University of Kentucky
UKnowledge

Statistics Faculty Publications

Statistics

2015

The Characteristic Imset Polytope of Bayesian Networks with Ordered Nodes

Jing Xi
University of Kentucky

Ruriko Yoshida
University of Kentucky, ruriko.yoshida@uky.edu

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Follow this and additional works at: https://uknowledge.uky.edu/statistics_facpub

Part of the Statistics and Probability Commons

The Characteristic Imset Polytope of Bayesian Networks with Ordered Nodes

THE CHARACTERISTIC IMSET POLYTOPE OF BAYESIAN NETWORKS WITH ORDERED NODES[*]

JING XI[†] AND RURIKO YOSHIDA[‡]

Abstract. In 2010, M. Studený, R. Hemmecke, and S. Lindner explored a new algebraic description of graphical models, called characteristic imsets. Compared with standard imsets, characteristic imsets have several advantages: they are still unique vector representatives of conditional independence structures, 0-1 vectors, and more intuitive in terms of graphs than standard imsets. After defining a characteristic imset polytope (cim-polytope) as the convex hull of all characteristic imsets with a given set of nodes, they also showed that a model selection in graphical models, which maximizes a quality criterion, can be converted into a linear programming problem over the cim-polytope. However, in general, for a fixed set of nodes, the cim-polytope can have exponentially many vertices over an exponentially high dimension. Therefore, in this paper, we focus on the family of directed acyclic graphs whose nodes have a fixed order. This family includes diagnosis models described by bipartite graphs with a set of m nodes and a set of n nodes for any $m, n \in \mathbb{Z}_+$. We first consider cim-polytopes for all diagnosis models and show that these polytopes are direct products of simplices. Then we give a combinatorial description of all edges and all facets of these polytopes. Finally, we generalize these results to the cim-polytopes for all Bayesian networks with a fixed underlying ordering of nodes with or without fixed (or forbidden) edges.

Key words. graphical model, characteristic imset polytope, diagnosis model, bipartite graph, directed acyclic graphs

AMS subject classifications. 51M20, 52B20, 13P25

DOI. 10.1137/130933848

1. Introduction. Bayesian networks (BNs), also known as belief networks, Bayes networks, Bayes(ian) models, or probabilistic directed acyclic graphical models (GMs), find applications in modeling knowledge in many areas, such as computational biology, bioinformatics (gene regulatory networks, protein structure, gene expression analysis [4], learning epistasis from genome-wide association studies data sets [5]), and medicine [15]. BNs are a part of the family of probabilistic GMs. These graphical structures represent knowledge about probabilistic structures for a statistical model. More precisely, each node in the graph represents a random variable and an edge between the nodes represents probabilistic dependencies among the random variables corresponding to the nodes adjacent to the edge [7]. BNs corresponding to GM structure are known as directed acyclic graphs (DAGs) defined by a set of nodes (vertices) and a set of directed edges.

In order to infer parameters from the observed data set, we first apply a model selection criterion called the *quality criterion*, which provides a way to construct highly predictive BN models from data by choosing the graph which gives the given criteria, such as Bayesian information criteria [10] or Akaike information criteria [1], maximum (see [12] for more details on quality criterions). Intuitively a quality criterion is a function, $\mathcal{Q}(G, D)$, which takes a DAG, G, and an observed data set, D, to evaluate how good the DAG G explains the observed data D. Note that different DAGs, G_1 and

[†]Mathematics Department, North Carolina State University, Raleigh, NC 27695-8205 (jing.xi.xj@gmail.com).

[‡]Statistics Department, University of Kentucky, Lexington, KY 40536-0082 (ruriko.yoshida@uky.edu).

G_2, may have the same conditional independences (CIs). In that case we say G_1, G_2 are *Markov equivalent*. When researchers wish to infer the conditional independences of the BN structure from the observed data set, each set of Markov equivalent graphs is described by one graph called the *essential graph* corresponding to the Markov equivalence class of DAGs [2]. In this paper we focus on quality criterions $\mathcal{Q}(G, D)$ such that $\mathcal{Q}(G_1, D) = \mathcal{Q}(G_2, D)$ if and only if G_1, G_2 are Markov equivalent.

Since in general there are superexponentially many essential graphs with a fixed set of nodes N, maximizing the quality criterion, $\mathcal{Q}(G, D)$, over all possible essential graphs with N is known to be NP-hard. Studený developed an algebraic representation of each essential graph G, called the *standard imset*, which is an integral vector representation of G in $\mathbb{R}^{2^{|N|}-|N|-1}$. From the viewpoint of this setting a criterion function $\mathcal{Q}(G, D)$ is a dot product of vectors in $\mathbb{R}^{2^{|N|}-|N|-1}$. In 2010, Studený, Vomlel, and Hemmecke showed that maximizing $\mathcal{Q}(G, D)$ over all essential graphs can be formulated as a linear programming problem over the convex hull of standard imsets for all possible essential graphs [14]. This gives us a systematic way to find the best graph with the optimality certificate rather than finding the best graph by brute-force search. Then Studený, Hemmecke, and Lindner explored an alternative vector representative of the BN structure, called *characteristic imset*. Compared with standard imsets, characteristic imsets have several advantages: they are still unique vector representative of conditional independence structures; they are 0-1 vectors; and they are more intuitive in terms of graphs than standard imsets [13].

In general, however, the dimension of the convex hull of the characteristic imsets with the fixed set of nodes N, called the *characteristic imset polytope* (cim-polytope), is exponentially large, and there are double exponentially many vertices as well as facets of the cim-polytope. Thus it is infeasible to optimize by software if $|N| > 6$. In order to solve the LP problem for a larger $|N|$, we need to understand the structure of the cim-polytope, such as a combinatorial description of edges and facets of the polytope, so that we might be able to apply a simplex method to find an optimal solution. However, in general, it is challenging because there are too many facets and too many edges of the polytope. Therefore here we start with a particular family of BN models, namely, *diagnosis models*.

In medical studies, researchers are often interested in probabilistic models in order to correctly diagnose a disease from a patient's symptoms. The diagnosis models, also known as quick medical reference diagnostic models, were introduced in [11] to diagnose a disease from a given set of symptoms of a patient. Therefore, here we focus on diagnosis models (e.g., [9]). Under this model, a DAG representing the model is a bipartite graph with two sets of nodes, one representing m diseases and one representing n symptoms, and a set of directed edges from nodes representing diseases to nodes representing symptoms (see Figure 1 for an example).

The main purpose of this paper is to understand the properties of the cim-polytopes for BNs, that is, the first steps toward characterizing explicitly the cim-polytopes for BNs. First, we are able to find an explicit combinatorial description of all edges of the cim-polytopes for diagnosis models with fixed m and n, that is, if G_1, G_2 are graphs representing two diagnosis models such that all symptoms have the same parents in G_1 and in G_2 except one symptom, then the characteristic imsets representing G_1, G_2 form an edge of the cim-polytope for diagnosis models. Then we prove that these cim-polytopes are direct products of n many $(2^m - 1)$ dimensional simplices, and an explicit description of all facets of them can be given based on this structure. Finally we generalize these results for the cim-polytopes for BNs with a fixed underlying ordering with or without fixed (or forbidden) edges.

This paper is organized as follows. In section 2 we introduce notation, and we state some definitions as well as propositions and their proofs. Section 3 gives a description of the cim-polytopes for diagnosis models, and section 4 gives a description of the cim-polytopes for BNs with a fixed underlying ordering. We end with a discussion of our future work in section 6.

2. Definitions and propositions for diagnosis models. In this section we state some notation and remind readers of some definitions.

DEFINITION 2.1. *A diagnosis model can be described by a bipartite graph whose nodes $N = \{a_1, \ldots, a_m\} \cup \{b_1, \ldots, b_n\}$ can be divided into disjoint sets $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$. Nodes in A can be interpreted as diseases and nodes in B can be interpreted as symptoms. Every single edge can be drawn only from a disease to a symptom. An example is given by Figure* 1.

For fixed A and B, where $|A| = m$ and $|B| = n$, we define $\mathcal{G}_{m,n} = \{$All possible directed bipartite graphs defined in Definition 2.1 *based on A and $B\}$.*
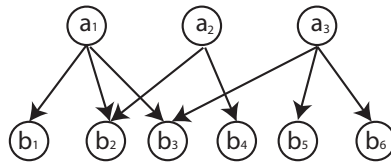

FIG. 1. *An example of a bipartite graph, $m = 3$, $n = 6$.*

Recall the definition of characteristic imsets.

DEFINITION 2.2. *Let G be an acyclic directed graph over N. The* characteristic imset *for G can be introduced as a zero-one vector c_G with components $c_G(S)$, where $S \subseteq N$, $|S| \geq 2$, given by*

$$c_G(S) = 1 \iff \exists\, i \in S \, such \, that \, j \in pa_G(i) \, for \, \forall\, j \in S\backslash\{i\},$$

where $j \in pa_G(i)$ means G includes the edge from j to i.

PROPOSITION 2.3. *Fix $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$. Assume $G \in \mathcal{G}_{m,n}$ and $|N| = m + n > 2$. Then $c_G(T)$ can take value 1 if and only if T has the form of $a_{i_1} \ldots a_{i_k} b_j$, where $1 \leq k \leq m$, $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.*

Proof. Notice that $\forall\, T \subseteq N$, $|T| \geq 2$, we can write T in the form

$$
\begin{aligned}
(2.1) \quad T = a_{i_1} \ldots a_{i_k} b_{j_1} \ldots b_{j_l}, \text{ where } \quad &0 \leq k \leq m, \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}, \\
&0 \leq l \leq n, \{j_1, \ldots, j_l\} \subseteq \{1, \ldots, n\}, \\
&k + l \geq 2,
\end{aligned}
$$

where l can be neither 0 nor greater than 1 because all a_i's must have empty parent sets (Definition 2.1), i.e., $l = 1$.   □

PROPOSITION 2.4. *Notation is adopted from Proposition* 2.3. *Suppose T has the form of $a_{i_1} \ldots a_{i_k} b_j$, where $1 \leq k \leq m$, $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$, then $c_G(T) = \prod_{s = i_1, \ldots, i_k} c_G(a_s b_j)$.*

Proof. Again by Definition 2.1, $\forall\, s, t \in \{i_1, \ldots, i_k\}$, $a_s \notin pa_G(a_t)$. Therefore,

$$
\begin{aligned}
(2.2) \quad c_G(T) = 1 \quad &\iff \{a_{i_1} \ldots a_{i_k}\} \subseteq pa_G(b_j) \\
&\iff a_s \in pa_G(b_j) \,\forall\, s = i_1, \ldots, i_k \\
&\iff c_G(a_s b_j) = 1 \,\forall\, s = i_1, \ldots, i_k.
\end{aligned}
$$

700 JING XI AND RURIKO YOSHIDA

Recall that $c_G(T)$ is binary. Thus $c_G(T) = \prod_{s=i_1,\ldots,i_k} c_G(a_s b_j)$. $\quad\square$

Remark 2.5. Proposition 2.4 implies that $\forall\ G \in \mathcal{G}_{m,n}$, c_G is determined by only $m \cdot n$ coordinates, $\{c_G(a_i b_j) : i = 1,\ldots,m, j = 1,\ldots,n\}$, i.e., the existence of directed edges $a_i \to b_j$, $i = 1,\ldots,m$ and $j = 1,\ldots,n$. Another way to see this property is that $\forall\ G \in \mathcal{G}_{m,n}$, G can be determined by $pa_G(b_j)$, $b_j \in B$. Thus if we consider a permutation of coordinates in c_G that corresponds to a permutation of T where T has the form in Proposition 2.3, then these coordinates can be broken into n parts:

$$\underline{a_1 b_1, \ldots, a_m b_1, \ldots, a_1 \ldots a_m b_1}, \underbrace{a_1 b_2, \ldots, a_m b_2, \ldots, a_1 \ldots a_m b_2}, \ldots, \underline{a_1 b_n, \ldots, a_1 \ldots a_m b_n},$$

where the sth part of coordinates $c_G(T)$, $T \in \{a_1 b_s, \ldots, a_m b_s, a_1 a_2 b_s, \ldots, a_1 \ldots a_m b_s\}$ only depends on $pa_G(b_s)$, and different parts are completely irrelevant in the sense that $pa_G(b_s)$, $b_s \in B$, can be decided separately.

PROPOSITION 2.6. *Fix m and n. The number of elements in $\mathcal{G}_{m,n}$ is 2^{mn}.*

Proof. This proof is trivial from Remark 2.5. $\quad\square$

PROPOSITION 2.7. *Suppose $G \in \mathcal{G}_{m,n}$. The number of nonzero coordinates in c_G is at most $n \cdot (2^m - 1)$.*

Proof. This result is straightforward from Proposition 2.3 by counting the number of coordinates $c_G(T)$, where T has the form shown in Proposition 2.3. $\quad\square$

DEFINITION 2.8. *Recall several definitions in elementary geometry (see [18] for more details on polyhedral geometry):*

- *a* closed convex polyhedron *(which will be called a* polyhedron *for short) in $\mathbb{R}^q$ can be defined by a system of linear inequalities,*
$$\{\mathbf{x} \in \mathbb{R}^q : A\mathbf{x} \leq \mathbf{b}\},$$
where A is a $p \times q$ matrix in $\mathbb{R}^{p \times q}$ and $\mathbf{b}$ is a vector in $\mathbb{R}^p$;
- *a* closed convex polytope *(which will be called a* polytope *for short) is defined as the convex hull of a finite set of points;*
- *if a polyhedron is bounded, then it is a polytope;*
- *for a polytope $\mathbf{P}$, we define $\text{vert}(\mathbf{P})$ as the set of vertices of $\mathbf{P}$;*
- *A* d-simplex *is a d-dimensional polytope which has exactly $d+1$ vertices. It is notated as Δ_d.*

Let $DAGs(N)$ be the set of all DAGs over N, and consider a class of graphs $\mathcal{G} \subseteq DAGs(N)$ that contains all graphs which we are interested in. We call the convex hull of $\{c_G : G \in \mathcal{G}\}$, $\mathbf{P}_{\mathcal{G}} = conv\{c_G : G \in \mathcal{G}\}$, the cim-polytope for $\mathcal{G}$. Note that it is obvious that $\text{vert}(\mathbf{P}_{\mathcal{G}}) = \{c_G : G \in \mathcal{G}\}$.

For fixed A and B in Definition 2.1, define $\mathbf{P}_{m,n} := \mathbf{P}_{\mathcal{G}_{m,n}}$. Proposition 2.7 implies that the dimension of $\mathbf{P}_{m,n}$ is at most $n \cdot (2^m - 1)$. We will show that the dimension of $\mathbf{P}_{m,n}$ is actually exactly $n \cdot (2^m - 1)$.

3. The cim-polytopes for diagnosis models.

3.1. Combinatorial description of edges on $\mathbf{P}_{m,n}$.

DEFINITION 3.1. *Consider a class of graphs $\mathcal{G} \subseteq DAGs(N)$. For all G, $H \in \mathcal{G}$, G and H are called* neighbors *if c_G and c_H form an edge in $\mathbf{P}_{\mathcal{G}}$, the cim-polytope for $\mathcal{G}$.*

LEMMA 3.2. *Fix m. Suppose G_1, $G_2 \in \mathcal{G}_{m,1}$ are arbitrary two distinct graphs in $\mathcal{G}_{m,1}$. Then G_1 and G_2 are neighbors, i.e., c_{G_1} and c_{G_2} form an edge in $\mathbf{P}_{m,1}$.*

Proof. Let $N = A \cup B$, where $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1\}$. We need to prove $\exists$ a cost vector w, such that $w \cdot c_{G_1} = w \cdot c_{G_2} > w \cdot c_{G_3}$, $\forall\ G_3 \in \mathcal{G}_{m,1}$ distinct from G_1 and G_2.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 12/15/15 to 160.36.192.7. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

By Remark 2.5, G_1 and G_2 are determined by $pa_{G_1}(b_1)$ and $pa_{G_2}(b_1)$, respectively. We will discuss by two scenarios of $pa_{G_1}(b_1)$ and $pa_{G_2}(b_1)$: one is a subset of the other, and neither one is a subset of the other.

(1) One is a subset of the other. Without loss of generality, suppose $pa_{G_1}(b_1) \subsetneq pa_{G_2}(b_1)$.

Define $A_1 = pa_{G_1}(b_1)$, $A_2 = pa_{G_2}(b_1)$, $A_{2\setminus 1} = pa_{G_2}(b_1)\setminus pa_{G_1}(b_1)$, and $A_{comp} = (pa_{G_2}(b_1))^c$ (i.e., the complement set of $pa_{G_2}(b_1)$). Note that $A_{2\setminus 1} \neq \emptyset$, and A_1 and A_{comp} can be $\emptyset$; A_1, $A_{2\setminus 1}$, and A_{comp} from a partition of N.

Consider a function $w : \mathcal{P}(N) \mapsto \mathbb{R}$, where $w(T) = 0$ if $|T| < 2$. Then similar to imsets, w can also be considered as a vector, and we assume that the permutations of coordinates in w and in characteristic imsets coincide.

(1.a) If $|A_{2\setminus 1}| > 1$, we define w as

$$
w(T) = \begin{cases}
c & \text{for} \quad T = a_i b_j, \ a_i \in A_1, \\
-c & \text{for} \quad T = a_i b_j, \ a_i \notin A_1, \\
|A_{2\setminus 1}| \cdot c & \text{for} \quad T = A_{2\setminus 1} \cup \{b_1\}, \\
0 & \text{for} \quad T \subset N, |T| > 2, \text{ and } T \neq A_{2\setminus 1} \cup \{b_1\},
\end{cases}
$$

where c is a positive number.
Then $\forall \ G_3 \in \mathcal{G}_{m,1}$, we have

$$
\begin{aligned}
w \cdot c_{G_3} &= |A_1 \cap pa_{G_3}(b_1)| \cdot c - |pa_{G_3}(b_1)\setminus A_1| \cdot c + |A_{2\setminus 1}| \cdot c \cdot c_{G_3}(A_{2\setminus 1} \cup \{b_1\}) \\
&= |A_1 \cap pa_{G_3}(b_1)| \cdot c - |pa_{G_3}(b_1) \cap A_{2\setminus 1}| \cdot c - |pa_{G_3}(b_1) \cap A_{comp}| \cdot c \\
&\quad + |A_{2\setminus 1}| \cdot c \cdot c_{G_3}(A_{2\setminus 1} \cup \{b_1\}).
\end{aligned}
$$

In this equation,
- $|A_1 \cap pa_{G_3}(b_1)| \cdot c \leq |A_1| \cdot c$, where "=" holds if and only if $A_1 \subset pa_{G_3}(b_1)$;
- $-|pa_{G_3}(b_1) \cap A_{2\setminus 1}| \cdot c + |A_{2\setminus 1}| \cdot c \cdot c_{G_3}(A_{2\setminus 1} \cup \{b_1\}) \leq 0$, where "=" holds if and only if $pa_{G_3}(b_1) \cap A_{2\setminus 1} = \emptyset$ or $A_{2\setminus 1}$;
- $-|pa_{G_3}(b_1) \cap A_{comp}| \cdot c \leq 0$, where "=" holds if and only if $pa_{G_3}(b_1) \cap A_{comp} = \emptyset$.

Therefore, $w \cdot c_{G_3} \leq |A_1| \cdot c$, where "=" holds if and only if $G_3 = G_1$ or G_2.

(1.b) If $|A_{2\setminus 1}| = 1$, we let $A_{2\setminus 1} = \{a_q\}$, and define w as

$$
w(T) = \begin{cases}
c & \text{for} \quad T = a_i b_j, \ a_i \in A_1, \\
-c & \text{for} \quad T = a_i b_j, \ a_i \notin A_2, \\
0 & \text{for} \quad T = a_q b_1, \\
0 & \text{for} \quad T \subset N, |T| > 2, \text{ and } T \neq A_{2\setminus 1} \cup \{b_1\},
\end{cases}
$$

where c is a positive number.
Then similar to (1.a), we can prove that $w \cdot c_{G_3} \leq |A_1| \cdot c$, where "=" holds if and only if $G_3 = G_1$ or G_2.

(2) Neither one is a subset of the other.

Define $A_1 = pa_{G_1}(b_1)$, $A_2 = pa_{G_2}(b_1)$, $A_{1\cap 2} = pa_{G_1}(b_1) \cap pa_{G_2}(b_1)$, $A_{1\setminus 2} = pa_{G_1}(b_1)\setminus pa_{G_2}(b_1)$, $A_{2\setminus 1} = pa_{G_2}(b_1)\setminus pa_{G_1}(b_1)$, $A_{1\cup 2} = pa_{G_1}(b_1) \cup pa_{G_2}(b_1)$, and $A_{comp} = (A_{1\cup 2})^c$. Note that $A_{1\setminus 2}$, $A_{2\setminus 1} \neq \emptyset$, $A_{1\cap 2}$, and A_{comp} can be $\emptyset$; $A_{1\cap 2}$, $A_{1\setminus 2}$, $A_{2\setminus 1}$, and A_{comp} is a partition of N.

Consider a function w similar to (1) that can also be considered as a vector with a proper permutation of coordinates.

(2.a) If $|A_{1\backslash 2}| > 1$ and $|A_{2\backslash 1}| > 1$, we define w as

$$w(T) = \begin{cases} c & \text{for} \quad T = a_i b_j, \ a_i \in A_{1\cap 2}, \\ -c & \text{for} \quad T = a_i b_j, \ a_i \notin A_{1\cap 2}, \\ -2c & \text{for} \quad T = A_{1\backslash 2} \cup A_{2\backslash 1} \cup \{b_1\}, \\ (|A_{1\backslash 2}| + 1) \cdot c & \text{for} \quad T = A_{1\backslash 2} \cup \{b_1\}, \\ (|A_{2\backslash 1}| + 1) \cdot c & \text{for} \quad T = A_{2\backslash 1} \cup \{b_1\}, \\ 0 & \text{for} \quad \text{other } T \subset N, |T| > 2, \end{cases}$$

where c is a positive number.
Then $\forall \, G_3 \in \mathcal{G}_{m,1}$, we have

$$\begin{aligned} w \cdot c_{G_3} = \ & |pa_{G_3}(b_1) \cap A_{1\cap 2}| \cdot c - |pa_{G_3}(b_1) \cap A_{1\backslash 2}| \cdot c \\ & - |pa_{G_3}(b_1) \cap A_{2\backslash 1}| \cdot c - |pa_{G_3}(b_1) \cap A_{comp}| \cdot c \\ & + (|A_{1\backslash 2}| + 1) \cdot c \cdot c_{G_3} \\ & \qquad \cdot (A_{1\backslash 2} \cup \{b_1\}) + (|A_{2\backslash 1}| + 1) \cdot c \cdot c_{G_3}(A_{2\backslash 1} \cup \{b_1\}) \\ & - 2c \cdot c_{G_3}(A_{1\backslash 2} \cup A_{2\backslash 1} \cup \{b_1\}) \\ = \ & |pa_{G_3}(b_1) \cap A_{1\cap 2}| \cdot c \\ & - |pa_{G_3}(b_1) \cap A_{1\backslash 2}| \cdot c + (|A_{1\backslash 2}| + 1) \cdot c \cdot c_{G_3}(A_{1\backslash 2} \cup \{b_1\}) \\ & - |pa_{G_3}(b_1) \cap A_{2\backslash 1}| \cdot c + (|A_{2\backslash 1}| + 1) \cdot c \cdot c_{G_3}(A_{2\backslash 1} \cup \{b_1\}) \\ & - 2c \cdot c_{G_3}(A_{1\backslash 2} \cup A_{2\backslash 1} \cup \{b_1\}) \\ & - |pa_{G_3}(b_1) \cap A_{comp}| \cdot c. \end{aligned}$$

In this equation,
- $|pa_{G_3}(b_1) \cap A_{1\cap 2}| \cdot c \le |A_{1\cap 2}| \cdot c$, where "=" holds if and only if $A_{1\cap 2} \subset pa_{G_3}(b_1)$;
- $-|pa_{G_3}(b_1) \cap A_{1\backslash 2}| \cdot c + (|A_{1\backslash 2}| + 1) \cdot c \cdot c_{G_3}(A_{1\backslash 2} \cup \{b_1\}) \le c$, where "=" holds if and only if $A_{1\backslash 2} \subset pa_{G_3}(b_1)$;
- $-|pa_{G_3}(b_1) \cap A_{2\backslash 1}| \cdot c + (|A_{2\backslash 1}| + 1) \cdot c \cdot c_{G_3}(A_{2\backslash 1} \cup \{b_1\}) \le c$, where "=" holds if and only if $A_{2\backslash 1} \subset pa_{G_3}(b_1)$;
- $-2c \cdot c_{G_3}(A_{1\backslash 2} \cup A_{2\backslash 1} \cup \{b_1\}) \le 0$, where "=" holds if and only if $(A_{1\backslash 2} \cup A_{2\backslash 1}) \nsubseteq pa_{G_3}(b_1)$;
- $-|pa_{G_3}(b_1) \cap A_{comp}| \cdot c \le 0$, where "=" holds if and only if $pa_{G_3}(b_1) \cap A_{comp} = \emptyset$.

The above conditions cannot be satisfied simultaneously, but notice that
- when $pa_{G_3}(b_1) = A_{1\cap 2}$, $w \cdot c_{G_3} = |A_{1\cap 2}| \cdot c + 0 + 0 + 0 + 0 = |A_{1\cap 2}| \cdot c$;
- when $pa_{G_3}(b_1) = A_1$, i.e., $G_3 = G_1$, $w \cdot c_{G_3} = |A_{1\cap 2}| \cdot c + c + 0 + 0 + 0 = (|A_{1\cap 2}| + 1) \cdot c$;
- when $pa_{G_3}(b_1) = A_2$, i.e., $G_3 = G_2$, $w \cdot c_{G_3} = |A_{1\cap 2}| \cdot c + 0 + c + 0 + 0 = (|A_{1\cap 2}| + 1) \cdot c$;
- when $pa_{G_3}(b_1) = A_{1\cup 2}$, $w \cdot c_{G_3} = |A_{1\cap 2}| \cdot c + c + c - 2c + 0 = |A_{1\cap 2}| \cdot c$.

Now it is obvious that $w \cdot c_{G_3} \le (|A_{1\cap 2}| + 1) \cdot c$, where "=" holds if and only if $G_3 = G_1$ or G_2.

(2.b) If only one of $|A_{1\backslash 2}|$ and $|A_{2\backslash 1}|$ is 1, suppose $|A_{1\backslash 2}| = 1$ and $|A_{2\backslash 1}| > 1$. We define w as

$$w(T) = \begin{cases} c & \text{for} \quad T = a_i b_j, \ a_i \in A_1, \\ -c & \text{for} \quad T = a_i b_j, \ a_i \notin A_1, \\ -2c & \text{for} \quad T = A_{1\backslash 2} \cup A_{2\backslash 1} \cup \{b_1\}, \\ (|A_{2\backslash 1}| + 1) \cdot c & \text{for} \quad T = A_{2\backslash 1} \cup \{b_1\}, \\ 0 & \text{for} \quad \text{other } T \subset N, |T| > 2, \end{cases}$$

where c is a positive number.

Then similar to (2.a), we can prove that $w \cdot c_{G_3} \leq (|A_{1 \cap 2}| + 1) \cdot c$, where "=" holds if and only if $G_3 = G_1$ or G_2.

(2.c) If $|A_{1 \setminus 2}| = |A_{2 \setminus 1}| = 1$, we define w as

$$
w(T) = \begin{cases}
c & \text{for} & T = a_i b_j, \ a_i \in A_{1 \cup 2}, \\
-c & \text{for} & T = a_i b_j, \ a_i \notin A_{1 \cup 2}, \\
-2c & \text{for} & T = A_{1 \setminus 2} \cup A_{2 \setminus 1} \cup \{b_1\}, \\
0 & \text{for} & \text{other } T \subset N, |T| > 2,
\end{cases}
$$

where c is a positive number.

Then similar to (2.a), we can prove that $w \cdot c_{G_3} \leq (|A_{1 \cap 2}| + 1) \cdot c$, where "=" holds if and only if $G_3 = G_1$ or G_2. $\square$

THEOREM 3.3. *Fix m and n. Two graphs, G_1, $G_2 \in \mathcal{G}_{m,n}$ are neighbors if and only if $\exists \, b_i \in B$ such that $pa_{G_1}(b_i) \neq pa_{G_2}(b_i)$ and $pa_{G_1}(b_j) = pa_{G_2}(b_j) \, \forall \, b_j \in B$ and $b_j \neq b_i$, i.e., all nodes but one have exactly the same parent sets in G_1 and G_2.*

Proof. We will prove "if" and "only if" separately.

(1) Proof of "if" part.

Suppose G_1, $G_2 \in \mathcal{G}_{m,n}$, and there exists $b_i \in B$ such that $pa_{G_1}(b_i) \neq pa_{G_2}(b_i)$ and $pa_{G_1}(b_j) = pa_{G_2}(b_j) \, \forall \, b_j \in B$, $b_j \neq b_i$. We need to prove G_1 and G_2 are neighbors.

Consider an arbitrary graph $G_3 \in \mathcal{G}_{m,n}$. We need to prove $\exists$ a cost vector w such that $w \cdot c_{G_1} = w \cdot c_{G_2} \geq w \cdot c_{G_3}$, where "=" holds if and only if $G_3 = G_1$ or G_2.

Define the following graphs (a graphical example will be given in Remark 3.4):

- G_1', G_2', $G_3' \in \mathcal{G}_{m,1}$ with symptom $B_{m,1} = \{b_i\}$ such that $pa_{G_1'}(b_i) = pa_{G_1}(b_i)$, $pa_{G_2'}(b_i) = pa_{G_2}(b_i)$, and $pa_{G_3'}(b_i) = pa_{G_3}(b_i)$;
- G_0, $G_3'' \in \mathcal{G}_{m,(n-1)}$ with symptoms $B_{m,(n-1)} = B \setminus \{b_i\}$ such that $pa_{G_0}(b_j) = pa_{G_1}(b_j) = pa_{G_2}(b_j)$, and $pa_{G_3''}(b_j) = pa_{G_3}(b_j) \, \forall \, b_j \in B_{m,(n-1)}$.

By Remark 2.5, with a proper permutation of coordinates, we can write the characteristic imsets of G_1, G_2, and G_3 in the form

$$(3.1) \qquad c_{G_1} = (c_{G_1'}, c_{G_0}),$$

$$(3.2) \qquad c_{G_2} = (c_{G_2'}, c_{G_0}),$$

$$(3.3) \qquad c_{G_3} = (c_{G_3'}, c_{G_3''}).$$

- By Lemma 3.2, G_1' and G_2' are neighbors, i.e., $\exists$ a cost vector w_1 such that $w_1 \cdot c_{G_1'} = w_1 \cdot c_{G_2'} \geq w_1 \cdot c_{G_3'} \, \forall \, G_3' \in \mathcal{G}_{m,1}$, where "=" holds if and only if $G_3' = G_1'$ or G_2'.
- Since $c_{G_0} \in vert(\mathbf{P}_{\mathcal{G}_{m,(n-1)},c})$, $\exists$ a cost vector w_2 such that $w_2 \cdot c_{G_0} \geq w_2 \cdot c_{G_3''} \, \forall \, G_3'' \in \mathcal{G}_{m,(n-1)}$, where "=" holds if and only if $G_3'' = G_0$.

Let $w = (w_1 \ w_2)$. We have

$$
\begin{aligned}
w \cdot c_{G_1} &= w_1 \cdot c_{G_1'} + w_2 \cdot c_{G_0} \\
&= w_1 \cdot c_{G_2'} + w_2 \cdot c_{G_0} = w \cdot c_{G_2} \\
&\geq w_1 \cdot c_{G_3'} + w_2 \cdot c_{G_3''} = w \cdot c_{G_3},
\end{aligned}
$$

where "=" holds if and only if (i) $G_3' = G_1'$ or G_2', and (ii) $G_3'' = G_0$, i.e., $G_3 = G_1$ or G_2.

(2) Proof of "only if" part.

Suppose G_1, $G_2 \in \mathcal{G}_{m,n}$ are neighbors, i.e., $\exists$ a cost vector w such that $w \cdot c_{G_1} = w \cdot c_{G_2} > w \cdot c_G \ \forall \ G \in \mathcal{G}_{m,n}$, $G \neq G_1$, G_2. We are going to prove this part by contradiction.

Suppose $\exists \ b_i, b_j \in B$ distinct, $pa_{G_1}(b_i) \neq pa_{G_2}(b_i)$, and $pa_{G_1}(b_j) \neq pa_{G_2}(b_j)$. Define the following graphs (a graphical example will be given in Remark 3.4):

- G_1', $G_2' \in \mathcal{G}_{m,1}$ with symptom $B_{m,1} = \{b_i\}$ such that $pa_{G_1'}(b_i) = pa_{G_1}(b_i)$ and $pa_{G_2'}(b_i) = pa_{G_2}(b_i)$;
- G_1'', $G_2'' \in \mathcal{G}_{m,1}$ with symptom $B_{m,1} = \{b_j\}$ such that $pa_{G_1''}(b_j) = pa_{G_1}(b_j)$ and $pa_{G_2''}(b_j) = pa_{G_2}(b_j)$;
- G_1''', $G_2''' \in \mathcal{G}_{m,(n-2)}$ with symptoms $B_{m,(n-2)} = B\backslash\{b_i, b_j\}$ such that $pa_{G_1'''}(b_k) = pa_{G_1}(b_k)$ and $pa_{G_2'''}(b_k) = pa_{G_2}(b_k) \ \forall \ b_k \in B_{m,(n-2)}$;
- $G_3 \in \mathcal{G}_{m,n}$ is all the same as G_1 but $pa_{G_3}(b_i) = pa_{G_2}(b_i)$;
- $G_4 \in \mathcal{G}_{m,n}$ is all the same as G_1 but $pa_{G_4}(b_j) = pa_{G_2}(b_j)$;
- $G_5 \in \mathcal{G}_{m,n}$ is all the same as G_2 but $pa_{G_5}(b_i) = pa_{G_1}(b_i)$ and $pa_{G_5}(b_j) = pa_{G_1}(b_j)$; notice that G_5 might be same as G_1.

Similarly to (1), with a proper permutation of coordinates, we can write the characteristic imsets of G_1 through G_5 in the following form:

$$c_{G_1} = (c_{G_1'}, c_{G_1''}, c_{G_1'''}),$$
$$c_{G_2} = (c_{G_2'}, c_{G_2''}, c_{G_2'''}),$$
$$c_{G_3} = (c_{G_2'}, c_{G_1''}, c_{G_1'''}),$$
$$c_{G_4} = (c_{G_1'}, c_{G_2''}, c_{G_1'''}),$$
$$c_{G_5} = (c_{G_1'}, c_{G_1''}, c_{G_2'''}).$$

With the same permutation of coordinates, w can be written as $w = (w_1, w_2, w_3)$. Thus we have

- $G_3 \neq G_1$ or G_2, which implies

$$w \cdot c_{G_1} = w_1 \cdot c_{G_1'} + w_2 \cdot c_{G_1''} + w_3 \cdot c_{G_1'''}$$
$$> w \cdot c_{G_3} = w_1 \cdot c_{G_2'} + w_2 \cdot c_{G_1''} + w_3 \cdot c_{G_1'''}$$
$$\Longrightarrow w_1 \cdot c_{G_1'} > w_1 \cdot c_{G_2'};$$

- similarly $G_4 \neq G_1$ or G_2 implies $w_2 \cdot c_{G_1''} > w_2 \cdot c_{G_2''}$.

There is a contradiction:

$$w \cdot c_{G_2} = w_1 \cdot c_{G_2'} + w_2 \cdot c_{G_2''} + w_3 \cdot c_{G_2'''}$$
$$< w_1 \cdot c_{G_1'} + w_2 \cdot c_{G_1''} + w_3 \cdot c_{G_2'''} = w \cdot c_{G_5}$$
$$\Longrightarrow w \cdot c_{G_2} < w \cdot c_{G_5}. \qquad \square$$

Remark 3.4. Two graphical examples will be given for a more intuitive view of the proof of Theorem 3.3.

- Part (1), the proof of the "if" statement. In Figure 2, $m = 4$, $n = 3$, and $b_i = b_1$.
- Part (2), the proof of the "only if" statement. In Figure 3, $m = 4$, $n = 3$, $b_i = b_1$, and $b_j = b_2$.

3.2. $\mathbf{P}_{m,n}$ is a direct product of simplices.

COROLLARY 3.5. *Fix m and n. For an arbitrary $G \in \mathcal{G}_{m,n}$, G has $n \cdot (2^m - 1)$ many neighbors.*
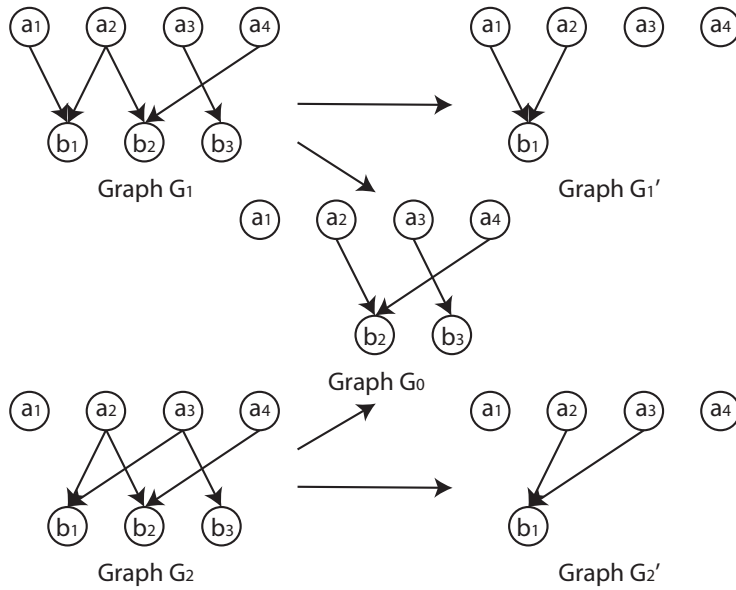
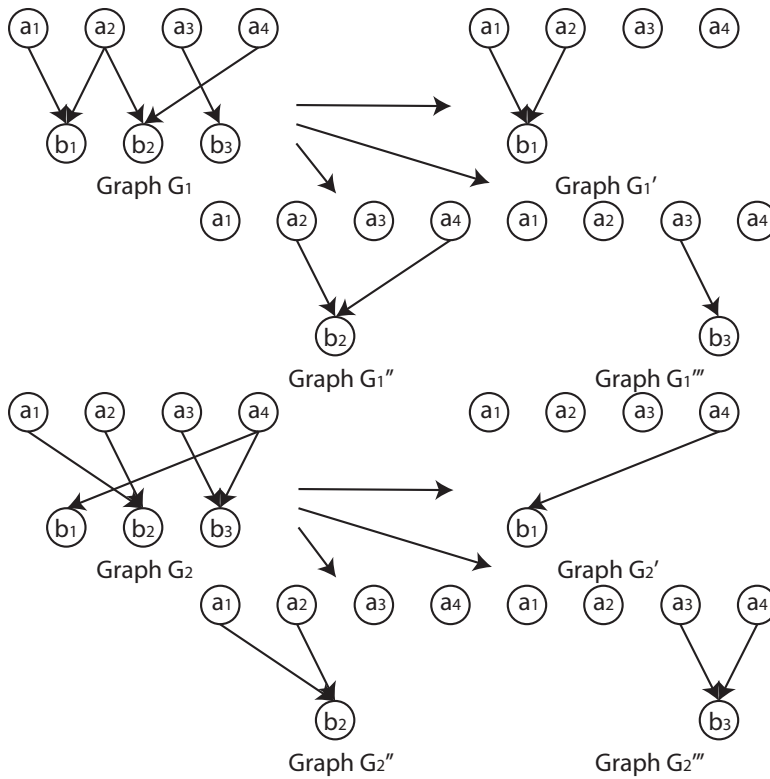FIG. 2. *An example for the proof of Theorem* 3.3, *part* (1).

FIG. 3. *An example for the proof of Theorem* 3.3 (2).

Proof. By Theorem 3.3, $\forall H \in \mathcal{G}_{m,n}$, G and H are neighbors if and only if $\exists\, b_k \in B$ such that $pa_G(b_k) \neq pa_H(b_k)$ and $pa_G(b_j) = pa_H(b_j)\ \forall\, b_j \in B$ and $b_j \neq b_k$. So for every $b_i \in B$, we let $H \in \mathcal{G}_{m,n}$ be a graph distinct from G but having the same subgraph for $N \backslash \{b_i\}$ as G; then H is a neighbor of G and by Proposition 2.6 there are $2^m - 1$ different choices of Hs. Since these neighbors are distinct, the total number of neighbors for G is $n \cdot (2^m - 1)$. $\quad\square$

Remark 3.6. Theorem 3.5 implies that every vertex of $\mathbf{P}_{m,1}$ has $(2^m - 1)$ neighbors. Since $|vert(\mathbf{P}_{m,1})| = 2^m$ (by Proposition 2.6), $\mathbf{P}_{m,1}$ is a simplex with dimension $(2^m - 1)$, i.e., $\mathbf{P}_{m,1} = \Delta_{2^m-1}$.

THEOREM 3.7. $\mathbf{P}_{m,n}$ *is the direct product of n many Δ_{2^m-1}, i.e.,*

$$\mathbf{P}_{m,n} = \underbrace{\Delta_{2^m-1} \times \Delta_{2^m-1} \times \cdots \times \Delta_{2^m-1}}_{n\ many}.$$

And the i_{th} simplex is $\mathbf{P}_{m,1}$ with the same diseases A and only one symptom $\{b_i\}$.

Proof. Fix m, we are going to prove the equality by induction on n.

- $n = 1$. See Remark 3.6.
- Fix $q \in \mathbb{Z}^+$. Suppose the equality holds for $\mathbf{P}_{m,n}\ \forall\, n < q$; then we need to prove that it also holds for $\mathbf{P}_{m,q}$. Recall that for $\mathcal{G}_{m,q}$, the symptoms are $B = \{b_1, b_2, \ldots, b_q\}$.
  First, we need to prove that $\mathbf{P}_{m,q} \subseteq \mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1}$.
  Similarly to the proof of Theorem 3.3, $\forall\, G \in \mathcal{G}_{m,q}$, we define graphs:
    – $G' \in \mathcal{G}_{m,(q-1)}$ is the subgraph of G for $A \cup (B \backslash \{b_q\})$, which implies $c_{G'} \in \mathbf{P}_{m,q-1}$;
    – $G'' \in \mathcal{G}_{m,1}$ is the subgraph of G for $A \cup \{b_q\}$, which implies $c_{G''} \in \mathbf{P}_{m,1}$.
  With a proper permutation of coordinates, we can write c_G in the form

$$c_G = (c_{G'}, c_{G''}).$$

Recall that $vert(\mathbf{P}_{m,q}) = \{c_G : G \in \mathcal{G}_{m,q}\}$, so $\forall x \in \mathbf{P}_{m,q}$, with the same permutation of coordinates, we have

$$(3.4) \qquad x = \sum_{G \in \mathcal{G}_{m,q}} \alpha_G c_G = \left(\sum_{G \in \mathcal{G}_{m,q}} \alpha_G c_{G'}\ ,\ \sum_{G \in \mathcal{G}_{m,q}} \alpha_G c_{G''} \right),$$

where $0 \leq \alpha_G \leq 1\ \forall G \in \mathcal{G}_{m,q}$ and $\sum_{G \in \mathcal{G}_{m,q}} \alpha_G = 1$. Note that $\sum_{G \in \mathcal{G}_{m,q}} \alpha_G c_{G'} \in \mathbf{P}_{m,q-1}$ and $\sum_{G \in \mathcal{G}_{m,q}} \alpha_G c_{G''} \in \mathbf{P}_{m,1}$. Equation (3.4) implies $x \in \mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1}$. Hence

$$\mathbf{P}_{m,q} \subseteq \mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1}.$$

Second, we need to prove that $\mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1} \subseteq \mathbf{P}_{m,q}$.
Let $\mathcal{G}_{m,q-1}$ have symptoms $B_{m,(q-1)} = B \backslash \{b_q\}$ and $\mathcal{G}_{m,1}$ have symptom $B_{m,1} = \{b_q\}$. Then $\forall\, G' \in \mathcal{G}_{m,(q-1)}$ and $G'' \in \mathcal{G}_{m,1}$, we can define $G \in \mathcal{G}_{m,q}$ such that $pa_G(b_i) = pa_{G'}(b_i)\ \forall\, b_i \in B_{m,(q-1)}$, and $pa_G(b_q) = pa_{G''}(b_q)$. c_G has the form $c_G = (c_{G'}, c_{G''})$.

Then $\forall\, x \in \mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1}$, x can be written as

$$
x = \left(\sum_{G' \in \mathcal{G}_{m,q-1}} \beta_{G'} c_{G'} \ , \ \sum_{G'' \in \mathcal{G}_{m,1}} \gamma_{G''} c_{G''} \right)
$$
$$
= \sum_{G' \in \mathcal{G}_{m,q-1}} \sum_{G'' \in \mathcal{G}_{m,1}} \beta_{G'} \gamma_{G''} \left(c_{G'} \ , \ c_{G''} \right)
$$
$$
= \sum_{G' \in \mathcal{G}_{m,q-1}} \sum_{G'' \in \mathcal{G}_{m,1}} (\beta_{G'} \gamma_{G''}) \ c_G \ ,
$$

where $0 \le \beta_{G'}, \gamma_{G''} \le 1$, $\forall G' \in \mathcal{G}_{m,q-1}$, $\forall G'' \in \mathcal{G}_{m,1}$, and $\sum_{G' \in \mathcal{G}_{m,q-1}} \beta_{G'} = 1$, $\sum_{G'' \in \mathcal{G}_{m,1}} \gamma_{G''} = 1$. Note that

$$
\sum_{G' \in \mathcal{G}_{m,q-1}} \sum_{G'' \in \mathcal{G}_{m,1}} (\beta_{G'} \gamma_{G''}) = \sum_{G' \in \mathcal{G}_{m,q-1}} \beta_{G'} \left(\sum_{G'' \in \mathcal{G}_{m,1}} \gamma_{G''} \right)
$$
$$
= \sum_{G' \in \mathcal{G}_{m,q-1}} \beta_{G'} = 1,
$$

which leads to $x \in \mathbf{P}_{m,q}$. Hence

$$
\mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1} \subseteq \mathbf{P}_{m,q}.
$$

Therefore,

$$
\mathbf{P}_{m,q} = \mathbf{P}_{m,q-1} \times \mathbf{P}_{m,1} = \underbrace{\Delta_{2^m-1} \times \cdots \times \Delta_{2^m-1}}_{q\text{-1 many}} \times \Delta_{2^m-1}
$$
$$
= \underbrace{\Delta_{2^m-1} \times \cdots \times \Delta_{2^m-1}}_{q\text{ many}}. \qquad \square
$$

Theorem 3.7 implies that $\mathbf{P}_{m,n}$ is a simple polytope with dimension $n \cdot (2^m - 1)$. In section 5, we will give another proof which uses linear algebra to show that $\mathbf{P}_{m,n}$ is simple and obtain its dimension.

3.3. Expression of facets of $\mathbf{P}_{m,n}$. Based on Theorem 3.7, we are going to show the expression of facets of $\mathbf{P}_{m,n}$ using the following lemma.

LEMMA 3.8 (see [18]). *Suppose $\mathbf{P}$ is the direct product of simplices $\Delta_{\alpha_1}, \ldots, \Delta_{\alpha_k}$. Then every facet of $\mathbf{P}$ has the form $\Delta_{\alpha_1} \times \cdots \times \Delta_{\alpha_{i-1}} \times F_{\alpha_i} \times \Delta_{\alpha_{i+1}} \times \cdots \times \Delta_{\alpha_k}$, where F_{α_i} is a facet of Δ_{α_i}.*

Remark 3.9. Lemma 3.8 implies that in order to study the facets of a direct product of simplices, we can simply study the facets of each simplex. As by Theorem 3.7 $\mathbf{P}_{m,n}$ is a direct product of n many $\mathbf{P}_{m,1}$, and our problem is simplified to studying the facets of $\mathbf{P}_{m,1}$. Thus we assume $B = \{b_1\}$ in the following content of this section.

Assume $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1\}$. By Proposition 2.7, the vertices of $\mathbf{P}_{m,1}$ have at most $2^m - 1$ many nonzero coordinates. We define the indeterminates, i.e., variables, $\{x_s,\ s \subseteq A,\ s \ne \emptyset\}$, where one indeterminate x_s for each coordinate $c_G(s \cup \{b_1\})$ in the characteristic imset c_G, $G \in \mathcal{G}_{m,1}$. Define the vector of indeterminates $x = \{x_s,\ s \subseteq A,\ s \ne \emptyset\}$. Suppose $A_m x \le b_m$ is the system of inequalities that defines $\mathbf{P}_{m,1}$. We can define a $2^m \times 2^m$ matrix: $D_m = [b_m| - A_m]$. Denote the elements in D_m by $(d_{st})_{s \subseteq A, t \subseteq A}$ so that we can rewrite the system of inequalities as

$d_{s\emptyset} + \sum_{t \subseteq A, t \neq \emptyset} d_{st} x_t \geq 0$, $s \subseteq A$. Then we have the expression of 2^m facets of $\mathbf{P}_{m,1}$ as follows:

$$F_s = \mathbf{P}_{m,1} \cap \Big\{ x : d_{s\emptyset} + \sum_{t \subseteq A, t \neq \emptyset} d_{st} x_t = 0 \Big\}, \ s \subseteq A,$$

where the elements d_{st}, s, $t \subseteq A$, can be obtained using Theorem 3.10.

THEOREM 3.10. *The elements in matrix D_m satisfies*
- $d_{st} \neq 0$ *if and only if* $s \subseteq t$;
- *if* $s \subseteq t$, *then* $d_{st} = (-1)^{|t|-|s|}$.

This implies that $\mathbf{P}_{m,1}$ has 2^m facets:

$$F_s = \mathbf{P}_{m,1} \cap \Big\{ x : d_{s\emptyset} + \sum_{t \subseteq A, t \neq \emptyset} d_{st} x_t = 0 \Big\}, \ s \subseteq A.$$

What's more, $\forall \, s \subseteq A$, $vert(\mathbf{P}_{m,1}) \backslash \{c_{G_s}\} \subset F_s$, where $pa_{G_s}(b_1) = s$.

Proof. For convenience, let $x_\emptyset \equiv 1$. Then $\forall s \subseteq A$, let $d_{s\cdot} = (d_{st})_{t \subseteq A}$ be the corresponding row of D_m and G_s be the graph in $\mathcal{G}_{m,1}$ such that $pa_{G_s}(b_1) = s$. Now we can rewrite the system of inequalities as

$$\sum_{t \subseteq A} d_{st} x_t = d_{s\cdot}(1 \ x)^T \geq 0 \ \forall \, s \subseteq A.$$

We are going to prove that $\forall s \subseteq A$, we can find $2^m - 1$ vertices on F_s that are linearly independent, and this implies that F_s is a facet of $\mathbf{P}_{m,1}$. In fact, we will prove that $\{c_{G_{s'}}, \ s' \subseteq A, \ s' \neq s\} \subset F_s$ and $c_{G_s} \notin F_s$, i.e., $d_{s\cdot}(1 \ c_{G_{s'}})^T = 0 \ \forall \, s' \subseteq A, \ s' \neq s$, and $d_{s\cdot}(1 \ c_{G_s})^T > 0$.

Notice that $\forall t \subseteq A$, $c_{G_{s'}}(t \cup \{b_1\}) \neq 0$ if and only if $t \subseteq pa_{c_{G_{s'}}}(b_1) = s'$, and $d_{st} \neq 0$ if and only if $s \subseteq t$. So,

$$d_{s\cdot}(1 \ c_{G_{s'}})^T = d_{s\emptyset} + \sum_{t \subseteq A, \ t \neq \emptyset} d_{st} c_{G_{s'}}(t \cup \{b_1\}) = d_{s\emptyset} + \sum_{s \subseteq t \subseteq s', \ t \neq \emptyset} d_{st} = \sum_{s \subseteq t \subseteq s'} d_{st}.$$

Therefore, we have
- if $s = s'$, then $d_{s\cdot}(1 \ c_{G_{s'}})^T = d_{ss} = 1 > 0$;
- if $s \subsetneq s'$, then $d_{s\cdot}(1 \ c_{G_{s'}})^T = \sum_{s \subseteq t \subseteq s'} (-1)^{|t|-|s|} = \sum_{t' \subseteq s' \backslash s} (-1)^{|t'|} = 0$;
- if $s \not\subseteq s'$, then $d_{s\cdot}(1 \ c_{G_{s'}})^T = 0$. □

Example 3.11 (facets of $\mathbf{P}_{2,1}$). The notation is adopted from Theorem 3.10. Fix $m = 2$ and $n = 1$.
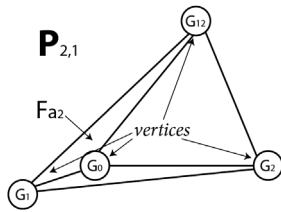
The matrix $D_2 = [b_2| - A_2]$:

$$\begin{pmatrix} c_{G_0} \\ c_{G_1} \\ c_{G_2} \\ c_{G_{12}} \end{pmatrix} = \begin{array}{cccc} T & a_1 b_1 & a_2 b_1 & a_1 a_2 b_1 \\ \left(\begin{array}{ccc} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{array} \right) \end{array}$$

All characteristic imsets are given as a matrix:

The system of inequalities that defines $\mathbf{P}_{2,1}$:

$$
D_2 = \begin{array}{c} s\backslash t \\ \emptyset \\ a_1 \\ a_2 \\ a_1a_2 \end{array} \begin{array}{cccc} \emptyset & a_1 & a_2 & a_1a_2 \\ \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}
$$

$$
\begin{array}{c} s\backslash t \\ \emptyset \\ a_1 \\ a_2 \\ a_1a_2 \end{array} \begin{array}{cccc} \emptyset & a_1 & a_2 & a_1a_2 \\ \begin{pmatrix} 1 & -x_{a_1} & -x_{a_2} & +x_{a_1a_2} \\ & x_{a_1} & & -x_{a_1a_2} \\ & & x_{a_2} & -x_{a_1a_2} \\ & & & x_{a_1a_2} \end{pmatrix} \end{array} \begin{array}{c} \geq 0 \\ \geq 0 \\ \geq 0 \\ \geq 0 \end{array}
$$

Vertices c_{G_0}, c_{G_1}, and $c_{G_{12}}$ are in the facet F_{a_2}, while c_{G_2} is not.

4. The cim-polytopes for BNs. The results in section 3 are limited to diagnosis models. In this section, we will generalize the results to all BNs with the same underlying order.

4.1. Underlying ordering of DAGs. For a set of random variables $N = \{a_1, \ldots, a_n\}$, where now n is the total number of nodes in N. For all $G \in DAGs(N)$, there exists an *underlying ordering* over N, $[n]_G = (a_{[1]}, \ldots, a_{[n]})$, such that if $a_{[i]} \to a_{[j]}$ in G, then $i < j$. We are are now interested in the class of graphs which share a specific underlying ordering $[n]$, i.e., $\mathcal{G}_{[n]} = \{G \in DAGs(N) : [n]_G = [n]\}$, and its cim-polytope $\mathbf{P}_{[n]} = \mathbf{P}_{\mathcal{G}_{[n]}}$.

Example 4.1 (underlying ordering of graphs). Let $N = \{a_1, a_2, a_3\}$. Consider an ordering over N, $[n] = (a_2, a_1, a_3)$, i.e., $a_{[1]} = a_2$, $a_{[2]} = a_1$, and $a_{[3]} = a_3$. Then $\forall G \in \mathcal{G}_{[n]}$, the only type of directed edges allowed in G are $a_{[i]} \to a_{[j]}$, where $i < j$. For instance, $a_2 \to a_1$ is allowed, while $a_1 \to a_2$ is not. Thus graph G_1 in Figure 4(a) and graph G_2 in Figure 4(b) are both in $\mathcal{G}_{[n]}$. Graph G_3 in Figure 4(c) is not in $\mathcal{G}_{[n]}$ since it has arrow $a_1 \to a_2$, and the underlying ordering for G_3, i.e., $[n]_{G_3}$, can either be (a_1, a_2, a_3) or (a_1, a_3, a_2).

Remark 4.2. For a specific ordering $[n]$ and an arbitrary $G \in \mathcal{G}_{[n]}$, we have the following proposition that is similar to Proposition 2.4:

- For all $T \subseteq N$, $|T| = k \geq 2$, we can order the elements in T according to $[n]$ and write T in the form of $a_{[i_1]}a_{[i_2]} \ldots a_{[i_k]}$, where $i_1 < i_2 < \cdots < i_k$. Then $c_G(T) = \prod_{s=i_1,\ldots,i_{k-1}} c_G(a_{[s]}a_{[i_k]})$. This property means that the whole c_G is determined by $\binom{n}{2}$ coordinates, $\{c_G(a_{[i]}a_{[j]}), i < j\}$, which can also be interpreted as the existence of the directed edges $a_{[i]} \to a_{[j]}$, $i < j$.

Another way to see this property is that $\forall G \in \mathcal{G}_{[n]}$, G can be determined by $pa_G(a_{[i]})$, $i = 2, \ldots, n$, since $pa_G(a_{[1]}) = \emptyset$. Similarly to Remark 2.5, we can consider a permutation of coordinates in c_G that corresponds to a permutation of T; then these
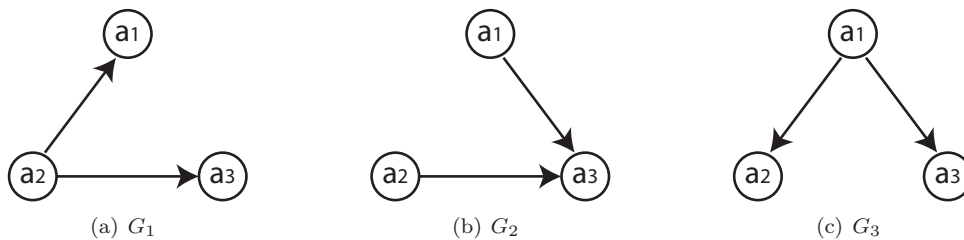

(a) G_1 (b) G_2 (c) G_3

Fig. 4. *Three graphs to illustrate the underlying ordering of graphs.*

coordinates can be broken into $n-1$ parts:

$$\underline{(12)}, \underbrace{(13), (23), (123)}, \underline{(14), (24), (34), \dots, (1234)}, \dots,$$

$$\underline{(1n), (2n), \dots, ((n-1)n), \dots, (12\dots n)},$$

where $(i_1 \dots i_k)$ stands for $T = a_{[i_1]}a_{[i_2]}\dots a_{[i_k]}$, $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$. The kth part of the coordinations $\{c_G(T): a_{[j]} \notin T \ \forall j > k\}$ only depends on $pa_G(a_{[k]})$, and different parts are completely irrelevant in the sense that $pa_G(a_{[k]})$, $a_{[k]} \in N$, can be decided separately.

4.2. Structure, edges and facets of $\mathbf{P}_{[n]}$.

THEOREM 4.3. *Suppose $n \geq 2$. $\mathbf{P}_{[n]}$ is a direct product of a sequence of simplices:*

$$\mathbf{P}_{[n]} = \underbrace{\Delta_{2^1-1} \times \Delta_{2^2-1} \times \cdots \times \Delta_{2^{n-1}-1}}_{n-1 \ simplices},$$

where the i_{th} simplex Δ_{2^i-1} is the same as the cim-polytope for diagnosis models, $\mathbf{P}_{i,1}$, with diseases $A = \{a_{[1]}, \dots, a_{[i]}\}$ and one symptom $\{a_{[i+1]}\}$.

Proof. The proof is similar to Theorem 3.7. Another version of the proof can also be found in [8, Theorem 3.1.1]. □

Remark 4.4. Two immediate results from Theorem 4.3 are that
- the dimension of $\mathbf{P}_{[n]}$ is $2^n - (n+1)$, and it is a simple polytope;
- the facets of $\mathbf{P}_{[n]}$ can be obtained by Lemma 3.8 and Theorem 3.10.

Remark 4.5. Note that the equality in Theorem 4.3 is actually $\mathbf{P}_{[n]} = \Delta_{2^0-1} \times \Delta_{2^1-1} \times \Delta_{2^2-1} \times \cdots \times \Delta_{2^{n-1}-1}$, where Δ_{2^0-1} is omitted as it has dimension 0 (a point). Theorem 4.3 and its proof also imply that $\forall x \in \mathbf{P}_{[n]}$, $x \in vert(\mathbf{P}_{[n]})$ if and only if with the permutation of coordinates in Remark 4.2, x can be written in the form of $x = (v_1, v_2, \dots, v_{n-1})$, where v_i is the vertex of Δ_{2^i-1}, $i = 1, \dots, n-1$. Suppose $x = c_G$, $G \in \mathcal{G}_{[n]}$; then $v_i = c_{G_i}$, where G_i is in $\mathcal{G}_{i,1}$ with diseases $N_{[i]}$ and symptom $a_{[i+1]}$, $i = 1, \dots, n-1$, and $pa_{G_i}(a_{[i+1]}) = pa_G(a_{[i+1]})$.

The following theorem will be stated in two forms which are equivalent by Theorem 4.3 and Lemma 3.2.

THEOREM 4.6. *Fix an underlying ordering $[n]$ over N.*
- *(From the viewpoint of graph theory.) Two graphs, G_1, $G_2 \in \mathcal{G}_{[n]}$, are neighbors in $\mathcal{G}_{[n]}$ if and only if $\exists\, a_{[i]} \in N$ such that $pa_{G_1}(a_{[i]}) \neq pa_{G_2}(a_{[i]})$ and $pa_{G_1}(a_{[j]}) = pa_{G_2}(a_{[j]}) \ \forall\, a_{[j]} \in N$ and $a_{[j]} \neq a_{[i]}$, i.e., all nodes but one have exactly the same parent sets in both G_1 and G_2.*
- *(From the viewpoint of polyhedral geometry.) For all $\mathbf{x} \in \mathbf{P}_{[n]}$, $\mathbf{x}$ is on an edge of $\mathbf{P}_{[n]}$ if and only if with the permutation of coordinates shown in Remark 4.2 $\mathbf{x}$ can be written in the form of $\mathbf{x} = (v_1, \dots, v_{i-1}, e_i, v_{i+1}, \dots, v_{n-1})$, where e_i belongs to an edge on Δ_{2^i-1}, $i \in \{1, \dots, n-1\}$, and $v_j \in vert(\Delta_{2^j-1})$, $j \in \{1, \dots, n-1\}\backslash\{i\}$.*

Proof. The proof from the view of graph theory is very similar to the proof of Theorem 3.3, and the proof from the view of polyhedral geometry is equivalent to proving "$\exists$ vertices v^1, $v^2 \in \mathbf{P}_{[n]}$ such that $\mathbf{x} = \beta v^1 + (1-\beta)v^2$, where $0 \leq \beta \leq 1$, and v^1, v^2 form an edge in $\mathbf{P}_{[n]}$" if and only if "$\mathbf{x}$ can be written in the form of $\mathbf{x} = (v_1, \dots, v_{i-1}, e_i, v_{i+1}, \dots, v_{n-1})$, $i \in \{1, \dots, n-1\}$." Details are omitted [17]. □

4.3. Graphs with forbidden (or fixed) edges. Fix an underlying ordering of nodes $[n]$ and consider $\mathcal{G}_{[n]}$. When a specific set of directed edges is forbidden in $\mathcal{G}_{[n]}$, we can define sets of nodes $\Omega = \{\Omega_i^0, i = 2, \dots, n\} \cup \{\Omega_i^1, i = 2, \dots, n\}$ such that

$\Omega_i^0 \subseteq \Omega_i^1 \subseteq \{a_{[1]}, \ldots, a_{[i-1]}\}$, and the class of graphs we are interested in becomes $\mathcal{G}_{[n],\Omega} = \{G \in DAGs(N) : [n]_G = [n], \ \Omega_i^0 \subseteq pa_G(a_{[i]}) \subseteq \Omega_i^1, \ i = 2, \ldots, n\}$, i.e., edges $\{a_{[j]} \to a_{[i]} : a_{[j]} \in \Omega_i^0, \ i = 2, \ldots, n\}$ are fixed edges, and edges $\{a_{[j]} \to a_{[i]} : a_{[j]} \in \{a_{[1]}, \ldots, a_{[i-1]}\}\setminus\Omega_i^1, \ i = 2, \ldots, n\}$ are forbidden edges. The cim-polytope for $\mathcal{G}_{[n],\Omega}$ is $\mathbf{P}_{\mathcal{G}_{[n],\Omega}}$. Using a similar strategy, we are able to show that $\mathbf{P}_{\mathcal{G}_{[n],\Omega}}$ is a direct product of a sequence of simplices.

THEOREM 4.7.

$$\mathbf{P}_{\mathcal{G}_{[n],\Omega}} = \mathbf{P}_{a_{[2]}} \times \ldots \times \mathbf{P}_{a_{[n]}}$$

$$(4.1) \quad = \underbrace{\Delta_{2^{|\Omega_2^1|-|\Omega_2^0|}-1} \times \cdots \times \Delta_{2^{|\Omega_2^1|-|\Omega_2^0|}-1}}_{2^{|\Omega_2^0|} \ many} \times \cdots \times \underbrace{\Delta_{2^{|\Omega_n^1|-|\Omega_n^0|}-1} \times \cdots \times \Delta_{2^{|\Omega_n^1|-|\Omega_n^0|}-1}}_{2^{|\Omega_n^0|} \ many},$$

where the ith polytope

$$\mathbf{P}_{a_{[i+1]}} = \underbrace{\Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1} \times \cdots \times \Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1}}_{2^{|\Omega_{i+1}^0|} \ many}$$

is a $(2^{|\Omega_{i+1}^1|}-2^{|\Omega_{i+1}^0|})$-face of $\mathbf{P}_{|\Omega_{i+1}^1|,1} = \Delta_{2^{|\Omega_{i+1}^1|}-1}$, where $\mathbf{P}_{|\Omega_{i+1}^1|,1}$ is the cim-polytope for diagnosis models with diseases $A = \Omega_{i+1}^1$ and one symptom $a_{[i+1]}$.

Proof. To prove

$$\mathbf{P}_{a_{[i+1]}} = \underbrace{\Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1} \times \cdots \times \Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1}}_{2^{|\Omega_{i+1}^0|} \ many},$$

we permute the coordinates in the following way:

$$(4.2) \qquad \{T : T \subseteq \Omega_{i+1}^0 \cup a_{[i+1]}\} \cup \bigcup_{\Omega_s \subseteq \Omega_{i+1}^0} \{T \subseteq \Omega_{i+1}^1 \cup a_{[i+1]} : T \cap \Omega_{i+1}^0 = \Omega_s\},$$

i.e., $c_G(T) \ \forall G \in \mathcal{G}_{[n],\Omega}$ can be split into the subvectors $(c_G(T), \text{ where } T \subseteq \Omega_{i+1}^0 \cup a_{[i+1]})$ and $(c_G(T), \text{ where } T \subseteq \Omega_{i+1}^1 \cup a_{[i+1]} \text{ and } T \cap \Omega_{i+1}^0 = \Omega_s) \ \forall \Omega_s \subseteq \Omega_{i+1}^0$.

Then using a strategy similar to the previous proofs, we can prove the following:
- $c_G(T), \ T \subseteq \Omega_{i+1}^0 \cup a_{[i+1]}$, are all fixed;
- $\forall \Omega_s \subseteq \Omega_{i+1}^0$, the convex hull of $\{(c_G(T), \text{ where } T \subseteq \Omega_{i+1}^1 \cup a_{[i+1]} \text{ and } T \cap \Omega_{i+1}^0 = \Omega_s) : \forall G \in \mathcal{G}_{[n],\Omega}\}$ is $\Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1}$ (see Example 4.8);
-

$$\mathbf{P}_{a_{[i+1]}} = \underbrace{\Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1} \times \cdots \times \Delta_{2^{|\Omega_{i+1}^1|-|\Omega_{i+1}^0|}-1}}_{2^{|\Omega_{i+1}^0|} \ many};$$

- Equation 4.1 holds. $\quad\square$

Example 4.8. Consider a DAG G which has seven nodes $\{a_1, \ldots, a_7\}$. After fixing an underlying ordering, we can write these nodes as $\{a_{[1]}, \ldots, a_{[7]}\}$, where $a_{[i]} \to a_{[j]}$ in G implies $i < j$. Suppose edges $a_{[1]} \to a_{[6]}$ and $a_{[2]} \to a_{[6]}$ are fixed and edge $a_{[5]} \to a_{[6]}$ is forbidden. Consider those T containing $a_{[6]}$. Then coordinates $c_G(T) = 0$

if $a_{[5]} \in T$, and other coordinates $c_G(T)$ where $a_{[j]} \notin T$, $\forall j > 6$, can be ordered as follows (values with respect to different DAGs are listed as a matrix):

$$T\backslash\{a_{[6]}\} \quad \begin{bmatrix} a_{[1]} & a_{[2]} & a_{[1]}a_{[2]} & a_{[3]} & a_{[4]} & a_{[3]}a_{[4]} & a_{[1]}a_{[3]} & a_{[1]}a_{[4]} & a_{[1]}a_{[3]}a_{[4]} \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\begin{matrix} a_{[2]}a_{[3]} & a_{[2]}a_{[4]} & a_{[2]}a_{[3]}a_{[4]} & a_{[1]}a_{[2]}a_{[3]} & a_{[1]}a_{[2]}a_{[4]} & a_{[1]}a_{[2]}a_{[3]}a_{[4]} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{matrix} \bigg],$$

where the four rows correspond to graphs G_i, $i = 1, \ldots, 4$, such that $pa_{G_1}(a_{[6]}) = \{a_{[1]}, a_{[2]}\}$, $pa_{G_1}(a_{[6]}) = \{a_{[1]}, a_{[2]}, a_{[3]}\}$, $pa_{G_1}(a_{[6]}) = \{a_{[1]}, a_{[2]}, a_{[4]}\}$, and $pa_{G_1}(a_{[6]}) = \{a_{[1]}, a_{[2]}, a_{[3]}, a_{[4]}\}$.

It is obvious that the cim-polytope for diagnosis models, $\mathbf{P}_{m,n}$, is a special case of $\mathbf{P}_{\mathcal{G}_{[n]},\Omega}$: the underlying ordering of nodes is $(a_1, \ldots, a_m, b_1, \ldots, b_n)$ (the ordering is not unique in the sense that any pair of two diseases or two symptoms can be exchanged), $\Omega_i^0 = \Omega_i^1 = \emptyset$ for $i = 1, \ldots, m$, while $\Omega_i^0 = \emptyset$ and $\Omega_i^1 = \{a_1, \ldots, a_m\}$ for $i = m + 1, \ldots, m + n$. Note that based on (4.1), all edges of $\mathbf{P}_{\mathcal{G}_{[n]},\Omega}$ can be found similarly to Theorem 4.6, and its facets can also be obtained by Lemma 3.8 and Theorem 3.10.

5. Prove $\mathbf{P}_{m,n}$ is simple using linear algebra. Recall that in section 3, we first proved that $\mathbf{P}_{m,1}$ is a simplex Δ_{2^m-1}, and then we proved that $\mathbf{P}_{m,n}$ is a direct product of n many Δ_{2^m-1}, which implies that $\mathbf{P}_{m,n}$ is a simple polytope with dimension $n \cdot (2^m - 1)$. Now we are going to show another flow to prove that $\mathbf{P}_{m,n}$ is simple.

First, we will use linear algebra to show that $\mathbf{P}_{m,n}$ has dimension $n \cdot (2^m - 1)$. We adopt the notation from section 3. Given N, by Propositions 2.3 and 2.7, we can define $\mathcal{S}_{m,n}$ as the support of $\{c_G : G \in \mathcal{G}_{m,n}\}$, i.e.:

$$\mathcal{S}_{m,n} = \{T : \exists\, G \in \mathcal{G}_{m,n} \text{ such that } c_G(T) = 1\} \subset \mathcal{P}(N),$$

where $\mathcal{P}(N)$ is the power set of N.

THEOREM 5.1. *Fix m and n. The dimension of $\mathbf{P}_{m,n}$ is exactly $n \cdot (2^m - 1)$.*

Proof. Similar to imsets, we can consider the standard basis $\mathbf{e_T}$, $T \subset N$, as functions $\mathbf{e_T} : \mathcal{P}(N) \mapsto \mathbb{Z}$ such that $\forall\, T_0 \subset N$, $\mathbf{e_T}(T_0) = 1$ if $T_0 = T$, and 0 otherwise. Each $\mathbf{e_T}$ can also be considered as a vector with coordinates $T_0 \subset N$.

It is obvious that (1) $\{c_G, G \in \mathcal{G}_{m,n}\} \subset \mathbb{R}^{2^{m+n}-(m+n+1)}$; (2) $\{\mathbf{e_T}, T \in \mathcal{S}_{m,n}\}$ is a basis of $\mathbb{R}^{n\cdot(2^m-1)}$ that is embedded in $\mathbb{R}^{2^{m+n}-(m+n+1)}$ (Proposition 2.7); and (3) $\{c_G, G \in \mathcal{G}_{m,n}\}$ can be written as a linear combination of $\{\mathbf{e_T}, T \in \mathcal{S}_{m,n}\}$. We are going to prove that $\{\mathbf{e_T}, T \in \mathcal{S}_{m,n}\}$ can be expressed as a linear combination of $\{c_G, G \in \mathcal{G}_{m,n}\}$. Notice that $\{\mathbf{e_T}, T \in \mathcal{S}_{m,n}\}$ is equivalent to $\{\mathbf{e_T}, T \subset N$ and T has the form of $a_{i_1} \ldots a_{i_k} b_j$, where $1 \le k \le m$, $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}\}$ (Proposition 2.3), and we can prove the statement by induction on $|T|$.

- When $|T| = 2$ (i.e., $k = 1$), i.e., $T = a_i b_j$, where $a_i \in A$ and $b_j \in B$, we know $c_G = \mathbf{e_T}$, where $G \in \mathcal{G}_{m,n}$ has only one edge $a_i \to b_j$.
- Suppose $\forall\, T$, T has the form in Proposition 2.3 and $|T| \le k$, $\mathbf{e_T}$ can be written as a linear combination of $\{c_G, G \in \mathcal{G}_{m,n}\}$. Now consider $T_k = a_{i_1} \ldots a_{i_k} b_j$, where $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

Let $G \in \mathcal{G}_{m,n}$ have k edges: $a_{i_l} \to b_j$, $l = 1 \ldots k$. Then

$$\mathbf{e_{T_k}} = c_G - \sum_{T_a \subset \{a_{i_1}, \ldots, a_{i_k}\}, 0 < |T_a| < k} \mathbf{e_{T_a \cup \{b_j\}}}.$$

Since $\forall\, T_a \subset \{a_{i_1}, \ldots, a_{i_k}\}$, $0 < |T_a| < k$ (i.e., $T_a \subsetneq \{a_{i_1}, \ldots, a_{i_k}\}$), $|T_a \cup b_j| \le k$, $\mathbf{e_{T_a \cup b_j}}$ can be expressed as a linear combination of $\{c_G, G \in \mathcal{G}_{m,n}\}$. Therefore, $\mathbf{e_{T_k}}$ can be written as a linear combination of $\{c_G, G \in \mathcal{G}_{m,n}\}$. □

A special case of $n = 1$ in Theorem 5.1 and Proposition 2.6 claims that $\mathbf{P}_{m,1}$ has 2^m vertices and dimension $2^m - 1$. This directly leads to Corollary 5.2.

COROLLARY 5.2. *Fix m, $\mathbf{P}_{m,1}$ is a simplex with dimension $2^m - 1$, i.e., $\mathbf{P}_{m,1} = \Delta_{2^m - 1}$.*

Lemma 3.2 is an immediate result of Corollary 5.2, while Theorems 3.5 and 3.7 can be obtained based on Lemma 3.2 and Corollary 5.2 using the same proofs in section 3. It is worth mentioning that Theorems 5.1 and 3.5 imply that $\mathbf{P}_{m,n}$ is a simple polytope with dimension $n \cdot (2^m - 1)$ because the number of neighbors for each vertex equals the dimension of the polytope. In 2000, Kaibel and Wolff proved that a zero-one polytope is simple if and only if it equals a direct product of zero-one simplices [6]. Recalling that cim-polytopes are zero-one polytopes, we are able to conclude that $\mathbf{P}_{m,n}$ is a direct product of zero-one simplices [6]. Our progress is that we proved an even strong result in Theorem 3.7 with an intuitive graphical interpretation of each simplex in the direct product.

6. Discussion.

6.1. Connection to K2 algorithm. If we consider a criterion $\mathcal{Q}$ which is a regular criterion, then it was proved that $\mathcal{Q}$ can be written as $\mathcal{Q}(G, D) = s(D) - \langle r_D, c_G \rangle$, where the entropy $s(D)$ and the data vector r_D only depends on the data D and c_G is the characteristic imset of G [12, 8]. Once the cim-polytope can be written as a direct product of simplices, we are able to find the optimal BN structure by maximizing a target function in each simplex: given data $D \in DATA(N, d)$,

$$(6.1) \qquad \max_{G \in \mathcal{G}_{[n],\Omega}} \mathcal{Q}(G, D) \implies \min_{\mathbf{x} \in \mathbf{P}_{\mathcal{G}_{[n],\Omega},c}} r_D^T \mathbf{x} = \sum_{i=2}^{n} \min_{\mathbf{x_i} \in \Delta_{2^{|\Omega_i|}-1}} r_{D,i}^T \mathbf{x_i},$$

where $\mathbf{x_i}$ contains the coordinates $\{T \subseteq \Omega_i \cup \{a_{[i]}\} : |T| \ge 2, a_{[i]} \in T, a_{[j]} \notin T, \forall j > i\}$ in $\mathbf{x}$, and the coordinates of $r_{D,i}^T$ match the coordinates of $\mathbf{x_i}$. This implies that we can find the optimal parent sets of $a_{[i]}$, $i = 2, \ldots, n$, sequentially until we obtain the whole BN structure, which will be exactly the optimal BN structure in $\mathcal{G}_{[n],\Omega}$.

Equation (6.1) gives a polyhedral geometric insight of the K2 algorithm [3], which is a well-known heuristic method in learning BNs. Recall that in the K2 algorithm, an ordering on the nodes is also fixed and parent sets of $a_{[i]}$, $i = 2, \ldots, n$, are also determined sequentially. However, in order to find the optimal BN, (6.1) claims that we need to find $G_i \in \mathcal{G}_{|\Omega_i|,1}$ such that $r_{D,i}^T c_{G_i} = \min_{\mathbf{x_i} \in \Delta_{2^{|\Omega_i|}-1}} r_{D,i}^T \mathbf{x_i}$, while the K2 algorithm obtains each parent set $pa_G(a_{[i]})$ by adding nodes to $\emptyset$ one by one (or removing nodes from $\{a_{[1]}, \ldots, a_{[i-1]}\}$ one by one), which cannot guarantee that the resulting parent sets are optimal (see Example 6.1 for a counterexample).

Example 6.1. Consider $\mathcal{G}_{3,1}$. The characteristic imsets of all possible graphs in $\mathcal{G}_{3,1}$ are listed as a matrix:

$$
\begin{pmatrix} c_{G_0} \\ c_{G_1} \\ c_{G_2} \\ c_{G_3} \\ c_{G_{12}} \\ c_{G_{23}} \\ c_{G_{13}} \\ c_{G_{123}} \end{pmatrix} =
\begin{array}{c}
\begin{array}{ccccccc} T & a_1b_1 & a_2b_1 & a_3b_1 & a_1a_2b_1 & a_1a_3b_1 & a_2a_3b_1 & a_1a_2a_3b_1 \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}
\end{array}.
$$

We are going to show that the resulting BN of the K2 algorithm is not the optimal solution.

- Forward selection: each parent set $pa_G([i])$ is obtained by adding nodes to $\emptyset$ one by one. Suppose $r_D^T = (-1, -2, -1, -3, -10, -4, 20)$ which satisfies $r_D^T c_{G_{13}} = -12 < r_D^T c_G$, $\forall G \in \mathcal{G}_{3,1}$, $G \neq G_{13}$, i.e., the optimal graph is G_{13}. In the K2 algorithm, we start from $pa_G(b_1) = \emptyset$. Next, a_2 is added to $pa_G(b_1)$ because $r_D^T c_{G_2} = -2 < r_D^T c_{G_1} = r_D^T c_{G_3} = -1$. Then a_3 is added to $pa_G(b_1)$ because $r_D^T c_{G_{23}} = -7 < r_D^T c_{G_{12}} = -6$. The procedure ends here because $r_D^T c_{G_{23}} = -7 < r_D^T c_{G_{123}} = -1$. The graph chosen by the K2 algorithm, G_{23}, is not the optimal graph.
- Backward selection: each parent set $pa_G(a_{[i]})$ is obtained by removing nodes from $\{a_{[1]}, \ldots, a_{[i-1]}\}$ one by one. Suppose $r_D^T = (-3, -1, -1, 3, 3, 0, 10)$ which satisfies $r_D^T c_{G_1} = -3 < r_D^T c_G$, $\forall G \in \mathcal{G}_{3,1}$, $G \neq G_1$, i.e., the optimal graph is G_1. In the K2 algorithm, we start from $pa_G(b_1) = \{a_1, a_2, a_3\}$. Next, a_1 is removed from $pa_G(b_1)$ because $r_D^T c_{G_{23}} = -2 < r_D^T c_{G_{12}} = r_D^T c_{G_{13}} = -1$. The procedure ends here because $r_D^T c_{G_{23}} = -2 < r_D^T c_{G_2} = r_D^T c_{G_3} = -1$. The graph chosen by the K2 algorithm, G_{23}, is not the optimal graph.

6.2. Open problems. Further work and open problems remain for this topic. As mentioned, the main purpose of studying the structure of cim-polytopes is reducing the time complexity of learning BNs by suggesting polyhedral geometry techniques. But the reality is that even though we have simplified our problem of learning BNs to LP problems over each simplex (see (6.1)) in the direct product shown in Theorem 4.3 and (4.1), and have described all edges and facets of these simplices (see section 3), if the number of nodes is large, the procedure of searching the optimal solution in each simplex may still be very time-consuming. In this sense, simulations and analysis on real datasets are necessary to compare the solution and time complexity of our method with other existing classifiers [16]. On the other hand, we also need to study the misspecification (i.e., the underlying ordering of nodes is misspecified) and data sensitivity problems of our method via simulations.

Another way to reduce the time complexity is to consider setting up a maximum number of parents to control the model complexity, especially when the number of nodes is too large. In this case, since the underlying ordering is fixed, the cim-polytope is still a direct product of simplices. Thus all edges of the cim-polytope can be found similarly to Theorem 4.6, but the expression of facets for each simplex is not clear.

Notice that all conclusions and discussions in this paper have been based on a fixed underlying ordering of nodes. However, in practice, it is often hard to decide such an ordering. One way to compromise is that we can fix the ordering of some of the nodes and consider every permutation of the rest of the nodes. For instance, when we use single-nucleotide polymorphism (SNP) data to examine phenotypes, we are more

interested in how genes affect phenotypes and how phenotypes affect each other. Thus we can consider DAGs where all edges between SNPs and edges from phenotypes to SNPs are forbidden, i.e., we only need to consider the permutation of phenotypes.

This paper focuses on the case that all random variables in N are finite random variables. How to generalize our method to the case that some or all of the random variables in N are continuous random variables remains an open problem.

REFERENCES

[1] H. AKAIKE, *Information theory and an extension of the maximum likelihood principle*, in Proceedings of the Second International Symposium on Information Theory, B. Petrox and F. Caski, eds., 1973, pp. 267–281.

[2] S. A. ANDERSSON, D. MADIGAN, AND M. D. PERLMAN, *A characterization of Markov equivalence classes for acyclic digraphs*, Ann. Statist., 25 (1997), pp. 505–541.

[3] G. F. COOPER, *A bayesian method for the induction of probabilistic networks from data*, Machine Learning, 9 (1992), pp. 309–347.

[4] N. FRIEDMAN, M. LINIAL, I. NACHMAN, AND D. PE'ER, *Using Bayesian networks to analyze expression data*, J. Comput. Biol., 7 (2000), pp. 601–620.

[5] X. JIANG, R. E. NEAPOLITAN, M. M. BARMADA, AND S. VISWESWARAN, *Learning genetic epistasis using Bayesian network scoring criteria*, BMC Bioinformatics, 12 (2011).

[6] V. KAIBEL AND M. WOLFF, *Simple 0/1-polytope*, Europ. J. Combin., 21 (2000), pp. 139–144.

[7] S. L. LAURITZEN, *Graphical Models*, Clarendon Press, Oxford, UK, 1996.

[8] S. LINDNER, *Discrete Optimisation in Machine Learning: Learning of Bayesian Network Structures and Conditional Independence Implication*, Ph.D. thesis, Technische Universität München, 2012.

[9] P. J. F. LUCAS, *Bayesian model-based diagnosis*, Internat. J. Approx. Reason., 27 (2001), pp. 99–119.

[10] G. SCHWARZ, *Estimating the dimension of a model*, Ann. Statist., 6 (1978), pp. 461–464.

[11] M. A. SHWE, D. E. HECKERMAN, M. HENRION, H. P. LEHMANN, AND G. F. COOPER, *Probabilistic diagnosis using a reformulation of the internist-1/QMR knowledge base: I. The probabilistic model and inference algorithms*, Methods of Information in Medicine, 30 (1991), pp. 241–255.

[12] M. STUDENÝ, *Probabilistic Conditional Independence Structures*, Springer, New York, 2005.

[13] M. STUDENÝ, R. HEMMECKE, AND S. LINDNER, *Characteristic imset: A simple algebraic representative of a Bayesian network structure*, in Proceedings of the 5th European Workshop on Probabilistic Graphical Models, 2010, pp. 257–264.

[14] M. STUDENÝ, J. VOMLEL, AND R. HEMMECKE, *A geometric view on learning Bayesian network structures*, Internat. J. Approx. Reason., 51 (2010), pp. 573–586.

[15] J. UEBERSAX, *Genetic Counseling and Cancer Risk Modeling: An Application of Bayes Nets*, Ravenpack International, Marbella, Spain, 2004.

[16] J. VOMLEL, H. KRUŽÍK, P. TŮMA, J. PŘEČEK, AND M. HUTYRA, *Machine learning methods for mortality prediction in patients with ST elevation myocardial infarction*, in Proceedings of WUPES, 2012, pp. 204–213.

[17] J. XI, *Polytopes Arising from Binary Multi-way Contingency Tables and Characteristic Imsets for Bayesian Networks*, Ph.D. thesis, University of Kentucky, Lexington, 2013.

[18] G. ZIEGLER, *Lectures on Polytopes*, Springer, New York, 1994.