Unveiling the Role of Randomization in Multiclass Adversarial Classification: Insights from Graph Theory

Lucas Gnecco Heredia[1], **Matteo Sammut**[1], **Muni Sreenivas Pydi**[1]
Rafael Pinot[2], **Benjamin Negrevergne**[1], **Yann Chevaleyre**[1]
[1] LAMSADE, Université Paris Dauphine - PSL, CNRS, Paris, France.
[2] Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, Paris, France.

Abstract

Randomization as a mean to improve the adversarial robustness of machine learning models has recently attracted significant attention. Unfortunately, much of the theoretical analysis so far has focused on binary classification, providing only limited insights into the more complex multiclass setting. In this paper, we take a step toward closing this gap by drawing inspiration from the field of graph theory. Our analysis focuses on discrete data distributions, allowing us to cast the adversarial risk minimization problems within the well-established framework of set packing problems. By doing so, we are able to identify three structural conditions on the support of the data distribution that are necessary for randomization to improve robustness. Furthermore, we are able to construct several data distributions where (contrarily to binary classification) switching from a deterministic to a randomized solution significantly reduces the optimal adversarial risk. These findings highlight the crucial role randomization can play in enhancing robustness to adversarial attacks in multiclass classification.

1 INTRODUCTION

Modern machine learning models such as neural networks are highly vulnerable to small (imperceptible) adversarial perturbations Biggio et al. (2013); Szegedy et al. (2014). Over the past decade, significant efforts have been made to develop strong attacks Goodfellow et al. (2014); Kurakin et al. (2016); Carlini and Wagner (2017); Croce and Hein (2020), practical defense mechanisms Madry et al. (2017); Moosavi-Dezfooli et al. (2019); Cohen et al. (2019); Croce and Hein (2020); Salman et al. (2019), and advance the theoretical understanding of this phenomenon Awasthi et al. (2021); Pydi and Jog (2023); Trillos et al. (2023b); Pydi and Jog (2023); Meunier et al. (2022); Frank and Niles-Weed (2024). Among existing defense mechanisms, a prominent class known as *randomized defenses* aims to enhance model robustness through the use of randomization. Initially explored in empirical studies Xie et al. (2017); Dhillon et al. (2018); Panousis et al. (2021), this approach has since gained significant attention in theoretical research Pinot et al. (2019, 2020, 2022); Meunier et al. (2021); Gnecco Heredia et al. (2024); Yang et al. (2022); Huang et al. (2022).

While the study of randomized defenses remains an open area of research, the extensive literature on adversarial example theory and on the specific role of randomization has provided valuable insights into how well randomized defenses might work, particularly in binary classification. Notably, recent work Gnecco Heredia et al. (2024) has shown that if the hypothesis class is sufficiently rich (e.g., when considering the set of all Borel measurable functions) then randomized strategies cannot enhance the robustness of the optimal classifier in binary classification. A similar result holds for Lebesgue measurable functions, when combining results from Meunier et al. (2021) and Pydi and Jog (2023).

The case of multiclass classification however remains largely understudied and missunderstood. One might assume that existing results in the binary classification framework would naturally extend to the multiclass setting, as is often the case in standard classification. However, perhaps unsurprisingly, this does not hold in the adversarial classification context. Recent work Trillos et al. (2023a); Dai et al. (2024) has indeed presented a counterexample demonstrating that, in the multiclass setting, there exist simple data distributions with finite

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258.

support where randomization improves the robustness of the optimal classifier, even when considering very rich hypothesis classes. The proof of this counterexample heavily relies on a symmetry argument, raising several follow-up questions about whether this example can be generalized. Specifically, we are interested in addressing the following question:

Can we characterize the data distributions for which randomization enhances the adversarial robustness in multiclass classification?

In this paper, we take a step toward providing a principled answer to this question. Specifically, we study finite discrete data distributions and analyze the value of the randomization gap (*i.e.*, the difference between the optimal adversarial risk with and without randomization) for Borel measurable functions.[1] To analyze the data distributions, we first present (in Section 3) a graph-theoretical characterization of their intrinsic vulnerabilities using the concepts of conflict graphs and conflict hypergraphs, first introduced in Dai et al. (2024). We then map the computation of their randomization gap to two well-known graph-theoretical problems: set packing and fractional set packing. This approach enables us to derive two important sets of results that characterize the randomization gap based on specific structural properties of the data distribution we consider. Our main contributions are as follows.

Contribution 1: Identification of necessary structures for a positive randomization gap. We demonstrate that the existence of a positive randomization gap (which indicates that randomization enhances robustness) in the multiclass adversarial classification problem is intrinsically linked to the presence of specific structural characteristics in the conflict hypergraph associated with the data distribution we study. Specifically, we show that a data distribution cannot exhibit a positive randomization gap unless its corresponding conflict hypergraph possesses one of the following features: it contains a hole, an anti-hole, or it misses to include all the cliques of its skeleton (as detailed in Section 4). This finding provides a necessary condition for a data distribution to exhibit a positive randomization gap, and thereby helps us gain deeper insights into the conditions under which randomization may be beneficial for robust multiclass classification.

Contribution 2: Improved counterexamples with arbitrarily large randomization gap. While our first contribution provides valuable insights, it does not constitute a sufficient condition for a positive randomization gap. Hence, it does not fully characterize the set of data distributions for which randomization improves robustness. In fact, we expect (see Section 4) that it is impossible to craft a sufficient condition which could be checked on a discrete distribution in polynomial time. Nevertheless, inspired by our structural conditions, we present (in Section 5) several novel counterexamples that generalize and enhance the initial results from Trillos et al. (2023a); Dai et al. (2024). Specifically, we identify a variety of data distributions and existence results for which we can establish that the randomization gap approaches $1/2$. This finding implies that, under certain conditions, the effectiveness of randomization can be arbitrary good.

Our work not only extends existing literature but also opens new avenues for exploring the relationship between data distribution structures and the effectiveness of randomized defenses in the multiclass setting.

2 PRELIMINARIES

Notations. We denote by $\mathbb{R}_+$ the set of non-negative real numbers, *i.e.*, $[0, \infty)$. For a vector $\boldsymbol{u}$, we denote by $\boldsymbol{u}^{(j)}$ the j-th component of $\boldsymbol{u}$. For a positive integer K, we use the notation $[K]$ to refer to the set $\{1, \dots, K\}$. We further denoted by Δ^K, the *probability simplex* in $\mathbb{R}^K$, *i.e.*, $\Delta^K := \{\boldsymbol{u} \in \mathbb{R}_+^K \mid \sum_{i=1}^K \boldsymbol{u}^{(i)} = 1\}$. We also denote $\mathbf{1}_n$ the vector in $\mathbb{R}^n$ with all components one.

Additionally, for any space $\mathcal{Z}$, let $\mathcal{P}(\mathcal{Z})$ represent the set of finite discrete distributions, *i.e.*, probability measures with finite support on $\mathcal{Z}$. A distribution $\mu \in \mathcal{P}(\mathcal{Z})$ is characterized by its support $s_\mu = (z_1, \dots, z_n)$ of size n, and its *probability vector* $\boldsymbol{\omega}_\mu \in \Delta^n$, which assigns a mass $\boldsymbol{\omega}_\mu^{(i)}$ to each point z_i in the support. Finally, for any discrete space $\mathcal{Z}$, we denote by $2^{\mathcal{Z}}$ the *power set* of $\mathcal{Z}$, *i.e.*, the set of all subsets of $\mathcal{Z}$.

Adversarial classification: deterministic setting. Let $\mathcal{X} \subseteq \mathbb{R}^d$, $\mathcal{Y} = [K]$ and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Solving a classification task in the deterministic setting involves finding a function f in the set of functions $\mathcal{F}_{\text{det}} := \{f : \mathcal{X} \to \mathcal{Y}\}$ that minimizes the expected error on μ (a.k.a. the *risk*). More formally, it involves solving the following optimization problem:

$$\inf_{f \in \mathcal{F}_{\text{det}}} \mathbb{E}_{(x,y)\sim\mu} \left[\mathbb{1}\{f(x) \neq y\}\right]. \tag{1}$$

However, the classifiers that are approximate solutions to (1) may be vulnerable to *adversarial examples*. Given a classifier $f \in \mathcal{F}_{\text{det}}$ and a data sample $(x, y) \in \mathcal{X} \times \mathcal{Y}$, an adversarial example is an input $x_{\text{adv}} \in \mathcal{X}$ that is perceptively indistinguishable from x but is misclassified by f, *i.e.*, $f(x_{\text{adv}}) \neq y$. Note that although the notion of perceptibility is a complicated concept that depends on human biology, it is common to measure the magnitude of an adversarial perturbation using an ℓ_p norm (with $p \in [1, \infty]$).

[1] Since we focus on discrete distributions, any well-defined classifier is Borel measurable.

Thus an adversarial example x_{adv} is considered indistinguishable from x when $\|x_{\text{adv}} - x\|_p \leq \epsilon$ with ϵ chosen empirically. To account for the existence of adversarial examples, we now define the problem of *adversarial classification* as follows:

$$\inf_{f \in \mathcal{F}_{\text{det}}} \mathbb{E}_{(x,y)\sim\mu} \left[\sup_{x_{\text{adv}} \in B_p(x,\epsilon)} \mathbb{1}\{f(x_{\text{adv}}) \neq y\} \right], \quad (2)$$

where $B_p(x,\epsilon) \coloneqq \{x_{\text{adv}} \in \mathcal{X} : \|x_{\text{adv}} - x\|_p \leq \epsilon\}$. The value of (2) is the called the *optimal adversarial risk for deterministic classifiers*, and we denote this risk by $\mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu,\epsilon)$ in the rest of this paper.

Adversarial classification: randomized setting. In the randomized setting, the goal is identical, except we now consider a wider set of functions $\mathcal{F}_{\text{rand}}$ that includes *randomized classifiers*. Formally $\mathcal{F}_{\text{rand}}$ can be defined as follows:

$$\mathcal{F}_{\text{rand}} \coloneqq \{f : \mathcal{X} \to \mathcal{P}(\mathcal{Y})\}. \quad (3)$$

There are several differences between the randomized and the deterministic setting. First, if f is a randomized classifier from $\mathcal{F}_{\text{rand}}$ and $x \in \mathcal{X}$ is an arbitrary input, then obtaining a class label for x using f now requires sampling a class label $z \sim f(x)$. Second, the error made by a randomized classifier needs to be computed in expectation with respect to this sampling procedure. By analogy with the deterministic setting, we can thus define the *adversarial classification problem* for *randomized classifiers* as follows:

$$\inf_{f \in \mathcal{F}_{\text{rand}}} \mathbb{E}_{(x,y)\sim\mu} \left[\sup_{x' \in B_p(x,\epsilon)} \mathbb{E}_{z\sim f(x')} \left[\mathbb{1}\{z \neq y\}\right] \right]. \quad (4)$$

The value of (4) is the *optimal adversarial risk for randomized classifiers*, which we denote $\mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu,\epsilon)$.

Randomization gap. Note that any deterministic classifier can be rewritten as a randomized one that only outputs Dirac measures. Hence, the set of deterministic classifiers $\mathcal{F}_{\text{det}}$ can be conceptually considered as a subset of randomized classifiers $\mathcal{F}_{\text{rand}}$. Accordingly, we always have that $\mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu,\epsilon) \leq \mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu,\epsilon)$. In this paper, we are interested in determining the characteristics of μ that may lead this inequality to either become an equality or a strict inequality. In other words, we study the notion of *randomization gap* of a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, defined as:

$$\text{rg}(\mu,\epsilon) \coloneqq \mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu,\epsilon) - \mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu,\epsilon). \quad (5)$$

We offer a novel framework for understanding, and measuring the randomization gap, by focusing on finite discrete distributions. Specifically, we take inspiration from a reformulation of the adversarial risk minimization first introduced in Dai et al. (2024) to express the randomization gap as a graph theoretical problem, as we describe in the next section.

3 COMPUTING THE RANDOMIZATION GAP USING GRAPH THEORETICAL TOOLS

To compute the randomization gap of a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ using graph theoretical tools, we rely on the concept of *conflict hypergraph*, a representation that captures the intrinsic vulnerabilities of the optimal classifier on μ, first introduced in Dai et al. (2024). Using this representation, we map the adversarial risk minimization problems (2) and (4) to well studied graph theoretical problems, known as set packing and fractional set packing. Before diving into the technical details, let us present some graph theoretical terminology.

Graph theoretical terminology. A simple undirected graph $G = (V, E)$ consists of a set of vertices V and a set of edges $E \subseteq \{e \in 2^V \text{ s.t. } |e| = 2\}$ denoting the connections between the vertices. We denote $|G|$ the number of vertices of G. A *hypergraph* $H = (V, \mathcal{E})$ is a generalization of a graph in which the hyperedges are subsets of V and can thus contain more than two vertices. We call k-hyperedges all hyperedges that connect exactly k vertices. Note that a hypergraph that only contains 2-hyperedges is a simple graph.

3.1 Conflict hypergraph, conflict graph

The notions of *conflict hypergraph* and *conflict graph* are both useful to characterize intrinsic vulnerabilities of a finite discrete distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ to adversarial attacks. They rely on the concept of a *conflict*: two points (x_1, y_1) and (x_2, y_2) from the support of μ are said to be in conflict if they are both within close range ($\|x_1 - x_2\|_p \leq 2\epsilon$) and belong to different classes ($y_1 \neq y_2$). Informally, configurations of two or more points in conflict are relevant to our problem because the adversarial risk has to be positive on at least one of these points, thus points involved in a conflict are vulnerable to adversarial attacks. To assess the vulnerability of a distribution μ, we record all the conflicts as hyperedges in a hypergraph that has a vertex for each element of the support μ. However, hypergraphs are sometime difficult to analyze, and thus it is also useful to consider the conflict graph of μ, which only records pairwise conflicts between points. The conflict hypergraph, and the conflict graph, can be defined as:

Definition 3.1 (Adapted from Dai et al. (2024))**.** *Let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with support $s_\mu = \{(x_i, y_i)\}_{i\in[n]}$. For any ℓ_p norm ($p \in (1, \infty]$) and any $\epsilon > 0$, the conflict hypergraph of μ at level ϵ is the hypergraph $H^\epsilon_{s_\mu} = (V, \mathcal{E})$ with $V = [n]$ and hyperedge set defined as follows. A set $e \in 2^V$ is a hyperedge of $H^\epsilon_{s_\mu}$ if and only if both the following assertions hold:*

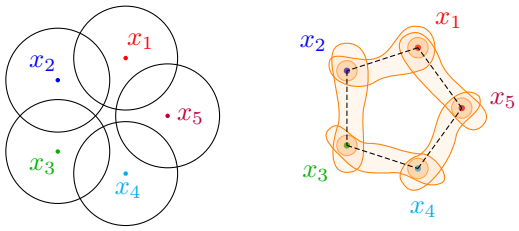

Figure 1

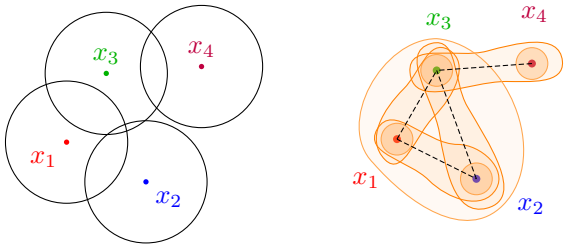

Figure 2

Figures 1,2: On the LHS of each figure, a set S of points of different classes with their ϵ-balls in the ℓ_2 norm. On the RHS, the conflict hypergraph H_S^ϵ represented as orange regions, and the conflict graph G_S^ϵ represented as black dashed lines. For the left figure, the set of hyperedges is $\mathcal{E} = \cup_{i\in[5]}\left[\{i\} \cup \{i, i+1_{\text{mod}5}\}\right]$, while for the right figure it is $\mathcal{E} = 2^{[3]} \cup \{3,4\} \cup \{4\}$. Best viewed in color.

1. *For any distinct $i, j \in e$ we have $y_i \neq y_j$,*

2. $\bigcap_{i\in e} B_p(x_i, \epsilon) \neq \emptyset$ *(i.e. the ϵ-balls overlap).*

Similarly, the conflict graph is the graph $G^\epsilon_{s_\mu} = (V, E)$ whose edge set is the largest subset of 2^V such that for all $e \in E$, both the previous conditions hold and $|e| = 2$.

The topology of these graphs and hypergraphs depends only on the support of the distribution μ. Hence, two distributions with the same support will have the same conflict graph and hypergraph. Clearly, the conflict graph[2] captures less information about the intrinsic adversarial vulnerability of μ than the conflict hypergraph. Nevertheless, as we will demonstrate in the subsequent sections, it remains sufficient to identify many of the key structural properties needed for a positive randomization gap to arise.

Illustration. We present in Figures 1 and 2 examples of finite discrete distributions and their corresponding conflict graphs and hypergraphs. In both cases, we assume a uniform distribution over the points of the support, all of which belong to different classes. On the LHS of each figure, we depict the support of the distributions and the ϵ-balls using the ℓ_2 norm. On the RHS of each figure, the dashed lines between the points represent the edges of the conflict graph $G^\epsilon_{s_\mu}$, and the orange regions represent the hyperedges of $H^\epsilon_{s_\mu}$.

Remark 1. *If no two points in the support of μ are conflicting, then the conflict hypergraph will only contain singletons. In this case, one can easily show that $\mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu, \epsilon) = \mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) = 0$. See e.g., (Meunier et al., 2021, Appendix E.4) for a proof when $K = 2$.*

[2]In graph theory, the conflict graph corresponds to the *2-section* of the conflict hypergraph (see e.g. Berge (1984)).

3.2 Set packing and adversarial risk for deterministic classifiers

Using the notion of conflict hypergraph for a given distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we now map the adversarial risk minimization problem (2) in the deterministic setting to a well studied graph theoretical problem known as the *set packing problem*.

Set packing problem. Let $H = (V, \mathcal{E})$ be a hypergraph and $\boldsymbol{\omega} \in \mathbb{R}_+^{|V|}$ a weight vector for the vertices. The *set packing problem* over $(H, \boldsymbol{\omega})$ seeks a packing (*i.e.*, a subset of vertices that intersects each hyperedge at most once) with maximum cumulative weight. We represent a packing $Q \subseteq V$ using its characteristic vector $q_Q \in \{0,1\}^n$, defined component-wise as $q_Q^{(i)} = \mathbb{1}\{i \in Q\}$ for all $i \in [n]$. Then, the set of possible packings for H can be written as

$$\mathcal{Q}(H) = \left\{ q \in \{0,1\}^n \mid \sum_{i\in e} q^{(i)} \leq 1, \ \forall e \in \mathcal{E} \right\}, \quad (6)$$

and the set packing problem can be formally put as finding a packing with maximal cumulative weight:[3]

$$\text{IP}(H, \boldsymbol{\omega}) := \max_{q\in\mathcal{Q}(H)} \boldsymbol{\omega}^T q. \quad (7)$$

In Theorem 3.1 below, we map the optimal adversarial risk for deterministic classifiers on a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ to the value of the set packing problem on the conflict hypergraph of μ. This result is essentially a byproduct of (Dai et al., 2024, Corollary 1). However, our analytical framework significantly differs from Dai et al. (2024). Hence, we present a proof for Theorem 3.1 in Appendix A for the sake of completeness.

[3]The set packing problem is traditionally defined with a weight vector $\boldsymbol{\omega}_\mu = \mathbf{1}_n$ (see e.g. Schrijver (1979)). For simplicity, we use the name *set packing problem* even though we considered the weighted version throughout the paper.

Theorem 3.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:*

$$1 - \mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu, \epsilon) = \text{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu).$$

Theorem 3.1 allows us to compute the optimal adversarial risk for deterministic classifiers using the set packing problem, which can be expressed as an integer linear problem and for which powerful open-source solvers are available Gurobi Optimization, LLC (2024). In simple cases, such as the examples shown in Figures 1 and 2, the set packing problem can be manually solved, as we present below.

When considering the packings of $H^\epsilon_{s_\mu}$ in Figure 1, we can see that including the point x_1 in the packing automatically excludes two points: x_2 and x_5. This same reasoning applies for all points, which leads to the conclusion that a packing of $H^\epsilon_{s_\mu}$ have size at most 2. We conclude that $\text{IP}(H^\epsilon_{s_\mu}, \frac{1}{5}\mathbf{1}_5) = 2/5$. Therefore, the optimal adversarial risk for deterministic classifiers over the distribution from Figure 1 is $3/5$. For the distribution in Figure 2, one can easily check that one maximal packing is $\{x_2, x_4\}$. Thus, $\text{IP}(H^\epsilon_{s_\mu}, \frac{1}{4}\mathbf{1}_4) = 2/4$ and the optimal (deterministic) adversarial risk is $1/2$.

3.3 Fractional set packings and adversarial risk for randomized classifiers

Similar to the deterministic setting, we can map the adversarial risk minimization problem (4) for randomized classifiers to a relaxed version of the set packing problem, called the *fractional set packing problem*, where the characteristic vector of a packing Q can be fractional, *i.e.*, $q_Q \in [0, 1]^n$. Specifically, given a hypergraph $H = (V, \mathcal{E})$ and weights $\boldsymbol{\omega} \in \mathbb{R}^{|V|}_+$, the fractional set packing problem over $(H, \boldsymbol{\omega})$ consists in finding the *fractional packing* with maximum cumulative weight:

$$\text{FP}(H, \boldsymbol{\omega}) \coloneqq \max_{q \in \mathcal{Q}^{\text{frac}}(H)} \boldsymbol{\omega}^T q, \tag{8}$$

where $\mathcal{Q}^{\text{frac}}(H) = \left\{ q \in [0, 1]^n \mid \sum_{i \in e} q^{(i)} \leq 1, \ \forall e \in \mathcal{E} \right\}$.

Then, similar to Theorem 3.1, we can show that the optimal adversarial risk for randomized classifiers is related to the value of the fractional set packing problem. The proof of Theorem 3.2 is deferred to Appendix A.

Theorem 3.2. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:*

$$1 - \mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) = \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu).$$

Theorem 3.2 provides a way to compute the optimal adversarial risk for randomized classifiers using linear programming, which can be done efficiently. For the simple examples shown in Figures 1 and 2, the fractional set packing can also be solved manually (See Appendix A.2). Hence, the optimal adversarial risks for randomized classifiers for the examples shown in Figures 1 and 2 are, in both cases, $1/2$. Finally, by combining Theorem 3.1 and 3.2 together, we can rewrite the randomization gap of any finite discrete distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ as follows.

Corollary 3.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:*

$$\text{rg}(\mu, \epsilon) = \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) - \text{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{9}$$

Since we have already solved both problems for the examples in Figure 1 and 2, we can apply Corollary 3.1 to compute the randomization gap for these two cases. For Figure 1, we find a positive gap of $1/2 - 2/5 = 1/10$, whereas for Figure 2 the randomization gap is 0. The fact that the randomization gap is positive in Figure 1 but zero in Figure 2, indicates a fundamental difference between the two distributions. However, it remains unclear how to determine which distributions may have a positive randomization gap. In the next section, we will characterize the structural properties of discrete distributions that exhibit a positive randomization gap. This is made possible by the reformulation in Corollary 3.1, which enables us to characterize these distributions by examining their conflict hypergraph representations.

4 LINK BETWEEN STRUCTURAL PROPERTIES OF THE CONFLICT HYPERGRAPH AND THE RANDOMIZATION GAP

In this section, building upon Corollary 3.1, we establish necessary conditions on the conflict hypergraph $H^\epsilon_{s_\mu}$ for a distribution μ to exhibit a positive randomization gap at level ϵ. We will be able to link a positive randomization gap with the presence of specific structures in the conflict hypergraph and the conflict graph of the distribution. Before diving in, let us introduce some graph theoretical terminology that we will use throughout the remainder of the section.

Different types of induced subgraphs. Let $G = (V, E)$ be an arbitrary graph. An *induced subgraph* of G is a graph $G' = (V', E')$ where $V' \subset V$, $E' = \{\{i, j\} \in E \mid i, j \in V'\}$. A *clique* of G is an induced subgraph of G in which every two distinct vertices are adjacent, and is said to be maximal if it cannot be extended by adding a vertex from V to V'. A *hole* of G is an induced subgraph of G consisting of a cycle of more than three vertices. Accordingly, we can define an *anti-hole* of G as an induced subgraph whose complement is a hole in the complement of G. Finally,

a graph is considered *perfect* if it contains neither an odd hole nor an odd anti-hole Chudnovsky et al. (2003).

4.1 Decomposition of the randomization gap using clique hypergraph

We begin by rewriting the randomization gap as the sum of two non-zero terms. Each term in this decomposition will help us uncover key structural features in the conflict hypergraph. To achieve this, we first need to introduce the concept of *clique hypergraph* that plays a central role in the rewriting.

Definition 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we define the clique hypergraph of μ at level ϵ as the hypergraph $C^{\epsilon}_{s_\mu} = (V, \mathcal{E})$, where $V = [n]$ and $\mathcal{E}$ is the set of all maximal cliques of $G^{\epsilon}_{s_\mu}$.*

Coming back to the examples from Figures 1 and 2, we can illustrate the concept of a clique hypergraph. Specifically, in Figure 1, the clique hypergraph has a hyperedge set $\{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{5,1\}\}$, while in Figure 2, the hyperedge set is $\{\{1,2,3\},\{3,4\}\}$. With this notion of clique hypergraph at hand, we now introduce the technical lemma that allows us to decompose the randomization gap as follows.

Lemma 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, $\epsilon > 0$, and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Then we have*

$$\mathrm{IP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{IP}(G^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \quad (10)$$

$$\mathrm{FP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \leq \mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \leq \mathrm{FP}(G^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \quad (11)$$

Combining Corollary 3.1 and Lemma 4.1, we can rewrite the randomization gap using the packing problems on the clique hypergraph $C^{\epsilon}_{s_\mu}$ and the conflict hypergraph $H^{\epsilon}_{s_\mu}$ of μ. Then by reintroducing the value of fractional packing problem on the clique hypergraph, we obtain our decomposition of the randomization gap, as follows:

$$\mathrm{rg}(\mu, \epsilon) = \mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \text{ - } \mathrm{FP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \quad (12)$$

$$+\mathrm{FP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \text{ - } \mathrm{IP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \quad (13)$$

The above decomposition holds true as $\mathrm{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$ can be replaced by $\mathrm{IP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$, thanks to (10). Furthermore, both (12) and (13) are non-negative. Indeed, (12) is non-negative due to (11) and (13) is non-negative because the value of the fractional set packing problem is always an upper bound of the set packing problem. By studying the conditions under which each one of these terms is positive, we can identify the necessary conditions on the conflict hypergraph $H^{\epsilon}_{s_\mu}$ for μ to exhibit a positive randomization gap.

4.2 Necessary structures in the distribution for a positive randomization gap

On the one hand, we note that (12) becomes zero if the conflict hypergraph $H^{\epsilon}_{s_\mu}$ coincides with the clique hypergraph $C^{\epsilon}_{s_\mu}$. Thanks to Berge (1984), we know that this occurs when every clique in the conflict graph $G^{\epsilon}_{s_\mu}$ is a hyperedge in $H^{\epsilon}_{s_\mu}$; in this case $H^{\epsilon}_{s_\mu}$ is said to be *conformal* Berge (1984). Hence, for (12) to be positive, one needs $H^{\epsilon}_{s_\mu}$ to be non-conformal. On the other hand, determining when (13) is non-zero involves using the characterization of perfect graphs Chudnovsky et al. (2003); Conforti et al. (2014). Specifically, (13) is zero if the conflict graph $G^{\epsilon}_{s_\mu}$ is perfect. To summarize, we can express conditions for the existence of a distribution with positive randomization gap as follows.

Theorem 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also $S = \{(x_i, y_i)\}_{i \in [n]}$ be an arbitrary set of points from $\mathcal{X} \times \mathcal{Y}$. There exists a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with support $s_\mu = S$ such that $\mathrm{rg}(\mu, \epsilon) > 0$ if and only if at least one of the following assertions holds true:*

a) H^{ϵ}_S is not conformal (i.e., there exists a clique in G^{ϵ}_S that is not a hyperedge in H^{ϵ}_S).

b) G^{ϵ}_S is not perfect (i.e., it contains at least one odd hole or one odd anti-hole).

Theorem 4.1 is a strong result providing both necessary and sufficient conditions for the *existence* of a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with positive randomization gap. However, it does not provide a testable condition for a *given* distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ to determine whether the randomization gap is positive. The following corollary provides (testable) necessary conditions on the support of μ for a positive randomization gap.

Corollary 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also consider $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. If $\mathrm{rg}(\mu, \epsilon) > 0$, then $G^{\epsilon}_{s_\mu}$ satisfies either assertion a) or assertion b) from Theorem 4.1.*

Hardness of verifying sufficient conditions. Given a distribution μ, Corollary 4.1 identifies a necessary condition for the randomization gap to be positive. Furthermore, this condition is checkable in polynomial time Cornuejols et al. (2003); Boros et al. (2023). However, we conjecture that any condition that is both *necessary and sufficient* cannot be checked in polynomial time, unless "P=NP"[4]. This conjecture is supported by the fact that, for a given discrete distribution μ and $\alpha \in [0, 1]$, checking if $\mathrm{rg}(\mu, \epsilon) \geq \alpha$ is co-NP-complete.

[4]It is widely believed that problems solvable in polynomial time (P) are not the same as problems whose solutions are verifiable in polynomial time (NP).

The proof, based on a reduction from the set packing problem, is in Appendix E.

Special case of $K = 2$. In binary classification, the conflict graph $G^{\epsilon}_{s_\mu}$ can be shown to be bipartite. Since any bipartite graph is perfect, we have that (13) is always zero. Additionally, the conflict hypergraph $H^{\epsilon}_{s_\mu}$ is always conformal, as a clique in $G^{\epsilon}_{s_\mu}$ is simply an edge and thus a hyperedge of $H^{\epsilon}_{s_\mu}$. This implies that (12) is also zero. Combining both arguments, the contrapositive of the implication stated in Corollary 4.1 tells us that the randomization gap is always zero in the binary case, which validates existing results from Bhagoji et al. (2019); Gnecco Heredia et al. (2024).

4.3 The special case of the ℓ_∞ norm

Interestingly, when considering the ℓ_∞ norm, we can prove that the only structures that matter in this case are those related to the perfect graph condition.

Corollary 4.2. *Consider the ℓ_∞ norm, and $\epsilon > 0$ and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with clique hypergraph $C^{\epsilon}_{s_\mu} = (V, \mathcal{E})$. Let $G' = (V', E')$ be an induced subgraph of $G^{\epsilon}_{s_\mu}$. If G' is a clique, then $V' \in \mathcal{E}$. Thus, for any $\boldsymbol{\omega} \in \mathbb{R}_+^{|V|}$:*

$$\mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}) = \mathrm{FP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}).$$

Corollary 4.2 shows that the first source of gap in (12) is always zero when using the ℓ_∞ norm. This implies that the only way for a distribution to exhibit a positive randomization gap is for the conflict graph $G^{\epsilon}_{s_\mu}$ to be non-perfect. In contrast, when using the ℓ_2 norm, such uncovered cliques can exist (see Figure 3). On the other hand, by Lemma C.5, both odd holes and anti-holes can exist when using any ℓ_p norm for $p \in (1, \infty]$ (see Example B.2).

5 THE RANDOMIZATION GAP CAN BE ARBITRARILY CLOSE TO $1/2$

We now identify a variety of discrete distributions for which we can establish that the randomization gap is arbitrarily close to $1/2$. Specifically, even though the structural conditions introduced in the last section are not sufficient, they provide a systematic way to construct distributions with a significantly large randomization gap, thus generalizing the initial results from Trillos et al. (2023a); Dai et al. (2024).

5.1 Large randomization gap based on the conformal condition

We start by designing a simple data distribution for which the randomization gap is close to $1/2$. Let us consider the discrete distribution μ that is uniformly distributed over the canonical basis of $\mathbb{R}^K$ and in which each vector is assigned to a distinct class. Specifically, let $(b_1, \ldots, b_K)$ denote the canonical basis of $\mathbb{R}^K$ and $\mathbf{1}_n$ denote the vector in $\mathbb{R}^n$ with all components to one. We set $s_\mu = \{(b_1, 1), \ldots, (b_K, K)\} \subset \mathbb{R}^K \times [K]$ and $\boldsymbol{\omega}_\mu = \frac{1}{K}\mathbf{1}_K$. Then for $\epsilon = 1/\sqrt{2}$ and considering the ℓ_2 norm, one can verify that the conflict hypergraph $H^{\epsilon}_{s_\mu}$ of μ contains all 2-hyperedges between every pair of points, with no hyperedges of larger size. In other words, the conflict hypergraph $H^{\epsilon}_{s_\mu}$ is a complete graph over K vertices, including self-loops. Furthermore, every set of vertices forms a clique in the conflict graph $G^{\epsilon}_{s_\mu}$, but none of the cliques of size greater than 2 are hyperedges in $H^{\epsilon}_{s_\mu}$. Therefore, $H^{\epsilon}_{s_\mu}$ is not conformal (See Figure 3).

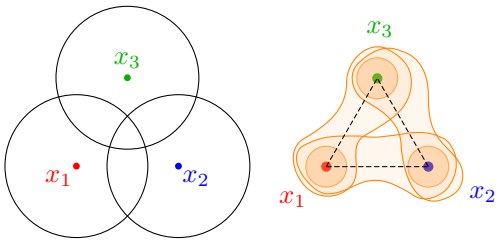

Figure 3: Example of a set of points with a non-conformal conflict hypergraph. The support is the canonical basis of $\mathbb{R}^3$, $\epsilon = 1.1/\sqrt{2}$, and $p = 2$. The LHS illustrates the points in the space together with their ϵ-balls, and the RHS shows $H^{\epsilon}_{s_\mu}$, which is not conformal because it does not contain the hyperedge $\{1, 2, 3\}$.

The vector $q = \frac{1}{2}\mathbf{1}_K$ is feasible for the fractional packing problem over $(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$, *i.e.*, $q \in \mathcal{Q}(H^{\epsilon}_{s_\mu})$, since the conflict hypergraph $H^{\epsilon}_{s_\mu}$ does not contain any hyperedges of size larger than 2. This implies that $\mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \geq \boldsymbol{\omega}_\mu^T q = \frac{1}{2}$. Furthermore, $H^{\epsilon}_{s_\mu}$ contains every possible 2-hyperedge, hence the only possible packings for $H^{\epsilon}_{s_\mu}$ are the singletons. This implies that $\mathrm{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \frac{1}{K}$. Combining these two arguments and using Corollary 3.1, the randomization gap of μ is greater than $\frac{1}{2} - \frac{1}{K}$. Therefore, we can make it arbitrarily close to $1/2$ by choosing K large enough.

The construction presented in this subsection generalizes the initial example from Trillos et al. (2023a); Dai et al. (2024) by scaling it to an arbitrary number of classes K. Nevertheless, it remains restrictive because the distribution μ we design is constrained to having one point per class. Additionally, such a construction is not feasible when considering the ℓ_∞ norm, as indicated by Corollary 4.2. For these reasons, we seek a procedure that enables the construction of less trivial distributions with a large randomization gap, valid for any ℓ_p-norm.

5.2 Large randomization gap based on the perfect graph condition

To design more general distributions, we focus on the perfect graph condition and leverage existing graph constructions from the graph theory literature Chung et al. (1993). Doing so, we obtain the following statement.

Theorem 5.1. *Fix any $\epsilon > 0$ and ℓ_p norm with $p \in (1, \infty]$. For any $\delta > 0$, there exist $d, K \in \mathbb{N}$ and a discrete distribution $\mu \in \mathcal{P}(\mathbb{R}^d \times [K])$ such that*

$$\mathrm{rg}(\mu, \epsilon) \geq 1/2 - \delta, \tag{14}$$

the conflict graph $G^{\epsilon}_{s_\mu}$ is not perfect and the conflict hypergraph $H^{\epsilon}_{s_\mu}$ is conformal. Furthermore, the number of classes satisfies $K \in \mathcal{O}\left(\sqrt{n/\log n}\right)$ with $n = |s_\mu|$.

Theorem 5.1 show that there exist non-trivial distributions (with $K \neq n$) for which the randomization gap is arbitrarily close to $1/2$. While Theorem 5.1 itself does not provide an example of such distribution, our proof is constructive and thus provide one. We outline below the main idea of this construction (the complete proof is provided in C.2).

Proof technique. The main idea of the proof consists in leveraging existing results in graph theory on the construction of large non-perfect, triangle-free graphs. Specifically, we aim to use the iterative procedure presented in Chung et al. (1993). Starting from a triangle-free graph G_0, Chung et al. (1993) provides a sequence of triangle-free graphs $\{G_t\}_{t\in\mathbb{N}}$ such that, for all $t \in \mathbb{N}$,

$$\mathrm{FP}(G_t, \boldsymbol{\omega}_t) - \mathrm{IP}(G_t, \boldsymbol{\omega}_t) \geq \frac{1}{2} - \left(\frac{2}{3}\right)^t |G_0|, \tag{15}$$

where $\boldsymbol{\omega}_t = \frac{1}{|G_t|}\mathbf{1}_{|G_t|}$ and $|G_0|$ is the size of G_0. Note that once the RHS of (15) becomes positive, the graphs G_t are guaranteed to be non-perfect. This result is arguably close to what we would like to demonstrate. Intuitively, we would like to design a sequence of distributions $\{\mu_t\}_{t\in\mathbb{N}}$ such that, for any $t \in \mathbb{N}$, the LHS of (15) can be represented as the randomization gap of μ_t. To design such a sequence, we first show in Lemma C.5 that for any graph G_t, there exists a distribution μ_t whose conflict graph $G^{\epsilon}_{s_{\mu_t}}$ is isomorphic to G_t. Second, we show that when G_t is triangle-free we can characterize the randomization gap of μ_t as

$$\mathrm{rg}(\mu_t, \epsilon) = \mathrm{FP}(G_t, \boldsymbol{\omega}_t) - \mathrm{IP}(G_t, \boldsymbol{\omega}_t). \tag{16}$$

This comes from the fact that the conflict graph $G^{\epsilon}_{s_{\mu_t}}$ is the loopless version of the conflict hypergraph $H^{\epsilon}_{s_{\mu_t}}$. Therefore, for any $\delta > 0$, one can choose a sufficiently large $t^* \in \mathbb{N}$ such that $(2/3)^{t^*} |G_0| < \delta$. Then, combining (15) and (16), the randomization gap of $\mu_{t^*} \in \mathcal{P}(\mathbb{R}^d \times [K])$ satisfies (14). Note that the dependency of K on the size of the support $n = |s_{\mu_{t^*}}|$ can be made explicit using the lemma C.5.

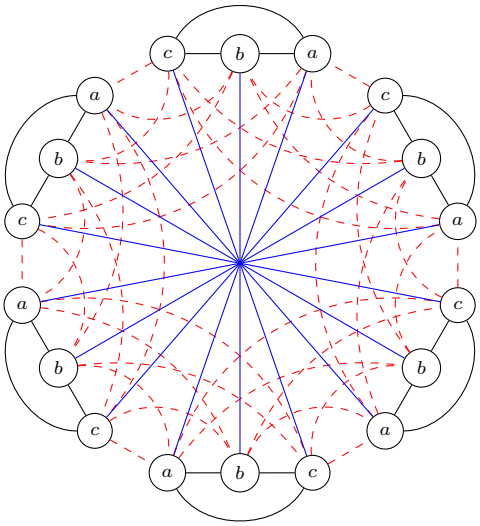

Figure 4: Example of the graph G_1 built in Chung et al. (1993) when the initial G_0 is the 3-cycle C_3. There are 6 copies of C_3 with nodes labeled a, b and c. Black edges are those within each copy, while red and blue edges are the ones added by the construction in Chung et al. (1993) between different copies of the initial graph.

Other possible constructions. To show the existence of non-trivial distributions for which the randomization gap is arbitrarily close to $1/2$, we could have used many other existing explicit constructions Erdös (1966); Graham et al. (1993); Alon (1995, 1994), or existence results Erdös (1961); Kim (1995) on non-perfect graphs. Those graphs share the property of being triangle-free. This implies they cannot contain any odd anti-hole, and thus their associated conflict hypergraph is conformal. Therefore, the only structures that induce a positive randomization gap are odd holes. We are not aware of any construction in which the graphs contain anti-holes. In remains an open question to determine whether we can design non-trivial distributions for which the randomization gap is arbitrarily close to $1/2$ using anti-hole based graph constructions.

6 DISCUSSIONS & RELATED WORK

In this paper, we present a step towards the characterization of data distributions for which randomization proves useful against adversarial examples. Recent work by Gnecco Heredia et al. (2024) demonstrated that in binary classification, the randomization gap (*i.e.*, the difference between the optimal adversarial

risk with and without randomization) is always zero for rich enough hypotheses classes. However, they left the multi-class question open. We provide the first (partial) characterization of distributions for which the randomization gap is positive.

Closely related work. Dai et al. (2024) recently reformulated the optimal adversarial risk as the value of a linear program. However, their objective was to provide lower bounds on this quantity, rather than addressing the difference between randomized and deterministic classifiers. They extend prior results from Bhagoji et al. (2019) for binary classification, where the authors had already mentioned the equivalence of their formulation to the König-Egévary theorem in the case of finite spaces. Trillos et al. (2023a) discussed in detail the example shown in Figure 3, which was also briefly mentioned in Dai et al. (2024). Our work provides further examples and presents a partial characterization of the distributions for which randomization can enhance robustness.

Open problems. While we provide necessary conditions for the randomization gap to be positive, a full characterization remains to be established. We also leave open the proof of our conjecture that any necessary and sufficient conditions cannot be checked in polynomial time, unless "P = NP". Although our study focuses on finite discrete distributions, extending the analysis to more general distributions is an interesting future direction. We hypothesize that infinite discrete distributions could be addressed by employing infinite linear programs and infinite graphs. Furthermore, extending the analysis to general Borel probability measures would likely require the application of optimal transport theory Trillos et al. (2023a).

Acknowledgments

This work was funded by the French National Research Agency (DELCO ANR-19-CE23-0016). This research was supported in part by the French National Research Agency under the France 2030 program, reference ANR-23-PEIA-0003. Rafael is partially supported by the French National Research Agency and the French Ministry of Research and Higher Education. Lucas would like to thank Denis Cornaz, Roland Grappe and Charles Nourry for fruitful discussions.

References

Alon, N. (1994). Explicit ramsey graphs and orthonormal labelings. *Electron. J. Comb.*, 1.

Alon, N. (1995). Tough ramsey graphs without short cycles. *J. Algebraic Comb.*, 4(3):189–195.

Awasthi, P., Frank, N. S., and Mohri, M. (2021). On the existence of the adversarial bayes classifier (extended version). *arXiv preprint arXiv:2112.01694*.

Berge, C. (1984). *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier.

Bhagoji, A. N., Cullina, D., and Mittal, P. (2019). Lower bounds on adversarial robustness from optimal transport. *Advances in Neural Information Processing Systems*, 32.

Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., Giacinto, G., and Roli, F. (2013). Evasion attacks against machine learning at test time. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 387–402. Springer.

Boros, E., Gurvich, V., Milanič, M., and Uno, Y. (2023). Dually conformal hypergraphs. *arXiv preprint arXiv:2309.00098*.

Carlini, N. and Wagner, D. (2017). Adversarial examples are not easily detected: Bypassing ten detection methods. In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, pages 3–14.

Chudnovsky, M., Robertson, N., Seymour, P. D., and Thomas, R. (2003). Progress on perfect graphs. *Mathematical Programming*, 97:405–422.

Chung, F. R., Cleve, R., and Dagum, P. (1993). A note on constructive lower bounds for the ramsey numbers r (3, t). *Journal of Combinatorial Theory, Series B*, 57(1):150–155.

Cohen, J., Rosenfeld, E., and Kolter, Z. (2019). Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning*, pages 1310–1320. PMLR.

Conforti, M., Cornuéjols, G., Zambelli, G., Conforti, M., Cornuéjols, G., and Zambelli, G. (2014). *Integer programming models*. Springer.

Cornuejols, G., Liu, X., and Vuskovic, K. (2003). A polynomial algorithm for recognizing perfect graphs. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, pages 20–27.

Croce, F. and Hein, M. (2020). Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *International conference on machine learning*, pages 2206–2216. PMLR.

Dai, S., Ding, W., Bhagoji, A. N., Cullina, D., Zheng, H., Zhao, B., and Mittal, P. (2024). Characterizing the optimal $0-1$ loss for multi-class classification with a test-time attacker. *Advances in Neural Information Processing Systems*, 36.

Davies, E. and Illingworth, F. (2022). The χ-ramsey problem for triangle-free graphs. *SIAM Journal on Discrete Mathematics*, 36(2):1124–1134.

Dhillon, G. S., Azizzadenesheli, K., Lipton, Z. C., Bernstein, J., Kossaifi, J., Khanna, A., and Anandkumar, A. (2018). Stochastic activation pruning for robust adversarial defense. *arXiv preprint arXiv:1803.01442*.

Erdös, P. and Hajnal, A. (1985). Chromatic number of finite and infinite graphs and hypergraphs. *Discrete Mathematics*, 53:281–285.

Erdös, P. (1961). Graph theory and probability. ii. *Canadian Journal of Mathematics*, 13:346–352.

Erdös, P. (1966). On the construction of certain graphs. *Journal of Combinatorial Theory*, 1(1):149–153.

Fishburn, P. C. (1983). On the sphericity and cubicity of graphs. *Journal of Combinatorial Theory, Series B*, 35(3):309–318.

Frank, N. S. and Niles-Weed, J. (2024). Existence and minimax theorems for adversarial surrogate risks in binary classification. *Journal of Machine Learning Research*, 25(58):1–41.

Gnecco Heredia, L., Pydi, M. S., Meunier, L., Negrevergne, B., and Chevaleyre, Y. (2024). On the role of randomization in adversarially robust classification. *Advances in Neural Information Processing Systems*, 36.

Goodfellow, I. J., Shlens, J., and Szegedy, C. (2014). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Graham, F. C., Cleve, R., and Dagum, P. (1993). A note on constructive lower bounds for the ramsey numbers r(3, t). *J. Comb. Theory B*, 57:150–155.

Gurobi Optimization, LLC (2024). Gurobi Optimizer Reference Manual.

Huang, Y., Yu, Y., Zhang, H., Ma, Y., and Yao, Y. (2022). Adversarial robustness of stabilized neural ode might be from obfuscated gradients. In *Mathematical and Scientific Machine Learning*, pages 497–515. PMLR.

Jensen, T. R. and Toft, B. (2011). *Graph coloring problems*. John Wiley & Sons.

Kim, J. H. (1995). The ramsey number r(3, t) has order of magnitude t2/log t. *Random Struct. Algorithms*, 7:173–208.

Kurakin, A., Goodfellow, I., and Bengio, S. (2016). Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., and Vladu, A. (2017). Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*.

Meunier, L., Ettedgui, R., Pinot, R., Chevaleyre, Y., and Atif, J. (2022). Towards consistency in adversarial classification. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 8538–8549. Curran Associates, Inc.

Meunier, L., Scetbon, M., Pinot, R. B., Atif, J., and Chevaleyre, Y. (2021). Mixed nash equilibria in the adversarial examples game. In *International Conference on Machine Learning*, pages 7677–7687. PMLR.

Moosavi-Dezfooli, S.-M., Fawzi, A., Uesato, J., and Frossard, P. (2019). Robustness via curvature regularization, and vice versa. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9078–9086.

Panousis, K., Chatzis, S., Alexos, A., and Theodoridis, S. (2021). Local competition and stochasticity for adversarial robustness in deep learning. In *International Conference on Artificial Intelligence and Statistics*, pages 3862–3870. PMLR.

Pinot, R., Ettedgui, R., Rizk, G., Chevaleyre, Y., and Atif, J. (2020). Randomization matters how to defend against strong adversarial attacks. In *International Conference on Machine Learning*, pages 7717–7727. PMLR.

Pinot, R., Meunier, L., Araujo, A., Kashima, H., Yger, F., Gouy-Pailler, C., and Atif, J. (2019). Theoretical evidence for adversarial robustness through randomization. *Advances in Neural Information Processing Systems*, 32.

Pinot, R., Meunier, L., Yger, F., Gouy-Pailler, C., Chevaleyre, Y., and Atif, J. (2022). On the robustness of randomized classifiers to adversarial examples. *Machine Learning*, 111(9):3425–3457.

Pydi, M. S. and Jog, V. (2023). The many faces of adversarial risk: An expanded study. *IEEE Transactions on Information Theory*.

Roberts, F. S. (1969). On the boxicity and cubicity of a graph. *Recent progress in combinatorics*, 1(1):301–310.

Salman, H., Li, J., Razenshteyn, I., Zhang, P., Zhang, H., Bubeck, S., and Yang, G. (2019). Provably robust deep learning via adversarially trained smoothed classifiers. *Advances in Neural Information Processing Systems*, 32.

Schrijver, A. (1979). Fractional packing and covering. In *Packing and covering in combinatorics*, volume 106, pages 201–274. Mathematisch Centrum Amsterdam.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2014). Intriguing properties of neural networks. iclr. 2014. *arXiv preprint arXiv:1312.6199*.

Trillos, N. G., Jacobs, M., and Kim, J. (2023a). The multimarginal optimal transport formulation of adversarial multiclass classification. *Journal of Machine Learning Research*, 24(45):1–56.

Trillos, N. G., Jacobs, M., and Kim, J. (2023b). On the existence of solutions to adversarial training in multiclass classification. *arXiv preprint arXiv:2305.00075*.

Xie, C., Wang, J., Zhang, Z., Ren, Z., and Yuille, A. (2017). Mitigating adversarial effects through randomization. *arXiv preprint arXiv:1711.01991*.

Yang, H., Wang, M., Yu, Z., and Zhou, Y. (2022). Rethinking feature uncertainty in stochastic neural networks for adversarial robustness. *arXiv preprint arXiv:2201.00148*.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [**Yes**] We do not include algorithms. Our mathematical framework is clearly explained in the Preliminaries section, and all our Theorems and Lemmas also state all the assumptions.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [**Yes**] Even if we do not have algorithms, we discuss the complexity of our problem and discuss why it is hard to find sufficient conditions in Section 4.
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [**Not Applicable**] We do not have experiments nor source code.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [**Yes**]
   (b) Complete proofs of all theoretical results. [**Yes**] All of them are in the Appendix to respect the space constraints. Most results used from previous work are restated, and always cited.
   (c) Clear explanations of any assumptions. [**Yes**] We do our best to explain both intuitively and rigorously all the assumptions and implications of all our results.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [**Not Applicable**]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [**Not Applicable**]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [**Not Applicable**]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [**Not Applicable**]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [**Not Applicable**]
   (b) The license information of the assets, if applicable. [**Not Applicable**]
   (c) New assets either in the supplemental material or as a URL, if applicable. [**Not Applicable**]
   (d) Information about consent from data providers/curators. [**Not Applicable**]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [**Not Applicable**]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [**Not Applicable**]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [**Not Applicable**]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [**Not Applicable**]

A SUPPLEMENTARY MATERIAL FOR SECTION 3: COMPUTING THE RANDOMIZATION GAP USING GRAPH THEORY

We adapt and simplify the proof presented in Dai et al. (2024) for both theorems. We prove Theorem 3.2, after which Theorem 3.1 will follow. We first introduce some definitions and establish two technical lemmas. Let us consider an ℓ_p norm with $p \in (1, \infty]$, $\epsilon > 0$, and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The adversarial risk of a classifier $f \in \mathcal{F}_{\text{rand}}$, denoted $\mathcal{R}(\mu, \epsilon, f)$, is defined as

$$\mathcal{R}(\mu, \epsilon, f) \coloneqq \mathbb{E}_{(x,y)\sim\mu}\left[\sup_{x' \in B_p(x,\epsilon)} \mathbb{E}_{z\sim f(x')}\left[\mathbb{1}\{z \neq y\}\right]\right].$$

The *adversarial accuracy* of a classifier $f \in \mathcal{F}_{\text{rand}}$, denoted $\mathcal{A}(\mu, \epsilon, f)$, is defined as

$$\mathcal{A}(\mu, \epsilon, f) \coloneqq \mathbb{E}_{(x,y)\sim\mu}\left[\inf_{x' \in B_p(x,\epsilon)} \mathbb{E}_{z\sim f(x')}\left[\mathbb{1}\{z = y\}\right]\right] = 1 - \mathcal{R}(\mu, \epsilon, f).$$

and we denote $\mathcal{A}^*_{\mathcal{F}}(\mu, \epsilon)$ the optimal adversarial accuracy over a family of classifier $\mathcal{F}$. Note that as deterministic classifiers can be represented as randomized ones that only output Dirac measures, $\mathcal{A}(\mu, \epsilon, f)$ and $\mathcal{R}(\mu, \epsilon, f)$ are also well-defined for $f \in \mathcal{F}_{\text{det}}$. We thus have that $\mathcal{A}^*_{\mathcal{F}}(\mu, \epsilon) = 1 - \mathcal{R}^*_{\mathcal{F}}(\mu, \epsilon)$ for both deterministic and randomized classifiers. Note that for any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with support $s_\mu = \{(x_i, y_i)\}_{i\in[n]}$ and probability vector $\boldsymbol{\omega}_\mu \in \Delta^n$, we can further rewrite the accuracy as follows:

$$\mathcal{A}(\mu, \epsilon, f) = \sum_{i\in[n]} \boldsymbol{\omega}_\mu^{(i)} \cdot \inf_{x' \in B_p(x_i,\epsilon)} f(x')^{(y_i)}. \tag{17}$$

Reader's note: In what follows, we make several simplifications to enhance readability. Let K be the number of classes, *i.e.* $\mathcal{Y} = [K]$. We use the natural identification between $\mathcal{P}(\mathcal{Y})$ and Δ^K, in which any vector $u \in \Delta^K$ represents a probability distribution over the K classes. Therefore, given a randomized classifier $f : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and a point $x \in \mathcal{X}$, we abuse the notation and call $f(x)$ the probability vector associated with the point x. On the other hand, and recalling that deterministic classifiers can be represented as randomized classifiers that only output Dirac measures over one class, we will often prove results for deterministic classifiers using this characterization.

A.1 Proofs of Theorem 3.1 and Theorem 3.2

We now link the adversarial classification problem with randomized classifiers with the fractional set packing problem. This is done via explicit constructions first introduced in Dai et al. (2024), which are presented in the proof of Lemma A.1. This lemma is the main tool used to prove both Theorem 3.1 and 3.2.

Lemma A.1 (Extended from Dai et al. (2024))**.** *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with support $s_\mu = \{(x_i, y_i)\}_{i\in[n]}$ and probability vector $\boldsymbol{\omega}_\mu \in \Delta^n$. The following statements hold true:*

a) For any fractional packing $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$, there exists a classifier $f_q \in \mathcal{F}_{\text{rand}}$ such that $\mathcal{A}(\mu, \epsilon, f_q) \geq \boldsymbol{\omega}_\mu^T q$.

b) For any classifier $f \in \mathcal{F}_{\text{rand}}$, there exists a fractional packing $q_f \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$ such that $\mathcal{A}(\mu, \epsilon, f) = \boldsymbol{\omega}_\mu^T q_f$.

Proof. Throughout the proof, we denote $\mathcal{E}$ the set of hyperedges of the conflict hypergraph $H^\epsilon_{s_\mu}$.

Proof of a). Let $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$ be a fractional packing for (8) over $(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu)$. The first part of the proof will be to define an auxiliary function g_q that will allow us, on a second stage, to define a classifier f_q with the properties that we seek.

Definition of the auxiliary function. Given $(x, y) \in \mathcal{X} \times \mathcal{Y}$, define $\mathcal{I}(x, y) = \{i \in [n] \mid x \in B_p(x_i, \epsilon) \text{ and } y_i = y\}$ the set of indices of neighbors of x on the set s_μ that are of class y. We denote $\mathcal{Y}_x = \{y \in \mathcal{Y} \mid \mathcal{I}(x, y) \neq \emptyset\}$ the set

of classes for which x has a neighbor in s_μ that belongs to that class. Denote $g_q : \mathcal{X} \to \mathbb{R}_+^K$ the function defined as follows:

$$\forall x \in \mathcal{X}, \quad g_q(x)^{(y)} = \begin{cases} \max_{i \in \mathcal{I}(x,y)} q^{(i)}, & \text{if } \mathcal{I}(x,y) \neq \emptyset. \\ 0, & \text{otherwise.} \end{cases}$$

Given that q is a fractional packing of $H^\epsilon_{s_\mu}$, we have that for any hyperedge $e \in \mathcal{E}$,

$$\sum_{i \in e} q^{(i)} \leq 1. \tag{18}$$

Recall that for every $i \in [n]$, $\{i\} \in \mathcal{E}$, which by (18) implies that

$$\forall i \in [n], \quad q^{(i)} \leq 1.$$

Thus, $0 \leq g_q(x)^{(y)} \leq 1$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Let us now consider a given $x \in \mathcal{X}$. If x is such that $\mathcal{Y}_x \neq \emptyset$, we can define for every $y \in \mathcal{Y}_x$ the index i_y as

$$i_y = \arg \max_{j \in \mathcal{I}(x,y)} q^{(j)}.$$

Then, for any x such that $\mathcal{Y}_x \neq \emptyset$, we can write $g_q(x)$ as

$$g_q(x)^{(y)} = \begin{cases} q^{(i_y)}, & \text{if } \mathcal{I}(x,y) \neq \emptyset. \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, let $e_x = \{i_y \in [n] \mid y \in \mathcal{Y}_x\}$. Note that $e_x \in \mathcal{E}$, because $x \in \bigcap_{j \in e} B_p(x_j, \epsilon)$. Given that $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$, the following inequality holds

$$\sum_{y \in \mathcal{Y}} g_q(x)^{(y)} = \sum_{y \in \mathcal{Y}_x} g_q(x)^{(y)} \quad = \sum_{i \in e_x} q^{(i)} \leq 1.$$

Definition of the classifier. We now define a classifier f_q as follows:

$$\forall x \in \mathcal{X}, \quad f_q(x)^{(y)} = \begin{cases} 1 - \sum_{k>1} g_q(x)^{(k)} & \text{if } y = 1, \\ g_q(x)^{(y)}, & \text{if } y > 1. \end{cases}$$

By construction, $\sum_{y \in \mathcal{Y}} f_q(x)^{(y)} = 1$, and also $f_q(x)^{(y)} \geq g_q(x)^{(y)}$ for any $(x, y) \in \mathcal{X} \times \mathcal{Y}$ with $\mathcal{Y}_x \neq \emptyset$[5]. The adversarial accuracy of f_q satisfies the following inequality:

$$\begin{aligned} \mathcal{A}(\mu, \epsilon, f_q) &= \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot \inf_{x' \in B_p(x_i, \epsilon)} f_q(x')^{(y_i)} \\ &\geq \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot \inf_{x' \in B_p(x_i, \epsilon)} \max_{j \in \mathcal{I}(x', y_i)} q^{(j)} \\ &\geq \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot \inf_{x' \in B_p(x_i, \epsilon)} q^{(i)} \qquad \text{(as } i \in \mathcal{I}(x', y_i)\text{)} \\ &\geq \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot q^{(i)} \end{aligned}$$

This completes the first part of the proof.

[5]Note that the isolated points, *i.e.* $\mathcal{Y}_x = \emptyset$, are not involved in the computation of the accuracy. Therefore, the classifier f_q can be defined in any way on those points. For this construction in particular, f_q will always predict class 1 on points x such that $\mathcal{Y}_x = \emptyset$.

Proof of b). Let us consider an arbitrary classifier $f \in \mathcal{F}_{\text{rand}}$, and denote q_f the vector defined as

$$\forall i \in [n], \quad q_f^{(i)} = \inf_{x' \in B_p(x_i, \epsilon)} f(x')^{(y_i)}.$$

Clearly, $q_f^{(i)} \geq 0$ for all $i \in [n]$. To see that q_f is feasible for (8) over $(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu)$, consider any hyperedge $e \in \mathcal{E}$ and take an arbitrary $x_e \in \bigcap_{i \in e} B_p(x_i, \epsilon)$. We have that

$$\begin{aligned}
\sum_{i \in e} q_f^{(i)} &= \sum_{i \in e} \inf_{x' \in B_p(x_i, \epsilon)} f(x')^{(y_i)} \\
&\leq \sum_{i \in e} f(x_e)^{(y_i)} && (x_e \in B_p(x_i, \epsilon) \text{ for all } i \in e) \\
&\leq \sum_{y \in \mathcal{Y}} f(x_e)^{(y)} && (y_i \neq y_j \text{ for any } i, j \in e) \\
&= 1 && (f \text{ is a valid classifier})
\end{aligned}$$

This holds for any hyperedge in $\mathcal{E}$, hence $q_f \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$. Finally, we can compute the accuracy of f:

$$\mathcal{A}(\mu, \epsilon, f) = \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot \inf_{x' \in B_p(x_i, \epsilon)} f(x')^{(y_i)} = \sum_{i \in [n]} \boldsymbol{\omega}_\mu^{(i)} \cdot q_f^{(i)} = \boldsymbol{\omega}_\mu^T q_f,$$

which concludes the proof. □

Theorem 3.2. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:*

$$1 - \mathcal{R}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) = \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu).$$

Proof. By the first item of Lemma A.1, we have that for any vector $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$, there is a classifier $f_q \in \mathcal{F}_{\text{rand}}$ for which

$$\mathcal{A}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) \geq \mathcal{A}(\mu, \epsilon, f_q) \geq \boldsymbol{\omega}_\mu^T q.$$

As this is true for any $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$, we have that

$$\mathcal{A}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) \geq \max_{q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})} \boldsymbol{\omega}_\mu^T q = \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{19}$$

By the second item of Lemma A.1, we have that for any classifier $f \in \mathcal{F}_{\text{rand}}$, there is a vector $q_f \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$ for which

$$\mathcal{A}(\mu, \epsilon, f) = \boldsymbol{\omega}_\mu^T q_f \leq \max_{q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})} \boldsymbol{\omega}_\mu^T q = \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu).$$

As this is true for any classifier $f \in \mathcal{F}_{\text{rand}}$, then by taking max over all classifiers we have that

$$\mathcal{A}^*_{\mathcal{F}_{\text{rand}}}(\mu, \epsilon) \leq \text{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{20}$$

Combining (19) and (20) concludes the proof.

□

Theorem 3.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. For any $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we have:*

$$1 - \mathcal{R}^*_{\mathcal{F}_{\text{det}}}(\mu, \epsilon) = \text{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu).$$

Proof. The reasoning is analogous to the proof of Theorem 3.2 but restricting to deterministic classifiers $\mathcal{F}_{\text{det}}$. Let $q \in \mathcal{Q}(H^\epsilon_{s_\mu})$ be an arbitrary packing of $H^\epsilon_{s_\mu}$. Note that $q \in \mathcal{Q}^{\text{frac}}(H^\epsilon_{s_\mu})$, so by Lemma A.1, there exists a classifier $f_q \in \mathcal{F}_{\text{rand}}$ such that $\mathcal{A}(\mu, \epsilon, f_q) \geq \boldsymbol{\omega}_\mu^T q$. By the construction of f_q in the proof of Lemma A.1, we know that f_q is deterministic, *i.e.* $f_q \in \mathcal{F}_{\text{det}}$. As this is true for any $q \in \mathcal{Q}(H^\epsilon_{s_\mu})$, we have that

$$\mathcal{A}^*_{\mathcal{F}_{\text{det}}}(\mu, \epsilon) \geq \max_{q \in \mathcal{Q}(H^\epsilon_{s_\mu})} \boldsymbol{\omega}_\mu^T q = \text{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{21}$$

On the other hand, for any deterministic classifier $f \in \mathcal{F}_{\text{det}}$, f can be represented as a randomized classifier, thus by Lemma A.1, there exists a vector $q_f \in \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$ for which $\mathcal{A}(\mu, \epsilon, f) = \boldsymbol{\omega}_\mu^T q_f$. Moreover, given the construction of q_f in the proof of Lemma A.1, we know that $q_f \in \{0,1\}^n$, which means that q_f is a packing, *i.e.*, $q_f \in \mathcal{Q}(H^{\epsilon}_{s_\mu})$, and therefore $\mathcal{A}(\mu, \epsilon, f) \leq \text{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$. As this is true for any classifier $f \in \mathcal{F}_{\text{det}}$, we have that

$$\mathcal{A}^*_{\mathcal{F}_{\text{det}}}(\mu, \epsilon) \leq \text{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{22}$$

Combining (21) and (22) concludes the proof.

□

A.2 Solving Examples of Fractional Set Packing Problems

Example presented in Figure 1. The fractional set packing problem for this case can be written as the following linear program

$$\begin{aligned} \max_{q \in [0,1]^5} \quad & \tfrac{1}{5}\mathbf{1}_5^T q \\ \text{s.t.} \quad & Bq \leq \mathbf{1}_{10} \end{aligned}$$

where $B \in \{0,1\}^{10\times 5}$ is the edge-incidence matrix of $H^{\epsilon}_{s_\mu}$. Note that as the largest hyperedges of $H^{\epsilon}_{s_\mu}$ have size 2, every row in B has at most two components equal to 1. This implies that the fractional packing with characteristic vector $q = \frac{1}{2}\mathbf{1}_5$ is feasible, *i.e.*, $q \in \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$, and it has a cumulative weight of $\frac{1}{2}$. Therefore, $\text{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$ is lower bounded by $\frac{1}{2}$. The dual problem of $\text{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$ is the following linear program:

$$\begin{aligned} \min_{z \in [0,1]^{10}} \quad & \mathbf{1}_{10}^T z \\ \text{s.t.} \quad & B^T z \geq \tfrac{1}{5}\mathbf{1}_5 \end{aligned} \tag{23}$$

The vector z can be interpreted as assigning weights to the hyperedges of $H^{\epsilon}_{s_\mu}$, while the vector $B^T z$ represents the total cumulative weight assigned to the vertices of $H^{\epsilon}_{s_\mu}$. Consequently, each constraint in (23) can be understood as a requirement that each vertex be covered by hyperedges with a cumulative weight of at least $\frac{1}{5}$.

Consider the vector z^*, where the components corresponding to the hyperedges associated with singletons are set to 0, and those corresponding to hyperedges of size 2 are set to $\frac{1}{10}$. As a result, z^* consists of five components equal to 0 and five components equal to $\frac{1}{10}$.

To verify that z^* is feasible, observe that each vertex is contained in exactly two hyperedges of size 2. Since all these hyperedges have a weight of $\frac{1}{10}$ in z^*, we can confirm that every vertex has a total cumulative weight of exactly $\frac{1}{5}$. The associated value of z^* in (23) is $\frac{1}{2}$. Therefore, the dual problem is upper bounded by $\frac{1}{2}$, and by duality, we conclude that $\text{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \frac{1}{2}$.

Example presented in Figure 2. Similarly, the fractional set packing problem for this case can be written as a linear program as follows

$$\begin{aligned} \max_{q \in [0,1]^4} \quad & \tfrac{1}{4}\mathbf{1}_4^T q \\ \text{s.t.} \quad & Bq \leq \mathbf{1}_9 \end{aligned}$$

where $B \in \{0,1\}^{9\times 4}$ is the edge-incidence matrix of $H^{\epsilon}_{s_\mu}$. In this case, the fractional packing with characteristic vector $q = (1/2, 1/2, 0, 1)$ is feasible, and it has a cumulative weight of $\frac{1}{2}$. Therefore, $\text{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$ is lower bounded by $\frac{1}{2}$.

The dual problem of $\text{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu)$ is the following linear program:

$$\begin{aligned} \min_{z \in [0,1]^{9}} \quad & \mathbf{1}_{9}^T z \\ \text{s.t.} \quad & B^T z \geq \tfrac{1}{4}\mathbf{1}_4 \end{aligned} \tag{24}$$

Consider the vector z^* that is all zeros, except for the component corresponding to the 3-hyperedge $\{1,2,3\}$ and the hyperedge $\{4\}$, both of which have value $\frac{1}{4}$. To verify that z^* is feasible, observe that the 3-hyperedge $\{1,2,3\}$

covers these three vertices with a weight of $\frac{1}{4}$. Additionally, the vertex $\{4\}$ is covered by its own hyperedge, so all vertices are properly covered. The associated value of z^* in (24) is $\frac{1}{4} \cdot 2 = \frac{1}{2}$. Therefore, the dual problem is upper bounded by $\frac{1}{2}$, and by duality, we conclude that $\mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \frac{1}{2}$.

B SUPPLEMENTARY MATERIAL FOR SECTION 4: LINK BETWEEN STRUCTURAL PROPERTIES OF THE CONFLICT HYPERGRAPH AND THE RANDOMIZATION GAP

The proof of Theorem 4.1 relies on decomposing the randomization gap into the sum of two non-negative terms. In Section B.1, we prove that this decomposition holds true thanks to Lemma 4.1. In Section B.2, we prove two lemmas adapted from (Chudnovsky et al., 2003, Theorem 4.1) that establish equivalences between these terms being zero and the presence of certain structures in the conflict hypergraph. Following this, we present the full proof of Theorem 4.1. In Section B.3, we prove Corollary 4.2, which states that one of the identified structures cannot appear when considering the ℓ_∞ norm. Finally, in Section B.4, we provide several examples of the existence of specific structures for different ℓ_p norms.

Reader's note: In this section, we will often use the notation $\mathcal{E}(H)$ to refer to the set of hyperedges of hypergraph H. For readability, we sometimes use the notation $E(G)$ to refer to the edge set of a graph G, particularly when handling multiple graphs and hypergraphs simultaneously. Moreover, given a graph $G = (V, E)$, we often say that $V' \subset V$ is a clique in G, meaning that the induced subgraph $G' = (V', E')$, where $E' = \{\{i, j\} \in E \mid i, j \in V'\}$, is a clique of G. Finally, given a graph $G = (V, E)$, we define the clique hypergraph of G as $H = (V, \mathcal{E})$, where $\mathcal{E}$ is the set of all maximal cliques of G.

B.1 Proof of Lemma 4.1

Lemma B.1 is a technical result used in the proof of Lemma 4.1. It can be interpreted as establishing a hierarchy between the conflict graph, the conflict hypergraph and the clique hypergraph.

Lemma B.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also $S = \{(x_i, y_i)\}_{i \in [n]}$ be an arbitrary set of points from $\mathcal{X} \times \mathcal{Y}$. The following assertion hold true*

a) For any edge $e \in E(G^{\epsilon}_S)$, there exists some hyperedge $e' \in \mathcal{E}(H^{\epsilon}_S)$ such that $e \subseteq e'$.

b) For any hyperedge $e \in \mathcal{E}(H^{\epsilon}_S)$, there exists some hyperedge $e' \in \mathcal{E}(C^{\epsilon}_S)$ such that $e \subseteq e'$.

Proof. **Proof of a).** This is true by definition of conflict hypergraph and conflict graph. In particular, by the fact that G^{ϵ}_S is the 2-section of H^{ϵ}_S.

Proof of b). Now consider an arbitrary hyperedge $e \in \mathcal{E}(H^{\epsilon}_S)$. By definition of conflict graph, this implies that

$$\forall i, j \in e \text{ with } i \neq j, \quad \{i, j\} \in E(G^{\epsilon}_S).$$

The fact that all the pairs $i, j \in e$ are edges of G^{ϵ}_S means that e constitutes a clique of G^{ϵ}_S. Then, by definition of the clique hypergraph as the one whose hyperedges are the maximal cliques of G^{ϵ}_S, we have that

$$\exists e' \in \mathcal{E}(C^{\epsilon}_S) \text{ such that } e \subseteq e'.$$

□

The hierarchy of the three hypergraphs, as established in Lemma B.1, corresponds to an ordering of the values in their respective fractional set packing problems. According to Lemma 4.1, the three hypergraphs represent different formulations of the same set packing problem, with their fractional versions providing upper bounds of varying tightness.

Lemma 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, $\epsilon > 0$, and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. Then we have*

$$\mathrm{IP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{IP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{IP}(G^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{10}$$

$$\mathrm{FP}(C^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \leq \mathrm{FP}(H^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu) \leq \mathrm{FP}(G^{\epsilon}_{s_\mu}, \boldsymbol{\omega}_\mu). \tag{11}$$

Proof. **Proof of** (11). We will begin by using Lemma B.1 to show that the set of fractional packings of the hypergraphs $C^{\epsilon}_{s_\mu}$, $H^{\epsilon}_{s_\mu}$ and $G^{\epsilon}_{s_\mu}$ satisfy the following relation:

$$\mathcal{Q}^{\text{frac}}(C^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(G^{\epsilon}_{s_\mu}).$$

This will prove that (11) holds, as the fractional packing problem is a maximization problem.

($\mathcal{Q}^{\text{frac}}(C^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$). Take an arbitrary $q \in \mathcal{Q}^{\text{frac}}(C^{\epsilon}_{s_\mu})$ and an arbitrary hyperedge $e \in \mathcal{E}(H^{\epsilon}_{s_\mu})$. By Lemma B.1, there exists a hyperedge $e' \in \mathcal{E}(C^{\epsilon}_{s_\mu})$ such that $e \subseteq e'$. We then have that

$$\sum_{i \in e} q^{(i)} \leq \sum_{i \in e'} q^{(i)} \leq 1,$$

where the last inequality holds because $q \in \mathcal{Q}^{\text{frac}}(C^{\epsilon}_{s_\mu})$ and $e' \in \mathcal{E}(C^{\epsilon}_{s_\mu})$. As this holds for every hyperedge $e \in \mathcal{E}(H)$, we conclude that q is a fractional packing of $H^{\epsilon}_{s_\mu}$, *i.e.* $q \in \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$.

($\mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(G^{\epsilon}_{s_\mu})$). Similarly, take an arbitrary $q \in \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$ and an arbitrary edge $e \in \mathcal{E}(G^{\epsilon}_{s_\mu})$. By Lemma B.1, there exists a hyperedge $e' \in \mathcal{E}(H^{\epsilon}_{s_\mu})$ such that $e \subseteq e'$. We then have that

$$\sum_{i \in e} q^{(i)} \leq \sum_{i \in e'} q^{(i)} \leq 1,$$

where the last inequality holds because $q \in \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu})$ and $e' \in \mathcal{E}(H^{\epsilon}_{s_\mu})$. As this holds for every hyperedge $e \in \mathcal{E}(G^{\epsilon}_{s_\mu})$, we conclude that q is a fractional packing of $G^{\epsilon}_{s_\mu}$, *i.e.* $q \in \mathcal{Q}^{\text{frac}}(G^{\epsilon}_{s_\mu})$. We thus have that $\mathcal{Q}^{\text{frac}}(C^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(H^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}^{\text{frac}}(G^{\epsilon}_{s_\mu})$, which proves that (11) holds.

Proof of (10)**.** We will prove that

$$\mathcal{Q}(C^{\epsilon}_{s_\mu}) = \mathcal{Q}(H^{\epsilon}_{s_\mu}) = \mathcal{Q}(G^{\epsilon}_{s_\mu}).$$

Note that the reasoning used to deal with (11) can be directly applied to prove that

$$\mathcal{Q}(C^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}(H^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}(G^{\epsilon}_{s_\mu}).$$

Therefore, it suffices to show that $\mathcal{Q}(G^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}(C^{\epsilon}_{s_\mu})$. Suppose by contradiction that this is not the case. Then there must exists some binary vector $q \in \{0,1\}^n$ such that $q \in \mathcal{Q}(G^{\epsilon}_{s_\mu})$ but $q \notin \mathcal{Q}(C^{\epsilon}_{s_\mu})$. In particular, the condition $q \notin \mathcal{Q}(C^{\epsilon}_{s_\mu})$ implies that

$$\exists e \in \mathcal{E}(C^{\epsilon}_{s_\mu}) \text{ such that } \sum_{i \in e} q^{(i)} > 1.$$

Given that q is a binary vector, this implies that

$$\exists i, j \in e,\ i \neq j \text{ such that } q^{(i)} = q^{(j)} = 1.$$

These two indices i, j form an edge in $G^{\epsilon}_{s_\mu}$, which in turn implies that q cannot be a packing of $G^{\epsilon}_{s_\mu}$. This contradiction allows us to conclude that $\mathcal{Q}(G^{\epsilon}_{s_\mu}) \subseteq \mathcal{Q}(C^{\epsilon}_{s_\mu})$, and therefore

$$\mathcal{Q}(C^{\epsilon}_{s_\mu}) = \mathcal{Q}(H^{\epsilon}_{s_\mu}) = \mathcal{Q}(G^{\epsilon}_{s_\mu}).$$

This concludes the proof. □

B.2 Proof of Theorem 4.1.

Lemmas B.2 and B.3 are adapted from (Chudnovsky et al., 2003, Theorem 4.1) and are used to prove Theorem 4.1. Each lemma presents an equivalence between the positivity of (12) or (13), and the existence of a particular structure in the conflict hypergraph.

Lemma B.2 (Conformal hypergraphs)**.** *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also $S = \{(x_i, y_i)\}_{i \in [n]}$ be an arbitrary set of points from $\mathcal{X} \times \mathcal{Y}$. The assertions a) and b) below are equivalent.*

a) $\forall \boldsymbol{\omega} \in \mathbb{R}_+^n, \quad \mathrm{FP}(H_S^\epsilon, \boldsymbol{\omega}) = \mathrm{FP}(C_S^\epsilon, \boldsymbol{\omega})$

b) Every clique in G_S^ϵ *is a hyperedge in* H_S^ϵ.

Proof. **b)** $\implies$ **a)**. Let $\boldsymbol{\omega} \in \mathbb{R}_+^n$. From Lemma 4.1, we have that $\mathrm{FP}(H_S^\epsilon, \boldsymbol{\omega}) \geq \mathrm{FP}(C_S^\epsilon, \boldsymbol{\omega})$. Furthermore, for any fractional packing of H_S^ϵ with characteristic vector $q \in \mathcal{Q}^{\mathrm{frac}}(H_S^\epsilon)$, assertion **b)** implies that $q \in \mathcal{Q}^{\mathrm{frac}}(C_S^\epsilon)$. Thus, $\mathcal{Q}^{\mathrm{frac}}(H_S^\epsilon) \subseteq \mathcal{Q}^{\mathrm{frac}}(C_S^\epsilon)$ and then $\mathrm{FP}(H_S^\epsilon, \boldsymbol{\omega}) \leq \mathrm{FP}(C_S^\epsilon, \boldsymbol{\omega})$ which conclude this first implication.

a) $\implies$ **b)**. By contradiction, suppose there exists a clique c' in G_S^ϵ such that c' is not a hyperedge in H_S^ϵ (*i.e.* $c' \notin \mathcal{E}(H_S^\epsilon)$). Let c be the maximal clique in G_S^ϵ containing c' (*i.e.* $c \in \mathcal{E}(C_S^\epsilon)$). Note that, as H_S^ϵ is downward closed, we have that $c \notin \mathcal{E}(H_S^\epsilon)$. Consider the probability vector $\boldsymbol{\omega} \in \Delta^n$ where $\boldsymbol{\omega}^{(i)} = \frac{1}{|c|}\mathbb{1}\{i \in c\}$ for all $i \in [n]$. Note that $\boldsymbol{\omega}$ is an optimal solution for the fractional set packing problem over $(C_S^\epsilon, \boldsymbol{\omega})$ as $\boldsymbol{\omega}^T\boldsymbol{\omega} = \frac{1}{|c|}$ and for any $q \in \mathcal{Q}^{\mathrm{frac}}(C_S^\epsilon)$ we have

$$\boldsymbol{\omega}^T q = \frac{1}{|c|}\sum_{i \in c} q^{(i)} \leq \frac{1}{|c|} \quad (\text{as } c \in \mathcal{E}(C_S^\epsilon)).$$

We now construct a feasible solution q for the fractional set packing problem over $(H_S^\epsilon, \boldsymbol{\omega})$ such that $\boldsymbol{\omega}^T q > \boldsymbol{\omega}^T\boldsymbol{\omega}$, which would imply that $\mathrm{FP}(H_S^\epsilon, \boldsymbol{\omega}) \geq \boldsymbol{\omega}^T q > \boldsymbol{\omega}^T\boldsymbol{\omega} = \mathrm{FP}(C_S^\epsilon, \boldsymbol{\omega})$, hence contradicting **a)**. Let $\mathcal{E}_c := \{e \in \mathcal{E}(H_S^\epsilon) : e \subseteq c\}$ and consider $e^* \in \mathrm{argmax}\{|e| : e \in \mathcal{E}_c\}$. Note that $e^* \subset c$ as $c \notin \mathcal{E}_c$, which implies

$$\sum_{i \in e^*} \boldsymbol{\omega}^{(i)} < \sum_{i \in c} \boldsymbol{\omega}^{(i)} = 1. \tag{25}$$

Let $i^* \in e^*$ and define $q \in [0,1]^n$ such that $q^{(i)} = \boldsymbol{\omega}^{(i)}$ for $i \neq i^*$ and $q_{i^*} = 1 - \sum_{i \in e^* \setminus \{i^*\}} \boldsymbol{\omega}^{(i)}$. Notice that

$$\begin{aligned}
\boldsymbol{\omega}^T q = \frac{1}{|c|}\sum_{i \in c} q^{(i)} &= \frac{1}{|c|}\left(\sum_{i \in c \setminus e^*} q^{(i)} + \sum_{i \in e^*} q^{(i)}\right) \\
&= \frac{1}{|c|}\left(\sum_{i \in c \setminus e^*} \boldsymbol{\omega}^{(i)} + 1\right) \\
&> \frac{1}{|c|}\left(\sum_{i \in c \setminus e^*} \boldsymbol{\omega}^{(i)} + \sum_{i \in e^*} \boldsymbol{\omega}^{(i)}\right) \quad \text{(by (25))} \\
&= \boldsymbol{\omega}^T\boldsymbol{\omega}.
\end{aligned}$$

In order to conclude the proof, we need to check that q is feasible, *i.e.* $q \in \mathcal{Q}^{\mathrm{frac}}(H_S^\epsilon)$. Let us consider an arbitrary $e \in \mathcal{E}(H_S^\epsilon)$. Note that the following is true given the definition of q:

$$\sum_{i \in e} q^{(i)} = \sum_{i \in e \cap c} q^{(i)}.$$

Now let us consider two cases, depending on whether i^* belongs to e or not. If $i^* \in e \cap c$, then

$$\begin{aligned}
\sum_{i \in e \cap c} q^{(i)} &= 1 - \sum_{i \in e^* \setminus \{i^*\}} q^{(i)} + \sum_{\substack{i \in e \cap c \\ i \neq i^*}} q^{(i)} \\
&= 1 - \frac{|e^*| - 1}{|c|} + \frac{|e \cap c| - 1}{|c|} \\
&= 1 - \frac{|e^*| - |e \cap c|}{|c|} \\
&\leq 1. \qquad \text{(By definition of } e^*\text{)}
\end{aligned}$$

Otherwise, if $i^* \notin e \cap c$, then

$$\sum_{i \in e \cap c} q^{(i)} = \frac{|e \cap c|}{|c|} \leq 1.$$

This concludes the proof. □

Lemma B.3 (Perfect graphs). *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also $S = \{(x_i, y_i)\}_{i \in [n]}$ be an arbitrary set of points from $\mathcal{X} \times \mathcal{Y}$. The assertions a) and b) below are equivalent.*

a) $\forall\ \boldsymbol{\omega} \in \mathbb{R}^n_+, \quad \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) = \mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega})$.

b) G^ϵ_S *is perfect.*

Proof. Let A be the $m \times n$ edge-incidence matrix of the hypergraph C^ϵ_S. Recall the linear program formulation of the fractional set packing problem

$$\begin{aligned} \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) = \quad & \max_{q \in [0,1]^n} \quad \boldsymbol{\omega}^T q \\ & \text{s.t.} \qquad Aq \leq \mathbf{1}_m \end{aligned} \tag{26}$$

By (Chudnovsky et al., 2003, Theorem 4.1), the linear program in (26) has an integral optimum solution *for every* objective function $\boldsymbol{\omega} \in \mathbb{R}^n_+$ if and only if the matrix A is the edge-incidence matrix of the clique hypergraph of a perfect graph.

As we already assumed that A is the incidence matrix of C^ϵ_S, which is the clique hypergraph of G^ϵ_S, then the only condition to ensure the existence of an integral optimum solution *for every* objective function $\boldsymbol{\omega} \in \mathbb{R}^n_+$ for $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega})$ is that G^ϵ_S is perfect.

□

Theorem 4.1. *Let us consider an ℓ_p norm with $p \in (1, \infty]$, and $\epsilon > 0$. Let also $S = \{(x_i, y_i)\}_{i \in [n]}$ be an arbitrary set of points from $\mathcal{X} \times \mathcal{Y}$. There exists a distribution $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with support $s_\mu = S$ such that $\mathrm{rg}(\mu, \epsilon) > 0$ if and only if at least one of the following assertions holds true:*

a) H^ϵ_S *is not conformal (i.e., there exists a clique in G^ϵ_S that is not a hyperedge in H^ϵ_S).*

b) G^ϵ_S *is not perfect (i.e., it contains at least one odd hole or one odd anti-hole).*

Proof. **First implication.** Let $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be the distribution with support s_μ and probability vector $\boldsymbol{\omega}_\mu$ such that $\mathrm{rg}(\mu, \epsilon) > 0$. By the decomposition of the randomization gap, one of either (12) or (13) has to be positive.

If (12) is positive, *i.e.* $\mathrm{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu)$ - $\mathrm{FP}(C^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) > 0$, then by Lemma B.2, there exists a clique in $G^\epsilon_{s_\mu}$ that is not a hyperedge of $H^\epsilon_{s_\mu}$. This means that $H^\epsilon_{s_\mu}$ is not conformal.

If (13) is positive, *i.e.* $\mathrm{FP}(C^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu)$ - $\mathrm{IP}(C^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) > 0$, then by Lemma B.3, the graph $G^\epsilon_{s_\mu}$ is not perfect, which means that it has either an odd hole or an odd anti-hole.

Second implication. If *a)* holds and there is a clique of G^ϵ_S that is not a hyperedge of H^ϵ_S, then by Lemma B.2 there must exist some vector $\boldsymbol{\omega}$ such that $\mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega}) \neq \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega})$. By Lemma 4.1, it is always true that $\mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega}) \geq \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega})$, which implies that $\mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega})$ - $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) > 0$.

If *b)* holds and G^ϵ_S is not perfect, by Lemma B.3, there must exist some vector $\boldsymbol{\omega}$ such that $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) \neq \mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega})$. Given that $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) \geq \mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega})$, we have that $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega})$ - $\mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega}) > 0$.

Define μ as the distribution with support on S and probability vector $\boldsymbol{\omega}$ such that either $\mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega})$ - $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) > 0$ or $\mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega})$ - $\mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega}) > 0$. Then, we have that:

$$\mathrm{rg}(\mu, \epsilon) = \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) \text{ - } \mathrm{IP}(C^\epsilon_S, \boldsymbol{\omega}) + \mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega}) \text{ - } \mathrm{FP}(C^\epsilon_S, \boldsymbol{\omega}) > 0.$$

□

B.3 The special case of the ℓ_∞ norm.

Having identified that the randomization gap is linked with the presence of three possible problematic structures in the conflict hypergraph, it is worth asking if these structures actually exist. It turns out that when considering the ℓ_∞ norm, the uncovered cliques related to the nonconformity of the conflict hypergraph can never occur.

Corollary 4.2. *Consider the ℓ_∞ norm, and $\epsilon > 0$ and $\mu \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with clique hypergraph $C^\epsilon_{s_\mu} = (V, \mathcal{E})$. Let $G' = (V', E')$ be an induced subgraph of $G^\epsilon_{s_\mu}$. If G' is a clique, then $V' \in \mathcal{E}$. Thus, for any $\boldsymbol{\omega} \in \mathbb{R}_+^{|V|}$:*

$$\mathrm{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}) = \mathrm{FP}(C^\epsilon_{s_\mu}, \boldsymbol{\omega}).$$

Proof. Let $\{(x_i, y_i)\}_{i\in[n]} \subset \mathcal{X} \times \mathcal{Y}$ be the support of μ such that $1, \dots, n'$ are the indices in V' with $n' = |V'|$. Suppose that G' is a clique. Recall that $\mathcal{X} \subseteq \mathbb{R}^d$. For any $j \in [d]$, consider a permutation $\sigma_j : [n'] \to [n']$ that satisfies $x^{(j)}_{\sigma_j(1)} \leq \cdots \leq x^{(j)}_{\sigma_j(n')}$. Then, given that $\{\sigma_j(1), \sigma_j(n')\} \in E'$, we have that $B_\infty(x_{\sigma_j(1)}, \epsilon) \cap B_\infty(x_{\sigma_j(n')}, \epsilon) \neq \emptyset$, which implies

$$I_j \coloneqq [x^{(j)}_{\sigma_j(n')} - \epsilon, x^{(j)}_{\sigma_j(1)} + \epsilon] \neq \emptyset, \qquad \text{for any } j \in [d].$$

Let $z \in \mathbb{R}^d$ be such that $z^{(j)} \in I_j$ for every $j \in [d]$. Given that for any $j \in [d]$, $z^{(j)}$ satisfies the inequality $x^{(j)}_{\sigma_j(n')} - \epsilon \leq z^{(j)} \leq x^{(j)}_{\sigma_j(1)} + \epsilon$ and that $x^{(j)}_{\sigma_j(1)} \leq \cdots \leq x^{(j)}_{\sigma_j(n')}$, we can conclude that:

$$\forall i \in V', \quad \forall j \in [d], \quad x^{(j)}_i - \epsilon \leq z^{(j)} \leq x^{(j)}_i + \epsilon.$$

In other words, we have that $z \in B_\infty(x_i, \epsilon)$ for all $i \in V'$. This implies that the intersection of the ϵ-balls is non-empty, *i.e.*, $\bigcap_{i\in V'} B_\infty(x_i, \epsilon) \neq \emptyset$. Since G' is a clique of the conflict graph, for any $i, j \in V'$ we have that $\{i, j\} \in E'$ and thus $y_i \neq y_j$. Therefore, V' is a hyperedge in $H^\epsilon_{s_\mu}$. □

B.4 Constructions of specific structures for different ℓ_p norms

Example B.1 (Non-conformal conflict hypergraph in the ℓ_2 norm.)**.** *Let us consider the ℓ_2 norm. Consider $\mu \in \mathcal{P}(\mathbb{R}^K \times [K])$ the uniform distribution over the canonical basis of $\mathbb{R}^K$, i.e., $s_\mu = \{(b_1, 1), \dots, (b_K, K)\} \subset \mathbb{R}^K \times [K]$ and $\boldsymbol{\omega}_\mu = \frac{1}{K}\mathbf{1}_K$.*

Intersection of ϵ-balls for subsets of points. *For any subset $S \subseteq [K]$ such that $|S| = m$, we denote by b_S the average of the canonical vectors indexed by S, i.e.,*

$$b_S \coloneqq \frac{1}{m} \sum_{i\in S} b_i.$$

Furthermore, we can prove the following equivalence (see below):

$$\bigcap_{i\in S} B_2(b_i, \epsilon) \neq \emptyset \iff b_S \in \bigcap_{i\in S} B_2(b_i, \epsilon). \tag{27}$$

Simple calculations tell us that, for any S of size m and $i \in [n]$, the Euclidean distance between b_S and b_i is exactly $\sqrt{\frac{m-1}{m}}$. Hence, using (27), we have

$$\bigcap_{i\in S} B_2(b_i, \epsilon) \neq \emptyset \iff \epsilon \geq \sqrt{\frac{m-1}{m}}. \tag{28}$$

The conflict hypergraph is not conformal. *Let us set $\epsilon = \frac{1}{\sqrt{2}}$. We have that*

$$\forall i, j \in [K], \quad \|b_i - b_j\|_2 = 2\epsilon,$$

and therefore $\{i, j\} \in \mathcal{E}(H^\epsilon_{s_\mu})$ for all $i, j \in [K]$. This means that the subset of vertices $[K]$ induces a clique in $G^\epsilon_{s_\mu}$.

On the other hand, for any subset $S \subset [K]$ of size m greater than 2, we have that $\sqrt{\frac{m-1}{m}} > \epsilon$, so by (28) we have that

$$\forall S \subset [K], \quad |S| > 2 \implies \bigcap_{i\in S} B_2(b_i, \epsilon) = \emptyset,$$

which implies that no subset of size greater than 2 can be a hyperedge of $H^\epsilon_{s_\mu}$. In particular, the subset $[K]$, which constitutes a clique in $G^\epsilon_{s_\mu}$, is not a hyperedge of $H^\epsilon_{s_\mu}$. Thus, $H^\epsilon_{s_\mu}$ is not conformal.

***Proof of* (27).** *Now we proceed to prove the result stated in (27). We will do so by contradicting the fact that the mean minimizes the sum of squared distances. One direction is obvious, so we are going to prove that if the intersection $\cap_{i\in S}B_2(b_i,\epsilon)$ is non-empty, then the average b_S must belong to it.*

Suppose by contradiction that $b_S \notin \cap_{i\in S}B_2(b_i,\epsilon)$ but that $\cap_{i\in S}B_2(b_i,\epsilon)\neq\emptyset$. Then, there exists an index $i^*\in S$ such that $b_S\notin B_2(b_{i^*},\epsilon)$. This implies that

$$\|b_S - b_{i^*}\|_2 > \epsilon. \tag{29}$$

Note that the distance from b_S to any b_i, $i\in S$ is exactly $\sqrt{\frac{m-1}{m}}$, so (29) implies the following:

$$\forall i\in S,\quad \|b_S - b_i\|_2 > \epsilon. \tag{30}$$

By (30), we have that

$$\sum_{i\in S}\|b_S - b_i\|_2^2 > m\epsilon^2. \tag{31}$$

Now, from the assumption that the intersection is non-empty, take any $\bar{x}\in\cap_{i\in S}B_2(b_i,\epsilon)\neq\emptyset$. Note that this implies that

$$\forall i\in S,\quad \|\bar{x} - b_i\|_2^2 \leq \epsilon^2. \tag{32}$$

Recall that the average b_S has the property of minimizing the sum of squared norms, i.e.

$$b_s \in \operatorname*{argmin}_{x\in\mathbb{R}^K}\sum_{i\in S}\|x - b_i\|_2^2. \tag{33}$$

However, by (31) and (32) we get that

$$\sum_{i\in S}\|\bar{x} - b_i\|_2^2 \leq m\epsilon^2 < \sum_{i\in S}\|b_S - b_i\|_2^2.$$

This contradicts (33).

Example B.2 (Odd anti-holes with the ℓ_∞ norm)**.** *The anti-hole of size 5 is isomorphic to the hole of size 5, which is always possible to build. Let us see that for sizes greater than 5, it remains possible to build odd anti-holes.*

Fix $d=7$, and consider the following points in $\mathbb{R}^7$:

$$\begin{array}{lllllllll}
x_1 = & (0 & ,0.2 & ,0.3 & ,0.4 & ,0.5 & ,0.6 & ,1 &) \\
x_2 = & (1 & ,0 & ,0.3 & ,0.4 & ,0.5 & ,0.6 & ,0.7 &) \\
x_3 = & (0.1 & ,1 & ,0 & ,0.4 & ,0.5 & ,0.6 & ,0.7 &) \\
x_4 = & (0.1 & ,0.2 & ,1 & ,0 & ,0.5 & ,0.6 & ,0.7 &) \\
x_5 = & (0.1 & ,0.2 & ,0.3 & ,1 & ,0 & ,0.6 & ,0.7 &) \\
x_6 = & (0.1 & ,0.2 & ,0.3 & ,0.4 & ,1 & ,0 & ,0.7 &) \\
x_7 = & (0.1 & ,0.2 & ,0.3 & ,0.4 & ,0.5 & ,1 & ,0 &)
\end{array}$$

Let $\epsilon = 0.5 - 0.01$. Then it can be seen that the following properties hold:

- $\forall i\in[7], B_\infty(x_i,\epsilon)\cap B_\infty(x_{(i-1)[\mathrm{mod}\,7]},\epsilon) = B_p(x_i,\epsilon)\cap B_p(x_{(i+1)[\mathrm{mod}\,7]},\epsilon)=\emptyset$
- $\forall i,j\in[7]$ *such that* $i<j$ *and* $j-i\neq 1[\mathrm{mod}\,7]$, $\quad B_p(x_i,\epsilon)\cap B_p(x_j,\epsilon)\neq\emptyset$

In other words, these points form an anti-hole of size 7. By an analogous reasoning, one can conclude that odd anti-holes of any size exist for a sufficiently large dimension d.

C SUPPLEMENTARY MATERIAL FOR SECTION 5: THE RANDOMIZATION GAP CAN BE ARBITRARILY CLOSE TO 1/2

The proof of Theorem 5.1 relies on two key elements: (i) the construction of a non-perfect graph for which the value of the set packing problem and its fractional counterpart differ significantly, and (ii) the representation of this difference as the randomization gap of a distribution. In Section C.1, we demonstrate through Lemma C.5 that any loopless graph is isomorphic to the conflict graph of a distribution. Furthermore, if the graph is triangle-free, the representation in (ii) is amenable. In Section C.2, we establish Corollary C.1, which provides a triangle-free construction for (i), based on the iterative procedure presented in Chung et al. (1993). Combining both results, we prove Theorem 5.1 in Section C.3.

C.1 Every graph is the conflict graph of a distribution

In Lemmas C.2 and C.4 we prove that for the ℓ_∞ and ℓ_p norms (with $p \in (1, \infty)$) respectively, any graph G can be constructed by considering the overlap of some ϵ-balls. This corresponds to the second condition in the definition of the conflict graph (Definition 3.1). Building on this, we utilize the chromatic number (Definition C.1) to restrict the overlap of ϵ-balls to points from different classes (satisfying the first condition in Definition 3.1). Consequently, we prove Lemma C.5, which states that G is isomorphic to the conflict graph $G^\epsilon_{s_\mu}$ of a distribution μ. Furthermore, if G is triangle-free, then it is isomorphic to the loopless version of the conflict hypergraph $H^\epsilon_{s_\mu}$. This implies that the values of their respective (fractional) set packing problems coincide.

Lemma C.1 establishes a straightforward equivalence that will be used in the proofs throughout this subsection.

Lemma C.1 (Restatement of intersecting neighborhoods)**.** *Let us consider an ℓ_p norm with $p \in (1, \infty]$ and $\epsilon > 0$. The following assertion holds true:*

$$\forall x, x' \in \mathcal{X}, \quad B_p(x, \epsilon) \cap B_p(x', \epsilon) \neq \emptyset \iff \|x - x'\|_p \leq 2\epsilon$$

Proof. First, suppose that $B_p(x, \epsilon) \cap B_p(x', \epsilon) \neq \emptyset$ and take any $z \in B_p(x, \epsilon) \cap B_p(x', \epsilon)$. Then, by the triangle inequality

$$\|x - x'\|_p = \|x - z + z - x'\|_p \leq \|x - z\|_p + \|z - x'\|_p \leq \epsilon + \epsilon = 2\epsilon$$

For the other direction, suppose that $\|x - x'\|_p \leq 2\epsilon$ and consider the point $z = \frac{x+x'}{2}$. Then,

$$\|x - z\|_p = \|x - \frac{x + x'}{2}\|_p = \|\frac{x - x'}{2}\|_p = \frac{1}{2}\|x - x'\|_p \leq \frac{1}{2}2\epsilon = \epsilon$$

Thus, $z \in B_p(x, \epsilon)$. An analogous argument exchanging x by x' yields that $z \in B_p(x', \epsilon)$. As $z \in B_p(x, \epsilon) \cap B_p(x', \epsilon)$, we conclude that $B_p(x, \epsilon) \cap B_p(x', \epsilon) \neq \emptyset$.

□

Now we restate and prove lemmas related to the cubicity and sphericty of graphs. Basically, these results show for any p-norm, $\epsilon > 0$ and graph G, we can think of G as the intersection graph of some set of $|G|$ points. This will be the first step in the construction of discrete distributions, as it provides the support of the distribution in such a way that the conflicts between points are exactly those represented in G.

Lemma C.2 (Cubicity, restated from Roberts (1969))**.** *Let $G = (V, E)$ be any graph with $n \in \mathbb{N}^*$ vertices. Let us consider the ℓ_∞ norm and $\epsilon > 0$. There exist $d \in \mathbb{N}^*$ and a set of points $x_1, \ldots, x_n$ from $\mathbb{R}^d$ such that the following holds:*

$$\forall i, j \in V, \quad \{i, j\} \in E \iff B_\infty(x_i, \epsilon) \cap B_\infty(x_j, \epsilon) \neq \emptyset$$

Proof. Following the remark given in (Roberts, 1969, Section 2), we build coordinate functions to embed the n vertices of G into $\mathbb{R}^n$. For any $i \in V$, consider the following vector $x_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(n)}) \in \mathbb{R}^n$, where:

$$\forall j \in V,\ x_i^{(j)} = \begin{cases} 0, & \text{if } j = i \\ 0.9\epsilon, & \text{if } i \neq j \text{ and } \{i,j\} \in E \\ 1.1\epsilon, & \text{if } i \neq j \text{ and } \{i,j\} \notin E \end{cases}$$

Now, we distinguish between two key cases.

Case 1: If $\{i,j\} \in E$, the distance $\|x_i - x_j\|_\infty$ is strictly less than ϵ. We need to verify that for all $k \in [n]$, $|x_i^{(k)} - x_j^{(k)}| < \epsilon$. Let us consider each coordinate k:

- If $k = i$, then:

$$|x_i^{(k)} - x_j^{(k)}| = |0 - x_j^{(i)}| = x_j^{(i)} = 0.9\epsilon,$$

- Similarly, if $k = j$, then:

$$|x_i^{(k)} - x_j^{(k)}| = |0 - x_i^{(j)}| = x_i^{(j)} = 0.9\epsilon,$$

- If $k \neq i$ and $k \neq j$, then both $x_i^{(k)}$ and $x_j^{(k)}$ are either 0.9ϵ or 1.1ϵ, depending on whether $\{i,k\} \in E$ and $\{j,k\} \in E$. In any case, we have:

$$|x_i^{(k)} - x_j^{(k)}| \leq |1.1\epsilon - 0.9\epsilon| = 0.2\epsilon < \epsilon.$$

Thus, for all coordinates $k \in [n]$, we have $|x_i^{(k)} - x_j^{(k)}| < \epsilon$. Therefore, $\|x_i - x_j\|_\infty < \epsilon$, as desired.

Case 2: If $\{i,j\} \notin E$, the distance $\|x_i - x_j\|_\infty$ is strictly greater than ϵ.

- If $k = i$ or $k = j$, then:

$$|x_i^{(k)} - x_j^{(k)}| = 1.1\epsilon > \epsilon$$

Thus, if $\{i,j\} \notin E$ then $\|x_i - x_j\|_\infty > \epsilon$.

Conclusion. We have shown that $\|x_i - x_j\|_\infty \leq \epsilon$ if and only if $\{i,j\} \in E$.

By replacing ϵ by 2ϵ and using Lemma C.1, we obtain that:

$$\{i,j\} \in E \iff B_\infty(x_i, \epsilon) \cap B_\infty(x_j, \epsilon) \neq \emptyset.$$

This concludes the proof.

□

We now prove a result related to the concept of sphericity Fishburn (1983) of a graph, but generalized to any p-norm with $p \in (1, \infty)$.

Lemma C.4 (Sphericity for any p-norm)**.** *Let $G = (V, E)$ be any graph with $n \in \mathbb{N}^*$ vertices. Let us consider an ℓ_p norm with $p \in (1, \infty)$ and $\epsilon > 0$. There exist $d \in \mathbb{N}^*$ and a set of points $x_1, \ldots, x_n$ from $\mathbb{R}^d$ such that for any $i, j \in V$ such that $i \neq j$, the following holds:*

$$\{i,j\} \in E \iff B_p(x_i, \epsilon) \cap B_p(x_j, \epsilon) \neq \emptyset. \tag{34}$$

Proof. Let $G = (V, E)$ be any graph with $n \in \mathbb{N}^*$ vertices and $m \in \left[\frac{(n-1)n}{2}\right]$ edges. For any $i \in V$, we denote by $\deg_i$ the degree of i, i.e., the number of neighbors of i in G. Set $d = n + m$, and define $x_1, \ldots x_n \in \mathbb{R}^{n+m}$ as follows.

Definition of the vectors. For any $i \in [n]$ and $k \in [d]$, define $x_i^{(k)}$ as

$$x_i^{(k)} = \begin{cases} 1, & \text{if } k \in [m] \text{ and } i \in e_k \\ (n - \deg_i)^{\frac{1}{p}}, & \text{if } k = m + i \\ 0, & \text{otherwise.} \end{cases}$$

Let us now consider $i, j \in V$ such that $i \neq j$, we can distinguish two cases.

Case 1. If $\{i,j\} \notin E$ then $\nexists k \in [m]$ such that $x_i^{(k)} = x_j^{(k)} = 1$. Furthermore, as $i \neq j$, we have $m+i \neq m+j$. Hence, we have $\|x_i - x_j\|_p = \left(\deg_i + \deg_j + (n - \deg_i) + \left(n - \deg_j\right)\right)^{\frac{1}{p}} = (2n)^{\frac{1}{p}}$.

Case 2. If $\{i,j\} \in E$, then $\exists! k \in [m]$ such that $x_i^{(k)} = x_j^{(k)} = 1$. Furthermore, $m+i \neq m+j$. Hence, we have $\|x_i - x_j\|_p = \left(\deg_i + \deg_j - 1 + (n - \deg_i) + \left(n - \deg_j\right)\right)^{\frac{1}{p}} = (2n-1)^{\frac{1}{p}}$.

Finally, multiplying each x_i by a constant $2\epsilon \cdot (2n-1)^{-\frac{1}{p}}$ we obtain that

$$\|x_i - x_j\|_p = \begin{cases} 2\epsilon \cdot \left(\frac{2n}{2n-1}\right)^{\frac{1}{p}}, & \text{if } \{i,j\} \notin E \\ 2\epsilon, & \text{if } \{i,j\} \in E \end{cases}$$

As $\left(\frac{2n}{2n-1}\right)^{\frac{1}{p}} \geq 1$ for any $n \in \mathbb{N}^*$ and $p \in (1, \infty)$, we finally get that

$$\|x_i - x_j\|_p \leq 2\epsilon \iff \{i,j\} \in E.$$

By Lemma C.1, we conclude that (34) holds. □

Now that we know that for any graph G we can find a set of points S such that $G \simeq G_S^\epsilon$, we only need to include the labels of the points to be able to create a discrete distribution for a classification task. Interestingly, the condition that all points in conflict must be from different classes has a direct counterpart in graph theory: *vertex colorings*. We define below the chromatic number before proving Lemma C.5.

Definition C.1 (Chromatic number $\chi(G)$)**.** *Given a graph $G = (V, E)$, a* coloring *of the vertices of G is a function $c : V \to \mathbb{N}$ such that*

$$\forall\ \{i,j\} \in E, \quad c(i) \neq c(j).$$

The chromatic number *of G, denoted $\chi(G)$, is the smallest number of colors needed to color the vertices of G. That is, $\chi(G)$ is the smallest M such that there exists a coloring $c : V \to [M]$.*

Lemma C.5. *Let $G = (V, E)$ be any loopless graph with $n \in \mathbb{N}^*$ vertices. Let us consider an ℓ_p norm with $p \in (1, \infty]$ and $\epsilon > 0$. There exist $d, K \in \mathbb{N}$ and $S = \{(x_i, y_i)\}_{i \in [n]} \subset \mathbb{R}^d \times [K]$, such that $G_S^\epsilon \simeq G$, where the relation $\simeq$ denotes graph isomorphism. Moreover, if G is triangle-free, then G is isomorphic to the loopless version of H_S^ϵ and the number of classes satisfies $K \in \mathcal{O}\left(\sqrt{n/\log n}\right)$, where $n = |S|$.*

Proof. Using either Lemma C.2 or Lemma C.4 for $p = \infty$ or $p \in (1, \infty)$ respectively, there exist $d \in \mathbb{N}^*$ and a set of points $\{x_i\}_{i \in [n]}$ in $\mathbb{R}^d$ such that for any $i, j \in V$ with $i \neq j$, the following holds:

$$\{i,j\} \in E \iff B_p(x_i, \epsilon) \cap B_p(x_j, \epsilon) \neq \emptyset. \tag{35}$$

For the number of classes, take $K = \chi(G)$, the chromatic number of G. By definition of the chromatic number, there exists a K-coloring of G, *i.e* a function $c : V \mapsto [K]$, such that for any $i, j \in V$

$$\{i,j\} \in E \implies c(i) \neq c(j). \tag{36}$$

We define $S = \{(x_i, c(i))\}_{i \in [n]}$ and show that G_S^ϵ is isomorphic to G.

Proving isomorphism. As G and G_S^ϵ share the same set of vertices, we have to show that

$$\{i,j\} \in E \iff \{i,j\} \in E(G_S^\epsilon).$$

If $\{i,j\} \in E$, as G is loopless we have $i \neq j$, and therefore by (35) and (36) we know that $B_p(x_i, \epsilon) \cap B_p(x_j, \epsilon) \neq \emptyset$ and $c(i) \neq c(j)$. These two conditions imply that $\{i,j\} \in E(G_S^\epsilon)$ by definition of the conflict graph.

On the other hand, if $\{i,j\} \in E(G_S^\epsilon)$, we have that in particular, the ϵ-balls intersect, *i.e.* $B_p(x_i, \epsilon) \cap B_p(x_j, \epsilon) \neq \emptyset$. As G_S^ϵ is loopless, we have $i \neq j$, and this implies that $\{i,j\} \in E$ by (35).

Adding the triangle-free assumption. Note that we do not know the full structure of the conflict hypergraph H_S^ϵ, apart from the fact that the corresponding conflict graph G_S^ϵ is isomorphic to the graph G. However, if we add the assumption that G is triangle-free, we can show that H_S^ϵ is the loopless version of G. Suppose by contradiction that H_S^ϵ has a hyperedge $e \in \mathcal{E}(H_S^\epsilon)$ of size greater than 2. Then there would be a triplet of points x_i, x_j, x_k such that

$$B_p(x_i, \epsilon) \cap B_p(x_j, \epsilon) \cap B_p(x_k, \epsilon) \neq \emptyset.$$

This would imply the existence of the three edges $\{i, j\}, \{i, k\}$ and $\{j, k\}$ in G_S^ϵ, and therefore in G, which form a triangle. As G is triangle-free, we can conclude that H_S^ϵ cannot contain any hyperedge of size greater than 2. Thus, by definition of conflict graph G_S^ϵ, we conclude that G is isomorphic to the loopless version of H_S^ϵ.

Regarding the number of classes when G is triangle-free, we know that $K = \chi(G) \in \mathcal{O}\left(\sqrt{\frac{n}{\log n}}\right)$ (See Davies and Illingworth (2022); Erdös and Hajnal (1985); Jensen and Toft (2011)).

□

C.2 Constructing non-perfect graphs with large randomization gap

Lemma C.5 ensures that any graph G can be materialized as the conflict graph of some discrete distribution μ. Moreover, if the G is triangle-free, then G is essentially the conflict hypergraph, which implies that we can compute the randomization gap of μ using only the (fractional) set packing problem over G. To prove Theorem 5.1 we thus only need to build a graph G for which we can control the gap between the value of the fractional set packing problem and the set packing problem. With this objective in mind, we go over the construction proposed in Chung et al. (1993) of a graph that is triangle-free and without large independent sets. We restate and adapt their main results to our notation for completeness. The construction in Chung et al. (1993) is iterative. Starting from a graph G, the authors propose to build a new, larger graph H called its *fibration*, using 6 copies of G and connecting vertices between copies in a particular manner. This construction will preserve the property of being triangle-free, will increase the number of nodes by a factor of 6, but more importantly, it will at most increase the size of the largest independent set by a factor of 4. Let us formalize this.

Definition C.2 (Independence number $\alpha(G)$)**.** *Given a graph G, the* independence number *of G, denoted $\alpha(G)$, is the size of a maximum independent set of G.*

Definition C.3 (Fibration Chung et al. (1993))**.** *For any graph $G = (V(G), E(G))$, the* fibration *of G is the graph $H = (V(H), E(H))$ defined as follows:*

1. $V(H) = V(G) \times [6]$.
2. $\forall i \in [6], \quad \{u, v\} \in E(G) \implies \{(u, i), (v, i)\} \in E(H)$.
3. $\forall i, j \in [6]$ *with* $j \equiv (i+1)[\text{mod } 6], \quad \{u, v\} \in E(G) \implies \{(u, i), (v, j)\} \in E(H)$ *and* $\{(u, j), (v, i)\} \in E(H)$.
4. $\forall i, j \in [6]$ *with* $j \equiv (i+3)[\text{mod } 6], \quad u \in V(G) \implies \{(u, i), (u, j)\} \in E(H)$.

In Figure 5 we show an example of the fibration of the cycle with three nodes C_3. There are two properties of the fibration that are important:

Lemma C.6 (Triangle-free property preservation, Lemma 1 in Chung et al. (1993))**.** *If G is triangle-free and H is the fibration of G, then H is triangle-free.*

Lemma C.7 (Bound on the maximum independent set, Lemma 2 in Chung et al. (1993))**.** *If H is the fibration of G, then $\alpha(H) \leq 4 \cdot \alpha(G)$.*

To construct our graph of interest, we will iteratively apply the fibration operation, starting from an initial graph. Let G_0 be a fixed triangle-free graph, and let G_t denote the graph obtained after applying the fibration operation t times to G_0. Then, by Lemmas C.6 and C.7, we have the following properties:

$$G_t \text{ is triangle-free}$$
$$|G_t| = 6^t \cdot |G_0| \tag{37}$$
$$\alpha(G_t) \leq 4^t \cdot \alpha(G_0) \tag{38}$$

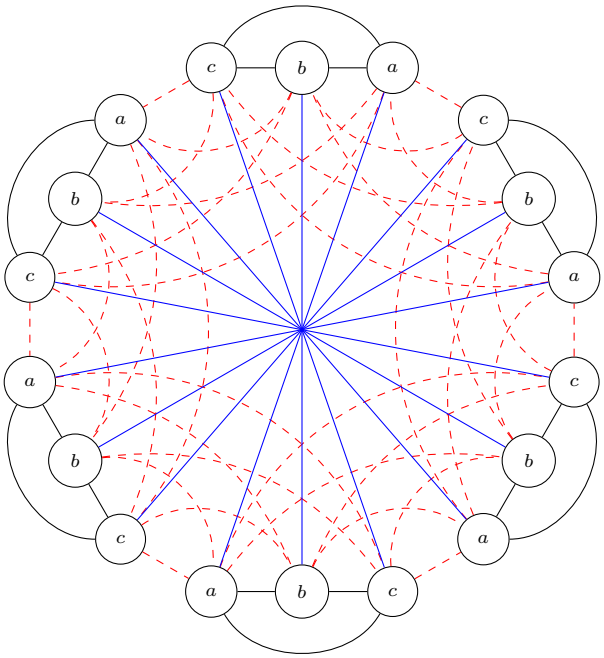

Figure 5: Example of the graph G_1 built in Chung et al. (1993) when the initial G_0 is the 3-cycle C_3. There are 6 copies of C_3 with nodes labeled a, b and c. Black edges are those within each copy, while red and blue edges are the ones added by the construction in Chung et al. (1993) between different copies of the initial graph.

For any graph G, the set packing problem with the weight vector $\mathbf{1}_{|G|}$ has a natural interpretation: it is exactly the size of the maximum independent set. In other words, for any graph G,

$$\mathrm{IP}(G, \mathbf{1}_{|G|}) = \alpha(G). \tag{39}$$

Coming back to the graph G_t, by (38) and (39), we can conclude that

$$\mathrm{IP}(G_t, \mathbf{1}_{|G_t|}) \leq 4^t \alpha(G_0). \tag{40}$$

On the other hand, for any graph G, the vector $\frac{1}{2}\mathbf{1}_{|G|}$ is feasible for the fractional set packing problem because there are no hyperedges of size larger than 2, so any constraint in the definition of fractional packing involves at most 2 vertices. In other words, for any graph G,

$$\mathrm{FP}(G, \mathbf{1}_{|G|}) \geq \frac{1}{2}\mathbf{1}_{|G|}^T \mathbf{1}_{|G|} \geq \frac{|G|}{2}. \tag{41}$$

Now let us consider the normalized vector $\frac{1}{|G_t|}\mathbf{1}_{|G_t|}$. Clearly, we have that

$$\mathrm{IP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) = \frac{1}{|G_t|}\mathrm{IP}(G_t, \mathbf{1}_{|G_t|}) \tag{42}$$

$$\mathrm{FP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) = \frac{1}{|G_t|}\mathrm{FP}(G_t, \mathbf{1}_{|G_t|}) \tag{43}$$

By (41) and (43), we can lower bound the value of the fractional set packing problem over G_t as follows:

$$\mathrm{FP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) \geq \frac{1}{2}. \tag{44}$$

Now we upper bound the value of the set packing problem for the t-fibration graph:

$$\begin{aligned}
\mathrm{IP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) &= \frac{1}{|G_t|}\mathrm{IP}(G_t, \mathbf{1}_{|G_t|}) && \text{(By (42))} && (45)\\
&\leq \frac{4^t}{|G_t|}\alpha(G_0) && \text{(By (40))} && (46)\\
&\leq \frac{4^t}{6^t}\alpha(G_0) && \text{(By (37))} && (47)
\end{aligned}$$

From (44) and (47) we can immediately deduce the following:

Corollary C.1. *For any graph G_0, let G_t be the graph obtained by applying the fibration operation t times, starting from G_0. Then, for any $\delta > 0$, the following holds:*

$$t > \frac{\log(\delta/\alpha(G_0))}{\log(2/3)} \implies \mathrm{FP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) - \mathrm{IP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) \geq \frac{1}{2} - \delta \tag{48}$$

Proof. If $t > \frac{\log\left(\frac{\delta}{\alpha(G_0)}\right)}{\log\left(\frac{2}{3}\right)}$ then we have

$$\begin{aligned}
t\log\frac{2}{3} &\leq \log\frac{\delta}{\alpha(G_0)}\\
\left(\frac{2}{3}\right)^t &\leq \frac{\delta}{\alpha(G_0)}\\
\left(\frac{2}{3}\right)^t \alpha(G_0) &\leq \delta
\end{aligned}$$

Thus, with (47), $t > \frac{\log\left(\frac{\delta}{\alpha(G_0)}\right)}{\log\left(\frac{2}{3}\right)}$ implies $\mathrm{IP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) \leq \delta$. Then, injecting (44), we obtain the desired result. □

C.3 Proof of Theorem 5.1

We are ready to state the main result of the section, that will use the fibration operation with Lemma C.5 to produce discrete distributions with randomization gap arbitrarily close to 1/2.

Theorem 5.1. *Fix any $\epsilon > 0$ and ℓ_p norm with $p \in (1, \infty]$. For any $\delta > 0$, there exist $d, K \in \mathbb{N}$ and a discrete distribution $\mu \in \mathcal{P}(\mathbb{R}^d \times [K])$ such that*

$$\mathrm{rg}(\mu, \epsilon) \geq 1/2 - \delta, \tag{14}$$

the conflict graph $G^\epsilon_{s_\mu}$ is not perfect and the conflict hypergraph $H^\epsilon_{s_\mu}$ is conformal. Furthermore, the number of classes satisfies $K \in \mathcal{O}\left(\sqrt{n/\log n}\right)$ with $n = |s_\mu|$.

Proof. For the fixed ϵ, p-norm and δ, take any triangle-free loopless graph G (for example, the 5-cycle C_5). Take any t that satisfies the condition in (48) from Corollary C.1 and consider G_t the t-fibration of G.

By Lemma C.6, G_t is triangle-free, and by Corollary C.1, we have that

$$\mathrm{FP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) - \mathrm{IP}(G_t, \frac{1}{|G_t|}\mathbf{1}_{|G_t|}) \geq \frac{1}{2} - \delta \tag{49}$$

By Lemma C.5, as G_t is loopless, there exist $d, K \in \mathbb{N}$ and $S = \{(x_i, y_i)\}_{i \in [|G_t|]} \subset \mathbb{R}^d \times [K]$ such that $G^\epsilon_S \simeq G_t$. Moreover, as G_t is triangle-free, we have that H^ϵ_S without considering loops is isomorphic to G_t. This implies[6] that

$$\forall \boldsymbol{\omega} \in \mathbb{R}^n_+, \quad \mathrm{FP}(H^\epsilon_S, \boldsymbol{\omega}) = \mathrm{FP}(G^\epsilon_S, \boldsymbol{\omega}).$$

[6]The constraints associated to loops are not considered in the definition of a fractional packing as $\mathcal{Q}^{\mathrm{frac}}(H^\epsilon_S) \subseteq [0,1]^n$.

Define μ as the discrete distribution with support $s_\mu = S$ and uniform probability vector $\boldsymbol{\omega}_\mu = \frac{1}{|G_t|}\mathbf{1}_{|G_t|}$. Given that $\mathrm{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{IP}(G^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu)$ by Corollary 4.1, we can conclude thanks to (49) that

$$\mathrm{rg}(\mu, \epsilon) = \mathrm{FP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) - \mathrm{IP}(H^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) = \mathrm{FP}(G^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) - \mathrm{IP}(G^\epsilon_{s_\mu}, \boldsymbol{\omega}_\mu) \geq \frac{1}{2} - \delta.$$

To check that $H^\epsilon_{s_\mu}$ is conformal, notice that as $G^\epsilon_{s_\mu}$ is triangle-free, the largest cliques of $G^\epsilon_{s_\mu}$ are the 2-edges, which are also hyperedges in $H^\epsilon_{s_\mu}$. Thus, all cliques from $G^\epsilon_{s_\mu}$, including the maximal ones, are hyperedges of $H^\epsilon_{s_\mu}$. Lastly, by Theorem 4.1, that $G^\epsilon_{s_\mu}$ is not perfect.

The fact that $K \in \mathcal{O}(\sqrt{\frac{n}{\log n}})$ where $n = |G_t|$ is also given by Lemma C.5 because G_t is triangle-free.

□

E SUPPLEMENTARY MATERIAL: ON THE HARDNESS OF COMPUTING THE RANDOMIZATION GAP.

Here, we formally redefine the optimization and decision problems discussed in the paper. Our focus is to express these problems in the context of hypergraph-based set packing, providing clear formulations for both their optimization and decision versions. We then prove the co-NP-completeness of the Randomization Gap decision problem.

E.1 Definition of the Optimization Problems

Optimization Problem: Weighted Set Packing $\mathrm{IP}(H, \boldsymbol{\omega})$

Input: A hypergraph $H = (V, \mathcal{E})$ and a weight vector $\boldsymbol{\omega} \in \mathbb{Q}_+^{|V|}$

Output: $\sup_{q \in \mathcal{Q}(H)} \boldsymbol{\omega}^T q$

Optimization Problem: *Fractional* Weighted Set Packing $\mathrm{FP}(H, \boldsymbol{\omega})$

Input: A hypergraph $H = (V, \mathcal{E})$ and a weight vector $\boldsymbol{\omega} \in \mathbb{Q}_+^{|V|}$

Output: $\sup_{q \in ConvexHull(\mathcal{Q}(H))} \boldsymbol{\omega}^T q$

E.2 Decision Problems Associated to these Optimization Problems

Decision Problem: Weighted Set Packing $\mathrm{IP}(H, \boldsymbol{\omega}, \alpha)$

Input: A hypergraph $H = (V, \mathcal{E})$ and a weight vector $\boldsymbol{\omega} \in \mathbb{Q}_+^{|V|}$, a rational number $\alpha \in \mathbb{Q}$

Output: `YES` if there exists a packing $q \in \mathcal{Q}(H)$ such that $\boldsymbol{\omega}^T q \geq \alpha$, otherwise `NO`

Decision Problem: Randomization Gap Problem $\mathrm{RG}(H, \boldsymbol{\omega}, \alpha)$

Input: A hypergraph $H = (V, \mathcal{E})$ and a weight vector $\boldsymbol{\omega} \in \mathbb{Q}_+^{|V|}$, a rational number $\alpha \in \mathbb{Q}$

Output: `YES` if and only if $\mathrm{FP}(H, \boldsymbol{\omega}) - \mathrm{IP}(H, \boldsymbol{\omega}) \geq \alpha$, otherwise `NO`

E.3 Reminder on NP-hardness and co-NP-hardness

Only *decision* problems are in **NP** or in **co-NP**.

- A decision problem is in **NP** if, given a "yes" instance[7] x, there exists a certificate (or witness) c that can be verified in polynomial time. Formally, the problem is in **NP** is there exists a polynomial-time verification algorithm $A(x, c)$ such that:
  - If x is a "yes" instance, then there exists a certificate c such that $A(x, c)$ outputs `yes`.
  - Reciprocally, if x is a "no" instance, then for any certificate c, $A(x, c)$ will output `no`.
- A decision problem is in **co-NP** if, given a "no" instance x, there exists a certificate c that can be verified in polynomial time. Formally, the problem is in **co-NP** iff there exists a polynomial-time verification algorithm $A(x, c)$ such that:
  - If x is a "no" instance, then there exists a certificate c such that $A(x, c)$ outputs `yes`.
  - Reciprocally, if x is a "yes" instance, then for any certificate c, $A(x, c)$ will output `no`.

A decision problem is NP-complete if it belongs to the class NP *and* it is NP-hard. Same for co-NP.

E.4 Hardness of the RG decision problem

Theorem E.1. *The* $\mathrm{RG}(H, \boldsymbol{\omega}, \alpha)$ *problem is co-NP complete*

Proof. To show a problem is co-NP complete, we must first show it belongs to co-NP, and then that it is co-NP hard.

The problem belongs to co-NP. To show it belongs to co-NP, we must show that there exist certificates for all NO instances. For all NO instances $(H, \boldsymbol{\omega}, \alpha)$ we have $\mathrm{FP}(H, \boldsymbol{\omega}) - \alpha < \mathrm{IP}(H, \boldsymbol{\omega})$. Here, there must exist a packing q (which will be our certificated) such that $\mathrm{FP}(H, \boldsymbol{\omega}) - \alpha < \boldsymbol{\omega}^T q$. So for any NO instance, it suffices to pick any packing q satisfying this inequality to convince a verification algorithm that our instance is in fact a NO instance. Thus, the problem belongs to co-NP.

The problem is co-NP hard. To show it is co-NP-hard, we build a trivial reduction from the $\mathrm{IP}(H, \boldsymbol{\omega}, \alpha)$ problem. Assume by contradiction that there exists a polynomial time algorithm $Alg(H, \boldsymbol{\omega}, \alpha)$ able to solve the RG problem. Let us show that this would immediately imply the existence of a polynomial time algorithm to solve the IP problem, which contradicts the well known NP-hardness of IP. More precisely, Let D be the least common denominator of $\boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_{|V|}$ and define the function $floor(z) = \frac{1}{D} \lfloor z \times D \rfloor$. Let $(H, \boldsymbol{\omega}, \alpha')$ be an arbitrary instance of the IP problem. Define $\alpha = \mathrm{FP}(H, \boldsymbol{\omega}) - floor(\alpha')$. Observe that these conditions are equivalent:

$$\begin{gathered}
\mathrm{IP}(H, \boldsymbol{\omega}, \alpha') = \mathrm{No} \\
\mathrm{IP}(H, \boldsymbol{\omega}) < \alpha' \\
\mathrm{IP}(H, \boldsymbol{\omega}) \leq floor(\alpha') \\
\mathrm{FP}(H, \boldsymbol{\omega}) - \mathrm{IP}(H, \boldsymbol{\omega}) \geq \mathrm{FP}(H, \boldsymbol{\omega}) - floor(\alpha') \\
\mathrm{RG}(H, \boldsymbol{\omega}, \alpha) = Yes
\end{gathered}$$

Thus, we can run $Alg(H, \boldsymbol{\omega}, \alpha)$ in polynomial time and return the opposite of the boolean answer to solve $\mathrm{IP}(H, \boldsymbol{\omega}, \alpha')$ also in polynomial time. This contradicts NP-hardness of IP. Thus, our problem is co-NP-Hard, thus co-NP-complete.

□

[7] a YES instance is a input on which a correct algorithm should output YES