

Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

Mode Poset Probability Polytopes

by

Guido Montúfar and Johannes Rauh

Preprint no.: 22

2015



Mode Poset Probability Polytopes

Guido Montúfar¹ and Johannes Rauh²

¹ Max Planck Institute for Mathematics in the Sciences,
Inselstraße 22, 04103 Leipzig, Germany,
`montufar@mis.mpg.de`,

² Leibniz Universität Hannover,
Welfengarten 1, 30167 Hannover, Germany,
`rauh@math.uni-hannover.de`

Abstract. A mode of a probability vector is a local maximum with respect to some vicinity structure on the set of elementary events. The mode inequalities cut out a polytope from the simplex of probability vectors. Related to this is the concept of strong modes. A strong mode of a distribution is an elementary event that has more probability mass than all its direct neighbors together. The set of probability distributions with a given set of strong modes is again a polytope. We study the vertices, the facets, and the volume of such polytopes depending on the sets of (strong) modes and the vicinity structures.

1 Introduction

Many probability models used in practice are given in a parametric form. Sometimes it is useful to also have an implicit description in terms of properties that characterize the probability distributions that belong to the model. Such a description can be used to check whether a given probability distribution lies in the model or, otherwise, to estimate how far it lies from the model. For example, if a given model has a parametrization by polynomial functions, then one can show that it has a *semialgebraic description*; that is, an implicit description as the solution set of polynomial equations and polynomial inequalities. Finding this description is known as the *implicitization* problem, which in general is very hard to solve completely. Even if it is not possible to give a full implicit description, it may be possible to confine the model by simple polynomial equalities and inequalities. Here we are interested in simple confinements, in terms of natural classes of linear equalities and inequalities.

We consider polyhedral sets of discrete probability distributions defined by prescribed sets of modes. A mode is a local maximum of a probability vector. Locality is with respect to a given vicinity structure in the set of coordinate indices; that is, x is a (strict) mode of a probability vector p if and only if $p_x > p_y$, for all neighbors y of x . The vicinity structure depends on the setting. For probability distributions on a set of fixed-length strings, it is natural to call two strings neighbors if and only if they have Hamming distance one. For probability distributions on integer intervals, it is natural to call two integers

neighbors if and only if they are consecutive. In general, a vicinity structure is just a graph with undirected edges.

Modes are important characteristics of probability distributions. In particular, the question whether a probability distribution underlying a statistical experiment has one or more modes is important in applications. Also, many statistical models consist of “nice” probability distributions that are “smooth” in some sense. Such probability distributions have only a limited number of modes. Another motivation for studying modes was given in [2], where it was observed that mode patterns are a practical way to differentiate between certain parametric model classes.

Besides from modes, we are also interested in the related concept of strong modes introduced in [2]. A point x is a (strict) strong mode of a probability distribution p if and only if $p_x > \sum_{y \sim x} p_y$, where the sum runs over all neighbors y of x . Strong modes offer similar possibilities as modes for studying models of probability distributions. While strong modes are more restrictive than modes, they are easier to study.

One observation is: Suppose that $p = \sum_{i=1}^k \lambda_i p^i$ is a mixture of k probability distributions. If p has a strict strong mode $x \in V$, then x must be a mode of one of the distributions p^i , because if $p^i(x) \leq p^i(y_i)$ for some neighbor y_i of x for all i , then $\sum_i \lambda_i p^i(x) \leq \sum_i \lambda_i p^i(y_i) \leq \sum_{y \sim x} \sum_i \lambda_i p^i(y)$. For example, a mixture of k uni-modal distributions has at most k strong modes. Surprisingly, the same statement is not true for modes: A mixture of k product distributions may have more than k modes [2]. Still, the number of modes of a mixture of product distributions is bounded, although this bound is not known in general. As another example, in [2] it was shown that a restricted Boltzmann machine with m hidden nodes and n visible nodes, where $m < n$ and m is even, does not contain probability distributions with certain patterns of 2^m strict strong modes.

In this paper we derive essential properties of (strong) mode polytopes, depending on the vicinity structures and the considered patterns of (strong) modes. In particular, we describe the vertices, the facets, and the volume of these polytopes. It is worth mentioning that mode probability polytopes are closely related to order and poset polytopes. We describe this relation at the end of Section 2.

This paper is organized as follows: In Section 2 we study the polytopes of modes and in Section 3 the polytopes of strong modes.

2 The polytope of modes

We consider a finite set of elementary events V and the set of probability distributions on this set, $\Delta(V)$. We endow V with a vicinity structure described by a graph. Let $G = (V, E)$ be a simple graph (i.e., no multiple edges and no loops). For any $x, y \in V$, if $(x, y) \in E$ is an edge in G , we write $x \sim y$. Since we assume that the graph is simple, $x \sim y$ implies $x \neq y$.

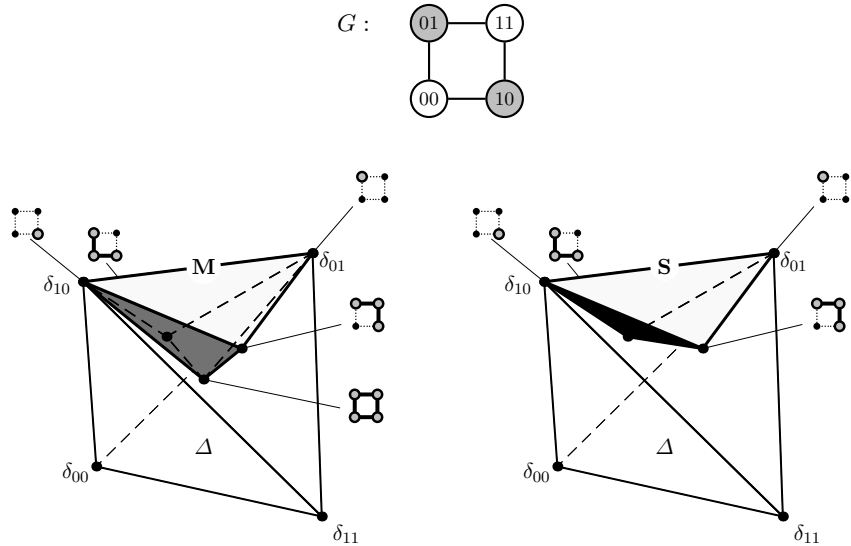


Fig. 1. Above: The graph G from Examples 1 and 2, with \mathcal{C} marked in gray. Below: The corresponding polytopes $\mathbf{M}(G, \mathcal{C})$ and $\mathbf{S}(G, \mathcal{C})$. Each vertex of these polytopes is a uniform distribution supported on a subset of G , as explained in Propositions 1 and 3.

Definition 1. A point $x \in V$ is a mode of a probability distribution $p \in \Delta(V)$ if $p_x \geq p_y$ for all $y \sim x$.

Definition 2. Consider a subset $\mathcal{C} \subseteq V$. The polytope of \mathcal{C} -modes in G is the set $\mathbf{M}(G, \mathcal{C})$ of all probability distributions $p \in \Delta(V)$ for which every $x \in \mathcal{C}$ is a mode.

The set $\mathbf{M}(G, \mathcal{C})$ is always non-empty, since it contains the uniform distribution. It is a polytope, because it is a closed convex set defined by finitely many linear inequalities and, as a subset of $\Delta(V)$, it is bounded. We are interested in the properties of this polytope, depending on G and \mathcal{C} .

Recall that a set of vertices of a graph is *independent*, if it does not contain two adjacent elements. If \mathcal{C} is not independent, then $\mathbf{M}(G, \mathcal{C})$ is not full-dimensional as a subset of $\Delta(V)$; that is, $\dim \mathbf{M}(G, \mathcal{C}) < \dim(\Delta(V)) = |V| - 1$. For, if $x, y \in \mathcal{C}$ are neighbors, then the defining equations of $\mathbf{M}(G, \mathcal{C})$ imply that $p_x \geq p_y \geq p_x$; that is, any $p \in \mathbf{M}(G, \mathcal{C})$ satisfies $p_x = p_y$. In the following we will ignore this degenerate case and assume that the set of modes is independent.

In some applications, for example those mentioned in the introduction, it is more natural to study *strict modes*; i.e. points $x \in V$ with $p_x > p_y$ for all $y \sim x$. A description of the set of distributions with prescribed strict modes is easy to obtain from a description of $\mathbf{M}(G, \mathcal{C})$.

Example 1. Let G be a square with vertices $V = \{00, 01, 10, 11\}$ and edges $E = \{(00, 01), (00, 10), (01, 11), (10, 11)\}$. The polytope $\mathbf{M}(G, \mathcal{C})$ for $\mathcal{C} = \{01, 10\}$ is given in Figure 1.

Vertices. We have defined $\mathbf{M}(G, \mathcal{C})$ by linear inequalities (H-representation). Next we determine its vertices (V-representation). For any non-empty $W \subseteq V \setminus \mathcal{C}$ and $y \in V$ write $y \sim W$ if $y \sim x$ for some $x \in W$. Moreover, let $N_{\mathcal{C}}(W) = \{y \in \mathcal{C} : y \sim W\}$ (this is the set of declared modes which are neighbors of W), and let $e_{\mathcal{C}}^W$ be the uniform distribution on $N_{\mathcal{C}}(W) \cup W$.

Proposition 1.

1. $\mathbf{M}(G, \mathcal{C})$ is the convex hull of $\{e_{\mathcal{C}}^W : \emptyset \neq W \subseteq V \setminus \mathcal{C}\} \cup \{\delta_x : x \in \mathcal{C}\}$, where δ_x denotes the point distribution concentrated on x .
2. For any $x \in \mathcal{C}$, the distribution δ_x is a vertex of $\mathbf{M}(G, \mathcal{C})$.
3. $e_{\mathcal{C}}^W$ is a vertex of $\mathbf{M}(G, \mathcal{C})$ iff for any $x, y \in W$, $x \neq y$, there is a path $x = x_0 \sim x_1 \sim \dots \sim x_r = y$ in G with $x_0, x_2, \dots \in W$ and $x_1, x_3, \dots \in N_{\mathcal{C}}(W)$.

Proof. Clearly, for every non-empty $W \subseteq V \setminus \mathcal{C}$, the vector $e_{\mathcal{C}}^W$ belongs to $\mathbf{M}(G, \mathcal{C})$, and the same is true for the vectors δ_x with $x \in \mathcal{C}$ (\mathcal{C} is independent). Next we show that each $p \in \mathbf{M}(G, \mathcal{C})$ can be written as a convex combination of $\{e_{\mathcal{C}}^W : \emptyset \neq W \subseteq V \setminus \mathcal{C}\} \cup \{\delta_x : x \in \mathcal{C}\}$. We do induction on the cardinality of $W := \text{supp}(p) \setminus \mathcal{C}$. If $|W| = 0$, then $p \in \Delta(\mathcal{C})$ is a convex combination of $\{\delta_x : x \in \mathcal{C}\}$. Now assume $|W| > 0$. Let $\lambda = \min\{p_x : x \in W\}$. Then, $p - \lambda e_{\mathcal{C}}^W \geq 0$ (component-wise) and $\sum_x (p_x - \lambda e_{\mathcal{C}}^W(x)) = (1 - \lambda)$. Therefore,

$$p' := \frac{1}{1 - \lambda}(p - \lambda e_{\mathcal{C}}^W) \in \Delta(V).$$

Moreover, one checks that $p' \in \mathbf{M}(G, \mathcal{C})$. By definition, $\text{supp}(p') \setminus \mathcal{C} \subsetneq \text{supp}(p) \setminus \mathcal{C}$. By induction, $\text{supp}(p')$ is a convex combination of $\{e_{\mathcal{C}}^W : \emptyset \neq W \subseteq V \setminus \mathcal{C}\} \cup \{\delta_x : x \in \mathcal{C}\}$, and so the same is true for p .

It remains to check which elements of $\{e_{\mathcal{C}}^W : \emptyset \neq W \subseteq V \setminus \mathcal{C}\} \cup \{\delta_x : x \in \mathcal{C}\}$ are vertices of $\mathbf{M}(G, \mathcal{C})$. Since δ_x is a vertex of $\Delta(V)$, it is also a vertex of $\mathbf{M}(G, \mathcal{C})$. Let $W \subset V \setminus \mathcal{C}$ be non-empty. Call a path such as in the statement of the proposition an *alternating path*. Suppose that there is no alternating path from x to y for some $x, y \in W$. Let $W_1 = \{z \in W : \text{There is an alternating path from } x \text{ to } z\}$ and let $W_2 = W \setminus W_1$. Then W_1, W_2 are non-empty, and $N_{\mathcal{C}}(W_1) \cap N_{\mathcal{C}}(W_2)$ is empty. Hence $e_{\mathcal{C}}^W$ is a convex combination of $e_{\mathcal{C}}^{W_1}$ and $e_{\mathcal{C}}^{W_2}$, and $e_{\mathcal{C}}^W$ is not a vertex.

Let W be a non-empty subset of $V \setminus \mathcal{C}$ such that any pair of elements of W is connected by an alternating path. To show that $e_{\mathcal{C}}^W$ is a vertex, for any different non-empty set $W' \subseteq V \setminus \mathcal{C}$ we need to find a face of $\mathbf{M}(G, \mathcal{C})$ that contains $e_{\mathcal{C}}^W$ but not $e_{\mathcal{C}}^{W'}$. If there exists $x \in W' \setminus W$, then $e_{\mathcal{C}}^{W'}(x) > 0 = e_{\mathcal{C}}^W(x)$. Hence, $e_{\mathcal{C}}^W$ lies on the face of $\mathbf{M}(G, \mathcal{C})$ defined by $p_x \geq 0$, but $e_{\mathcal{C}}^{W'}$ does not. Otherwise, $W' \subsetneq W$. Let $x' \in W \setminus W'$ and $y' \in W' \neq \emptyset$. By assumption, there exists an alternating path from x' to y' in W . On this path, there exist $x \in W \setminus W'$ and $y \in \mathcal{C}$ with $y \sim x$ and $y \in N_{\mathcal{C}}(W')$. Therefore, $e_{\mathcal{C}}^{W'}(y) - e_{\mathcal{C}}^{W'}(x) > 0 = e_{\mathcal{C}}^W(y) - e_{\mathcal{C}}^W(x)$. \square

Corollary 1. $\mathbf{M}(G, \mathcal{C})$ is a full-dimensional sub-polytope of $\Delta(V)$.

Proof. The convex hull of $\{\delta_x : x \in \mathcal{C}\} \cup \{e_{\mathcal{C}}^{\{y\}} : y \in V \setminus \mathcal{C}\}$ is a $(|V| - 1)$ -simplex and a subset of $\mathbf{M}(G, \mathcal{C})$. \square

Facets. $\mathbf{M}(G, \mathcal{C})$ is defined, as a subset of $\Delta(V)$, by the inequalities

$$\begin{aligned} p_x &\geq 0, & \text{for all } x \in V, & & (\text{positivity inequalities}) \\ p_x &\geq p_y, & \text{for all } x \in \mathcal{C} \text{ and } y \sim x. & & (\text{mode inequalities}) \end{aligned}$$

Next we discuss, which of these inequalities define facets.

Proposition 2.

1. For any $x \in V \setminus \mathcal{C}$, the positivity inequality $p_x \geq 0$ defines a facet.
2. If $x \in \mathcal{C}$, then $p_x \geq 0$ defines a facet iff x is isolated in G .
3. For any $x \in \mathcal{C}$ and $y \sim x$, the mode inequality $p_x \geq p_y$ defines a facet.

Proof. 1. The inequality $p_x \geq 0$ defines a facet of the subsimplex from the proof of Corollary 1, and hence also of $\mathbf{M}(G, \mathcal{C})$.

2. If x is isolated, then x is a mode of any distribution. Therefore, $\mathbf{M}(G, \mathcal{C}) = \mathbf{M}(\mathcal{C} \setminus \{x\})$, and the statement follows from 1.

Otherwise, suppose there exists $y \in V$ with $x \sim y$. Since \mathcal{C} is independent, $y \notin \mathcal{C}$. Then $p_x = (p_x - p_y) + p_y$; that is, the inequality $p_x \geq 0$ is implied by the inequalities $p_x \geq p_y$ and $p_y \geq 0$, and $p_x \geq 0$ defines a sub-face of the facet $p_y \geq 0$, which is a strict sub-face, since it does not contain δ_x . Therefore, $p_x \geq 0$ does not define a facet itself.

3. Let $W := \{z \in \mathcal{C} : z \sim y\} \setminus \{x\}$. The uniform distribution on $W \cup \{y\}$ satisfies all defining inequalities of $\mathbf{M}(G, \mathcal{C})$, except $p_x \geq p_y$. \square

Triangulation and volume. The polytope $\mathbf{M}(G, \mathcal{C})$ has a natural triangulation that comes from a natural triangulation of $\Delta(V)$. Let $N = |V|$ be the cardinality of V . For any bijection $\sigma : \{1, \dots, N\} \rightarrow V$ let

$$\Delta_\sigma = \{p \in \Delta(V) : p_{\sigma(i)} \leq p_{\sigma(i+1)} \text{ for } i = 1, \dots, N-1\}.$$

Clearly, the Δ_σ form a triangulation of $\Delta(V)$. In particular, $\Delta(V) = \bigcup_\sigma \Delta_\sigma$ and $\text{vol}(\Delta_\sigma \cup \Delta_{\sigma'}) = \text{vol}(\Delta_\sigma) + \text{vol}(\Delta_{\sigma'})$ whenever $\sigma \neq \sigma'$.

Lemma 1. Let $\Sigma(G, \mathcal{C})$ be the set of all bijections $\sigma : \{1, \dots, N\} \rightarrow V$ that satisfy $\sigma^{-1}(x) < \sigma^{-1}(y)$ for all $y \in \mathcal{C}$ and $x \sim y$. Then $\mathbf{M}(G, \mathcal{C}) = \bigcup_{\sigma \in \Sigma(G, \mathcal{C})} \Delta_\sigma$.

Proof. If $\sigma \in \Sigma$ and $p \in \Delta_\sigma$, then $p \in \mathbf{M}(G, \mathcal{C})$ by definition. Conversely, let $p \in \mathbf{M}(G, \mathcal{C})$. Choose a bijection $\sigma : \{1, \dots, N\} \rightarrow V$ that satisfies the following:

1. $p_{\sigma(i+1)} \geq p_{\sigma(i)}$ for $i = 1, \dots, N-1$,
2. If $x \in \mathcal{C}$ and $y \sim x$, then $\sigma^{-1}(x) \leq \sigma^{-1}(y)$.

Clearly, $\sigma \in \Sigma$, and $p \in \Delta_\sigma$. \square

Corollary 2. $\text{vol}(\mathbf{M}(G, \mathcal{C})) = \frac{|\Sigma|}{|V|!} \text{vol}(\Delta(V))$.

Proof. All simplices Δ_σ have the same volume. Moreover, $\text{vol}(\Delta_\sigma \cap \Delta_{\sigma'}) = 0$ for $\sigma \neq \sigma'$. Thus, $\text{vol}(\mathbf{M}(G, \mathcal{C})) = |\Sigma| \text{vol}(\Delta_\sigma)$ and $\text{vol}(\Delta(V)) = |V|! \text{vol}(\Delta_\sigma)$. \square

It remains to compute the cardinality of $\Sigma(G, \mathcal{C})$. It is not difficult to enumerate $\Sigma(G, \mathcal{C})$ by iterating over the set V . However, $\Sigma(G, \mathcal{C})$ may be a very large, and so, enumerating it can take a very long time. In fact, this is a special instance of the problem of counting the number of linear extensions of a partial order (see below); a problem which in many cases is known to be $\#P$ -complete [1]. In our case, a simple lower bound is $|\Sigma(G, \mathcal{C})| \geq |\mathcal{C}|!|V \setminus \mathcal{C}|!$ (equality holds only when G is a complete bipartite graph and \mathcal{C} is one of the maximal independent sets).

Relation to order polytopes. The results in this section can also be derived from results about order polytopes. To explain this, it is convenient to slightly generalize our settings. Instead of looking at a graph G and an independent subset \mathcal{C} of nodes, consider a partial order \succeq on V and let

$$\mathbf{M}(\succeq) := \{p \in \Delta(V) : p_x \geq p_y \text{ whenever } x \succeq y\}.$$

The polytope $\mathbf{M}(G, \mathcal{C})$ arises in the special case where \succeq is defined by

$$x \succeq y \quad :\iff \quad x \sim y \text{ and } x \in \mathcal{C}.$$

The relation \succeq defined in this way from G and \mathcal{C} is a partial order precisely if \mathcal{C} is independent. Our results about vertices, facets and volumes directly generalize to $\mathbf{M}(\succeq)$. We omit further details at this point.

The *order polytope* of a partial order arises by looking at subsets of the unit hypercube instead of subsets of the probability simplex (see [3] and references):

$$\mathcal{O}(\succeq) := \{p \in [0, 1]^V : p_x \geq p_y \text{ whenever } x \succeq y\}.$$

One can show that $\mathbf{M}(\succeq)$ is the vertex figure of $\mathcal{O}(\succeq)$ at the vertex 0. This observation allows to transfer the results from [3] to $\mathbf{M}(G, \mathcal{C})$.

3 The polytope of strong modes

Definition 3. A point $x \in V$ is a *strong mode* of a probability distribution $p \in \Delta(V)$ if $p_x \geq \sum_{y \sim x} p_y$.

Definition 4. Consider a subset $\mathcal{C} \subseteq V$. The polytope of strong \mathcal{C} -modes in G is the set $\mathbf{S}(G, \mathcal{C})$ all probability distributions $p \in \Delta(V)$ for which every $x \in \mathcal{C}$ is a strong mode.

Again, in applications one may be interested in *strict strong modes* that are characterized by strict inequalities of the form $p_x > \sum_{y \sim x} p_y$.

If $x \sim y$ for two strong modes of $p \in \Delta(V)$, then $p_x = p_y$ and $p_z = 0$ for all other neighbors z of x or y . In order to avoid such pathological cases, in the following we always assume that \mathcal{C} is an independent subset of G .

Example 2. Consider the graph from Example 1. For $\mathcal{C} = \{01, 10\}$, the polytope $\mathbf{S}(G, \mathcal{C})$ is given in Figure 1.

Again, we are interested in the vertices of the polytope $\mathbf{S}(G, \mathcal{C})$. For any $x \in V$ let $N_{\mathcal{C}}(x) = \{y \in \mathcal{C} : y \sim x\}$ (this is the set of strong modes which are neighbors of x) and let $f_{\mathcal{C}}^x$ be the uniform distribution on $N_{\mathcal{C}}(x) \cup \{x\}$.

Proposition 3. *If \mathcal{C} is independent, then $\mathbf{S}(G, \mathcal{C})$ is a $(|V| - 1)$ -simplex with vertices $f_{\mathcal{C}}^x$, $x \in V$.*

Proof. To see that $\{f_{\mathcal{C}}^x : x \in V\}$ is linearly independent, observe that the matrix with columns $f_{\mathcal{C}}^x$ is in tridiagonal form when V is ordered such that the vertices in \mathcal{C} come before the vertices in $V \setminus \mathcal{C}$. Therefore, the probability distributions $f_{\mathcal{C}}^x$ span a $(|V| - 1)$ -dimensional simplex.

It is easy to check that $f_{\mathcal{C}}^x \in \mathbf{S}(G, \mathcal{C})$ for any $x \in V$. It remains to prove that any $p \in \mathbf{S}(G, \mathcal{C})$ lies in the convex hull of $\{f_{\mathcal{C}}^x : x \in V\}$. We do induction on the cardinality of $W := \text{supp}(p) \setminus \mathcal{C}$. If $|W| = 0$, then $p \in \Delta(\mathcal{C})$ is a convex combination of $\{\delta_x : x \in \mathcal{C}\} = \{f_{\mathcal{C}}^x : x \in \mathcal{C}\}$. Otherwise, let $x \in W$. Then

$$p' := \frac{1}{1 - p_x}(p - p_x f_{\mathcal{C}}^x) \in \Delta(V),$$

since $p \in \mathbf{M}(G, \mathcal{C})$. Moreover, $p' \in \mathbf{M}(G, \mathcal{C})$. The statement now follows by induction, since $\text{supp}(p') \setminus \mathcal{C} = W \setminus \{x\}$. \square

Proposition 4. *The facets of $\mathbf{S}(G, \mathcal{C})$ are $p_x \geq \sum_{y \sim x} p_y$ for all $x \in \mathcal{C}$ and $p_x \geq 0$ for all $x \in V \setminus \mathcal{C}$.*

Proof. It is easy to verify that each of the faces defined by these inequalities contains $|V| - 1$ vertices. \square

Proposition 5. $\text{vol}(\mathbf{S}(G, \mathcal{C})) = \left(\prod_{x \in V} \frac{1}{|N_{\mathcal{C}}(x)| + 1} \right) \text{vol}(\Delta(V))$.

Proof. After rearrangement of columns, the matrix

$$(f_{\mathcal{C}}^x)_{x \in V} = \left((\delta_x)_{x \in \mathcal{C}}, \left(\frac{1}{|N_{\mathcal{C}}(x)| + 1} \mathbf{1}_{N_{\mathcal{C}}(x)} \right)_{x \in V \setminus \mathcal{C}, x \sim \mathcal{C}}, (\delta_x)_{x \in V \setminus \mathcal{C}, x \not\sim \mathcal{C}} \right)$$

is in upper triangular form, with diagonal elements $\frac{1}{|N_{\mathcal{C}}(x)| + 1}$, $x \in V$. The statement now follows from the next Lemma 2. \square

Lemma 2. *Let $\Delta = \text{conv}\{e_0, \dots, e_d\}$ be the standard d -simplex in \mathbb{R}^{d+1} and let $s_0, \dots, s_d \in \Delta$. Then the d -volume of $S = \text{conv}\{s_0, \dots, s_d\}$ satisfies*

$$\text{vol}(S) = |\det(s_0, \dots, s_d)| \text{vol}(\Delta).$$

Proof. The $(d + 1)$ -volume of the parallelepiped spanned by $s_0, \dots, s_d \in \mathbb{R}^{d+1}$ is $|\det(s_0, \dots, s_d)|$. The volume of an n -simplex with vertices v_0, \dots, v_n in \mathbb{R}^n is $\frac{1}{n!} |\det(v_1 - v_0, \dots, v_n - v_0)|$. Hence the volume of the $(d + 1)$ -simplex P with vertices $(0, s_0, \dots, s_d)$ is $\text{vol}(P) = \frac{1}{(d+1)!} |\det(s_0, \dots, s_d)|$. Note that P is a pyramid over S of height $h = \frac{1}{\sqrt{d+1}}$. Thus $\text{vol}(P) = \frac{h}{d+1} \text{vol}(S)$. The volume of the regular d -simplex is $\text{vol}(\Delta) = \frac{\sqrt{d+1}}{d!}$. The statement follows by combining these formulas. \square

Example 3. Generalizing Examples 1 and 2, let G be the edge graph of an n -cube, such that $V = \{0, 1\}^n$ and two points are adjacent if their Hamming distance is one.

a) If $\mathcal{C} \subseteq V$ has cardinality $|\mathcal{C}| = k$ and minimum distance 3, then \mathbf{S} has 2^n vertices and volume $\text{vol}(\mathbf{S}) = 2^{-kn} \text{vol}(\Delta)$, whereas \mathbf{M} has $k(2^n - 1) + 2^n - kn$ vertices and volume $\text{vol}(\mathbf{M}) = \frac{|\Sigma|}{2^n} \text{vol}(\Delta) \geq k!2^{-kn} \text{vol}(\Delta)$.

b) If \mathcal{C} is the set of all even-parity strings, then \mathbf{S} has 2^n vertices and volume $\text{vol}(\mathbf{S}) = (n + 1)^{-2^{n-1}} \text{vol}(\Delta)$, whereas \mathbf{M} has $2^{2^{n-1}} - 1 + 2^{n-1}$ vertices and volume $\text{vol}(\mathbf{M}) = \frac{|\Sigma|}{2^n} \text{vol}(\Delta) \geq \binom{2^n}{2^{n-1}}^{-1} \text{vol}(\Delta)$. For $n = 2$ and $n = 3$ we have $|\Sigma| = 4$ and $|\Sigma| = 720$. The next open case is $n = 4$.

References

1. G. Brightwell and P. Winkler. Counting linear extensions. *Order*, 8(3):225–242, 1991.
2. G. Montúfar and J. Morton. When does a mixture of products contain a product of mixtures? *SIAM Journal on Discrete Mathematics*, 29:321–347, 2015.
3. R. Stanley. Two poset polytopes. *Discrete Comput. Geom.*, 1:9–23, 1986.