Taylor & Francis
Taylor & Francis Group

Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Compression schemes for concept classes induced by three types of discrete undirected graphical models

Tingting Luo & Benchong Li

To cite this article: Tingting Luo & Benchong Li (2023) Compression schemes for concept classes induced by three types of discrete undirected graphical models, Statistical Theory and Related Fields, 7:4, 287-295, DOI: 10.1080/24754269.2023.2260046

To link to this article: https://doi.org/10.1080/24754269.2023.2260046

STATISTICAL THEORY AND RELATED FIELDS
2023, VOL. 7, NO. 4, 287–295
https://doi.org/10.1080/24754269.2023.2260046

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Compression schemes for concept classes induced by three types of discrete undirected graphical models

Tingting Luo and Benchong Li

School of Mathematics and Statistics, Xidian University, Xi'an, People's Republic of China

ABSTRACT
Sample compression schemes were first proposed by Littlestone and Warmuth in 1986. Undirected graphical model is a powerful tool for classification in statistical learning. In this paper, we consider labelled compression schemes for concept classes induced by discrete undirected graphical models. For the undirected graph of two vertices with no edge, where one vertex takes two values and the other vertex can take any finite number of values, we propose an algorithm to establish a labelled compression scheme of size VC dimension of associated concept class. Further, we extend the result to other two types of undirected graphical models and show the existence of labelled compression schemes of size VC dimension for induced concept classes. The work of this paper makes a step forward in solving sample compression problem for concept class induced by a general discrete undirected graphical model.

ARTICLE HISTORY
Received 8 November 2022
Revised 26 June 2023
Accepted 14 July 2023

KEYWORDS
Discrete undirected graphical models; concept classes; VC dimension; sample compression schemes

1. Introduction

Compression is an important facet of learning. In statistical learning, excellent and appropriate compression can guarantee a small generalization error (Blumer et al., 1987). A significant notion of compression is sample compression scheme, which was introduced by Littlestone and Warmuth (1986). There are two types of sample compression schemes: labelled and unlabelled. Such compression schemes are approaches for encoding a finite sample in a small subsample.

A labelled sample compression scheme of size k for a concept class consists of a compression function g and a reconstruction function h. Compression function g compresses a finite sample to a small labelled subsample of size at most k that is called compression set, and reconstruction function h aims to construct a concept consistent with the original sample set by the compression set. A crucial fact called compactness lemma was given by Ben-David and Litman (1998), which essentially says that if a finite class admits a compression scheme so does an infinite class. Vapnik-Chervonenkis (VC) dimension is a measure of the complexity of function classes. In statistical learning, VC dimension is deeply related to probably approximately correct learning (PAC) and sample compression scheme (Blumer et al., 1989; Floyd & Warmuth, 1995). In particular, for any concept class of VC dimension d the size of its sample compression scheme must be at least d (Bollobás & Radcliffe, 1995). Furthermore, Floyd and Warmuth (1995) and Warmuth (2003) proposed a sample compression conjecture, that is, for every class of VC dimension d, there exists a sample compression scheme of size $O(d)$. Up to now, this question has not been solved. However, it has been proved for many families of concept classes which are natural and important in statistical learning. Floyd and Warmuth (1995) showed that there is a labelled sample compression scheme of size d for any maximum class of VC dimension d. Moran and Warmuth (2016) presented the labelled sample compression schemes of size d for ample classes which are a natural generalization of maximum classes based on Sandwich Lemma (Anstee et al., 2002; Bollobás & Radcliffe, 1995). One noteworthy result is that Moran and Yehudayoff (2016) proved the existence of labelled compression scheme of size $\exp(d)$ for arbitrary class of VC dimension d. In a latest study, Chepoi et al. (2021) showed that the topes of a complex of oriented matroids (COM) have a labelled sample compression scheme of size VC dimension d.

Another type of compression scheme is unlabelled. An unlabelled sample compression scheme is that the compression set is unlabelled. In other words, the compression function compresses each given sample to a subset of the domain of the sample. For some specific concept classes, unlabelled sample compression schemes have been explored by several researchers and they have achieved a series of results (Chalopin et al., 2019; Helmbold et al., 1990; Kuzmin & Warmuth, 2007; Marc, 2022; Pálvölgyi & Tardos, 2020; Rubinstein & Rubinstein, 2012).

CONTACT Benchong Li libenchong@xidian.edu.cn School of Mathematics and Statistics, Xidian University, Xi'an 710126, People's Republic of China

Figure 1. A simple UG G_1.

Undirected graphical model (UGM), also known as Markov network, is one of the research hot points in the field of statistical learning. In the past a few decades, it has been successfully applied to many real problems (Friedman, 2004; Lauritzen, 1996). We know that a statistical model is a family of probability distributions, which is a (part of) real algebraic variety from the viewpoint of algebraic geometry (Pistone et al., 2001). In particular, considering discrete data, a statistical model is the set of all solutions of some polynomials in the probability simplex (Geiger et al., 2001; Settimi & Smith, 2000).

Given an undirected graph (UG) and sample space, one has an algebraic geometric characterization of the corresponding discrete UGM (Geiger et al., 2006). Based on the work in Geiger et al. (2006), for general discrete UGMs, Li and Yang (2018) developed an algebraic method for computing VC dimensions of corresponding concept classes. Their work lays a theoretical foundation for this paper.

A natural question is whether the sample compression conjecture holds for the concept classes induced by discrete UGMs. In this paper, we consider concept classes induced by three types of discrete UGMs. Using the quadratic binomials associated with a positive UGM, we construct a labelled compression scheme of size VC dimension d. That is, the compression conjecture is correct for these three concept classes. It is an open question whether the concept class induced by a general discrete UGM has a sample compression scheme of size VC dimension.

The remainder of this paper is organized as follows. Section 2 introduces main definitions and notations. Section 3 discusses a property of one-inclusion graphs of concept classes induced by general UGMs and establishes the existence of labelled sample compression schemes of size equal to their VC dimensions for classes induced by three kinds of discrete UGMs. Section 4 summarizes the work of this paper.

2. Preliminaries

In this section, we introduce some formal definitions and notations that appear in this paper.

2.1. Discrete UGMs

We consider UGMs whose underlying UGs are simple. A UG $G = (V, E)$ contains a set of vertices V and a set of unordered pairs vertices (undirected edges) E. A graph is complete if there exists an edge between every pair of vertices. Given a UG G, a clique is a maximal complete subgraph of G, and we use κ_G to denote the set of all cliques of G. Let $X_1, X_2, \ldots, X_n$ be discrete random variables (vertices in G), where $X_i \in [k_i] \doteq \{0, 1, \ldots, k_i - 1\}$, $k_i \in \mathbb{N}$, $k_i \geq 2$. Let $X = (X_1, X_2, \ldots, X_n)$ be an n-dimension vector, and then the sample space $\mathcal{X} = \prod_{i=1}^{n}[k_i]$. A UGM $\mathscr{P}$ is a family of probability distributions, and each element (probability distribution on $\mathcal{X}$) in $\mathscr{P}$ has the form

$$p_{x_1x_2\cdots x_n} \doteq P(x_1, x_2, \ldots, x_n) = P(x) \propto \prod_{K\in\kappa_G} \psi_K(x), \tag{1}$$

where $x = (x_1, x_2, \ldots, x_n) \in \mathcal{X}$, and $\psi_K(x)$ is a potential function that depends on $\mathcal{X}$ only through the values of variables in K. We focus on the collection of positive probability distributions in $\mathscr{P}$, denoted as $\mathscr{P}^+$.

Example 2.1: The graph G_1 in Figure 1 is a simple UG, $V = \{X_1, X_2\}$, $E = \{(X_1, X_2)\}$ and $\kappa_{G_1} = \{\{X_1\}, \{X_2\}\}$. If $X_1, X_2 \in [2]$, $\mathcal{X}$ has four values: $x_{(1)} = (0, 0)$, $x_{(2)} = (1, 1)$, $x_{(3)} = (0, 1)$, $x_{(4)} = (1, 0)$.

2.2. Algebraic geometry of discrete UGMs

We work in the real polynomial ring $\mathbb{R}[\mathcal{X}]$ whose indeterminates are elementary probabilities $p_{x_1x_2\cdots x_n}$. A subset $I \subseteq \mathbb{R}[\mathcal{X}]$ is an ideal if it satisfies the following conditions: (1) $0 \in I$, (2) if $f, g \in I$, then $f + g \in I$, and (3) if $f \in I$ and $h \in \mathbb{R}[\mathcal{X}]$, then $hf \in I$. Every ideal I in $\mathbb{R}[\mathcal{X}]$ associates a variety

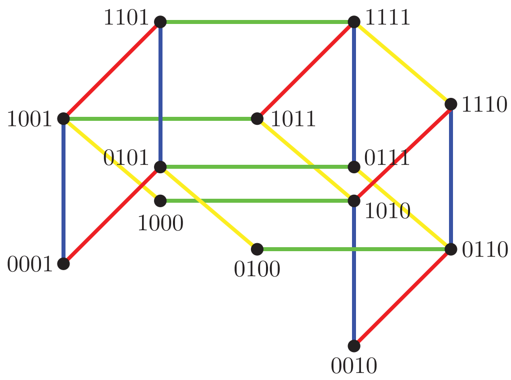
$$X_I^{\mathbb{R}_{>0}} = \{y \in \mathbb{R}^m_{>0} : f(y) = 0 \text{ for every } f \in I\},$$

where $m = \prod_{i=1}^{n} k_i$, $\mathbb{R}^m_{>0}$ is the set of m-dimensional vectors whose components are positive real numbers.

Given a UG G, we say X_i is independent of X_j given $\{X_1, X_2, \ldots, X_n\}\backslash\{X_i, X_j\}$ if $(X_i, X_j) \notin E$. Denote it by $X_iX_j|\{X_1, X_2, \ldots, X_n\}\backslash\{X_i, X_j\}$ and call it a pairwise conditional independence statement. Each X_iX_j

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$
0	0	0	1
0	0	1	0
0	1	0	0
0	1	0	1
0	1	1	0
0	1	1	1
1	0	0	0
1	0	0	1
1	0	1	0
1	0	1	1
1	1	0	1
1	1	1	0
1	1	1	1

(a)



(b)

Figure 2. (a) $\mathcal{C}_1$ and (b) its one-inclusion graph.

$|\{X_1, X_2, \dots, X_n\} \backslash \{X_i, X_j\}$ corresponds to a set of quadratic binomials: $\forall\ x_i, x_i' \in [k_i]$, $\forall\ x_j, x_j' \in [k_j]$, $\forall\ z \in \prod_{l=1, l\neq i, l\neq j}^{n} [k_l]$,

$$p_{x_i x_j z} p_{x_i' x_j' z} - p_{x_i x_j' z} p_{x_i' x_j z} \quad \text{holds.}$$

Let $I_{\text{pairwise}(G)}$ be the ideal in $\mathbb{R}[\mathcal{X}]$ generated by the quadratic binomials corresponding to all the pairwise conditional independence statements. The well-known Hammersley–Clifford theorem shows that $X^{\mathbb{R}_{>0}}_{I_{\text{pairwise}(G)}} = \mathscr{P}^+$ (Geiger et al., 2006; Lauritzen, 1996).

Example 2.2: Consider the same UG and sample space given in Example 2.1. Since $X_1 X_2$, the unique quadratic binomial associated with this discrete UGM is

$$p_{x_{(1)}} \cdot p_{x_{(2)}} - p_{x_{(3)}} \cdot p_{x_{(4)}}. \tag{2}$$

2.3. Concept classes, VC dimension, sample compression schemes

A concept class $\mathcal{C}$ over domain $\mathcal{X}$ (also denoted as dom($\mathcal{C}$)) is a family of functions of the form $c : \mathcal{X} \to \{0, 1\}$. Alternatively, $\mathcal{C}$ can be represented by the one-inclusion graph (Haussler et al., 1994). The vertices of the graph are all the concepts in $\mathcal{C}$ and two concepts have an edge if they differ at a single point. Each function $c \in \mathcal{C}$ is said to be a concept, and it can also be viewed as a subset of $\mathcal{X}$ where $x \in c$ if and only if $x \in \mathcal{X}$ and $c(x) = 1$. For some subset $A \subseteq \mathcal{X}$, the restriction of $\mathcal{C}$ on A is the class $\mathcal{C}|A = \{c \cap A : c \in \mathcal{C}\}$. If $|\mathcal{C}|A| = 2^{|A|}$ that means for every binary vector $b \in \{0, 1\}^{|A|}$ there is a concept c such that $c(A) = b$, then A is said to be shattered by $\mathcal{C}$. The VC dimension of $\mathcal{C}$ is given by

$$\text{VCdim}(\mathcal{C}) = \sup\{m | A \subseteq \mathcal{X} \text{ shattered by } \mathcal{C} \text{ and } |A| = m\}.$$

The sign function is defined as

$$\text{sign}\,(z) = \begin{cases} 1, & \text{if } z \geq 0, \\ 0, & \text{if } z < 0, \end{cases} \tag{3}$$

where $z \in \mathbb{R}$. The concept class induced by a discrete UGM is the set of Boolean functions over $\mathcal{X}$ of the form $\text{sign}(\log(P(x)/Q(x)))$ for $P, Q \in \mathscr{P}^+$, where $\mathscr{P}^+$ is a distribution class. Hence, $\text{sign}(\log(P(x)/Q(x))) = 1$ if $P(x) \geq Q(x)$ and $\text{sign}(\log(P(x)/Q(x))) = 0$ otherwise. A concept class $\mathcal{C}$ is called maximum if its size is $\binom{n}{\leq d} = \sum_{i=0}^{d} \binom{n}{i}$, where $n = |\text{dom}(\mathcal{C})|$ and $d = \text{VCdim}(\mathcal{C})$.

Example 2.3: According to Example 2.2., clearly, there is no pair of distributions $P, Q \in \mathscr{P}_1^+$ such that $(\text{sign}(\log(\frac{p_{00}}{q_{00}})), \text{sign}(\log(\frac{p_{11}}{q_{11}})), \text{sign}(\log(\frac{p_{01}}{q_{01}})), \text{sign}(\log(\frac{p_{10}}{q_{10}})))$ take the values $(1, 1, 0, 0)$ and $(0, 0, 1, 1)$. Since $p_{00} + p_{01} + p_{10} + p_{11} = q_{00} + q_{01} + q_{10} + q_{11} = 1$, the value $(0, 0, 0, 0)$ cannot appear. The concept class $\mathcal{C}_1$ is shown in Figure 2(a). Note that $\text{VCdim}(\mathcal{C}_1) = 3$, and it is not a maximum class. The one-inclusion graph of $\mathcal{C}_1$ is given in Figure 2(b).

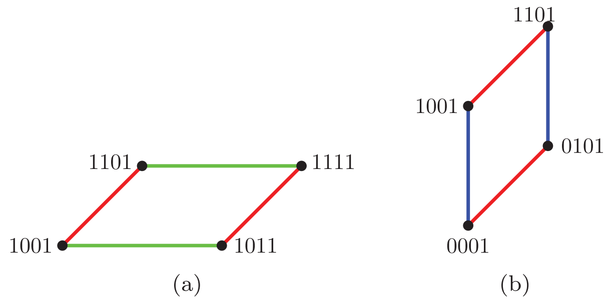


Figure 3. Two antipodal subgraphs of the graph in Figure 2(b).

A labelled sample is a set $s = \{(x_{(1)}, y_1), \ldots, (x_{(m)}, y_m)\}$, where $x_{(i)} \in \mathcal{X}$, $y_i \in \{0, 1\}$. A labelled sample compression scheme of a concept class $\mathcal{C}$ is a pair of functions (g, h). Function g is the compression function for mapping a labelled realizable finite sample that comes from some concept to a labelled subsample as a compression set. h is the reconstruction function for mapping the subsample to a hypothesis on $\mathcal{X}$ consistent with the entire sample. The size of the compression scheme is the size of the largest compression set. A sample compression scheme is proper if $h(g(s)) \in \mathcal{C}$ for any realizable sample s, otherwise, it is called improper.

3. Labelled sample compression schemes

In this section, we point out that one-inclusion graph of the concept class induced by a discrete UGM is not the tope graph of a COM if the underlying graph is incomplete and construct labelled sample compression schemes of size VC dimension for classes corresponding to three types of discrete UGMs.

3.1. One-inclusion graphs of concept classes induced by general discrete UGMs

For a discrete UGM, vertices in its underlying UG represent random variables. In this section, each vertex in the graph under consideration represents an object, which is not a random variable (the object is a concept in one-inclusion graph).

Let $G' = (V', E')$ be a finite, connected and simple graph. The distance $d(u, v) \doteq d_{G'}(u, v)$ between vertices u and v is defined as the length of a shortest (u, v)-path. We say that $G' = (V', E')$ is isometrically embeddable into a graph $H = (W, F)$ if there is a map $\varphi : V' \to W$ such that $d_H(\varphi(u), \varphi(v)) = d(u, v)$ for all $u, v \in V'$. A graph G' is a partial cube if it allows an isometric embedding into a hypercube graph Q_n.

It is significant for us to introduce two types of metric subgraphs. For a subgraph H, H is even if for every vertex v there exists unique vertex v' such that $d(v, v') = \mathrm{diam}(G')$, where $\mathrm{diam}(G')$ is the diameter of G'. An even subgraph is symmetric if $d(u, v) + d(u, v') = \mathrm{diam}(G')$ for all u, v in H. A symmetric-even subgraph H of G' is also called an antipodal subgraph. Two antipodal subgraphs of the graph Figure 2(b) are shown in Figure 3. A subgraph H is gated if for each vertex v in G' there exists a gate x in H, such that there is a shortest path from v through x to y for each vertex $y \in H$. In a partial cube that means there exists a path from v to H that does not use any color that appears in the edges of H. We say that a graph is antipodally gated if all of its antipodal subgraphs are gated.

Considering the subgraph Figure 3(b), it is an antipodal subgraph, while the vertex '0010' does not have gate. For vertex '0010', there is not a path from '0010' to Figure 3(b) that does not use red or blue edge. Similarly, the subgraph induced by vertex set {'0010','0110','1110','1010'} is a non-gated antipodal subgraph. Therefore, Figure 2(b) is not antipodally gated. By Theorem 1.1 in Knauer and Marc (2020), it is not the tope graph of a COM. Furthermore, each one-inclusion graph of concept class induced by a UGM whose underlying graph is incomplete is not the tope graph of a COM.

3.2. Sample compression schemes for the concept classes induced by discrete UGMs

Given a sample space, if a UG is complete, the corresponding concept class is a maximum class (Li & Yang, 2018), and its one-inclusion graph is the tope graph of a COM (Knauer & Marc, 2020). Now, we consider labelled compression scheme of $\mathcal{C}_1$.

Proposition 3.1: *$\mathcal{C}_1$ has a labelled compression scheme of size* 3.

Proof: Let $X_1, X_2 \in \{0, 1\}$, and consider G_2 in Figure 4. X_1 and X_2 are dependent. Note that the induced concept class $\mathcal{C}_2$ contains 15 concepts, and it is a maximum class. Then we can get a labelled compression scheme for $\mathcal{C}_2$

Figure 4. A UG G_2.

Table 1. A labelled compression scheme for $\mathcal{C}_1$.

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	Labelled compression sets
0	0	0	1	$\{(x_{(1)},0),(x_{(2)},0),(x_{(3)},0)\}$
0	0	1	0	$\{(x_{(1)},0),(x_{(2)},0),(x_{(4)},0)\}$
0	1	0	0	$\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},0)\}$
0	1	0	1	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},1)\}$
0	1	1	0	$\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},0)\}$
0	1	1	1	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	0	0	0	$\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},0)\}$
1	0	0	1	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},0)\}$, $\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},1)\}$
1	0	1	0	$\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},0)\}$, $\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},0)\}$
1	0	1	1	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},1)\}$, $\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},1)\}$, $\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	1	0	1	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},1),(x_{(3)},0),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},0),(x_{(4)},1)\}$
1	1	1	0	$\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},0)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},0)\}$
1	1	1	1	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},1)\}$

Table 2. A labelled compression scheme for a maximum class.

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	Labelled compression sets
0	0	0	0	$\{(x_{(1)},0),(x_{(2)},0),(x_{(3)},0)\}$, $\{(x_{(1)},0),(x_{(2)},0),(x_{(4)},0)\}$
0	0	0	1	$\{(x_{(1)},0),(x_{(2)},0),(x_{(4)},1)\}$
0	0	1	0	$\{(x_{(1)},0),(x_{(2)},0),(x_{(3)},1)\}$
0	1	0	0	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},0)\}$
0	1	0	1	$\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},1)\}$
0	1	1	0	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},0)\}$
0	1	1	1	$\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	0	0	0	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},0)\}$, $\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},0)\}$, $\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},0)\}$
1	0	0	1	$\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},1)\}$, $\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},1)\}$
1	0	1	0	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},1)\}$, $\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},0)\}$
1	0	1	1	$\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	1	0	0	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},1),(x_{(3)},0),(x_{(4)},0)\}$, $\{(x_{(2)},1),(x_{(3)},0),(x_{(4)},0)\}$
1	1	0	1	$\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},1),(x_{(3)},0),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},0),(x_{(4)},1)\}$
1	1	1	0	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},0)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},0)\}$
1	1	1	1	$\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},1)\}$

(Floyd & Warmuth, 1995; Moran & Warmuth, 2016), and a labelled compression scheme of size 3 for $\mathcal{C}_1$ as shown in Table 1 due to $\mathcal{C}_1 \subset \mathcal{C}_2$ and $\mathrm{VC\,dim}(\mathcal{C}_1) = \mathrm{VC\,dim}(\mathcal{C}_2) = 3$.

■

Remark 3.1: For G_1 in Figure 1, let $X_1 \in [k_1], X_2 \in [k_2]$, where $k_1, k_2 \in \mathbb{N}, k_1, k_2 \geq 2$ excluding the case $k_1 = k_2 = 2$. Then we can give the algebraic characterization of this discrete UGM:

$$p_{ij} \cdot p_{lk} - p_{ik} \cdot p_{lj}, \tag{4}$$

where $i, l \in [k_1]$, $i < l$, $j, k \in [k_2]$, $j < k$. It is easy to know that only $(0,0,1,1)$ and $(1,1,0,0)$ cannot occur on $\{(i,j),(l,k),(i,k),(l,j)\}$, that is, there are 14 functions, denoted by $\mathcal{C}_3$.

Note that $\mathrm{VCdim}(\mathcal{C}_3) = 3$. A natural idea would be to embed it into a maximum class of the same VC dimension that contains $\mathcal{C}_3$. As we all know there are 16 functions on 4 domain points in total, and then we can delete $(0,0,1,1)$ to get a maximum class of VC dimension 3 which contains $\mathcal{C}_3$. The labelled compression scheme for this maximum class is shown in Table 2. The concept $(1,1,0,0) \notin \mathcal{C}_3$, and it has four compression sets. If we know $\{((i,j),1),((l,k),1),((i,k),0)\}$, this evidence cannot be used to predict the label of (l,j) in $\mathcal{C}_3$ though the fact that it comes from $(1,1,0,1)$ is obvious. The key point is that compression sets corresponding to $(1,1,0,1)$ do not contain $\{((i,j),1),((l,k),1),((i,k),0)\}$ in the maximum class $\mathcal{C}_3 \cup (1,1,0,0)$. Thus we can assign the labelled compression set $\{((i,j),1),((l,k),1),((i,k),0)\}$ to $(1,1,0,1)$, $\{((i,j),1),((l,k),1),((l,j),0)\}$ to $(1,1,1,0)$, $\{((i,j),1),((i,k),0),((l,j),0)\}$ to $(1,0,0,0)$, and $\{((l,k),1),((i,k),0),((l,j),0)\}$ to $(0,1,0,0)$. Then one can obtain a new compression scheme for $\mathcal{C}_3$ as shown in Table 3. This compression scheme is crucial in this paper.

Table 3. A new labelled compression scheme for $\mathcal{C}_3$.

$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$x_{(4)}$	Labelled compression sets
0	0	0	0	$\{(x_{(1)},0),(x_{(2)},0),(x_{(3)},0)\}$, $\{(x_{(1)},0),(x_{(2)},0),(x_{(4)},0)\}$
0	0	0	1	$\{(x_{(1)},0),(x_{(2)},0),(x_{(4)},1)\}$
0	0	1	0	$\{(x_{(1)},0),(x_{(2)},0),(x_{(3)},1)\}$
0	1	0	0	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},0)\}$, $\{(x_{(2)},1),(x_{(3)},0),(x_{(4)},0)\}$
0	1	0	1	$\{(x_{(1)},0),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},0),(x_{(3)},0),(x_{(4)},1)\}$
0	1	1	0	$\{(x_{(1)},0),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},0)\}$
0	1	1	1	$\{(x_{(1)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	0	0	0	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},0)\}$, $\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},0)\}$, $\{(x_{(1)},1),(x_{(3)},0),(x_{(4)},0)\}$, $\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},0)\}$
1	0	0	1	$\{(x_{(1)},1),(x_{(2)},0),(x_{(4)},1)\}$, $\{(x_{(2)},0),(x_{(3)},0),(x_{(4)},1)\}$
1	0	1	0	$\{(x_{(1)},1),(x_{(2)},0),(x_{(3)},1)\}$, $\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},0)\}$
1	0	1	1	$\{(x_{(2)},0),(x_{(3)},1),(x_{(4)},1)\}$
1	1	0	1	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},0)\}$, $\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},1)\}$, $\{(x_{(1)},1),(x_{(3)},0),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},0),(x_{(4)},1)\}$
1	1	1	0	$\{(x_{(1)},1),(x_{(2)},1),(x_{(3)},1)\}$, $\{(x_{(1)},1),(x_{(2)},1),(x_{(4)},0)\}$, $\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},0)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},0)\}$
1	1	1	1	$\{(x_{(1)},1),(x_{(3)},1),(x_{(4)},1)\}$, $\{(x_{(2)},1),(x_{(3)},1),(x_{(4)},1)\}$

Remark 3.2: Consider formula (4). If we know arbitrary three of the four values $p_{ij}, p_{lk}, p_{ik}, p_{lj}$, then the value of the unknown probability can be obtained. However, according to Table 3, we must abide by the following rules: if we want to reconstruct $(0,0,0,0)$, $(0,1,0,1)$, $(0,1,1,0)$, $(1,0,0,1)$, $(1,0,1,0)$, $(1,1,1,1)$ we must know $\{((i,j),0),((l,k),0)\}$, $\{((i,j),0),((l,j),1)\}$, $\{((i,j),0),((i,k),1)\}$, $\{((l,k),0),\ ((l,j),1)\}$, $\{((l,k),0),((i,k),1)\}$, $\{((i,k),1,((l,j),1)\}$, respectively. Similarly, if we want to reconstruct concepts $(0,0,0,1)$, $(0,0,1,0)$, $(0,1,1,1)$, $(1,0,1,1)$, we need to know the specified three points $\{((i,j),0),((l,k),0),((l,j),1)\}$, $\{((i,j),0),((l,k),0),((i,k),1)\}$, $\{((i,j),0),((i,k),1),((l,j),1)\}$, $\{((l,k),0),((i,k),1),((l,j),1)\}$, respectively. The remaining four concepts in $\mathcal{C}_3$ need not be specific $0-1$ values.

Next we will show that when $X_1 \in [2]$ and X_2 takes any finite number of values, each concept in the induced class can be compressed to a labelled set of size VC dimension.

Lemma 3.2: *For G_1 in Figure 1, let $X_1 \in [2]$, $X_2 \in [k_2]$, $k_2 \in \mathbb{N}$ and $k_2 \geq 2$, let $\mathcal{C}_4$ denote the class induced by this discrete UGM, and then for each concept $c \in \mathcal{C}_4$, there exists a compression set of size k_2+1 which can reconstruct c exactly.*

Proof: For $k_2 = 2$, we have $\mathcal{C}_4 = \mathcal{C}_1$, and then the lemma holds. For $k_2 > 2$, $\mathcal{C}_4$ has $2k_2$ domain points, denoted as $(0,0), (1,0), \ldots, (1,k_2-1)$. The quadratic binomials associated with this UGM are as follows:

$$p_{0i} \cdot p_{1j} - p_{0j} \cdot p_{1i}, \tag{5}$$

where $i \in [k_2-1]$, $i < j \in [k_2]$. We know that $\mathrm{VCdim}(\mathcal{C}_4) = k_2+1$ (Li & Yang, 2018). $\forall\ c \in \mathcal{C}_4$, there are four cases to discuss.

Case (1): $\{c|\{(0,m),(1,m)\} :\ m \in [k_2]\} = (1,1)$.

Note that $c|\{(0,i),(1,j),(0,j),(1,i)\} = (1,1,1,1)$, so by Remark 3.2, we must retain $((0,j),1)$ and $((1,i),1)$ to reconstruct '1111'. Therefore, for any m we first retain $((0,m),1)$ and $((1,m),1)$. Then for any $n \in [k_2]\backslash m$, if $n<m$ retain $((1,n),1)$; if $n>m$ retain $((0,n),1)$. Then we can get the compression set of size k_2+1.

Case (2): $\{c|\{(0,m),(1,m)\} :\ m \in [k_2]\} = \{(1,1),(0,0)\}$. There are two subcases as follows.

Case (2a): $c|\{(0,m),(1,m)\} = (0,0)$ and $c|\{(0,n),(1,n)\} = (1,1)$ for each $n<m$. We can first retain $((0,m),0)$ and $((1,m),0)$. Then according to reconstruction rules, for $n<m$ we retain $((1,n),1)$. For $n>m$, if $c|\{(0,n),(1,n)\} = (1,1)$ then retain $((0,n),1)$; if $c|\{(0,n),(1,n)\} = (0,0)$ then retain $((1,n),0)$.

Case (2b): $c|\{(0,m),(1,m)\} = (1,1)$ and $c|\{(0,n),(1,n)\} = (0,0)$ for each $n<m$. We can first retain $((0,m),1)$, $((1,m),1)$ and $((0,n),0)$ for $n<m$. For $n>m$, if $c|\{(0,n),(1,n)\} = (1,1)$ then retain $((0,n),1)$; if $c|\{(0,n),(1,n)\} = (0,0)$ then retain $((1,n),0)$.

Case (3): $\forall\ m \in [k_2]$, $c|\{(0,m),(1,m)\} \in \{(1,1),(0,0),(0,1)\}$, and there exists at least one $m \in [k_2]$ such that $c|\{(0,m),(1,m)\} = (0,1)$.

Let $k = \min\{m; c|\{(0,m),(1,m)\} = (0,1)\}$. We first retain $((0,k),0)$ and $((1,k),1)$. Then by the rule shown in Table 3, for any $n<k$ if $c|\{(0,n),(1,n)\} = (0,0)$, we retain $((0,n),0)$ or $((1,n),0)$; if $c|\{(0,n),(1,n)\} = (1,1)$, we retain $((0,n),1)$ or $((1,n),1)$. For any $n>k$ if $c|\{(0,n),(1,n)\} = (0,0)$, we retain $((1,n),0)$; if $c|\{(0,n),(1,n)\} = (0,1)$, retain $((0,n),0)$ or $((1,n),1)$; if $c|\{(0,n),(1,n)\} = (1,1)$, retain $((0,n),1)$. This means that k_2+1 points are reserved.

Case (4): $\forall\ m \in [k_2], c|\{(0,m),(1,m)\} \in \{(1,1),(0,0),(1,0)\}$, and there exists at least one $m \in [k_2]$ such that $c|\{(0,m),(1,m)\} = (1,0)$.

Let $l = \max\{m; c|\{(0,m),(1,m)\} = (1,0)\}$. First, we retain $((0,l),1)$ and $((1,l),0)$. Then using similar techniques as in Case (3), it follows that for any $n < l$ if $c|\{(0,n),(1,n)\} = (0,0)$, we retain $((0,n),0)$; if $c|\{(0,n),(1,n)\} = (1,1)$, we retain $((1,n),1)$; if $c|\{(0,n),(1,n)\} = (1,0)$, we retain $((0,n),1)$ or $((1,n),0)$. For any $n > l$ if $c|\{(0,n),(1,n)\} = (0,0)$, we retain $((0,n),0)$ or $((1,n),0)$; if $c|\{(0,n),(1,n)\} = (1,1)$, we retain $((0,n),1)$ or $((1,n),1)$.

In summary, we can get a compression set A of size k_2+1 for any concept c. Clearly, the compression process ensures that we can reconstruct $c|(\mathrm{dom(c)}\backslash A)$ correctly only by points in the compression set A. ■

In fact, in the proof of Lemma 3.2, Case (3) contains the following four subcases: $\{c|\{(0,m),(1,m)\} : m \in [k_2]\} = \{(0,1)\}, \{(0,1),(0,0)\}, \{(0,1),(1,1)\}, \{(1,1),(0,0),(0,1)\}$. Case (4) also has four subcases: $\{c|\{(0,m),(1,m)\} : m \in [k_2]\} = \{(1,0)\}, \{(1,0),(0,0)\}, \{(1,0),(1,1)\}, \{(1,1),(0,0),(1,0)\}$.

On a high level, our goal is to construct a sample compression scheme for $\mathcal{C}_4$ which compresses any sample s of $\mathcal{C}_4$ to a subsample of size at most k_2+1, and then this subsample can be used to reconstruct a concept consistent with s. Next we will give an algorithm which generates such a compression scheme. The input is a labelled sample s of $\mathcal{C}_4$. The output is a subsample s' of size at most k_2+1 that represents some hypothesis consistent with s.

The Algorithm for Constructing a Labelled Compression Scheme.

- The compression algorithm: The input is a finite sample s of $\mathcal{C}_4$. Let $A_1, A_2, \ldots, A_n$ denote all domain sets corresponding to quadratic binomials that $2k_2$ domain points need to satisfy, $n = \binom{k_2}{2}$ and further let $A = \{A_t :\ |s|A_t| = 4, t = 1,\ldots,n\}$. If $A = \emptyset$, then $s' = s$. Otherwise, let $A' = \bigcup A_t$ where $A_t \in A$. Then compress $s|A'$ by the way used in the proof of Lemma 3.2 and denote the compression set for $s|A'$ by s'', and then $s' = s'' \bigcup (s|\mathrm{dom}(s)\backslash A')$.
- The reconstruction algorithm: The input is a subsample s' of size at most k_2+1, and reconstruction function is asked to predict the labels of elements in $\mathrm{dom}(\mathcal{C}_4)\backslash\mathrm{dom}(s')$. For any element $x \in \mathrm{dom}(\mathcal{C}_4)\backslash\mathrm{dom}(s')$, if there is a quadratic binomial that contains x and the other three points are in the subsample s', then predict the label of x by the scheme in Table 3. Let the set of such x be B, and then for elements in $\mathrm{dom}(\mathcal{C}_4)\backslash(\mathrm{dom}(s') \bigcup B)$, assign any set of possible $0-1$ values associated binomial (5) as their labels.

The following theorem (3.3) shows that the algorithm produces a correct compression set of size at most k_2+1.

Theorem 3.3: *Let s be a sample labelled consistently with some concept of $\mathcal{C}_4$, $1 \le |s| = m \le 2k_2$. Then the algorithm produces a compression set of size at most k_2+1 and the reconstructed hypothesis is consistent with the original sample s.*

Proof: We first show that the size of the compression set s' is at most k_2+1. If $A = \emptyset$, there is at most one pair of points (x_{0i}, y_{0i}), (x_{1i}, y_{1i}) in the sample s, where $y_{0i}, y_{1i} \in \{0,1\}$. Therefore, $|s'| = |s| \le 2 + (k_2 - 1) = k_2 + 1$. If $A \ne \emptyset$, there exist at least two pairs of points (x_{0i}, y_{0i}), (x_{1i}, y_{1i}) and (x_{0j}, y_{0j}), (x_{1j}, y_{1j}) in sample s where $i \ne j$. Let the set of these points be D, $|D| = 2n$, where $2 \le n \le k_2$. By Lemma 3.2, we can compress the labelled set D to s'' of size $n+1$. Then $|s'| = |s''| + |s - D| = n + 1 + m - 2n \le k_2 + 1$.

From the algorithm and Lemma 3.2, it follows immediately that reconstructed hypothesis is consistent with s on $\mathrm{dom}(s)\backslash\mathrm{dom}(s')$, and consequently, the sample set s. This ends the proof. ■

Example 3.4: Consider the graph G_1. If $X_1 \in [2], X_2 \in [3]$, let $\mathcal{C}_5$ denote the concept class induced by this discrete UGM, and then $\mathrm{dom}(\mathcal{C}_5) = \{(0,0),(1,0),(0,1),(1,1),(0,2),(1,2)\}$. The associated quadratic binomials are:

(1) $p_{00}\cdot p_{11} - p_{01}\cdot p_{10}$; (2) $p_{00}\cdot p_{12} - p_{02}\cdot p_{10}$; (3) $p_{01}\cdot p_{12} - p_{02}\cdot p_{11}$.

It is easy to know the VC dimension of $\mathcal{C}_5$ is 4 (Li & Yang, 2018). Consider the sample $s = \{((0,0),1), ((1,0),1), ((0,1),1), ((1,1),1), ((0,2),0), ((1,2),1)\}$ which is a concept in $\mathcal{C}_5$ (e.g. let $P = (0.12, 0.28, 0.12, 0.28, 0.06, 0.14)$, $Q = (0.09, 0.01,\ 0.09, 0.01, 0.72, 0.08)$, $s = \mathrm{sign}(\log(P(x)/Q(x)))$, $x \in \mathrm{dom}(\mathcal{C}_5)$). Using the proposed algorithm, we can compress s to $\{((0,0),1), ((0,1),1), ((0,2),0), ((1,2),1)\}$. The reconstruction function predicts the label '1' for $(1,0)$, '1' for $(1,1)$ by binomials (2) and (3), respectively. That is, the reconstructed hypothesis we obtain is exactly s.

If $s = \{((0,0),1), ((0,1),1), ((1,1),1), ((1,2),1)\}$, by the proposed algorithm, s itself is the labelled sample compression set. Then $(1,0)$ and $(0,2)$ are both labelled by '0'. This concept does not belong to $\mathcal{C}_5$, which means that the labelled compression scheme we presented is improper.

Naturally, we can extend Lemma 3.2 to a family of concept classes whose corresponding graph has three vertices and two edges.

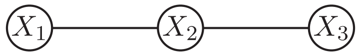


Figure 5. A UG of three vertices with two edges.

Theorem 3.5: *For the graph shown in Figure 5, let $X_1 \in [2]$, $X_i \in [k_i]$, where $k_i \in \mathbb{N}$, $k_i \geq 2$, $i = 2, 3$. Let $\mathcal{C}_6$ denote the class induced by this UGM. Then there exists a labelled compression scheme of size $k_2(k_3+1)$ for $\mathcal{C}_6$.*

Proof: There are $2 \cdot k_2 \cdot k_3$ domain points in $\mathcal{C}_6$, and $\mathrm{VCdim}(\mathcal{C}_6) = k_2(k_3+1)$ (Li & Yang, 2018). For the UG in Figure 5, $X_1X_3|X_2$, and one can use the sample compression scheme of $\mathcal{C}_4$ to construct scheme for $\mathcal{C}_6$.

For any sample set s, we can divide it into k_2 parts according to the value of X_2, denoted by $s_1, s_2, \ldots, s_{k_2}$. Note that for the elements in each part s_i, X_2 takes the same value $(i-1) \in [k_2]$. Let the compression sets of $s_1, s_2, \ldots, s_{k_2}$ be $s'_1, s'_2, \ldots, s'_{k_2}$, respectively, where $|s'_n| \leq k_3 + 1$, $n = 1, 2, \ldots, k_2$. Then the compression set of s is $\bigcup s'_n$ and $|\bigcup s'_n| \leq k_2(k_3+1)$. This ends the proof of Theorem 3.5. ■

Immediately, we have the following result.

Corollary 3.6: *For any UGM whose underlying graph has two cliques $K_1 = \{X_1, X_2, \ldots, X_{n_1}\}$, $K_2 = \{X_2, X_3, \ldots, X_n\}$ with $X_1 \in [2]$, then there is a labelled sample compression scheme of size VC dimension of the corresponding induced concept class.*

Proof: Note that we number the vertex that only in K_1 first, then vertices in $K_1 \cap K_2$, and $K_2 \backslash K_1$ finally. Suppose $X_i \in [k_i]$, where $k_i \in \mathbb{N}, k_i \geq 2, i = 2, \ldots, n$.

If $n_1 = 1$ ($K_1 \cap K_2 = \emptyset$), the two cliques can be viewed as two isolated vertices taking 2 and $\prod_{i=2}^{n} k_i$ values respectively, and this case reduces to the case of Lemma 3.2. If $n_1 \geq 2$ ($K_1 \cap K_2 = \{X_2, \ldots, X_{n_1}\}$), this UG can be viewed as the UG in Figure 5 taking 2, $\prod_{i=2}^{n_1} k_i$ and $\prod_{i=n_1+1}^{n} k_i$ values respectively, and this case reduces to the setting of Theorem 3.5. The conclusion is confirmed. ■

4. Conclusion and discussion

UGMs have become one of the popular models used for classification in statistical learning. In this paper, we focus on the question whether there exists a sample compression scheme of size VC dimension for the concept class induced by a discrete UGM. We show that for three types of discrete UGMs the answers are positive. The construction of these labelled compression schemes utilizes algebraic characterization of discrete UGMs.

A natural question is whether there are unlabelled sample compression schemes of size VC dimension for these three types of induced concept classes. For a general discrete UGM whether there exists a labelled sample compression scheme of size equal to VC dimension of the induced concept class remains open.

Acknowledgments

We would like to thank the Editor, Managing Editor, the Associate Editor and the reviewer for constructive comments and helpful suggestions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by National Natural Science Foundation of China (Research Plan in Shaanxi Province [Grant Number 12171382]), the Natural Science Basic Research Plan in Shaanxi Province [Grant Number 2020JM-188], and the Fundamental Research Funds for the Central Universities [Grant Number QTZX23002].

References

Anstee, R. P., Rányai, L., & Sali, A. (2002). Shattering news. *Graphs and Combinatorics*, *18*(1), 59–73. https://doi.org/10.1007/s003730200003.

Ben-David, S., & Litman, A. (1998). Combinatorial variability of Vapnik–Chervonenkis classes with applications to sample compression schemes. *Discrete Applied Mathematics*, *86*(1), 3–25. https://doi.org/10.1016/S0166-218X(98)00000-6 .

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information Processing Letters*, *24*(6), 377–380. https://doi.org/10.1016/0020-0190(87)90114-1 .

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1989). Learnability and the Vapnik–Chervonenkis dimension. *Journal of the ACM*, *36*(4), 929–965. https://doi.org/10.1145/76359.76371 .
Bollobás, B., & A. J. Radcliffe (1995). Defect sauer results. *Journal of Combinatorial Theory, Series A*, *72*(2), 189–208. https://doi.org/10.1016/0097-3165(95)90060-8.
Chalopin, J., Chepoi, V., Moran, S., & Warmuth, M. K. (2019). Unlabeled sample compression schemes and corner peelings for ample and maximum classes. *ICALP*, *34*, 1–15. https://doi.org/10.4230/LIPIcs.ICALP.2019.34 .
Chepoi, V., Knauer, K., & Philibert, M. (2021). Labeled sample compression schemes for complexes of oriented matroids. *arXiv:2110.15168v1*.
Floyd, S., & Warmuth, M. K. (1995). Sample compression, learnability, and the Vapnik–Chervonenkis dimension. *Machine Learning*, *21*(3), 269–304. https://doi.org/10.1023/A:1022660318680 .
Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science*, *303*(5659), 799–805. https://doi.org/10.1126/science.1094068.
Geiger, D., Heckerman, D., King, H. P., & Meek, C. (2001). Stratified exponential families: graphical models and model selection. *The Annals of Statistics*, *29*(2), 505–529. https://doi.org/10.1214/aos/1009210550.
Geiger, D., Meek, C., & Sturmfels, B. (2006). On the toric algebra of graphical models. *Annals of Statistics*, *34*(3), 1463–1492. https://doi.org/10.1214/009053606000000263.
Haussler, D., Sloan, R., & Warmuth, M. K. (1994). Predicting $\{0,1\}$ functions on randomly drawn points. *Inf. Comput.*, *115*(2), 248–292. https://doi.org/10.1006/inco.1994.1097.
Helmbold, D., Sloan, R., & Warmuth, M. K. (1990). Learning nested differences of intersection-closed concept classes. *Machine Learning*, *5*, 165–196. https://doi.org/10.1023/A:1022696716689 .
Knauer, K., & Marc, T. (2020). On tope graphs of complexes of oriented matroids. *Discrete & Computational Geometry*, *63*(2), 377–417. https://doi.org/10.1007/s00454-019-00111-z.
Kuzmin, D., & Warmuth, M. K. (2007). Unlabeled compression schemes for maximum classes. *Journal of Machine Learning Research*, *8*, 2047–2081. https://doi.org/10.1007/11503415_40.
Lauritzen, S. L. (1996). *Graphical models*. Oxford University Press.
Li, B., & Yang, Y. (2018). Complexity of concept classes induced by discrete Markov networks and Bayesian networks. *Pattern Recognition*, *82*, 31–37. https://doi.org/10.1016/j.patcog.2018.04.026.
Littlestone, N., & Warmuth, M. K. (1986). *Relating data compression and learnability* [Unpublished manuscript]. http://www.cse.ucsc.edu/manfred.
Marc, T. (2022). Unlabeled sample compression schemes for oriented matroids. *arXiv:2203.11535v2*.
Moran, S., & Warmuth, M. K. (2016). Labeled compression schemes for extremal classes. In *ALT* (pp. 34–39). Springer. https://doi.org/10.1007/978-3-319-46379-7_3
Moran, S., & Yehudayoff, A. (2016). Sample compression schemes for VC classes. In *ITA* (pp. 1–14). IEEE. https://doi.org/10.1109/ITA.2016.7888187
Pálvölgyi, D., & Tardos, G. (2020). Unlabeled compression schemes exceeding the VC-dimension. *Discrete Applied Mathematics*, *276*, 102–107. https://doi.org/10.1016/j.dam.2019.09.022.
Pistone, G., Riccomagno, E., & Wynn, H. P. (2001). *Algebraic statistics: Computational commutative algebra in statistics*. Chapman and Hall/CRC.
Rubinstein, B. I. P., & Rubinstein, J. H. (2012). A geometric approach to sample compression. *Journal of Machine Learning Research*, *13*(42), 1221–1261.
Settimi, R., & Smith, J. Q. (2000). Geometry, moments and conditional independence trees with hidden variables. *Annals of Statistics*, *28*(4), 1179–1205. https://doi.org/10.1214/aos/1015956712.
Warmuth, M. K. (2003). Compressing to VC dimension many points. In *Proceedings of the 16th Annual Conference on Learning Theory* (pp. 743–744). Springer. https://doi.org/10.1007/978-3-540-45167-9_60