frac{1}{m(h)} = \frac{2d-4}{2d-1}$ for the minimal preimage h of $G_{a,d}$, so Theorem 4.3 implies that $G_{a,d}$ is the optimal ambiguous graph in the context of the theorem.

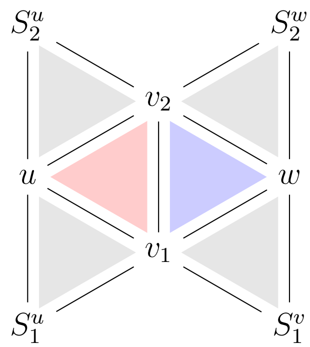

Figure 4.1: This figure displays $G_{a,3}$. One minimal preimage consists of the gray and red triangles while the other consists of the gray and blue triangles. For general $d \geq 3$, the set $\{v_1, v_2\}$ is replaced with $\{v_1, \ldots, v_{d-1}\}$. Then, the red and blue triangles are replaced by d-cliques. Furthermore, the sets S_i^u and S_i^w are replaced by $(d-2)$-cliques for $1 \leq i \leq d-1$.

4.2 Optimization result

Lemma 4.5. *Suppose $d \geq 3$ and $\gamma \in [\frac{d-1}{d+1}, \infty)$. Assume that h is a d-uniform hypergraph. Assume that the set of edges of h is $E_h \sqcup I$, where E_h and I are disjoint. Suppose U is a set of vertices of h such that each hyperedge in E_h is a subset of U. Suppose $\mathcal{P}$ is a set of edges such that for all $\{a,b\} \in \mathcal{P}$, $\{a,b\} \subset U$ and there exists $i \in I$ such that $\{a,b\} \subset i$. If*

$$\frac{(d-1)|E_h| - |U| + |\mathcal{P}|}{|E_h| + |\mathcal{P}|} \geq \gamma,$$

then $d - 1 - \frac{1}{m(h)} \geq \min(\gamma, 1)$.

Proof. Assume that

$$\frac{(d-1)|E_h| - |U| + |\mathcal{P}|}{|E_h| + |\mathcal{P}|} \geq \gamma.$$

For all $i \in I$, let $x_i = |i \cap U|$. Let Z be the set of $i \in I$ such that $x_i \leq 1$. Let V be the set of vertices in U or some hyperedge in $I \backslash Z$; that is, $V = U \bigcup_{i \in I \backslash Z} i$. Suppose h' is the subgraph of h that is induced by V.

Start with the vertex set U. The number of vertices is $|U|$. After adding the hyperedges $i \in I \backslash Z$, the number of vertices $|V|$ satisfies

$$|V| \leq |U| + \sum_{i \in I \backslash Z} |i \backslash U| = |U| + \sum_{i \in I \backslash Z} (d - x_i).$$

Therefore,

$$\alpha(h') \geq \frac{|E_h| + |I \backslash Z|}{|V|} \geq \frac{|E_h| + |I \backslash Z|}{|U| + \sum_{i \in I \backslash Z} (d - x_i)}.$$

Note that

$$d - 1 - \frac{1}{m(h)} \geq d - 1 - \frac{1}{\alpha(h')}$$

$$
\begin{aligned}
&\geq d-1-\frac{|U|+\sum_{i\in I\backslash Z}(d-x_i)}{|E_h|+|I\backslash Z|}\\
&=\frac{(d-1)|E_h|-|U|+(d-1)|I\backslash Z|-\sum_{i\in I\backslash Z}(d-x_i)}{|E_h|+|I\backslash Z|}\\
&=\frac{(d-1)|E_h|-|U|-|I\backslash Z|+\sum_{i\in I\backslash Z}x_i}{|E_h|+|I\backslash Z|}.
\end{aligned}
$$

Hence, it suffices to prove that

$$
\frac{(d-1)|E_h|-|U|-|I\backslash Z|+\sum_{i\in I\backslash Z}x_i}{|E_h|+|I\backslash Z|}\geq \min(\gamma,1).
$$

Next we use the technique from Section 2.1 to finish the proof. Observe that

$$
\sum_{i\in I\backslash Z}\binom{x_i}{2}\geq |\mathcal{P}|
$$

from the definition of $\mathcal{P}$. For all $i\in I\backslash Z$, let $y_i=\binom{x_i}{2}$. Note that

$$
x_i=\frac{1+\sqrt{1+8y_i}}{2}
$$

for all $i\in I\backslash Z$. Because $x_i\geq 2$ for all $i\in I\backslash Z$, $1\leq y_i\leq\binom{d}{2}$ for all $i\in I\backslash Z$. We have that

$$
\begin{aligned}
&\frac{(d-1)|E_h|-|U|-|I\backslash Z|+\sum_{i\in I\backslash Z}x_i}{|E_h|+|I\backslash Z|}\\
&\geq \min_{\substack{1\leq y_i\leq\binom{d}{2},\, i\in I\backslash Z,\\ \sum_{i\in I\backslash Z}y_i\geq|\mathcal{P}|}}\frac{(d-1)|E_h|-|U|-|I\backslash Z|+\sum_{i\in I\backslash Z}\frac{1+\sqrt{1+8y_i}}{2}}{|E_h|+|I\backslash Z|}.
\end{aligned}
$$

Hence, it suffices to prove that

$$
\min_{\substack{1\leq y_i\leq\binom{d}{2},\, i\in I\backslash Z,\\ \sum_{i\in I\backslash Z}y_i\geq|\mathcal{P}|}}\frac{(d-1)|E_h|-|U|-|I\backslash Z|+\sum_{i\in I\backslash Z}\frac{1+\sqrt{1+8y_i}}{2}}{|E_h|+|I\backslash Z|}\geq\min(\gamma,1).
$$

Assume that $M=|I\backslash Z|$ and replace $I\backslash Z$ with $\{1,\dots,M\}$, for simplicity. Furthermore, let $\mathcal{R}_M=\{(y_i)_{1\leq i\leq M}:1\leq y_i\leq\binom{d}{2},1\leq i\leq M,\sum_{i=1}^M y_i\geq|\mathcal{P}|\}$ for $M\geq 1$.
Case 1: $M=0$

If $M=0$, then $|\mathcal{P}|=0$, so

$$
\begin{aligned}
\frac{(d-1)|E_h|-|U|-M+\sum_{i=1}^M\frac{1+\sqrt{1+8y_i}}{2}}{|E_h|+M}&=\frac{(d-1)|E_h|-|U|}{|E_h|}\\
&=\frac{(d-1)|E_h|-|U|+|\mathcal{P}|}{|E_h|+|\mathcal{P}|}\geq\gamma.
\end{aligned}
$$

Case 2: $1 \le M \le |\mathcal{P}|$

Suppose $1 \le M \le |\mathcal{P}|$. Observe that

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M}$$

is concave in $(y_i)_{1\le i\le M}$, so the function is minimized over $\mathcal{R}_M$ at a vertex.

At the vertex, $M-1$ of the values must be elements of $\{1, \binom{d}{2}\}$. Without loss of generality, assume that $y_j \in \{1, \binom{d}{2}\}$ for $1 \le j \le M-1$. Assume that A of these values equal $\binom{d}{2}$ and $M - A - 1$ equal 1. Then, since $\sum_{i=1}^{M} y_i \ge |\mathcal{P}|$,

$$y_M - 1 + A\left(\binom{d}{2} - 1\right) \ge |\mathcal{P}| - M \Rightarrow (d-2)A + \frac{2(y_M - 1)}{d+1} \ge \frac{2(\mathcal{P} - M)}{d+1}. \tag{4.1}$$

Furthermore,

$$\begin{aligned}
&\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} \\
&= \frac{(d-1)|E_h| - |U| - M + Ad + 2M - 2A - 2 + \frac{1+\sqrt{1+8y_M}}{2}}{|E_h| + M} \\
&= \frac{(d-1)|E_h| - |U| + |\mathcal{P}| - (|\mathcal{P}| - M) + Ad - 2A + \frac{-3+\sqrt{1+8y_M}}{2}}{|E_h| + |\mathcal{P}| - (|\mathcal{P}| - M)}.
\end{aligned}$$

Let $X = (d-1)|E_h| - |U| + |\mathcal{P}|$, $Y = |E_h| + |\mathcal{P}|$, $W = |\mathcal{P}| - M$, and $\Delta = Ad - 2A + \frac{-3+\sqrt{1+8y_M}}{2}$. Then,

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} = \frac{X - W + \Delta}{Y - W},$$

and we know that $\frac{X}{Y} \ge \gamma \ge \frac{d-1}{d+1}$. Observe that

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} - \frac{X}{Y} = \frac{W(X-Y) + \Delta Y}{(Y-W)Y} = \frac{W(\frac{X-Y}{Y}) + \Delta}{Y - W}.$$

Because $\frac{X-Y}{Y} \ge \gamma - 1 \ge -\frac{2}{d+1}$,

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} - \frac{X}{Y} \ge \frac{\Delta - \frac{2W}{d+1}}{Y - W}.$$

The goal is to prove that $\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} \ge \frac{X}{Y}$. For this, it suffices to prove that $\Delta \ge \frac{2W}{d+1}$.

Since $\Delta = (d-2)A + \frac{-3+\sqrt{1+8y_M}}{2}$, in order to prove that $\Delta \ge \frac{2W}{d+1}$, using (4.1) gives that it suffices to prove that

$$\frac{-3 + \sqrt{1 + 8y_M}}{2} \ge \frac{2(y_M - 1)}{d+1},$$

where $1 \leq y_M \leq \binom{d}{2}$. We have that $f(x) = \frac{-3+\sqrt{1+8x}}{2} - \frac{2(x-1)}{d+1}$ is concave, so $f(x)$ is minimized over the interval $[1, \binom{d}{2}]$ at its endpoints. Since $f(1) = f(\binom{d}{2}) = 0$, $f(x) \geq 0$ over $[1, \binom{d}{2}]$, which shows that $\Delta \geq \frac{2W}{d+1}$. Thus,

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} \geq \frac{X}{Y} \geq \gamma.$$

Case 3: $M > |\mathcal{P}|$

Next, suppose $M > |\mathcal{P}|$. Then,

$$\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M}$$

is minimized over $\mathcal{R}_M$ at $y_i = 1$ for $1 \leq i \leq M$. Therefore, over $\mathcal{R}_M$ we have that

$$\begin{aligned}\frac{(d-1)|E_h| - |U| - M + \sum_{i=1}^{M} \frac{1+\sqrt{1+8y_i}}{2}}{|E_h| + M} &\geq \frac{(d-1)|E_h| - |U| + M}{|E_h| + M} \\ &= \frac{(d-1)|E_h| - |U| + |\mathcal{P}| + (M - |\mathcal{P}|)}{|E_h| + |\mathcal{P}| + (M - |\mathcal{P}|)} \\ &\geq \min(\gamma, 1).\end{aligned}$$

We are done. ■

4.3 Unweighted

Suppose h and g are hypergraphs such that $\text{Proj}(h) = \text{Proj}(g)$, $h \neq g$, and $e(h) \geq e(g)$. Note that $G := \text{Proj}(h)$ is not necessarily ambiguous (despite the title of this chapter) and g is not necessarily minimal. In this chapter we derive lower bounds for $d - 1 - \frac{1}{m(h)}$ for different values of d.

Let E_h be the set of hyperedges in h but not g and E_g be the set of hyperedges in g but not h. Furthermore, let I be the set of hyperedges in both h and g.

Let $\mathcal{E}_h$ be the set of edges of $\text{Proj}(E_h)$ and $\mathcal{E}_g$ be the set of edges of $\text{Proj}(E_g)$. Furthermore, let $\mathcal{E}$ be the set of edges of $\text{Proj}(I)$.

Lemma 4.6. *The set $\mathcal{E}_h$ is a subset of $\mathcal{E} \cup \mathcal{E}_g$ and the set $\mathcal{E}_g$ is a subset of $\mathcal{E} \cup \mathcal{E}_g$.*

Proof. Suppose $\{i, j\} \in \mathcal{E}_h$. Then, $\{i, j\}$ is an edge of $\text{Proj}(h) = G$. Hence, $\{i, j\}$ is an edge of $\text{Proj}(g) = G$, which implies that $\{i, j\} \in \mathcal{E} \cup \mathcal{E}_g$. ■

Let $\mathcal{P}$ be the symmetric difference of $\mathcal{E}_h$ and $\mathcal{E}_g$. From Lemma 4.6, $\mathcal{P} \subset \mathcal{E}$. Let U be the set of vertices in some hyperedge in $E_h \cup E_g$. Observe that for all $\{a, b\} \in \mathcal{P}$, $\{a, b\} \subset U$ and because $\mathcal{P} \subset \mathcal{E}$, there exists $i \in I$ such that $\{a, b\} \subset i$. Hence, E_h, I, U, and $\mathcal{P}$ satisfy the conditions of Lemma 4.5.

Let V_h be the set of vertices that are in E_h but not E_g, V_g be the set of vertices that are in E_g but not in E_h, and V_I be the set of vertices that are in both E_g and E_h. Observe that

$$U = V_h \sqcup V_g \sqcup V_I.$$

Suppose $v \in U$. Let $d_h(v)$ be the number of elements of E_h that contain v and $d_h^*(v)$ be the number of $u \in U$ such that $\{u, v\} \in \mathcal{E}_h \backslash \mathcal{E}_g$. Similarly, let $d_g(v)$ be the number of elements of E_g that contain v and $d_g^*(v)$ be the number of $u \in U$ such that $\{u, v\} \in \mathcal{E}_g \backslash \mathcal{E}_h$.

Suppose $v \in V_I$. Let $k(v)$ be the largest positive integer k such that there exists $i_h \in E_h$ and $i_g \in E_g$ such that $v \in i_h, i_g$ and $|i_h \backslash i_g| = |i_g \backslash i_h| = k$. Assume that $i_h \in E_h$ and $i_g \in E_g$ satisfy $v \in i_h, i_g$ and $|i_h \backslash i_g| = k(v)$. We have that $i_h \neq i_g$ so $i_h \backslash i_g$ and $i_g \backslash i_h$ are nonempty. Let $i_h(v) = i_h$ and $i_g(v) = i_g$. If there are multiple choices for (i_h, i_g), we can select one choice randomly. Let $n_h(v)$ be the number of $w \in i_h \backslash i_g$ such that $\{v, w\} \in \mathcal{E}_h \backslash \mathcal{E}_g$ and $n_g(v)$ be the number of $w \in i_g \backslash i_h$ such that $\{v, w\} \in \mathcal{E}_g \backslash \mathcal{E}_h$.

The following lemma is implied by Lemma 4.9. However, we include its proof since its contents motivate later methods.

Lemma 4.7. *Suppose $d \geq 5$. Then $d - 1 - \frac{1}{m(h)} \geq \frac{d-1}{d+1}$.*

Proof. From Lemma 4.5, it suffices to prove that

$$\frac{(d-1)|E_h| - |U| + |\mathcal{P}|}{|E_h| + |\mathcal{P}|} \geq \frac{d-1}{d+1}.$$

Thus, it suffices to prove that

$$d(d-1)|E_h| + 2|\mathcal{P}| \geq (d+1)|U|. \tag{4.2}$$

We have that

$$d|E_h| = \sum_{v \in U} d_h(v) = \sum_{v \in V_h \cup V_I} d_h(v) \text{ and } d|E_g| = \sum_{v \in V_g \cup V_I} d_g(v).$$

Furthermore,

$$2|\mathcal{P}| = \sum_{v \in U} d_h^*(v) + d_g^*(v),$$

so

$$2|\mathcal{P}| = \sum_{v \in V_h} d_h^*(v) + \sum_{v \in V_g} d_g^*(v) + \sum_{v \in V_I} (d_h^*(v) + d_g^*(v)). \tag{4.3}$$

Suppose $v \in V_I$. For simplicty, let $i_h = i_h(v)$ and $i_g = i_g(v)$.

Assume that $n_h(v) = 0$. Suppose $w \in i_h \backslash i_g$. Then, there exists $i \in E_g$ such that $\{u, w\} \in i$ because $\{u, w\} \in \mathcal{E}_h$ and $\{u, w\} \notin \mathcal{E}_h \backslash \mathcal{E}_g$. Because $w \notin i_g$, $i \neq i_g$. As $v \in i, i_g$, $d_g(v) \geq 2$.

Assume that $n_h(v) > 0$. Then, there exists $w \in i_h \backslash i_g$ such that $\{u, w\} \in \mathcal{E}_h \backslash \mathcal{E}_g$, so $d_h^*(v) \geq 1$. Observe that $d_g(v) \geq 1$ because $v \in i_g$.

Hence,

$$d_g(v) \geq 1 + \mathbf{1}_{n_h(v)=0} \text{ and } d_h^*(v) \geq 1 - \mathbf{1}_{n_h(v)=0}.$$

We similarly have that

$$d_h(v) \geq 1 + \mathbf{1}_{n_g(v)=0} \text{ and } d_g^*(v) \geq 1 - \mathbf{1}_{n_g(v)=0}.$$

Let

$$N_h = \sum_{v \in V_I} \mathbf{1}_{n_h(v)=0} \text{ and } N_g = \sum_{v \in V_I} \mathbf{1}_{n_g(v)=0}.$$

We have that $d_h(v) \geq 1$ for all $v \in V_h$ and $d_g(v) \geq 1$ for all $v \in V_g$. Thus,

$$d|E_h| = \sum_{v \in V_h \cup V_I} d_h(v) \geq |V_h| + \sum_{v \in V_I} (1 + \mathbf{1}_{n_g(v)=0}) \geq |V_h| + |V_I| + N_g \tag{4.4}$$

and similarly,

$$d|E_g| \geq |V_g| + |V_I| + N_h. \tag{4.5}$$

Adding (4.4) and (4.5) gives that

$$\begin{aligned} d(|E_h| + |E_g|) &\geq |V_h| + |V_g| + \sum_{v \in V_I} (2 + \mathbf{1}_{n_h(v)=0} + \mathbf{1}_{n_g(v)=0}) \\ &= |V_h| + |V_g| + 2|V_I| + N_h + N_g \\ &= |U| + |V_I| + N_h + N_g. \end{aligned}$$

Since $|E_h| \geq |E_g|$ because g is minimal,

$$d|E_h| \geq \frac{|U| + |V_I| + N_h + N_g}{2}. \tag{4.6}$$

Suppose $v \in V_h$. Suppose $v \in i$ for $i \in h$. We have that for all $w \in i$ such that $v \neq w$, $\{v, w\} \in \mathcal{E}_h \backslash \mathcal{E}_g$ because $v \in V_h$. Therefore, $d_h^*(v) \geq d - 1$. Similarly, if $v \in V_g$, $d_g^*(v) \geq d - 1$. Hence, (4.3) gives that

$$\begin{aligned} 2|\mathcal{P}| &\geq (d-1)|V_h| + (d-1)|V_g| + \sum_{v \in V_I} (2 - \mathbf{1}_{n_h(v)=0} - \mathbf{1}_{n_g(v)=0}) \\ &= (d-1)(|V_h| + |V_g|) + 2|V_I| - N_h - N_g. \end{aligned} \tag{4.7}$$

We have that

$$\begin{aligned} d(d-1)|E_h| + 2|\mathcal{P}| &\geq \frac{(d-1)|U|}{2} + \frac{(d+3)|V_I|}{2} + (d-1)(|V_h| + |V_g|) + \frac{d-3}{2}(N_h + N_g) \\ &= \frac{(d-1)|U|}{2} + \frac{(d+3)|U|}{2} + \frac{d-5}{2}(|V_h| + |V_g|) + \frac{d-3}{2}(N_h + N_g) \\ &= (d+1)|U| + \frac{d-5}{2}(|V_h| + |V_g|) + \frac{d-3}{2}(N_h + N_g) \\ &\geq (d+1)|U|. \end{aligned}$$

This proves that (4.2) is true, which finishes the proof. ■

Lemma 4.8. *Suppose $d = 4$. Then, $d - 1 - \frac{1}{m(h)} \geq \frac{2d-4}{2d-1} = \frac{4}{7}$.*

Proof. For the sake of contradiction, that $d - 1 - \frac{1}{m(h)} < \frac{4}{7} < \frac{3}{5}$, where $\frac{d-1}{d+1} = \frac{3}{5}$. From Lemma 4.5,

$$\frac{(d-1)|E_h| - |U| + |\mathcal{P}|}{|E_h| + |\mathcal{P}|} < \frac{3}{5}.$$

This is equivalent to

$$12|E_h| + 2|\mathcal{P}| < 5|U|. \tag{4.8}$$

For the sake of contradiction, assume that $|V_h| = 0$. Using (4.6) gives that

$$4|E_h| \geq \frac{|U| + |V_I| + N_h + N_g}{2}$$

and using (4.7) gives that

$$2|\mathcal{P}| \geq 3|V_g| + 2|V_I| - N_h - N_g = 2|U| + |V_g| - N_h - N_g.$$

Hence,

$$\begin{aligned} 12|E_h| + 2|\mathcal{P}| &\geq \frac{12}{8}(|U| + |V_I| + N_h + N_g) + 2|U| + |V_g| - N_h - N_g \\ &= \frac{7}{2}|U| + (|V_I| + |V_g|) + \frac{1}{2}(|V_I| + N_h + N_g) \\ &= \frac{9}{2}|U| + \frac{1}{2}(|V_I| + N_h + N_g). \end{aligned}$$

Therefore, (4.8) implies that

$$\frac{9}{2}|U| + \frac{1}{2}(|V_I| + N_h + N_g) < 5|U|,$$

so

$$|V_I| + N_h + N_g < |U|. \tag{4.9}$$

Additionally, using (4.5) gives that,

$$4|E_h| \geq 4|E_g| \geq |V_g| + |V_I| + N_h = |U| + N_h.$$

We therefore have that

$$\begin{aligned} 12|E_h| + 2|\mathcal{P}| &\geq 3(|U| + N_h) + 2|U| + |V_g| - N_h - N_g \\ &= 5|U| + |V_g| + 2N_h - N_g. \end{aligned}$$

Thus, (4.8) gives that

$$5|U| + |V_g| + 2N_h - N_g < 5|U| \Rightarrow 2N_h + |V_g| < N_g.$$

Substituting this in (4.9) implies that

$$|V_I| + 3N_h + |V_g| < |U|,$$

which is a contradiction to $|U| = |V_I| + |V_g|$. Thus, $|V_h| > 0$.

Next we prove that $m(h) \geq \frac{1}{d-1-\frac{2d-4}{2d-1}} = \frac{7}{17}$. Suppose $v \in V_h$. Assume that $i \in h$ and $v \in i$. Suppose $i = \{v, u_1, u_2, u_3\}$. Observe that because $v \in V_h$, $\{v, u_1\}, \{v, u_2\}, \{v, u_3\} \in \mathcal{E}_h \backslash \mathcal{E}_g$. Hence, each of these edges is contained in an element of I by Lemma 4.6. Let

$$\mathcal{I} = \{\{v, u_1\}, \{v, u_2\}, \{v, u_3\}\}.$$

Step 1: Covering the edges of $\mathcal{I}$

We cannot cover the edges in $\mathcal{I}$ with one element of I. Suppose we cover the edges with $a, b \in I$, where $\{v, u_1, u_2\} \subset a$ and $\{v, u_3\} \subset b$ without loss of generality. Then, if κ is the subgraph of h induced by the vertices of a, b, and i, $e(\kappa) \geq 3$ and $v(\kappa) \leq 7$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{3}{7} > \frac{7}{17}.$$

Hence, $m(h) > \frac{7}{17}$.

Suppose we cover the edges in $\mathcal{I}$ with $a, b, c \in I$, where $\{v, u_1\} \subset a$, $\{v, u_2\} \subset b$, and $\{v, u_3\} \subset c$. Let $\mathfrak{f}(a) = \{v, u_1\}$, $\mathfrak{f}(b) = \{v, u_2\}$, $\mathfrak{f}(c) = \{v, u_3\}$, and $\mathfrak{f}(i) = \{v, u_1, u_2, u_3\}$. For any two distinct elements $x, y \in \{a, b, c, i\}$, $\mathfrak{f}(x) \cap \mathfrak{f}(y) \subset x \cap y$ since $\mathfrak{f}(x) \subset x$ and $\mathfrak{f}(y) \subset y$. Suppose there exists two distinct elements $x, y \in \{a, b, c, i\}$ such that $\mathfrak{f}(x) \cap \mathfrak{f}(y)$ is a strict subset of $x \cap y$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, and i, then $e(\kappa) \geq 4$ and $v(\kappa) \leq 9$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{4}{9} > \frac{7}{17}.$$

Assume that for any two distinct elements $x, y \in \{a, b, c, i\}$, $\mathfrak{f}(x) \cap \mathfrak{f}(y) = x \cap y$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, and i, $e(\kappa) \geq 4$ and $v(\kappa) = 10$.

Step 2: Covering $\{u_1, u_2\}$, $\{u_2, u_3\}$, and $\{u_3, u_1\}$

We have that the edges $\{u_1, u_2\}$, $\{u_2, u_3\}$, and $\{u_3, u_1\}$ must be covered by elements of $E_g \cup I$ by Lemma 4.6.

Assume that $d \in (E_h \cup I) \backslash \{a, b, c, i\}$ and $|d \cap \{u_1, u_2, u_3\}| \geq 2$. If κ is the subgraph of h induced by the vertices of a, b, c, d, and i, then

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{5}{12} > \frac{7}{17}.$$

Next, assume the condition $(*2)$ that there does not exist $d \in (E_h \cup I) \backslash \{a, b, c, i\}$ such that $|d \cap \{u_1, u_2, u_3\}| \geq 2$. In particular, this implies that the edges $\{u_1, u_2\}$, $\{u_2, u_3\}$, and $\{u_3, u_1\}$ must be covered by elements of E_g.

Let $\mathcal{V}$ be the set of vertices of a, b, c, and i; observe that $|\mathcal{V}| = 10$.

Step 2.1: Covering $\{u_1, u_2\}$

Suppose $d \in E_g$ and $\{u_1, u_2\} \subset d$. The two cases are $d \cap \{u_1, u_2, u_3\}$ equals $\{u_1, u_2\}$ or $\{u_1, u_2, u_3\}$.

Step 2.1.1: $d \cap \{u_1, u_2, u_3\} = \{u_1, u_2, u_3\}$

Assume that $d \cap \{u_1, u_2, u_3\} = \{u_1, u_2, u_3\}$. Suppose $d = \{w, u_1, u_2, u_3\}$. By Lemma 4.6, all edges in the set

$$\mathcal{S} = \{\{w, u_1\}, \{w, u_2\}, \{w, u_3\}\}$$

must be covered by an element of $E_h \cap I$.

Assume that $w \in \mathcal{V}$. For the sake of contradiction, assume that all edges in $\mathcal{S}$ are covered by an element of $\{a, b, c, i\}$. Then, $\{w, u_1\}$ is covered by some element of $\{a, b, c, i\}$. Thus, $w \in a \cup i$ since $u_1 \notin b, c$. Since $w \notin i$, $w \in a$. Similarly, $w \in b, c$. Hence, $w \in a \cap b \cap c = \{v\}$, which is a contradiction to $d \neq i$. Suppose $e \in E_h \cup I$ is not an element of $\{a, b, c, i\}$ and covers some element of $\mathcal{S}$. If κ is the subgraph of h induced by the vertices of a, b, c, e, and i, $e(\kappa) \geq 5$ and $v(\kappa) \leq 12$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{5}{12} > \frac{7}{17}.$$

Assume that $w \notin \mathcal{V}$. Then, no element of $\mathcal{S}$ is covered by an element of $\{a, b, c, i\}$. By ($*2$), each of the elements of $\mathcal{S}$ must be covered by a distinct element of $(E_h \cup I) \backslash \{a, b, c, i\}$, otherwise two elements of $\{u_1, u_2, u_3\}$ will be contained in a single element of $(E_h \cup I) \backslash \{a, b, c, i\}$. Hence, this is the only case that we consider.

Suppose the elements of $\mathcal{S}$ can be covered by three elements of $(E_h \cup I) \backslash \{a, b, c, i\}$ but not less than three. Then, there exists $e, f, j \in E_h \cup I$ such that $\{w, u_1\} \subset e$, $\{w, u_2\} \subset f$, and $\{w, u_3\} \subset j$. If κ is the subgraph of h induced by the vertices a, b, c, e, f, j, and i, $e(\kappa) \geq 7$ and $v(\kappa) \leq 17$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{7}{17}.$$

Step 2.1.2: $d \cap \{u_1, u_2, u_3\} = \{u_1, u_2\}$

Suppose $d \cap \{u_1, u_2, u_3\} = \{u_1, u_2\}$. Suppose $d = \{u_1, u_2, w_1, w_2\}$. Observe that the five edges in the set

$$\mathcal{S} = \{\{u_1, w_1\}, \{u_1, w_2\}, \{u_2, w_1\}, \{u_2, w_2\}, \{w_1, w_2\}\}$$

must be covered by elements of $E_h \cup I$ by Lemma 4.6.

Assume that $w_1, w_2 \in \mathcal{V}$. For the sake of contradiction, assume that all edges in $\mathcal{S}$ are covered by some element of $\{a, b, c, i\}$. Then, $\{w_1, u_1\}$ must be covered by a or i, so $w_1 \in a \cup i$. Also, $\{w_1, u_2\}$ must be covered by b or i, so $w_1 \in b \cup i$. Observe that $(a \cup i) \cap (b \cup i) = i$, so $w_1 \in i$. Similarly, $w_2 \in i$. This is a contradiction to $d \neq i$. Thus, some edge in $\mathcal{S}$ is not covered by some element of $\{a, b, c, i\}$. This edge must be covered by $e \in E_h \cup I$. Note that because $d \subset \mathcal{V}$ and $|d \cap e| \geq 2$, $|e \backslash \mathcal{V}| \leq 2$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, e, and i, $e(\kappa) \geq 5$ and $v(\kappa) \leq 12$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{5}{12} > \frac{7}{17}.$$

Assume that $|\{w_1, w_2\} \cap \mathcal{V}| = 1$. Without loss of generality, assume that $w_1 \in \mathcal{V}$ and $w_2 \notin \mathcal{V}$. We have that the edges $\{w_2, w_1\}$, $\{w_2, u_1\}$, and $\{w_2, u_2\}$ in $\mathcal{S}$ are not covered by elements of $\{a, b, c, i\}$. Let

$$\mathcal{Q} = \{\{w_2, w_1\}, \{w_2, u_1\}, \{w_2, u_2\}\}.$$

The cases that we must consider are that the elements of $\mathcal{Q}$ are covered by two or three elements of $E_h \cup I$. By ($*2$), they cannot be covered by one element of $E_h \cup I$.

Suppose the elements of $\mathcal{Q}$ can be covered by two elements of $E_h \cup I$. There exists $e, f \in E_h \cup I$ such that $w_2 \in e, f$ and $\{w_1, u_1, u_2\} \subset e \cup f$. Suppose $e, f \in E_h \cup I$ such that $\{w_2, u_1, u_2\} \subset e$ and $\{w_2, w_1\} \subset f$, without loss of generality; note that the vertices $\{w_1, u_1, u_2\}$ can be considered to be equivalent for the purposes of this computation. If κ is the subgraph of h induced by the vertices of a, b, c, e, f, and i, $e(\kappa) \geq 6$ and $v(\kappa) \leq 14$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{6}{14} > \frac{7}{17}.$$

Suppose the elements of $\mathcal{Q}$ can be covered by three elements of $E_h \cup I$ but not less than three elements. Then, there exists $e, f, j \in E_h \cup I$ such that $\{w_2, u_1\} \subset e$ and $\{w_2, u_2\} \subset f$, and $\{w_2, w_1\} \subset j$. Assume that e, f, and j satisfy this condition. If κ is the subgraph of h induced by the vertices of a, b, c, e, f, j, and i, $e(\kappa) \geq 7$ and $v(\kappa) \leq 17$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{7}{17}.$$

Next, assume that $w_1, w_2 \notin \mathcal{V}$. Then, none of the elements of $\mathcal{S}$ are covered by elements of $\{a, b, c, i\}$. The cases we consider are when the elements of $\mathcal{S}$ are covered by at least 2 and at most 5 elements of $E_h \cup I$. Furthermore, recall that there does not exist an element of $(E_h \cup I)\backslash\{a, b, c, i\}$ that contains $\{u_1, u_2\}$ by ($*2$).

Suppose the elements of $\mathcal{S}$ can be covered by two elements of $E_h \cup I$. There exists $e, f \in E_h \cup I$ such that $\{w_1, w_2, u_1\} \subset e$ and $\{w_1, w_2, u_2\} \subset f$. Assume that e and f satisfy this condition. If κ is the subgraph of h induced by the vertices of a, b, c, e, f, and i, then $e(\kappa) \geq 6$ and $v(\kappa) \leq 14$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{6}{14} > \frac{7}{17}.$$

Suppose the elements of $\mathcal{S}$ can be covered by three elements of $E_h \cup I$ but not less than three elements. Then, there exists $e, f, j \in E_h \cup I$ such that either $\{w_1, w_2, u_1\} \subset e$, $\{w_1, u_2\} \subset f$, and $\{w_2, u_2\} \subset j$ or $\{w_1, w_2, u_2\} \subset e$, $\{w_1, u_1\} \subset f$, and $\{w_2, u_1\} \subset j$. Without loss of generality, suppose $e, f, j \in E_h \cup I$ satisfy the condition that $\{w_1, w_2, u_1\} \subset e$, $\{w_1, u_2\} \subset f$, and $\{w_2, u_2\} \subset j$. If κ is the subgraph of h induced by the vertices of a, b, c, e, f, j, and i, then $e(\kappa) \geq 7$ and $v(\kappa) \leq 17$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{7}{17}.$$

Next, suppose the elements of $\mathcal{S}$ can be covered by four elements of $E_h \cup I$ but not less than four elements. Suppose $e, f, j, k \in E_h \cup I$ cover the elements of $\mathcal{S}$. Since $|\mathcal{S}| = 5$, at least one of e, f, j, k must contain three elements of $\{u_1, u_2, w_1, w_2\}$. Without loss of generality, suppose e satisfies this condition. Since e cannot contain $\{u_1, u_2\}$, suppose $\{w_1, w_2, u_1\} \subset e$, without loss of generality. The uncovered edges of $\mathcal{S}$ are $\{w_1, u_2\}$ and $\{w_2, u_2\}$. There exists two elements of $\{f, j, k\}$ that covers both of these edges, so the elements of $\mathcal{S}$ can be covered by at most three elements, which is a contradiction.

Suppose the elements of $\mathcal{S}$ are covered by five elements of $E_h \cup I$ but not less than five elements. Suppose $e, f, j, k, l \in E_h \cup I$ and $\{u_1, w_1\} \subset e$, $\{u_1, w_2\} \subset f$, $\{u_2, w_1\} \subset j$, $\{u_2, w_2\} \subset k$, and $\{w_1, w_2\} \subset l$. Let $\mathfrak{f}(e) = \{u_1, w_1\}$, $\mathfrak{f}(f) = \{u_1, w_2\}$, $\mathfrak{f}(j) = \{u_2, w_1\}$, $\mathfrak{f}(k) = \{u_2, w_2\}$, and $\mathfrak{f}(l) = \{w_1, w_2\}$. For any two distinct elements $x, y \in \{a, b, c, e, f, j, k, l, i\}$, $\mathfrak{f}(x) \cap \mathfrak{f}(y) \subset x \cap y$ since $\mathfrak{f}(x) \subset x$ and $\mathfrak{f}(y) \subset y$. Suppose there exists two distinct elements $x, y \in \{a, b, c, e, f, j, k, l, i\}$ such that $\mathfrak{f}(x) \cap \mathfrak{f}(y)$ is a strict subset of $x \cap y$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, e, f, j, k, l, and i, then $e(\kappa) \geq 9$ and $v(\kappa) \leq 21$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{9}{21} > \frac{7}{17}.$$

Assume that for any two distinct elements $x, y \in \{a, b, c, e, f, j, k, l, i\}$, $\mathfrak{f}(x) \cap \mathfrak{f}(y) = x \cap y$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, e, f, j, k, l, and i, $e(\kappa) \geq 9$ and $v(\kappa) = 22$.

Step 2.2: Covering $\{u_3, u_1\}$

There exists $d' \in E_g$ such that $\{u_3, u_1\} \subset d'$. Assume that $d' \cap \{u_1, u_2, u_3\} = \{u_3, u_1\}$; we have already considered the case $d' \cap \{u_1, u_2, u_3\} = \{u_1, u_2, u_3\}$ in the case $d \cap \{u_1, u_2, u_3\} = \{u_1, u_2, u_3\}$. Suppose $d' = \{u_3, u_1, w'_1, w'_2\}$. By symmetry, we have that the only case we must consider is if $w'_1, w'_2 \notin \mathcal{V}$ and the edges in the set

$$\{\{u_3, w'_1\}, \{u_3, w'_2\}, \{u_1, w'_1\}, \{u_1, w'_2\}, \{w'_1, w'_2\}\}$$

are covered by five elements e', f', j', k', and l' of $E_h \cap I$. Suppose $\{u_3, w'_1\} \subset e'$, $\{u_3, w'_2\} \subset f'$, $\{u_1, w'_1\} \subset j'$, $\{u_1, w'_2\} \subset k'$, and $\{w'_1, w'_2\} \subset l'$. Let $\mathfrak{f}(e') = \{u_3, w'_1\}$, $\mathfrak{f}(f') = \{u_3, w'_2\}$, $\mathfrak{f}(j') = \{u_1, w'_1\}$, $\mathfrak{f}(k') = \{u_1, w'_2\}$, and $\mathfrak{f}(l') = \{w'_1, w'_2\}$. By symmetry, we may further assume that for any two distinct elements $x, y \in \{a, b, c, e', f', j', k', l', i\}$, $\mathfrak{f}(x) \cap \mathfrak{f}(y) = x \cap y$.

Let $\mathcal{V}'$ be the set of vertices of elements of $\{a, b, c, e, f, j, k, l, i\}$; observe that $|\mathcal{V}'| = 22$.

Suppose $w'_1, w'_2 \in \mathcal{V}' \setminus \mathcal{V}$. Note that the edges $\{w'_1, u_3\}$ and $\{w'_2, u_3\}$ are not covered by any element of $\{e, f, j, k, l\}$ since no element of $\{e, f, j, k, l\}$ contains u_3. Because, $w'_1, w'_2 \notin \mathcal{V}$, $\{w'_1, u_3\}$ and $\{w'_2, u_3\}$ are not covered by any element of $\{a, b, c, i\}$. Since $\{w'_1, u_3\}$ is covered by e' and $\{w'_2, u_3\}$ is covered by f', we have that $e', f' \notin \{a, b, c, e, f, j, k, l, i\}$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, e, e', f, f', j, k, l, and i, $e(\kappa) \geq 11$ and $v(\kappa) \leq 26$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{11}{26} > \frac{7}{17}.$$

Assume that $|\{w'_1, w'_2\} \cap \mathcal{V}' \setminus \mathcal{V}| = 1$. Without loss of generality, assume that $w'_1 \in \mathcal{V}' \setminus \mathcal{V}$ and $w'_2 \notin \mathcal{V}'$. Using the argument from the previous case gives that $\{w'_1, u_3\}$ is not covered by any element of $\{a, b, c, e, f, j, k, l, i\}$, so $e' \notin \{a, b, c, e, f, j, k, l, i\}$. Because $w'_2 \notin \mathcal{V}'$, any element of $\{e', f', j', k', l'\}$ that contains w'_2 is not an element of $\{a, b, c, e, f, j, k, l, i\}$. Thus, $f', k', l' \notin \{a, b, c, e, f, j, k, l, i\}$. Then, if κ is the subgraph of h induced by the vertices of a, b, c, e, e', f, f', j, k, k', l, l', and i, $e(\kappa) \geq 13$ and $v(\kappa) \leq 31$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{13}{31} > \frac{7}{17}.$$

Assume that $w'_1, w'_2 \notin \mathcal{V}'$. Then, $e', f', j', k', l' \notin \{a, b, c, e, f, j, k, l, i\}$. Thus, if κ is the subgraph of h induced by the vertices of a, b, c, e, e', f, f', j, j', k, k', l, l', and i, $e(\kappa) \geq 14$ and $v(\kappa) \leq 34$ so

$$\frac{e(\kappa)}{v(\kappa)} \geq \frac{14}{34} = \frac{7}{17}.$$

This proves that $m(h) \geq \frac{7}{17}$. ■

Lemma 4.9. *Suppose $d \geq 5$. Then, $d - 1 - \frac{1}{m(h)} \geq \frac{2d-4}{2d-1}$.*

Proof. For the sake of contradiction, assume that $d - 1 - \frac{1}{m(h)} < \frac{2d-4}{2d-1}$. Let $\gamma = \frac{2d-4}{2d-1}$. We must have that $d - 1 - \frac{1}{m(h)} < \gamma$. Because $\gamma \geq \frac{d-1}{d+1}$, Lemma 4.5 gives that

$$\frac{(d-1)|E_h| - |U| + |\mathcal{P}|}{|E_h| + |\mathcal{P}|} < \gamma.$$

This is equivalent to

$$(d-1-\gamma)|E_h| + (1-\gamma)|\mathcal{P}| < |U|.$$

Observe that

$$|E_h| \geq \frac{1}{2d}(|V_h| + |V_g| + \sum_{v \in V_I}(d_h(v) + d_g(v)))$$

and from (4.3),

$$|\mathcal{P}| \geq \frac{1}{2}((d-1)(|V_h| + |V_g|) + \sum_{v \in V_I}(d_h^*(v) + d_g^*(v))).$$

Hence,

$$\begin{aligned} &\frac{d-1-\gamma}{2d}(|V_h| + |V_g| + \sum_{v \in V_I}(d_h(v) + d_g(v))) \\ &+ \frac{1-\gamma}{2}((d-1)(|V_h| + |V_g|) + \sum_{v \in V_I}(d_h^*(v) + d_g^*(v))) < |U|. \end{aligned}$$

For $v \in V_I$, let

$$f(v) = \frac{d-1-\gamma}{2d}(d_h(v) + d_g(v)) + \frac{1-\gamma}{2}(d_h^*(v) + d_g^*(v)).$$

We have that

$$\left(\frac{d-1-\gamma}{2d} + \frac{(d-1)(1-\gamma)}{2}\right)(|V_h| + |V_g|) + \sum_{v \in V_I} f(v) < |U|. \tag{4.10}$$

Claim 4.10. *Suppose $v \in V_I$. If $n_h(v) < k(v)$ then $d_g(v) > 1$ and if $n_g(v) < k(v)$ then $d_h(v) > 1$.*

Proof. Suppose $v \in V_I$, $i_g = i_g(v)$, and $i_h = i_h(v)$. Assume that $n_h(v) < k(v)$; the case $n_g(v) < k(v)$ follows similarly. Then, there exists $w \in i_h \backslash i_g$ such that $\{v, w\} \in \mathcal{E}_h$ and $\{v, w\} \notin \mathcal{E}_h \backslash \mathcal{E}_g$. Assume that w satisfies this condition. Then, there exists $i \in E_g$ such that $i \neq i_g$ and $\{v, w\} \subset i$, so $d_g(v) \geq 2$. ■

Let $\mathcal{V}$ be the set of $v \in V_I$ such that:

1. $d_h(v) = d_g(v) = 1$.

2. $k(v) = n_h(v) = n_g(v) = 1$.

Observe that if $v \in \mathcal{V}$, $i_h(v)$ and $i_g(v)$ are deterministic since $d_h(v) = 1$ and $d_g(v) = 1$, respectively. Furthermore, the formulation of $\mathcal{V}$ corresponds to the ambiguous graph $G_{a,d}$ from [5]. For the sake of contradiction, assume the condition ($*3$) that there does not exist $i_h \in E_h$ and $i_g \in E_g$ such that $|i_h \backslash i_g| = 1$ and $i_h \cap i_g \subset \mathcal{V}$.

Suppose $v \in \mathcal{V}$. Since $d_h(v) = d_g(v) = 1$, $d_h^*(v) \geq n_h(v) = 1$, and $d_g^*(v) \geq n_g(v) = 1$,

$$f(v) \geq \frac{d-1-\gamma}{d} + (1-\gamma). \tag{4.11}$$

Let $C = \frac{d-1-\gamma}{d} + (1-\gamma)$.

Suppose $v \in V_I \backslash \mathcal{V}$. If $n_h(v) < k(v)$ or $n_g(v) < k(v)$, then using Claim 4.10 gives that

$$f(v) \geq C + \frac{d-1-\gamma}{2d} - \frac{1-\gamma}{2}.$$

If $k(v) = n_h(v) = n_g(v) > 1$, then,

$$f(v) \geq C + (1-\gamma).$$

If $d_h(v) > 1$ or $d_g(v) > 1$ and $k(v) = n_h(v) = n_g(v) = 1$, then

$$f(v) \geq C + \frac{d-1-\gamma}{2d}.$$

Suppose $\beta \in \mathbb{R}$ such that $\frac{d-1-\gamma}{d} \geq (1+2\beta)(1-\gamma)$ and $0 \leq \beta \leq 1$. Then, since $\frac{d-1-\gamma}{2d} - \frac{1-\gamma}{2} \geq \beta(1-\gamma)$ and $\beta \leq 1$,

$$f(v) \geq C + \beta(1-\gamma) = \frac{d-1-\gamma}{d} + (1+\beta)(1-\gamma). \tag{4.12}$$

Note that $\frac{d-1-\gamma}{d} \geq (1+2\beta)(1-\gamma)$ if and only if

$$\beta \leq \frac{1}{2}\left(\frac{d-1-\gamma}{d(1-\gamma)} - 1\right) = \frac{1}{2}\left(\frac{2d}{3} - \frac{8}{3} + \frac{5}{3d}\right). \tag{4.13}$$

Let $\mathcal{A}$ be the set of $v \in V_I$ such that $d_h(v) + d_g(v) \geq 3$ and there exists $i_h \in E_h$ and $i_g \in E_g$ such that $v \in i_h \cap i_g$ and $|i_h \backslash i_g| = 1$. Let $\mathcal{S}$ be the set of $v \in \mathcal{A}$ such that $d_h(v) + d_g(v) = 3$ and $\mathcal{T}$ be the set of $v \in \mathcal{A}$ such that $d_h(v) + d_g(v) > 3$.

Let $\mathcal{V}^*$ be the set of $v \in \mathcal{V}$ such that $i_h(v) \cap i_g(v) \cap \mathcal{S}$ is nonempty. Furthermore, let $\mathcal{V}^{**}$ be the set of $v \in \mathcal{V}$ such that $i_h(v) \cap i_g(v) \cap \mathcal{T}$ is nonempty.

Suppose $v \in \mathcal{V}$. By ($*$3), there exists $u \in i_h(v) \cap i_g(v)$ such that $u \notin \mathcal{V}$. Assume that u satisfies this condition. We have that there exists $i \in E_h \cup E_g \backslash \{i_h(v), i_g(v)\}$ such that $u \in i$ because $u \notin \mathcal{V}$, so $d_g(u) + d_h(u) \geq 3$ and $u \in \mathcal{A}$. This implies that $\mathcal{V}^* \cup \mathcal{V}^{**} = \mathcal{V}$.

Claim 4.11. $(d-1)(\sum_{v\in\mathcal{S}} d_h(v) + d_g(v)) \geq 3|\mathcal{V}^*|$.

Proof. Suppose $v \in \mathcal{S}$. Without loss of generality, assume that $d_h(v) = 2$ and $d_g(v) = 1$. Suppose $i_h^1, i_h^2 \in E_h$, $i_h^1 \neq i_h^2$, $i_g \in E_g$, $|i_h^1 \backslash i_g| = 1$, and $v \in i_h^1 \cap i_h^2 \cap i_g$. Let

$$s(v) = ((i_h^1 \cap i_g) \cup (i_h^2 \cap i_g)) \cap \mathcal{V}^*.$$

Note that $s(v) \subset i_g \backslash \{v\}$ so $|s(v)| \leq d-1$. Hence,

$$(d-1)(d_h(v) + d_g(v)) = 3(d-1) \geq 3|s(v)|.$$

This implies that

$$(d-1)\sum_{v\in\mathcal{S}}(d_h(v) + d_g(v)) \geq 3\sum_{v\in\mathcal{S}}|s(v)|.$$

Suppose $u \in \mathcal{V}^*$ and suppose $v \in i_h(u) \cap i_g(u) \cap \mathcal{S}$; note that $i_h(u) \cap i_g(u) \cap \mathcal{S}$ is nonempty by the definition of $\mathcal{V}^*$. We have that $u \in s(v)$. Hence, $\mathcal{V}^* \subset \bigcup_{v\in\mathcal{S}} s(v)$ so $|\mathcal{V}^*| \leq \sum_{v\in\mathcal{S}} |s(v)|$. This finishes the proof. ∎

Claim 4.12. $(d-1)(\sum_{v\in\mathcal{T}} d_h(v)+d_g(v)) \geq 2|\mathcal{V}^{**}|$.

Proof. Suppose $v \in \mathcal{T}$. Define

$$t(v) = \left(\bigcup_{\substack{i_h\in E_h, i_g\in E_g,\\ v\in i_h\cap i_g,\\ |i_h\setminus i_g|=1}} (i_h \cap i_g) \right) \cap \mathcal{V}^{**}.$$

Suppose $i_h \in E_h$ and $v \in i_h$. Note that $i_h \cap t(v) \subset i_h\setminus\{v\}$ so $|i_h \cap t(v)| \leq d-1$. We have that

$$|t(v)| \leq \sum_{i_h\in E_h: v\in i_h} |i_h \cap t(v)| \leq (d-1)d_h(v).$$

Similarly, $|t(v)| \leq (d-1)d_g(v)$. Hence,

$$(d-1)(d_h(v)+d_g(v)) \geq 2|t(v)|.$$

We then have that

$$(d-1)\sum_{v\in\mathcal{T}}(d_h(v)+d_g(v)) \geq 2\sum_{v\in\mathcal{T}}|t(v)|.$$

Suppose $u \in \mathcal{V}^{**}$. We have that $i_h(u)\cap i_g(u)\cap\mathcal{T}$ is nonempty by the definition of $\mathcal{V}^{**}$. If $v \in i_h(u)\cap i_g(u)\cap\mathcal{T}$ then $u \in t(v)$. Hence, $\mathcal{V}^{**} \subset \bigcup_{v\in\mathcal{T}} t(v)$, so $|\mathcal{V}^{**}| \leq \sum_{v\in\mathcal{T}}|t(v)|$. We are done. ■

Observe that (4.10) implies that

$$\left(-\frac{d-1-\gamma}{2d}+\frac{(d-1)(1-\gamma)}{2}\right)(|V_h|+|V_g|)+\sum_{v\in V_I}\left(f(v)-\frac{d-1-\gamma}{d}\right) < \frac{1+\gamma}{d}|U|.$$

Note that $-\frac{d-1-\gamma}{2d}+\frac{(d-1)(1-\gamma)}{2} \geq \frac{1+\gamma}{d}$, so

$$\frac{1+\gamma}{d}(|V_h|+|V_g|)+\sum_{v\in V_I}\left(f(v)-\frac{d-1-\gamma}{d}\right) < \frac{1+\gamma}{d}|U|.$$

This implies that

$$\sum_{v\in V_I}\left(f(v)-\frac{d-1-\gamma}{d}\right) < \frac{1+\gamma}{d}|V_I|. \tag{4.14}$$

If $v \in \mathcal{V}$, then (4.11) gives that

$$f(v)-\frac{d-1-\gamma}{d} = 1-\gamma.$$

Hence, using (4.12) gives that

$$\begin{aligned}
&\sum_{v\in V_I}(f(v)-\frac{d-1-\gamma}{d})\\
&\geq (1-\gamma)|\mathcal{V}|+\sum_{v\in\mathcal{S}\cup\mathcal{T}}(f(v)-\frac{d-1-\gamma}{d})+\sum_{v\in V_I\backslash(\mathcal{V}\cup\mathcal{S}\cup\mathcal{T})}(f(v)-\frac{d-1-\gamma}{d})\\
&\geq (1-\gamma)|\mathcal{V}|+\sum_{v\in\mathcal{S}\cup\mathcal{T}}(f(v)-\frac{d-1-\gamma}{d})+(|V_I|-|\mathcal{V}|-|\mathcal{S}\cup\mathcal{T}|)(1+\beta)(1-\gamma)\\
&= (1-\gamma)|\mathcal{V}|+\sum_{v\in\mathcal{S}\cup\mathcal{T}}(f(v)-\frac{d-1-\gamma}{d}-(1+\beta)(1-\gamma))+(|V_I|-|\mathcal{V}|)(1+\beta)(1-\gamma).
\end{aligned} \tag{4.15}$$

Suppose $v\in\mathcal{S}$. Then,

$$\frac{d-1-\gamma}{2d}(d_h(v)+d_g(v))\geq 3\frac{d-1-\gamma}{2d}.$$

Without loss of generality, assume that $d_h(v)=2$ and $d_g(v)=1$. Suppose $i_h\in E_h$, $i_g\in E_g$, and $v\in i_h\cap i_g$. Observe that for all $w\in i_h\backslash i_g$, $\{w,v\}\in\mathcal{E}_h\backslash\mathcal{E}_g$ since $d_g(v)=1$. Since $|i_h\backslash i_g|\geq 1$, $d_h^*(v)\geq 1$. Similarly, if $d_h(v)=1$ and $d_g(v)=2$, $d_g^*(v)\geq 1$. It follows that

$$d_h^*(v)+d_g^*(v)\geq 1.$$

Hence,

$$\sum_{v\in\mathcal{S}}(f(v)-\frac{d-1-\gamma}{d}-(1+\beta)(1-\gamma))\geq|\mathcal{S}|\cdot(\frac{d-1-\gamma}{2d}-(\frac{1}{2}+\beta)(1-\gamma)).$$

Furthermore, using Claim 4.11 gives that

$$\begin{aligned}
&\sum_{v\in\mathcal{S}}(f(v)-\frac{d-1-\gamma}{d}-(1+\beta)(1-\gamma))\\
&\geq\frac{d-1-\gamma}{2d}\sum_{v\in\mathcal{S}}(d_h(v)+d_g(v))-(\frac{d-1-\gamma}{d}+(\frac{1}{2}+\beta)(1-\gamma))|\mathcal{S}|\\
&\geq\frac{d-1-\gamma}{2d}\cdot\frac{3|\mathcal{V}^*|}{d-1}-(\frac{d-1-\gamma}{d}+(\frac{1}{2}+\beta)(1-\gamma))|\mathcal{S}|.
\end{aligned}$$

Next, we consider a linear combination of these two inequalities. If

$$c^*=\frac{\frac{d-1-\gamma}{2d}-(\frac{1}{2}+\beta)(1-\gamma)}{\left(\frac{d-1-\gamma}{2d}-(\frac{1}{2}+\beta)(1-\gamma)\right)+\left(\frac{d-1-\gamma}{d}+(\frac{1}{2}+\beta)(1-\gamma)\right)}=\frac{\frac{d-1-\gamma}{2d}-(\frac{1}{2}+\beta)(1-\gamma)}{3\frac{d-1-\gamma}{2d}},$$

then

$$\sum_{v\in\mathcal{S}}(\frac{d-1-\gamma}{2d}(d_h(v)+d_g(v))-\frac{d-1-\gamma}{d}-(1+\beta)(1-\gamma))$$

$$
\begin{aligned}
&\geq (1-c^*)|\mathcal{S}| \cdot (\frac{d-1-\gamma}{2d} - (\frac{1}{2}+\beta)(1-\gamma)) \\
&+ c^*(\frac{d-1-\gamma}{2d} \cdot \frac{3|\mathcal{V}^*|}{d-1} - (\frac{d-1-\gamma}{d} + (\frac{1}{2}+\beta)(1-\gamma))|\mathcal{S}|) \\
&= (\frac{d-1-\gamma}{2d(d-1)} - \frac{(\frac{1}{2}+\beta)(1-\gamma)}{d-1})|\mathcal{V}^*|.
\end{aligned}
$$

Observe that $c^* \geq 0$ because $\frac{d-1-\gamma}{d} \geq (1+2\beta)(1-\gamma)$.

Suppose $v \in \mathcal{T}$. Then,

$$f(v) \geq 2\frac{d-1-\gamma}{d}.$$

Hence,

$$\sum_{v\in\mathcal{T}}(f(v) - \frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)) \geq |\mathcal{T}| \cdot (\frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)).$$

Furthermore, using Claim 4.12 gives that

$$
\begin{aligned}
&\sum_{v\in\mathcal{T}}(f(v) - \frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)) \\
&\geq \frac{d-1-\gamma}{2d}\sum_{v\in\mathcal{T}}(d_h(v)+d_g(v)) - (\frac{d-1-\gamma}{d} + (1+\beta)(1-\gamma))|\mathcal{T}| \\
&\geq \frac{d-1-\gamma}{d} \cdot \frac{|\mathcal{V}^{**}|}{d-1} - (\frac{d-1-\gamma}{d} + (1+\beta)(1-\gamma))|\mathcal{T}|.
\end{aligned}
$$

Next, we consider a linear combination of these two inequalities. If

$$c^{**} = \frac{\frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)}{\left(\frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)\right) + \left(\frac{d-1-\gamma}{d} + (1+\beta)(1-\gamma)\right)} = \frac{\frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)}{2\frac{d-1-\gamma}{d}},$$

then

$$
\begin{aligned}
&\sum_{v\in\mathcal{S}}(f(v) - \frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)) \geq (1-c^{**})|\mathcal{T}| \cdot (\frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)) \\
&+ c^{**}(\frac{d-1-\gamma}{d} \cdot \frac{|\mathcal{V}^{**}|}{d-1} - (\frac{d-1-\gamma}{d} + (1+\beta)(1-\gamma))|\mathcal{T}|) \\
&= (\frac{d-1-\gamma}{2d(d-1)} - \frac{(1+\beta)(1-\gamma)}{2(d-1)})|\mathcal{V}^{**}|.
\end{aligned}
$$

Similarly, $c^{**} \geq 0$ because $\frac{d-1-\gamma}{d} \geq (1+\beta)(1-\gamma)$.

Let $\beta^* = \frac{1}{2} + \beta$. Because $\frac{d-1-\gamma}{d} \geq (1+2\beta)(1-\gamma)$, $\frac{d-1-\gamma}{d} \geq 2\beta^*(1-\gamma)$. Furthermore, $\frac{1}{2} + \beta \geq \frac{1+\beta}{2}$ since $\beta \geq 0$. It follows that

$$\sum_{v\in\mathcal{S}\cup\mathcal{T}}(f(v) - \frac{d-1-\gamma}{d} - (1+\beta)(1-\gamma)) \geq (\frac{d-1-\gamma}{2d(d-1)} - (\frac{1}{2}+\beta)\frac{1-\gamma}{d-1})|\mathcal{V}^*|$$

$$
\begin{aligned}
&+(\frac{d-1-\gamma}{2d(d-1)}-\frac{(1+\beta)(1-\gamma)}{2(d-1)})|\mathcal{V}^{**}| \\
\geq&(\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})(|\mathcal{V}^{*}|+|\mathcal{V}^{**}|) \\
\geq&(\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})|\mathcal{V}|.
\end{aligned}
$$

Using this inequality and (4.15) gives that

$$\sum_{v\in V_I}(f(v)-\frac{d-1-\gamma}{d})\geq(1-\gamma+\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})|\mathcal{V}|+(|V_I|-|\mathcal{V}|)(1+\beta)(1-\gamma).$$

Afterwards, (4.14) implies that

$$(1-\gamma+\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})|\mathcal{V}|+(|V_I|-|\mathcal{V}|)(1+\beta)(1-\gamma)<\frac{1+\gamma}{d}|V_I|.$$

From (4.13), we can set β to be $\frac{1}{3}$. Let $\beta=\frac{1}{3}$ and $\beta^*=\frac{1}{2}+\beta=\frac{5}{6}$. Furthermore, let

$$\ell(x)=(1-\gamma+\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})x+(|V_I|-x)(1+\beta)(1-\gamma).$$

We have that $\ell(|\mathcal{V}|)<\frac{1+\gamma}{d}|V_I|$. Note that ℓ is a linear function and $0\leq|\mathcal{V}|\leq|V_I|$. Then, $\ell(|\mathcal{V}|)<\frac{1+\gamma}{d}|V_I|$ implies that $\min(\ell(0),\ell(|V_I|))<\frac{1+\gamma}{d}|V_I|$.

First, observe that

$$\ell(0)-\frac{1+\gamma}{d}|V_I|=((1+\beta)(1-\gamma)-\frac{1+\gamma}{d})|V_I|=\frac{1}{2d-1}(3(1+\beta)-\frac{4d-5}{d})|V_I|.$$

Since

$$3(1+\beta)-\frac{4d-5}{d}=\frac{5}{d}>0,$$

$\ell(0)\geq\frac{1+\gamma}{d}|V_I|$.

Furthermore,

$$\ell(|V_I|)=(1-\gamma+\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1})|V_I|.$$

We have that

$$1-\gamma+\frac{d-1-\gamma}{2d(d-1)}-\beta^*\frac{1-\gamma}{d-1}-\frac{1+\gamma}{d}=\frac{7d-5-6d\beta^*}{2d(d-1)(2d-1)}.$$

Since $\beta^*=\frac{5}{6}$,

$$7d-5-6d\beta^*=2d-5>0.$$

Thus, $\ell(|V_I|)\geq\frac{1+\gamma}{d}|V_I|$. This is a contradiction to $\min(\ell(0),\ell(|V_I|))<\frac{1+\gamma}{d}|V_I|$.

Therefore, there exists $i_h\in E_h$ and $i_g\in E_g$ such that $|i_h\backslash i_g|=1$ and $i_h\cap i_g\subset\mathcal{V}$. Assume that i_h and i_g satisfy this condition. Suppose $i_h=(i_h\cap i_g)\cup\{v_h\}$ and $i_g=(i_h\cap i_g)\cup\{v_g\}$.

Because $i_h \cap i_g \subset \mathcal{V}$, $d_h(v) = d_g(v) = 1$ for all $v \in i_h \cap i_g$, which implies that $\{v_h, v\} \in \mathcal{E}_h \backslash \mathcal{E}_g$ and $\{v_g, v\} \in \mathcal{E}_g \backslash \mathcal{E}_h$ for all $v \in i_h \cap i_g$.

Let $E'_h = \{i_h\}$, $E'_g = \{i_g\}$, $U' = i_h \cup i_g$, and $\mathcal{P}' = \{\{v_h, v\} : v \in i_h \cap i_g\} \cup \{\{v_g, v\} : v \in i_h \cap i_g\}$. Let I' be the set of $i \in I$ such that there exists $e \in \mathcal{P}'$ such that $e \subset i$. Because $\mathcal{P}' \subset \mathcal{P}$, each edge in $\mathcal{P}'$ is contained in some element of I'. Let V' be the set of vertices in i_h, i_g, or some hyperedge in I'. Let h' be the subgraph of h induced by V'. Furthermore, let h'' be the graph with vertex set V' and edge set $E'_h \cup I'$. We have that

$$d - 1 - \frac{1}{m(h)} \geq d - 1 - \frac{1}{\alpha(h')} \geq d - 1 - \frac{1}{\alpha(h'')}$$

so it suffices to prove that

$$d - 1 - \frac{1}{\alpha(h'')} \geq \frac{2d-4}{2d-1}.$$

Using Lemma 4.5 gives that to prove that $d - 1 - \frac{1}{\alpha(h'')} \geq \frac{2d-4}{2d-1}$, it suffices to prove that

$$\frac{(d-1)|E'_h| - |U'| + |\mathcal{P}'|}{|E'_h| + |\mathcal{P}'|} \geq \frac{2d-4}{2d-1}.$$

Observe that

$$\frac{(d-1)|E'_h| - |U'| + |\mathcal{P}'|}{|E'_h| + |\mathcal{P}'|} = \frac{d - 1 - (d+1) + 2(d-1)}{1 + 2(d-1)} = \frac{2d-4}{2d-1},$$

which finishes the proof. ∎

4.4 Weighted ambiguous graphs

For recovery after observing the weighted projection, we consider when two hypergraphs have the same weighted projections, which would imply that they have the same projection. First, we define weighted-ambiguous graphs, which are analogous to ambiguous graphs.

Definition 4.13. *A weighted-minimal preimage of a graph G is a hypergraph H such that $\mathrm{Proj}_W(H) = G$ and $|e(H)|$ is minimal. A graph G is weighted-ambiguous if there exists two distinct minimal weighted-preimages of G.*

We use similar conditions as the previous section: suppose h and g are hypergraphs such that $\mathrm{Proj}_W(h) = \mathrm{Proj}_W(g)$; observe that this immediately implies that $e(h) = e(g)$. The goal is to prove the following result:

Theorem 4.14. *Suppose $d \geq 3$. Then $d - 1 - \frac{1}{m(h)} \geq \frac{d}{2} - 1 \geq \frac{d-1}{d+1}$.*

First, observe that we can remove all hyperedges in both h and g; afterwards, the condition $\mathrm{Proj}_W(h) = \mathrm{Proj}_W(g)$ will still be satisfied, and $d - 1 - \frac{1}{m(h)}$ will be decreased. Hence, assume that $E(h)$ and $E(g)$ are disjoint. Since $h \neq g$, both $E(h)$ and $E(g)$ are nonempty.

Lemma 4.15. *Suppose $v \in V$. Then, $d_h(v) \neq 1$.*

Proof. For the sake of contradiction, assume that $d_h(v) = 1$. Suppose i is the hyperedge of h that contains v. Then, the weight of $\{v, u\}$ for all $u \in i\backslash\{v\}$ in $\mathrm{Proj}_W(h)$ equals one. Since $\mathrm{Proj}_W(g) = \mathrm{Proj}_W(h)$, each edge $\{v, u\}$ for $u \in i\backslash\{v\}$ is contained in a hyperedge i_u of g. If all of the i_u are equal, then they must all equal i, which is a contradiction to $E(h)$ and $E(g)$ being disjoint. Therefore, some two of the i_u are distinct, which implies that $d_g(v) \geq 2$. However, $\mathrm{Proj}_W(h) = \mathrm{Proj}_W(g)$ implies that $d_g(v) = d_h(v) = 1$, which is a contradiction. ■

Proof of Theorem 4.14. Let V' be the set of $v \in V$ such that $d_h(v) \geq 1$. Let h' be the subgraph of h induced by V'. Using Lemma 4.15 implies that

$$de(h') = de(h) = \sum_{v \in V'} d_h(v) \geq 2v(h'),$$

so $m(h) \geq \alpha(h') \geq \frac{2}{d} \Rightarrow d-1-\frac{1}{m(h)} \geq \frac{d}{2}-1$. It is straightforward to verify that $\frac{d}{2}-1 \geq \frac{d-1}{d+1}$. ■

The following result is an immediate implication of Theorem 4.14,

Corollary 4.16. *Suppose G is a weighted-ambiguous graph. For a preimage h of G, $d - 1 - \frac{1}{m(h)} \geq \frac{d}{2} - 1$.*

Observe that threshold of $\frac{d}{2} - 1$ is greater than the threshold of $\frac{2d-4}{2d-1}$ for the unweighted projection in Theorem 4.3. The reason for this is that the graph $G_{a,d}$ defined in Definition 4.2 is not ambiguous under the weighted projection. In fact, we show that the threshold in Theorem 4.14 is tight by defining a weighted-ambiguous graph that achieves it.

Definition 4.17. *Define the hypergraph $H = (V, E)$ as follows:*

- *$V = S_1 \sqcup S_2 \sqcup \{w_1, w_2, w_3, w_4\}$, where $|S_1| = |S_2| = d - 2$.*
- *E consists of $S_1 \cup \{w_1, w_2\}$, $S_1 \cup \{w_3, w_4\}$, $S_2 \cup \{w_2, w_3\}$, and $S_2 \cup \{w_4, w_1\}$.*

Then, define $G^w_{a,d} := \mathit{Proj}_w(H)$.

Suppose H' has the same vertex set as H and edge set $S_2 \cup \{w_1, w_2\}$, $S_2 \cup \{w_3, w_4\}$, $S_1 \cup \{w_2, w_3\}$, and $S_1 \cup \{w_4, w_1\}$. Then, H and H' are two distinct minimal preimages of $G^w_{a,d}$, so $G^w_{a,d}$ is a weighted-ambiguous graph. Since each vertex of H is contained in two hyperedges, we also observe that H achieves the lower bound in Theorem 4.14. For an example, when $d = 3$, H corresponds to an octahedron.

Chapter 5

Exact Recovery

5.1 Two-connected components

First, we introduce the notion of two-connected components from [5] in the following definition, which we use in the proof of Theorem 1.7.

Definition 5.1. *Suppose H is a hypergraph. Two hyperedges a and b of H are two-connected if there exists a sequence of hyperedges $(h_i)_{1\leq i\leq m}$ such that if $h_0 = a$ and $h_{m+1} = b$ then $|h_i \cap h_{i+1}| \geq 2$ for $0 \leq i \leq m$. A two-connected component of H is a set C of hyperedges of H such that any two elements of C are two-connected and no element of $E(H)\backslash C$ is two-connected to an element of C.*

From [5, Lemma 32], if $\delta < \frac{d-1}{d+1}$ then all two-connected components of $\mathcal{H}_c$ have $O(1)$ vertices with probability $1 - o_n(1)$. We refer to $\frac{d-1}{d+1}$ as the *two-connectivity threshold.*

The papers [10] and [9] use a notion of connectivity that generalizes two-connectivty. The random hypergraph models the papers consider are the same as $\mathcal{H}$ and the papers bound the sizes of the largest connected components given the probability a hyperedge is present. However, the results of the papers are not directly applicable to the two-connected components of $\mathcal{H}_c$ since the hyperedges in $\mathcal{H}_c$ are not independent. Because $\mathcal{H}_c$ is the set of d-cliques in Ψ, its hyperedges are positively correlated.

5.2 Proof of Theorem 1.7

From [5, Theorem 4], the probability of exact recovery is $1 - o_n(1)$ if $d = 3$ and $\delta < \frac{2d-4}{2d-1}$. In this section we address the remaining cases of Theorem 1.7. First, we address the lower bound of the exact recovery thresholds for $d \geq 4$. We state two lemmas.

Lemma 5.2 ([5, Lemma 17])**.** *Suppose K is a fixed d-uniform hypergraph. Then,*

$$\Pr[K \subset \mathcal{H}] = \begin{cases} o_n(1) & \text{if } p = o_n(n^{-\frac{1}{m(K)}}), \\ 1 - o_n(1) & \text{if } p = \omega_n(n^{-\frac{1}{m(K)}}), \\ \Omega_n(1) & \text{if } p = \Theta_n(n^{-\frac{1}{m(K)}}). \end{cases}$$

Lemma 5.3 ([5, Lemma 19])**.** *Suppose $\delta < \frac{d-1}{d+1}$. If for all finite ambiguous graphs G_a,*

$$\Pr[\mathit{Cli}(G_a) \textit{ is a two-connected component of } \mathcal{H}_c] = o_n(1),$$

then $\Pr[\mathcal{A}^*(\Psi) = \mathcal{H}] \geq 1 - o_n(1)$.

Theorem 5.4. *Suppose $d = 4$. If $\delta < \frac{2d-4}{2d-1}$ then the probability of exact recovery is $1 - o_n(1)$.*

Proof. Suppose $\delta < \frac{2d-4}{2d-1}$. From Lemma 5.3, it suffices to prove that for all ambiguous graphs G_a, the probability that $\mathrm{Cli}(G_a)$ is a 2-connected component of $\mathcal{H}_c$ is $o_n(1)$. Suppose G_a is an ambiguous graph. Let P be the set of hypergraphs h such that $\mathrm{Proj}(h) = G_a$. Then,

$$\Pr[\mathrm{Cli}(G_a) \text{ is a two-connected component of } \mathcal{H}_c] \leq \sum_{h \in P} \Pr[h \subset \mathcal{H}]. \tag{5.1}$$

Suppose $h \in P$. Using Theorem 4.3, or more specifically Lemma 4.8, gives that

$$-d + 1 + \delta < -d + 1 + \frac{2d-4}{2d-1} \leq -\frac{1}{m(h)}.$$

Then, Lemma 5.2 implies that $\Pr[h \subset \mathcal{H}] = o_n(1)$. Using (5.1) finishes the proof because P is finite. ∎

Theorem 5.5. *Suppose $d \geq 5$. If $\delta < \frac{d-1}{d+1}$ then the probability of exact recovery is $1 - o_n(1)$.*

Proof. We can prove this result using Theorem 4.3, or more specifically Lemma 4.9, and the same argument as the proof of Theorem 5.4. ∎

Remark 5.6. It suffices to the weaker result Lemma 4.7 rather than Lemma 4.9 in the proof of Theorem 5.5. Using the lemma gives that if $h \in P$, then

$$\delta < \frac{d-1}{d+1} \leq d - 1 - \frac{1}{m(h)},$$

and similarly Lemma 5.2 implies that $\Pr[h \subset \mathcal{H}] = o_n(1)$.

Next we address the upper bound of the exact recovery threshold. Observe that it suffices to prove that if $\delta > \min(\frac{d-1}{d+1}, \frac{2d-4}{2d-1})$, then the probability of exact recovery is $o_n(1)$. From Theorem 1.5, the partial recovery loss is $o_n(1)$ if $\delta > \frac{d-1}{d+1}$, which implies that the probability of exact recovery is $o_n(1)$ in this regime. Proving the following result completes the proof of Theorem 1.7.

Theorem 5.7. *Suppose $d \geq 3$. If $\delta = \frac{2d-4}{2d-1}$ then the probability of exact recovery is $1 - \Omega_n(1)$ and if $\delta > \frac{2d-4}{2d-1}$ then the probability of exact recovery is $o_n(1)$.*

Remark 5.8. The case where $\delta \geq \frac{2d-4}{2d-1}$ implies exact recovery having $1 - \Omega_n(1)$ probability has been proved in [5, Appendix A] for $3 \leq d \leq 5$ using two-connected components. Furthermore, when $d \geq 5$ the probability of exact recovery is $o_n(1)$ if $\delta > \frac{2d-4}{2d-1} \geq \frac{d-1}{d+1}$ from Theorem 1.5. Hence the main contribution of this result is proving that the probability of exact recovery is $o_n(1)$ if $\delta > \frac{2d-4}{2d-1}$ and $d = 3, 4$.

Proof of Theorem 5.7. Suppose $\delta \geq \frac{2d-4}{2d-1}$. It suffices to prove that $\mathcal{A}^*$ fails with probability $\Omega_n(1)$ and $1 - o_n(1)$ if $\delta > \frac{2d-4}{2d-1}$.

We consider a hypergraph that is a minimal preimage of $G_{a,d}$ repeated over many two-connected components. For the definition of $G_{a,d}$ and one of its minimal preimages, see Definition 4.2. Suppose $m \geq 1$ and h is a d-uniform hypergraph with vertex set V. Suppose $V = \bigsqcup_{i=1}^{m} V_i$. For $1 \leq i \leq m$ suppose

$$V_i = \{v_j^i : 1 \leq j \leq d+1\} \bigsqcup_{1 \leq j \leq d-1} S_j^{1;i} \bigsqcup_{1 \leq j \leq d-1} S_j^{2;i},$$

where $|S_j^{1;i}| = |S_j^{2;i}| = d-2$ for $1 \leq j \leq d-1$. Furthermore, suppose that for $1 \leq i \leq m$ the hyperedges of the subgraph of h induced by V_i are $\{v_1^i, \ldots, v_d^i\}$, $\{v_d^i, v_j^i, S_j^{1;i}\}$ for $j \in [d-1]$, and $\{v_{d+1}^i, v_j^i, S_j^{2;i}\}$ for $j \in [d-1]$. Additionally assume that h has no other hyperedges.

We have that

$$\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H} | h \subset \mathcal{H}, \mathcal{H} \text{ is not minimal}] = 1.$$

For all $G \in \{0,1\}^{\binom{[n]}{2}}$, let $S(G)$ be the set of hypergraphs H such that $\mathrm{Proj}(H) = G$, $h \subset H$, and H is minimal. Let $\mathcal{G}^*$ be the set of G such that $|S(G)| \geq 1$.

Suppose $G \in \mathcal{G}^*$. Suppose $H \in S(G)$. For $1 \leq i \leq m$, let H^i be the hypergraph obtained from H after the hyperedge for $\{v_1^i, \ldots, v_d^i\}$ is removed and the hyperedge for $\{v_1^i, \ldots, v_{d-1}^i, v_{d+1}^i\}$ is added. Note that H and H^i are distinct elements of $S(G)$ for $1 \leq i \leq m$ so $|S(G)| \geq m+1$.

Furthermore, for all $G \in \mathcal{G}^*$, let $P(G)$ be $\Pr[\mathcal{H} = H]$ for a hypergraph H such that $\mathrm{Proj}(H) = G$ and H is minimal. We have that

$$\begin{aligned}
\Pr[h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] &= \sum_{h \subset H, H \text{ is minimal}} \Pr[\mathcal{H} = H] \\
&= \sum_{G \in \mathcal{G}^*} \sum_{H \in S(G)} \Pr[\mathcal{H} = H] \\
&= \sum_{G \in \mathcal{G}^*} P(G) |S(G)|.
\end{aligned}$$

Furthermore,

$$\Pr[\mathcal{A}^*(\Psi) = \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] = \sum_{G \in \mathcal{G}^*} \sum_{H \in S(G)} \Pr[\mathcal{A}(\Psi) = H, \mathcal{H} = H].$$

Suppose $G \in \mathcal{G}^*$. We have that

$$\begin{aligned}
\sum_{H \in S(G)} \Pr[\mathcal{A}^*(\Psi) = H, \mathcal{H} = H] &= \sum_{H \in S(G)} \Pr[\mathcal{A}^*(\Psi) = H | \mathcal{H} = H] \Pr[\mathcal{H} = H] \\
&= P(G) \sum_{\substack{\mathrm{Proj}(H) = G, h \subset H, \\ H \text{ is minimal}}} \Pr[\mathcal{A}^*(\Psi) = H | \Psi = G] \\
&\leq P(G).
\end{aligned}$$

Therefore,

$$\Pr[\mathcal{A}^*(\Psi) = \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] \leq \sum_{G \in \mathcal{G}^*} P(G).$$

Since $|S(G)| \geq m$ for all $G \in \mathcal{G}^*$,

$$\Pr[\mathcal{A}^*(\Psi) = \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] \leq \sum_{G \in \mathcal{G}^*} P(G) \leq \frac{1}{m+1} \sum_{G \in \mathcal{G}^*} P(G)|S(G)|$$

$$= \frac{1}{m+1} \Pr[h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}].$$

It follows that

$$\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] \geq \frac{m}{m+1} \Pr[h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}].$$

Thus,

$$\begin{aligned}
\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}] \geq &\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is not minimal}] \\
&+ \Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}, h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] \\
\geq &\Pr[h \subset \mathcal{H}, \mathcal{H} \text{ is not minimal}] + \frac{m}{m+1} \Pr[h \subset \mathcal{H}, \mathcal{H} \text{ is minimal}] \\
\geq &\frac{m}{m+1} \Pr[h \subset \mathcal{H}].
\end{aligned}$$

Because

$$-\frac{1}{m(h)} = -d + 1 + \frac{2d-4}{2d-1} \leq -d + 1 + \delta,$$

$\Pr[h \subset \mathcal{H}] = \Omega(1)$ so $\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}] = \Omega(1)$. Assume that $\delta > \frac{2d-4}{2d-1}$. Then, $-\frac{1}{m(h)} < -d + 1 + \delta$ so $\Pr[h \subset \mathcal{H}] = 1 - o_n(1)$. It follows that for all $m \geq 1$,

$$\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}] \geq \frac{m}{m+1}(1 - o_n(1)).$$

Therefore $\Pr[\mathcal{A}^*(\Psi) \neq \mathcal{H}] = 1 - o_n(1)$. ∎

5.3 Proof of Theorem 1.9

Observe that the probability of exact recovery being $o_n(1)$ if $\delta > \frac{d-1}{d+1}$ follows from the partial recovery loss being $1 - o_n(1)$ if $\delta > \frac{d-1}{d+1}$ by Theorem 1.6. Thus it suffices to prove that the probability of exact recovery is $1 - o_n(1)$ if $\delta < \frac{d-1}{d+1}$.

First note that the analog of Lemma 5.3 is true with projections replaced by weighted projections. Afterwards we can use Theorem 4.14 and the same argument as the proof of Theorem 5.4.

5.4 Efficient algorithms for partial and exact recovery

For partial recovery, consider the algorithm that returns $\mathcal{H}_c$. It is clear that this algorithm is efficient, since there are $\binom{n}{d}$ hyperedges that must be considered to compute $\mathcal{H}_c$ from Ψ. In the regime $\delta < \frac{d-1}{d+1}$ where partial recovery is possible, this algorithm achieves a partial recovery loss of $o_n(1)$, because $\mathcal{H}_c$ outputs $q\binom{n}{d}$ hyperedges on average and $q = (1 + o_n(1))p$ by Corollary 3.5.

Furthermore, in the regime $\delta < \frac{d-1}{d+1}$, the MAP algorithm is efficiently computable because each of the two-connected components has size $O_n(1)$ with high probability, see [5, Theorem 10]. Since the exact recovery threshold is always at most $\frac{d-1}{d+1}$, this implies that we can achieve $1 - o_n(1)$ probability of exact recovery with an efficient algorithm when it is possible.

Chapter 6

Phase transition

In the main results in Section 1.3, we do not consider when δ equals the exact recovery threshold $\min\left(\frac{d-1}{d+1}, \frac{2d-4}{2d-1}\right)$ while analyzing exact recovery. In this chapter, we analyze this case by exhibiting regimes for p which are below the exact recovery threshold $n^{-d+1+\min\left(\frac{d-1}{d+1}, \frac{2d-4}{2d-1}\right)}$ by a polylogarithmic factor and at which exact recovery has probability $1 - o_n(1)$. Note that we no longer consider when $p \propto (1 + o_n(1))n^{-d+1+\delta}$ in this chapter.

The first result that we prove bounds the sizes of the two-connected components with high probability, similarly to [5, Lemma 12]. The difference is that we consider when $\delta = \frac{d-1}{d+1}$ rather than when $\delta < \frac{d-1}{d+1}$.

Theorem 6.1. *Suppose $\delta = \frac{d-1}{d+1}$. Assume that t satisfies $\lim_{n\to\infty} \log_n(t!) = \infty$. Furthermore, assume that $p \le at^{-d}n^{-d+1+\delta}$ for some $a > 0$. All 2-connected components in $\mathcal{H}_c$ have at most $O_n(t)$ vertices with probability $1 - o_n(1)$.*

Proof. We use the methods of Appendix B of the paper [5] in this proof; in particular, we follow the proof of Lemma 34 of the paper. Suppose $H' \in \mathrm{Grow}(H)$; that is, $E(H) \subsetneq E(H')$ and there exists a d-clique h that is two-connected to H such that:

- $\mathrm{Proj}(h) \subset \mathrm{Proj}(H')$.
- $|h' \cap h| \ge 2$ for all $h' \in E(H')\backslash E(H)$, and the sets $h' \cap h$ for $h' \in E(H')\backslash E(H)$ are disjoint.

Note that h can be any d-clique; for example, h can be a hyperedge of H. Furthermore, in contrast with [5], we allow $\mathrm{Proj}(h' \cap h)$ to be a subset of $\mathrm{Proj}(H)$, to account for when $\mathrm{Proj}(H') = \mathrm{Proj}(H)$ or similar cases. We have that all hypergraphs whose projection is a two-connected component is isomorphic to a hypergraph obtained from applying the grow operation to $[d]$ a finite number of times.

Let X_H and $X_{H'}$ be the expected number of appearances of H and H' in $\mathcal{H}$ as non-induced subgraphs. In this case, we only require the hyperedges of H to be a subset of the hyperedges of $\mathcal{H}$, so there does not have to be a bijection between the non-hyperedges. Note that H and H' do not have any disconnected vertices, so

$$\frac{|\mathrm{aut}(H')|\mathbb{E}[X_{H'}]}{|\mathrm{aut}(H)|\mathbb{E}[X_H]} = \frac{|v(H')|!\binom{n}{|v(H')|}p^{|e(H')|}}{|v(H)|!\binom{n}{|v(H)|}p^{|e(H)|}} \le \frac{|v(H')|!\binom{n}{|v(H')|}p^{|e(H')|}}{|v(H)|!\binom{n}{|v(H)|}p^{|e(H)|}}.$$

Therefore,

$$\frac{|\mathrm{aut}(H')|\mathbb{E}[X_{H'}]}{|\mathrm{aut}(H)|\mathbb{E}[X_H]} \le \frac{n^{-|v(H')|}|v(H')|!\binom{n}{|v(H')|}(at^{-d})^{|e(H')|}}{n^{-|v(H)|}|v(H)|!\binom{n}{|v(H)|}(at^{-d})^{|e(H)|}} \cdot n^{|v(H')|-|v(H)|+(-d+1+\delta)(|e(H')|-|e(H)|)}.$$

From [5, Lemma 35],

$$|v(H')| - |v(H)| + (-d+1+\delta)(|e(H')| - |e(H)|) \le \delta - \frac{d-1}{d+1} = 0. \tag{6.1}$$

Additionally, $|e(H')| - |e(H)| \le 2^d$, so $|v(H')| - |v(H)| \le (d-1)(|e(H')| - |e(H)|) \le (d-1)2^d$. Hence,

$$\begin{aligned}\frac{|\mathrm{aut}(H')|\mathbb{E}[X_{H'}]}{|\mathrm{aut}(H)|\mathbb{E}[X_H]} &\le \frac{n^{-|v(H')|}|v(H')|!\binom{n}{|v(H')|}(at^{-d})^{|e(H')|}}{n^{-|v(H)|}|v(H)|!\binom{n}{|v(H)|}(at^{-d})^{|e(H)|}} = \left(\prod_{i=|v(H)|}^{|v(H')|-1} \frac{n-i}{n}\right)\left(at^{-d}\right)^{|e(H')|-|e(H)|} \\ &\le \left(at^{-d}\right)^{|e(H')|-|e(H)|}.\end{aligned} \tag{6.2}$$

Suppose $H = [d]$ and $H' \in \mathrm{Grow}^{(T)}([d])$. Repeating (6.2) with $|\mathrm{aut}([d])| = d!$ gives that

$$|\mathrm{aut}(H')|\mathbb{E}[X_{H'}] \le d!\left(at^{-d}\right)^{|e(H')|-1}\mathbb{E}[X_H].$$

Furthermore, $|v(H')| \le d + T(d-1)2^d$. For all $k \ge 1$, the number of $H' \in \mathrm{Grow}^{(T)}([d])$ with k hyperedges is at most $\mathbf{1}_{k\ge t}\binom{\binom{d+T(d-1)2^d}{d}}{k}$; note that $k \ge T$ because we add at least one hyperedge at each step. Hence,

$$\begin{aligned}\sum_{H'\in\mathrm{Grow}^{(T)}([d])} |\mathrm{aut}(H')|\mathbb{E}[X_{H'}] &\le d!\sum_{k\ge T}\binom{\binom{d+T(d-1)2^d}{d}}{k}\left(at^{-d}\right)^{k-1}\mathbb{E}[X_H] \\ &\le \frac{d!}{T!}\sum_{k\ge T}\left(\frac{(d+T(d-1)2^d)^d}{d!}\right)^k\left(at^{-d}\right)^{k-1}\mathbb{E}[X_H].\end{aligned}$$

Suppose $T = \frac{t}{a^{\frac{1}{d}}(d-1)2^d}$. Then,

$$\left(\frac{(d+T(d-1)2^d)^d}{d!}\right)at^{-d} = \frac{(1+da^{\frac{1}{d}}t^{-1})^d}{d!} = \frac{1+o_n(1)}{d!}.$$

Let this quantity be ϵ. Note that c does not depend on n. Then,

$$\sum_{H'\in\mathrm{Grow}^{(T)}([d])} |\mathrm{aut}(H')|\mathbb{E}[X_{H'}] \le \frac{d!t^d}{aT!}\sum_{k\ge T}\epsilon^k\mathbb{E}[X_H] \le \frac{d!t^d}{aT!}\frac{\epsilon^T}{1-\epsilon}\cdot n^{1+\delta}\frac{at^{-d}}{d!} = \frac{n^{1+\delta}\epsilon^T}{T!(1-\epsilon)}. \tag{6.3}$$

Observe that $(T!)^{2a^{\frac{1}{d}}(d-1)2^d} > t!$ as $n \to \infty$. Therefore, since $\lim_{n\to\infty}\log_n(t!) = \infty$, $\lim_{n\to\infty}\log_n(T!) = \infty$. It follows that

$$\sum_{H'\in\mathrm{Grow}^{(T)}([d])} |\mathrm{aut}(H')|\mathbb{E}[X_{H'}] = o_n(1).$$

This implies that the probability that there exists $H' \in \text{Grow}^{(T)}([d])$ such that H' is a non-induced subgraph of $\mathcal{H}$ is $o_n(1)$. Since all elements of $\text{Grow}^{(T')}([d])$ for $T' \geq T$ must contain an element of $\text{Grow}^{(T)}([d])$ as a non-induced subgraph, all 2-connected components in $\mathcal{H}_c$ are isomorphic to the projection of an element of $\text{Grow}^{(T')}([d])$ for some $T' < T$ with probability $1 - o_n(1)$. Therefore, all 2-connected have $O_n(T) = O_n(t)$ vertices with probability $1 - o_n(1)$. ∎

Note that it is possible to refine the previous result to require $\lim_{n\to\infty} t! > C$, for some constant C that depends on a and d. For simplicity, we omit this calculation. Consider the following example of t such that $\lim_{n\to\infty} \log_n(t!) = \infty$.

Lemma 6.2. *Suppose $t = \frac{\log(n)}{\log(\log(\log(n)))}$. Then, $\lim_{n\to\infty} \log_n(t!) = \infty$.*

Proof. Using Stirling's approximation for $t!$ gives that $t! > (\frac{t}{e})^t$. Observe that

$$\log_n(e^t) = \frac{t}{\log(n)} = o_n(1).$$

Thus, it suffices to prove that $\lim_{n\to\infty} \log_n(t^t) = \infty$. Observe that

$$\log_n(t^t) = \frac{\log(n)}{\log(\log(\log(n)))} \cdot \frac{1}{\log(n)} \cdot (\log(\log(n)) - \log(\log(\log(\log(n))))).$$

Since $\lim_{n\to\infty} \frac{\log(\log(n))}{\log(\log(\log(n)))} = \infty$, $\lim_{n\to\infty} \log_n(t^t) = \infty$, which finishes the proof. ∎

Theorem 6.3. *Suppose $d \geq 6$, $\delta = \frac{d-1}{d+1}$, $a > 0$, and $p \leq at^{-d}n^{-d+1+\delta}$, where $\lim_{n\to\infty} \log_n(t!) = \infty$ and $\lim_{n\to\infty} \log_n(t) = 0$. Then the exact recovery probability is $1 - o_n(1)$.*

Proof. From (A.5) it suffices to prove that the probability that $\mathcal{H}$ is not a unique minimal preimage is $o_n(1)$; if $\mathcal{H}$ is a unique minimal preimage, then the MAP algorithm applied to its projected graph outputs $\mathcal{H}$. Suppose H is not a unique minimal preimage. Then, there exists a 2-connected component of $\text{Proj}(H)$ that H is not the unique minimal preimage for. That is, there exists a set of vertices $\mathcal{S}$ such that $\mathcal{S}$ induces a 2-connected component in $\text{Proj}(H)$ and if h is the subgraph of H induced by $\mathcal{S}$ then there exists a hypergraph g such that $\text{Proj}(g) = \text{Proj}(h)$ and $e(g) \leq e(h)$. From Lemma 4.9, $d - 1 - \frac{1}{m(h)} \geq \frac{2d-4}{2d-1}$. Then, H has a subgraph h' such that $d - 1 - \frac{1}{\alpha(h')} \geq \frac{2d-4}{2d-1}$. Furthermore, Theorem 6.1 gives that the sum of $\Pr[\mathcal{H} = H]$ for H such that $|\mathcal{S}| > Ct$ is $o_n(1)$ for some $C > 0$.

Let $\mathcal{N}$ be the set of H such that H contains a subgraph h' such that $d - 1 - \frac{1}{\alpha(h')} \geq \frac{2d-4}{2d-1}$ and $|v(h')| \leq Ct$. Since the probability that $\mathcal{H}$ is not uniquely minimal and $\mathcal{H} \notin \mathcal{N}$ is $o_n(1)$, it suffices to prove that $\Pr[\mathcal{H} \in \mathcal{N}] = o_n(1)$.

Let $\mathcal{R}$ be the set of hypergraphs h' such that $|v(h')| \leq Ct$ and $d - 1 - \frac{1}{\alpha(h')} \geq \frac{2d-4}{2d-1}$. It suffices to prove that the probability that an element of $\mathcal{R}$ appears as a subgraph of $\mathcal{H}$ is $o_n(1)$ to prove that $\Pr[\mathcal{H} \in \mathcal{N}] = o_n(1)$.

Suppose that the hypergraph h' satisfies $d - 1 - \frac{1}{\alpha(h')} \geq \frac{2d-4}{2d-1}$. Then,

$$\mathbb{E}[X_{h'}] \leq n^{|v(h')|}p^{|e(h')|} \leq (at^{-d})^{|e(h')|}n^{|e(h')|(\frac{1}{\alpha(h')}-(d-1-\delta))} \leq (at^{-d})^{|e(h')|}n^{-(\frac{2d-4}{2d-1}-\delta)|e(h')|}.$$

Observe that $\frac{2d-4}{2d-1} - \delta > 0$ since $d \geq 6$. Furthermore, the number of hypergraphs h' with v vertices and k hyperedges is at most $\binom{\binom{v}{d}}{k}$. Hence,

$$
\begin{aligned}
\sum_{h' \in \mathcal{R}} \mathbb{E}[X_{h'}] &\leq \sum_{\substack{d \leq v \leq Ct, \\ k \geq \frac{1}{d-1-\frac{2d-4}{2d-1}} v}} \binom{\binom{v}{d}}{k} (at^{-d})^k n^{-(\frac{2d-4}{2d-1}-\delta)k} \\
&\leq \sum_{\substack{d \leq v \leq Ct, \\ k \geq \frac{1}{d-1-\frac{2d-4}{2d-1}} v}} v^{dk} (at^{-d})^k n^{-(\frac{2d-4}{2d-1}-\delta)k} \\
&\leq \sum_{d \leq v \leq Ct} \frac{v^d at^{-d} n^{-(\frac{2d-4}{2d-1}-\delta)}}{1 - v^d at^{-d} n^{-(\frac{2d-4}{2d-1}-\delta)}} \\
&\leq Ct \frac{(Ct)^d at^{-d} n^{-(\frac{2d-4}{2d-1}-\delta)}}{1 - (Ct)^d at^{-d} n^{-(\frac{2d-4}{2d-1}-\delta)}} \\
&= o_n(1).
\end{aligned} \tag{6.4}
$$

Observe that we have used $\lim_{n\to\infty} \log_n(t) = 0$. This implies that $\sum_{h' \in \mathcal{R}} \Pr[h' \in \mathcal{H}] \leq \sum_{h' \in \mathcal{R}} \mathbb{E}[X_{h'}] = o_n(1)$. Applying union bound proves that the probability that an element of $\mathcal{R}$ appears in $\mathcal{H}$ is $o_n(1)$. ■

Theorem 6.4. *Suppose $d = 5$, $\delta = \frac{2d-4}{2d-1} = \frac{d-1}{d+1}$, and $p = o_n(t^{-d-1} n^{-d+1+\delta})$, where $\lim_{n\to\infty} \log_n(t!) = \infty$. Then the exact recovery probability is $1 - o_n(1)$.*

Proof. This theorem can be proved using the proof of Theorem 6.3. Similarly, we analyze when the two-connected components have size $O_n(t)$ with probability $1 - o_n(1)$. The only difference is in (6.4). ■

Theorem 6.5. *Suppose $d = 3, 4$, $\delta = \frac{2d-4}{2d-1}$, and $p = o_n(n^{-d+1+\delta})$. Then the exact recovery probability is $1 - o_n(1)$.*

Proof. This theorem can be proved using the fact that the 2-connected components have at most $O_n(1)$ vertices from [5, Lemma 32] and the proof of Theorem 6.3. ■

From Theorem 5.7 or [5, Appendix A], when $p \propto (1 + o_n(1)) n^{-d+1+\frac{2d-4}{2d-1}}$ for $d = 3, 4, 5$, then the probability of exact recovery is $1 - \Omega_n(1)$. Then, we see that Theorem 6.5 describes the phase transition in the exact recovery probability very sharply when $d = 3, 4$. Theorem 6.4 also captures the phase transition in the exact recovery probability for $d = 5$, but the transition is not as sharp.

Appendix A

Entropy and exact recovery

For simplicity, assume that the logarithms in entropy are base e. In this section we mainly analyze the entropy of $\mathrm{Proj}(\mathcal{H})$ and hence the conditional entropy $H(\mathcal{H}|\mathrm{Proj}(\mathcal{H})) = H(\mathcal{H}) - H(\mathrm{Proj}(\mathcal{H}))$.

Lemma A.1. *Suppose $\delta > \frac{d-1}{d+1}$, $c > 0$, and $p = (c + o_n(1))n^{-d+1+\delta}$. Then there exists $C > 0$ such that C does not depend on c and $H(\mathcal{H}|\Psi) \geq (C + o_n(1))H(\mathcal{H})$.*

Proof. Since $H(\mathcal{H}|\Psi) = H(\mathcal{H}) - H(\Psi)$, it suffices to prove that $H(\mathcal{H}) - H(\Psi) = \Omega_n(H(\mathcal{H}))$ if $\delta > \frac{d-1}{d+1}$.

First observe that

$$H(\mathcal{H}) = -\binom{n}{d} H_B(p) = c\frac{d-1-\delta}{d!}\log(n)n^{1+\delta} + o_n(\log(n)n^{1+\delta}). \tag{A.1}$$

For all $i, j \in [n]$ such that $i \neq j$, let $X_{\{i,j\}} = \mathbf{1}_{\{i,j\}\in e(\Psi)}$. We have that

$$H(\Psi) = H(X_{\{i,j\}} : 1 \leq i < j \leq n) \leq \sum_{1\leq i<j\leq n} H(X_{\{i,j\}}) = \binom{n}{2} H_B(1 - (1-p)^{\binom{n-2}{d-2}}). \tag{A.2}$$

Bernoulli's inequality implies that

$$1 - (1-p)^{\binom{n-2}{d-2}} \leq \binom{n-2}{d-2} p$$

and for sufficiently large n we have that $\binom{n-2}{d-2}p \leq \frac{1}{e}$. Hence, for sufficiently large n we have that

$$-(1 - (1-p)^{\binom{n-2}{d-2}})\log(1 - (1-p)^{\binom{n-2}{d-2}}) \leq -\binom{n-2}{d-2}p\log(\binom{n-2}{d-2}p)$$

since $-x\log(x)$ increases as x increases over $[0, \frac{1}{e}]$. Then,

$$\begin{aligned} H_B(1-(1-p)^{\binom{n-2}{d-2}}) &\leq -\binom{n-2}{d-2}p\log(\binom{n-2}{d-2}p) - (1-p)^{\binom{n-2}{d-2}}\log((1-p)^{\binom{n-2}{d-2}}) \\ &= (1-\delta)\binom{n-2}{d-2}p\log(n) + o_n(\log(n)n^{\delta-1}). \end{aligned}$$

Using this inequality and (A.2) gives that

$$H(\Psi) \leq c\frac{1-\delta}{2(d-2)!}\log(n)n^{1+\delta} + o_n(\log(n)n^{\delta+1}).$$

Using this inequality and (A.1) then gives that it suffices to prove that

$$\frac{1-\delta}{2(d-2)!} < \frac{d-1-\delta}{d!};$$

afterwards we can let $C = \frac{\frac{d-1-\delta}{d!} - \frac{1-\delta}{2(d-2)!}}{\frac{d-1-\delta}{d!}}$, which does not depend on c. This inequality is true because $\delta > \frac{d-1}{d+1}$, which finishes the proof. ■

Firstly observe that $H(\mathcal{H}|\Psi) = H(\mathcal{H}) - H(\Psi)$ and

$$H(\Psi) \leq H(\mathcal{H}) = \Theta_n(n^{1+\delta}\log(n)).$$

We analyze $H(\mathcal{H}|\Psi)$ and $H(\Psi)$ in the next result.

Theorem A.2. $H(\Psi) = \Theta_n(n^{1+\delta}\log(n))$. *Furthermore*

$$H(\mathcal{H}|\Psi) = \begin{cases} o_n(n^{1+\delta}) & \text{if } \delta < \frac{d-1}{d+1}, \\ O_n(n^{1+\delta}) & \text{if } \delta = \frac{d-1}{d+1}, \\ \Theta_n(n^{1+\delta}\log(n)) & \text{if } \delta > \frac{d-1}{d+1}. \end{cases}$$

Proof. We have that

$$\begin{aligned} H(\Psi) = I(\mathcal{H};\Psi) &\geq \binom{n}{d} I(\mathbf{1}_{[d]\in e(\mathcal{H})};\Psi) \\ &= \binom{n}{d}\left(H_B(p) - \sum_{G\in\mathcal{G}} \Pr[\Psi = G] H(\mathbf{1}_{[d]\in e(\mathcal{H})}|\Psi = G)\right) \\ &= \binom{n}{d}\left(H_B(p) - \sum_{G\in\mathcal{G}, \binom{[d]}{2}\subset e(G)} \Pr[\Psi = G] H(\mathbf{1}_{[d]\in e(\mathcal{H})}|\Psi = G)\right). \end{aligned}$$

Observe that $q = \sum_{G\in\mathcal{G}, \binom{[d]}{2}\subset E(G)} \Pr[\Psi = G]$. Then Jensen's inequality implies that

$$\sum_{G\in\mathcal{G}, \binom{[d]}{2}\subset e(G)} \Pr[\Psi = G] H(\mathbf{1}_{[d]\in e(\mathcal{H})}|\Psi = G) \leq qH_B\left(\frac{p}{q}\right).$$

Therefore

$$\begin{aligned} H(\Psi) &\geq \binom{n}{d}\left(H_B(p) - qH_B\left(\frac{p}{q}\right)\right) \\ &= \binom{n}{d}\left(-(1-p)\ln(1-p) - p\ln(q) + (q-p)\ln(1-\frac{p}{q})\right). \end{aligned} \tag{A.3}$$

Assume that $\delta < \frac{d-1}{d+1}$. Then Corollary 3.5 gives that $q = (1+o_n(1))p$ so $qH_B(\frac{p}{q}) = o_n(p)$ and (A.3) gives that $H(\Psi) = H(\mathcal{H}) - o_n(n^{1+\delta})$

Next assume that $\delta = \frac{d-1}{d+1}$. We have that $q = (1+\Theta_n(1))p$ from Corollary 3.5 so $qH_B(\frac{p}{q}) = O_n(p)$ and (A.3) gives that $H(\Psi) = H(\mathcal{H}) - O_n(n^{1+\delta})$.

Suppose $\delta > \frac{d-1}{d+1}$. Then $q = \omega_n(p)$ from Corollary 3.5 so if n is sufficiently large then Lemma A.5 gives that

$$(q-p)\ln(1-\frac{p}{q}) \geq -2(q-p)\frac{p}{q} = \Omega_n(p)$$

and using (A.3) gives that $H(\Psi) = \Omega_n(n^{1+\delta}\log(n))$. Furthermore Lemma A.1 gives that $H(\mathcal{H}|\Psi) = \Theta_n(H(\mathcal{H}))$. ■

Corollary A.3. *The conditional entropy $H(\mathcal{H}|\Psi)$ is $o_n(H(\mathcal{H}))$ if $\delta \leq \frac{d-1}{d+1}$.*

Proof. This follows from Theorem A.2. ■

Theorem A.4. *If n is sufficiently large then*

$$H(\Psi) \geq \left(1 - \frac{\binom{n}{d-2}p}{1-\binom{n}{d-2}p}\right)\sum_{i=2}^{n-d+2}(i-1)H_B(1-(1-p)^{\binom{n-i}{d-2}}) = \Omega(n^{1+\delta}\log(n)).$$

Proof. Assume that n is sufficiently large throughout the proof. Suppose the set of vertices in $\mathcal{H}$ is $[n]$. We construct a sequence of edges S. Initialize S as the empty sequence. From $i=2$ to $i=n$, add the edges $\{i,j\}$ to S from $j=1$ to $j=i-1$. Note that S contains each edge between two vertices of $[n]$ exactly once. It follows that

$$H(\Psi) = \sum_{a=1}^{\binom{n}{2}} H(X_{S_a}|X_{S_b}, 1 \leq b \leq a-1).$$

Suppose $1 \leq a \leq \binom{n}{2}$. Assume that $S_a = \{i,j\}$ where $1 \leq j < i \leq n$. Observe that the random variables $\mathbf{1}_{h\in\mathcal{H}}$ for $h \in \binom{([n]\setminus[i-1])\cup\{j\}}{d}$ are independent of the random variables X_{S_b} for $1 \leq b \leq a-1$. This is because no hyperedge in $\binom{([n]\setminus[i-1])\cup\{j\}}{d}$ contains S_b as an edge for $1 \leq b \leq a-1$. The probability that a hyperedge in $\binom{([n]\setminus[i-1])\cup\{j\}}{d}$ that contains S_a is present in $\mathcal{H}$ is

$$1-(1-p)^{\binom{n-i}{d-2}}.$$

Suppose $x_b \in \{0,1\}$ for $1 \leq b \leq a-1$. Hence,

$$\Pr[X_{S_a} = 1|X_{S_b} = x_b, 1 \leq b \leq a-1] \geq 1-(1-p)^{\binom{n-i}{d-2}}.$$

Note that $1-(1-p)^{\binom{n-i}{d-2}} \leq \binom{n-i}{d-2}p \leq \frac{1}{2}$ if n is sufficiently large. Thus, if

$$\Pr[X_{S_a} = 1|X_{S_b} = x_b, 1 \leq b \leq a-1] \leq (1-p)^{\binom{n-i}{d-2}}$$

as well, then

$$
\begin{aligned}
H(X_{S_a}|X_{S_b}=x_b, 1\le b\le a-1) &= H_B(\Pr[X_{S_a}=1|X_{S_b}=x_b, 1\le b\le a-1]) \\
&\ge H_B(1-(1-p)^{\binom{n-i}{d-2}}).
\end{aligned} \tag{A.4}
$$

We have that

$$\mathbb{E}_{X_{S_b},1\le b\le a-1}[\Pr[X_{S_a}=1|X_{S_b}=x_b, 1\le b\le a-1]] = \Pr[X_{S_a}=1] \le \binom{n}{d-2}p.$$

Therefore, Markov's inequality implies that

$$\Pr_{X_{S_b},1\le b\le a-1}[\Pr[X_{S_a}=1|X_{S_b}=x_b, 1\le b\le a-1] > (1-p)^{\binom{n-i}{d-2}}] \le \frac{\binom{n}{d-2}p}{(1-p)^{\binom{n-i}{d-2}}}.$$

Observe that $\frac{\binom{n}{d-2}p}{(1-p)^{\binom{n-i}{d-2}}} < \frac{\binom{n}{d-2}p}{1-\binom{n}{d-2}p}$. Using (A.4) then gives that

$$
\begin{aligned}
&H(X_{S_a}|X_{S_b}, 1\le b\le a-1) = \mathbb{E}_{X_{S_b},1\le b\le a-1}[H(X_{S_a}|X_{S_b}=x_b, 1\le b\le a-1)] \\
&\ge (1-\Pr_{X_{S_b},1\le b\le a-1}[\Pr[X_{S_a}=1|X_{S_b}=x_b, 1\le b\le a-1] \le (1-p)^{\binom{n-i}{d-2}}]) \\
&\cdot H_B(1-(1-p)^{\binom{n-i}{d-2}}) \\
&\ge (1-\frac{\binom{n}{d-2}p}{1-\binom{n}{d-2}p})H_B(1-(1-p)^{\binom{n-i}{d-2}}).
\end{aligned}
$$

Summing this inequality from $a=1$ to $a=\binom{n}{2}$ gives

$$H(\Psi) \ge (1-\frac{\binom{n}{d-2}p}{1-\binom{n}{d-2}p})\sum_{i=2}^{n-d+2}(i-1)H_B(1-(1-p)^{\binom{n-i}{d-2}}).$$

Observe that we use the fact that if $i > n-d+2$ then $H_B(1-(1-p)^{\binom{n-i}{d-2}}) = 0$. It is straightforward to check that this lower bound on $H(\Psi)$ is $\Omega_n(n^{1+\delta}\log(n))$. ■

A.1 Exact recovery

Lemma A.5. *$e^{-p} \ge 1-p$ for all $n\ge 1$ and $e^{-2p} \le 1-p$ if n is sufficiently large.*

Theorem A.6. *Suppose $\mathcal{X}$ is a finite set. Suppose $\mathcal{S} \subset \mathbb{Z}_{\ge 1}$ contains infinitely many elements. Suppose $n \in \mathcal{S}$. Suppose $f_n : \mathcal{X}^n \to \mathcal{Y}_n$ and $\mathcal{A}_n : \mathcal{Y}_n \to \mathcal{X}^n$ are functions such that for all $y \in \mathcal{Y}_n$, $f_n(\mathcal{A}_n(y)) = y$. Suppose $X^n \in \mathcal{X}^n$ is a random variable such that X_i, $i \ge 1$ are independent and identically distributed with distribution p_n. Assume that $x^- \in \mathcal{X}$ and $\lim_{n\to\infty} p_n(x) = 0$ and $\lim_{n\to\infty} np_n(x) = \infty$ for all $x \in \mathcal{X}^- := \mathcal{X}\backslash\{x^-\}$, where the limits are over $\mathcal{S}$. Also assume that $\sup_{y\in\mathcal{Y}_n}\Pr[f_n(X^n)=y] \le \frac{1}{e}$ if n is sufficiently large. The probability that $\mathcal{A}_n(f_n(X^n))$ is not X^n is at least $1-\frac{H(f_n(X^n))}{H(X^n)}+o_n(1)$, where the asymptotic term $o_n(1)$ does not depend on f_n.*

Proof. Suppose $n \in \mathcal{S}$. Let $\mathcal{U}_n$ be the set of $x^n \in \mathcal{X}^n$ such that there does not exist $y \in \mathcal{Y}_n$ such that $\mathcal{A}_n(y) = x^n$. Observe that

$$\Pr[\mathcal{A}_n(f_n(X^n)) \neq X^n] = \Pr[X^n \in \mathcal{U}_n] \tag{A.5}$$

In order to prove this, it suffices to prove that $\mathcal{A}_n(f_n(x^n)) = x^n \Leftrightarrow x^n \notin \mathcal{U}_n$. Assume that $\mathcal{A}_n(f_n(x^n)) = x^n$; then, it is clear that $x^n \notin \mathcal{U}_n$. Next, assume that $x^n \notin \mathcal{U}_n$. Suppose $x^n = \mathcal{A}_n(y)$ for $y \in \mathcal{Y}_n$. We have that $f_n(x^n) = f_n(\mathcal{A}_n(y)) = y$, so $\mathcal{A}_n(f_n(x^n)) = \mathcal{A}_n(y) = x^n$.

First observe that

$$\begin{aligned} H(X^n) = &- \sum_{y \in \mathcal{Y}_n} \Pr[X^n = \mathcal{A}_n(y)] \log(\Pr[X^n = \mathcal{A}_n(y)]) \\ &- \sum_{x^n \in \mathcal{U}_n} \Pr[X^n = x^n] \log(\Pr[X^n = x^n]). \end{aligned} \tag{A.6}$$

Suppose $x^n \in \mathcal{U}_n$. Then,

$$\log(\Pr[X^n = x^n]) = (n - \sum_{x \in \mathcal{X}^-} N_x(x^n)) \log(p_n(x^-)) + \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)).$$

Using Lemma A.5 and the fact that $\lim_{n\to\infty} p_n(x) = 0$ for $x \in \mathcal{X}^-$ gives that

$$0 \geq (n - \sum_{x \in \mathcal{X}^-} N_x(x^n)) \log(p_n(x^-)) \geq -2n \sum_{x \in \mathcal{X}^-} p_n(x)$$

if n is sufficiently large. Observe that

$$H(X^n) \geq -n \sum_{x \in \mathcal{X}^-} p_n(x) \log(p_n(x))$$

and $-\log(p_n(x)) = \Omega_n(1)$ since $p_n(x) = o_n(1)$ for all $x \in \mathcal{X}^-$. Hence

$$(n - \sum_{x \in \mathcal{X}^-} N_x(x^n)) \log(p_n(x^-)) = o_n(H(X^n)) \tag{A.7}$$

so

$$\log(\Pr[X^n = x^n]) = \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) + o_n(H(X^n)).$$

This implies that

$$\begin{aligned} &- \sum_{x^n \in \mathcal{U}_n} \Pr[X^n = x^n] \log(\Pr[X^n = x^n]) \\ &= o_n(H(X^n)) - \sum_{x^n \in \mathcal{U}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)). \end{aligned} \tag{A.8}$$

Let $\mathcal{W}_n$ be the set of $x^n \in \mathcal{X}^n$ such that

$$N_x(x^n) \in [np_n(x) - (np_n(x))^{\frac{3}{4}}, np_n(x) + (np_n(x))^{\frac{3}{4}}]$$

for all $x \in \mathcal{X}^-$. For all $x \in \mathcal{X}^-$ the random variable variable $|N_x(X^n)|$ has mean $np_n(x)$ and variance $np_n(x)(1-p_n(x))$. Because $p_n(x) = o_n(1)$ for all $x \in \mathcal{X}^-$, Chebyshev's inequality implies that $X^n \in \mathcal{W}_n$ with probability $1 - o_n(1)$. Therefore, $\Pr[X^n \in \mathcal{U}_n \backslash \mathcal{W}_n] = o_n(1)$. Observe that

$$\begin{aligned}
&\sum_{x^n \in \mathcal{U}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) \\
&= \sum_{x^n \in \mathcal{U}_n \cap \mathcal{W}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) \\
&+ \sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) \\
&= \Pr[X^n \in \mathcal{U}_n] \left(n \sum_{x \in \mathcal{X}^-} p_n(x) \log(p_n(x)) \right) \\
&+ \sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) + o_n(H(X^n)).
\end{aligned}$$

Suppose $x \in \mathcal{X}^-$. We have that

$$\begin{aligned}
\sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] N_x(x^n) &\leq \Big(\sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] \Big)^{\frac{1}{2}} \Big(\sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] N_x(x^n)^2 \Big)^{\frac{1}{2}} \\
&\leq \Big(\sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] \Big)^{\frac{1}{2}} \big(n^2 p_n(x)^2 + np_n(x)(1-p_n(x)) \big)^{\frac{1}{2}} \\
&= o_n(np_n(x)).
\end{aligned}$$

Hence

$$\sum_{x^n \in \mathcal{U}_n \backslash \mathcal{W}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) = o_n(H(X^n)).$$

Therefore using (A.7) with $x^n = (x^-)^n$ gives that

$$\begin{aligned}
&\sum_{x^n \in \mathcal{U}_n} \Pr[X^n = x^n] \sum_{x \in \mathcal{X}^-} N_x(x^n) \log(p_n(x)) \\
&= \Pr[X^n \in \mathcal{U}_n] \left(n \sum_{x \in \mathcal{X}^-} p_n(x) \log(p_n(x)) \right) + o_n(H(X^n)) \\
&= \Pr[X^n \in \mathcal{U}_n] H(X^n) + o_n(H(X^n)).
\end{aligned}$$

Then, using (A.6) gives that

$$H(X^n) = -\sum_{y \in \mathcal{Y}_n} \Pr[X^n = \mathcal{A}_n(y)] \log(\Pr[X^n = \mathcal{A}_n(y)]) + \Pr[X^n \in \mathcal{U}_n] H(X^n) + o_n(H(X^n)). \tag{A.9}$$

For all $y \in \mathcal{Y}_n$ we have that $\Pr[f_n(X^n) = y] \leq \frac{1}{e}$ if n is sufficiently large and

$$\Pr[f_n(X^n) = y] \geq \Pr[X^n = \mathcal{A}_n(y)].$$

Afterwards using the fact that the function $-x\log(x)$ increases as x increases over $[0, \frac{1}{e}]$ gives that

$$H(f_n(X^n)) \geq -\sum_{y\in\mathcal{Y}_n} \Pr[X^n = \mathcal{A}(y)]\log(\Pr[X^n = \mathcal{A}(y)]) \tag{A.10}$$

if n is sufficiently large. Afterwards using (A.9) and (A.10) gives that

$$H(X^n) - H(f_n(X^n)) \leq \Pr[X^n \in \mathcal{U}_n]H(X^n) + o_n(H(X^n)).$$

This finishes the proof after using (A.5). ■

Lemma A.7. *Suppose $\delta < 1$. Then*

$$\max_{G\in\mathcal{G}} \Pr[\Psi = G] = o_n(1).$$

Proof. Suppose G has at least $2\binom{d}{2}p\binom{n}{d} = \Theta(n^{1+\delta})$ edges. Then the number of hyperedges in a preimage H of G has at least $\frac{2\binom{d}{2}p\binom{n}{d}}{\binom{d}{2}} = 2p\binom{n}{d}$ hyperedges. Since $e(\mathcal{H})$ has mean $p\binom{n}{d}$ and variance $o_n(n^{2+2\delta})$,

$$\Pr[\Psi = G] \leq \frac{o_n(n^{2+2\delta})}{(p\binom{n}{d})^2} = o_n(1).$$

from Chebyshev's inequality.

Next suppose G has less than $2\binom{d}{2}p\binom{n}{d}$ edges. For some $m \geq 1$ there exists G_i, $1 \leq i \leq m$ which are distinct isomorphic copies of G such that each element of $\binom{[n]}{2}$ is an edge of G_i for some i. We have that

$$m > \frac{\binom{n}{2}}{2\binom{d}{2}p\binom{n}{d}} = \Omega_n(n^{1-\delta}).$$

Since the G_i are distinct and $\Pr[\Psi = G_i] = \Pr[\Psi = G]$ for $1 \leq i \leq m$,

$$\Pr[\Psi = G] \leq \frac{1}{m} = O_n(n^{\delta-1}) = o_n(1).$$

■

Remark A.8. If $\delta > 1$ then $\Pr[\mathrm{Proj}(\mathcal{H}) = \binom{[n]}{2}] = 1 - o_n(1)$.

We give an example of an application of Theorem A.6. Observe that the following corollary is weaker than Theorem 1.7, which proves that in the same regime, the probability of exact recovery is $o_n(1)$.

Corollary A.9. *Suppose $d \geq 5$ and $\delta < \frac{d-1}{d+1}$. Then, the probability of exact recovery is $1 - \Omega_n(1)$.*

Proof. By Lemma A.7, we can apply Theorem A.6 to get that the probability of exact recovery is at most $\frac{H(\Psi)}{H(\mathcal{H})} - o_n(1)$. By Lemma A.1, this quantity is $1 - \Omega_n(1)$. ■

References

[1] E. Abbe. "Community Detection and Stochastic Block Models: Recent Developments". *Journal of Machine Learning Research* 18.177 (2018), pp. 1–86.

[2] E. Abbe and A. Montanari. "Conditional Random Fields, Planted Constraint Satisfaction, and Entropy Concentration". *Theory of Computing* 11.17 (2015), pp. 413–443.

[3] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová. "Optimal errors and phase transitions in high-dimensional generalized linear models". *Proceedings of the National Academy of Sciences* 116.12 (2019), pp. 5451–5460.

[4] J. Barbier, N. Macris, and C. Rush. "All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 14915–14926.

[5] G. Bresler, C. Guo, and Y. Polyanskiy. "Thresholds for Reconstruction of Random Hypergraphs From Graph Projections". In: *Proceedings of Thirty Seventh Conference on Learning Theory*. Vol. 247. Proceedings of Machine Learning Research. PMLR, 2024, pp. 632–647.

[6] G. Bresler, Y. Polyanskiy, C. Guo, and A. Yao. "Partial and Exact Recovery of a Random Hypergraph from its Graph Projection". *arXiv preprint arXiv:2502.14988* (2025).

[7] Y. Chen, C. Fang, Z. Lin, and B. Liu. "Relational learning in pre-trained models: a theory from hypergraph recovery perspective". In: *Proceedings of the 41st International Conference on Machine Learning*. ICML'24. Vienna, Austria: JMLR.org, 2024.

[8] S. Cole and Y. Zhu. "Exact recovery in the hypergraph stochastic block model: A spectral algorithm". *Linear Algebra and its Applications* 593 (2020), pp. 45–73.

[9] O. Cooley, M. Kang, and C. Koch. "The Size of the Giant Component in Random Hypergraphs: a Short Proof". *The Electronic Journal of Combinatorics* 26.3 (2019).

[10] O. Cooley, M. Kang, and Y. Person. "Largest Components in Random Hypergraphs". *Combinatorics, Probability and Computing* 27.5 (2018), pp. 741–762.

[11] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. "Inference and Phase Transitions in the Detection of Modules in Sparse Networks". *Phys. Rev. Lett.* 107 (6 Aug. 2011), p. 065701.

[12] V. Feldman, W. Perkins, and S. Vempala. “On the Complexity of Random Satisfiability Problems with Planted Solutions”. In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC ’15. New York, NY, USA: Association for Computing Machinery, 2015, pp. 77–86.

[13] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth. “A Unifying Tutorial on Approximate Message Passing”. *Foundations and Trends® in Machine Learning* 15.4 (2022), pp. 335–536.

[14] J. Gaudio and N. Joshi. “Community Detection in the Hypergraph SBM: Exact Recovery Given the Similarity Matrix”. In: *Proceedings of Thirty Sixth Conference on Learning Theory*. Vol. 195. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 469–510.

[15] C. Kim, A. S. Bandeira, and M. X. Goemans. “Stochastic Block Model for Hypergraphs: Statistical limits and a semidefinite programming approach”. *arXiv preprint arXiv:1807.02884* (2018).

[16] F. Krzakala and L. Zdeborová. “Hiding Quiet Solutions in Random Constraint Satisfaction Problems”. *Phys. Rev. Lett.* 102 (23 June 2009), p. 238701.

[17] J. Lee, D. Kim, and H. W. Chung. “Robust Hypergraph Clustering via Convex Relaxation of Truncated MLE”. *IEEE Journal on Selected Areas in Information Theory* 1.3 (2020), pp. 613–631.

[18] Y. Polyanskiy and Y. Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2025.

[19] G. Reeves, J. Xu, and I. Zadik. “The All-or-Nothing Phenomenon in Sparse Linear Regression”. In: *Proceedings of the Thirty-Second Conference on Learning Theory*. Vol. 99. Proceedings of Machine Learning Research. PMLR, June 2019, pp. 2652–2663.

[20] Y. Wang and J. Kleinberg. “From Graphs to Hypergraphs: Hypergraph Projection and its Reconstruction”. In: *The Twelfth International Conference on Learning Representations*. 2024.

[21] Y. Wu, J. Xu, and S. H. Yu. “Settling the Sharp Reconstruction Thresholds of Random Graph Matching”. *IEEE Transactions on Information Theory* 68.8 (2022), pp. 5391–5417.

[22] J.-G. Young, G. Petri, and T. P. Peixoto. “Hypergraph reconstruction from network data”. *Communications Phyiscs* 4.135 (2021).