Noname manuscript No.
(will be inserted by the editor)

Sparse $(0,1)$ arrays and perfect phylogeny

Yaokun Wu · Yanzhen Xiong

Received: date / Accepted: date

Abstract Each family of partitions of a set X, say $\pi_1, \ldots, \pi_n$, naturally corresponds to an n-dimensional $(0,1)$ array of size $a_1 \times \cdots \times a_n$, where a_i is the number of parts of π_i, of which the $(t_1, \ldots, t_n)$-entry is 1 if and only if there is an element of X lying in the t_ith part of π_i for all $i \in \{1, \ldots, n\}$. Demonstrating a close link between sparsity and treelikeness, we report in the paper that the existence of perfect phylogeny for a partition system is equivalent to some sparsity requirements on its associated $(0,1)$ array. We propose a simple algorithm to construct all minimum displaying trees for a sparse $(0,1)$ array; we also show that all these minimum displaying trees can be directly read from the set of spanning trees of the Buneman graph of the array. From a $(0,1)$ array, what we really see are often its various subarrays and projections, namely its local shadows. Given a family of $(0,1)$ arrays, in view of their sparsity behavior, how can we tell if they can be local shadows of some common $(0,1)$ array? Measuring the sparsity in a way compatible with the perfect phylogeny problem, we reveal some interesting combinatorial patterns about the global geometry of a $(0,1)$ array.

Keywords character, partition lattice · phylogenetic X-tree · sparse completion · sparse tensor

Mathematics Subject Classification (2010) 05A18 · 05C05 · 05C65 · 05C85 · 15A69 · 65F50 · 68T10 · 92D15

Supported by NSFC (11671258) and STCSM (17690740800).

Yaokun Wu
School of Mathematical Sciences and MOE-LCS, Shanghai Jiao Tong University, Shanghai 200240, China
E-mail: ykwu@sjtu.edu.cn

Yanzhen Xiong
School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China
E-mail: xyz920805@sjtu.edu.cn

Contents

1 Introduction

Contingency tables are a staple of quantitative science; $(0,1)$ arrays, also known as Boolen tensors, are their qualitative counterpart and so the focus of combinatorial data analysis. It is well known that the set of n-uniform hypergraphs is essentially the set of n-dimensional $(0,1)$ symmetric tensors while the set of n-partitite hypergraphs [AHJ19] can be identified with the set of n-dimensional $(0,1)$ tensors. It may be less observed that partition systems and $(0,1)$ arrays are also closely related. Indeed, all the possible mutual intersection relationships among families of n partitions on a common ground set is naturally recorded by the set of all n-dimensional $(0,1)$ arrays. Partitions are nothing but equivalence relations and these seemingly trivial structures are underlying a host of thought-provoking questions in a qualitative study of information and uncertainty [KRY09, Section 1.4]. As predicted by Gelfand, lattice theory will play a leading role in the mathematics of the twenty-first century [Rot97, p. 1445]. Theorems of Whitman and Pudlák-Tůma [KRY09, Exercise 1.4.3] state that each (finite) lattice occurs as sublattice of a (finite) partition lattice. This means that the role of partition lattices in lattice theory is like that of symmetric groups in group theory. Therefore, although equivalence relations and $(0,1)$ arrays are so ubiquitous in everyday life, we should not be surprised to see that they can still bear fruit in many unexpected ways.

Analyzing categorical data requires us understand joint, marginal, and conditional frequencies. For a two-dimensional $(0,1)$ array, namely a Boolean matrix, we are led to examine its submatrices and row/column (Boolean) sum vectors. Similarly, for a high-dimensional array, we need to deal with its subarrays of different sizes and its margin distributions of various dimensions. How do these local shadows interact with each other? How do we know if they are really pieces from a same global picture? How do we construct a 'smallest' Boolean array whose corresponding local shadows line up with what we have in hand? How to visualize a multidimensional configuration? This kind of local-global problems are of intrinsic mathematics interest and have been studied for many practical purposes as well. For example, the study of perfect phylogeny is essentially to see how to incoporate all observed local information, say some partitions of the taxas, into a common evolutionary tree.

Trees are among those most fundamental structures in mathematics and in sciences. They stand at the boardline of "solvable problems"; when a correct perspective reveals that one object allows a tree to be its good approximation, we can visualize many hidden secrets of this object. As a graph, a tree T is characterized by its sparse property: Every subgraph of T induced by k vertices can have at most $k-1$ edges and equality holds when k equals the total number of vertices in T. This provides the intuitive and useful idea that every sparse structure should be viewed as a tree-like object [NOdM12]. As seen in compressed sensing, phylogenetics, and other applications, sparse data with a bit of tree-structure often allows us to get exact recovery from a small set of observations. A good deal of investigation has been devoted to understanding the meaning and usefulness of sparsity [ABC+18]; meanwhile, we have now seen very rich mathematics about tree-like structures. Note that a central problem in the study of sparse structure is to turn the vague idea of sparsity into accurate definitions which can match our intuition and lead to interesting mathematics. For example, the concept of (k, ℓ)-sparse hypergraph is proposed in rigidity theory [ST11, Whi89] and the aforementioned characterization of trees coincides with the claim that trees are just 2-uniform $(1, 1)$-tight hypergraphs [JK15, p. 94].

We say that a partition system $\mathcal{C}$ possesses a perfect phylogeny if there is a tree T whose set of leaves contains X such that for each partition π in $\mathcal{C}$ we can chop up the tree T in a way that different parts of π appear in different subtrees. This paper proposes a new way to measure the sparseness of a partition system $\mathcal{C}$ and its corresponding tensor M (see Definitions 1 and 2 and Example 4); you will find its similarity with the description of tree sparsity given in last paragraph. We manifest how the sparsity properties of various marginalizations/contractions of the array M are related to each other and we specify those sparsity requirements imposed on M which allow $\mathcal{C}$ to have a perfect phylogeny. Note that a deviation from our sparsity measures can be seen in some concrete numbers and so we can rigoriously discuss how tree-like a structure is even if that structure may not be accompanied by a tree. It is a challenging task to get an intuitive representation for a high-dimensional $(0, 1)$ array. Our effort in connecting perfect phylogeny and sparsity properties of corresponding tensors provides a tree representation of general sparse $(0, 1)$ arrays.

To tell the full story in a more precise language, we address different aspects of this work in the following six subsections, including the basic mathematics structures of our concern (partition system and $(0, 1)$ tensor), some background of the perfect phylogeny problem, our definition of sparse properties, how the sparsity behaviours at different levels are entangled with each other, how the existence of a perfect phylogeny is seen from our sparsity measures, namely how sparse properties characterize trees from the viewpoint of perfect phylogeny, and how to display sparse $(0, 1)$ arrays with trees of smallest sizes.

1.1 Partition system and $(0,1)$ tensor

For any positive integer n, we use $[n]$ to denote the set $\{1,\dots,n\}$. For a set, its size is the number of distinct elements in it; for a multiset, its size is the sum of the multiplicity of its elements. For instance, the size of the multiset $\{\{x,y,y,z\}\}$ is four while the size of the set $\{x,y,y,z\}=\{x,y,z\}$ is three. Let X be a finite set. A *character* on the *ground set* X is a multiset of disjoint subsets of X. One should notice that a character may contain several empty sets. Let φ be a character on a set X. We always use $\|\varphi\|$ for its size as a multiset. For any $Y\subseteq X$, the *trace of* φ *at* Y, denoted by $\varphi[Y]$, is the character on Y of the same size with φ that consists of all those sets of the form $Y\cap A$, $A\in\varphi$. We call each nonempty element of φ a *state* of it and we denote by $|\varphi|_{\mathrm{s}}$ the number of its states. We say that φ is an r*-state* character if $|\varphi|_{\mathrm{s}}\le r$. The character φ on X naturally corresponds to an $X\times X$ $(0,1)$ matrix R_φ, called its *relation matrix* [Bai17], such that $\mathrm{R}_\varphi(x,y)=1$ if and only if x and y fall into the same state of φ. It is easy to see that $|\varphi|_{\mathrm{s}}=\mathrm{Rank}(\mathrm{R}_\varphi)$. For ease of reference, we record the elements of φ as $\varphi^{-1}(1),\dots,\varphi^{-1}(t)$, where $t\doteq\|\varphi\|$, which allows us to think of φ as a map with domain $\cup_{A\in\varphi}A$, codomain $[t]$, and an image of size $|\varphi|_{\mathrm{s}}\le\|\varphi\|=t$, sending elements of $\varphi^{-1}(i)$ to i for all $i\in[t]$. We often write the character φ as $\varphi^{-1}(1)|\cdots|\varphi^{-1}(t)$. Two characters φ and ψ of the same size t should be viewed as the same whenever there is a permutation σ of $[t]$ such that $\varphi=\sigma\circ\psi$. The set of all characters on X is denoted by $\mathrm{C}(X)$. There is a natural poset structure on $\mathrm{C}(X)$, in which $\varphi\preceq\psi$ if and only if $\|\varphi\|-|\varphi|_{\mathrm{s}}\ge\|\psi\|-|\psi|_{\mathrm{s}}$ and it holds $\mathrm{R}_\varphi\le\mathrm{R}_\psi$, namely $\mathrm{R}_\psi-\mathrm{R}_\varphi$ is a nonnegative matrix. Note that the poset $\big(\mathrm{C}(X),\preceq\big)$ forms a lattice, which we call the *character lattice* on X. For every pair of characters φ and ψ on X, their *meet* and *join* in $\mathrm{C}(X)$ are designated by $\varphi\wedge\psi$ and $\varphi\vee\psi$, respectively. A multiset of characters on X is called a *character system* on X. We may regard a character system $\mathcal{C}$ on X of size n as a map from $[n]$ to $\mathrm{C}(X)$. Two character systems $\mathcal{C}$ and $\mathcal{D}$ of size n are considered the same if there exists a permutation σ of $[n]$ such that $\mathcal{C}=\mathcal{D}\circ\sigma$.

Let X be a finite set. A *partition* of X is a *full* character on X, namely a set of disjoint subsets of X whose union equals X. The *equality partition* [Bai17] of X, designated by I_X, is the partition such that every part of it is a singleton set. Note that the relation matrix of a character is an identity matrix if and only if the character is obtained from an equality partition by adding some empty sets. A multiset of partitions of X is called a *partition system* on X. If you think of a partition of X as a linear clustering of X, a partition system of size n on X provides an n-dimensional coordinate system on X, bringing to it a grid cell topology. A 2-state partition of X is also called a *split* on X [DHK+12]. A multiset of splits on X is known as a *split system* on X. We use $\mathrm{P}(X)$ for the set of all partitions of X. Observe that $\big(\mathrm{P}(X),\preceq\big)$ forms a sublattice of $\big(\mathrm{C}(X),\preceq\big)$, called the *partition lattice* on X.

Given positive integers $a_1,\dots,a_n$, a map $M\in\{0,1\}^{[a_1]\times\cdots\times[a_n]}$ is called an n-dimensional $(0,1)$ *tensor/array* (or $(0,1)$ n-tensor/array) of size $a_1\times\cdots\times a_n$ [Tar16]. Let M be an n-dimensional $(0,1)$ array of size $a_1\times\cdots\times a_n$. For any nonempty sets $I_i \subseteq [a_i], i \in [n]$, the restriction of M to $I_1 \times \cdots \times I_n$ is a *subarray* of M. Pick $d \in [n]$ and $\{i_1, \ldots, i_d\} \in \binom{[n]}{d}$. Let $M^{i_1,\ldots,i_d}$ be the d-dimensional $(0,1)$ array of size $a_{i_1} \times \cdots \times a_{i_d}$ such that

$$M^{i_1,\ldots,i_d}(\alpha_{i_1}, \ldots, \alpha_{i_d}) = 0 \quad \text{if and only if} \quad \sum_{\substack{\alpha_j \in [a_j] \\ j \notin \{i_1,\ldots,i_d\}}} \mathrm{M}(\alpha_1, \ldots, \alpha_n) = 0.$$

We call $M^{i_1,\ldots,i_d}$ the d-dimensional *projection/marginalization* of M to $\{i_1, \ldots, i_d\}$. People also say $M^{i_1,\ldots,i_d}$ is obtained by contracting the index set $[n] \setminus \{i_1, \ldots, i_d\}$ from M, namely by multiplying a corresponding all-ones Boolean tensor indexed by $[n] \setminus \{i_1, \ldots, i_d\}$.

Let n be a positive integer and let $\varphi_1, \ldots, \varphi_n$ be characters on a common ground set. The *intersection array* of $\varphi_1, \ldots, \varphi_n$, denoted by $\mathrm{M}_{\varphi_1,\ldots,\varphi_n}$, is the $(0,1)$ n-tensor in the set $\{0,1\}^{[\|\varphi_1\|] \times \cdots \times [\|\varphi_n\|]}$ such that

$$\mathrm{M}_{\varphi_1,\ldots,\varphi_n}(\alpha_1, \ldots, \alpha_n) = \begin{cases} 1, \text{ if } \varphi_1^{-1}(\alpha_1) \cap \cdots \cap \varphi_n^{-1}(\alpha_n) \neq \emptyset; \\ 0, \text{ otherwise.} \end{cases}$$

Remark 1 It is easy to see that an n-dimensional $(0,1)$ array M can be realized as $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ for some partition system $\{\{\pi_1, \ldots, \pi_n\}\}$ of size n. This means that many results can be stated in terms of partition systems and $(0,1)$ tensors interchangeably. For every $i \in [n]$, let us note that π_i does not contain the empty set as a part if and only if the corresponding one-dimensional projection of $M = \mathrm{M}_{\pi_1,\ldots,\pi_n}$ obtained by contracting $[n] \setminus \{i\}$ is an all-ones vector.

Remark 2 Let $\{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system. Pick $\{i_1, \ldots, i_d\} \in \binom{[n]}{d}$, one can easily check that $\mathrm{M}^{i_1,\ldots,i_d}_{\pi_1,\ldots,\pi_n} = \mathrm{M}_{\pi_{i_1},\ldots,\pi_{i_d}}$.

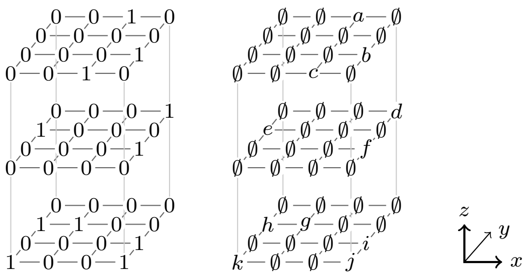

Fig. 1: A 3-dimensional array of size $4 \times 4 \times 3$ and a corresponding partition system.

Example 1 Let M be the 3-dimensional $(0,1)$ array on the left of Fig. 1. By replacing each 1-entry in M by a letter as in the middle of Fig. 1, we see that $M = \mathrm{M}_{\pi_x,\pi_y,\pi_z}$, where $\pi_x = ehk|g|ac|bdfij, \pi_y = ckj|bfi|egh|ad$, and $\pi_z = ghijk|def|abc$ are partitions of $X = \{a,b,c,d,e,f,g,h,i,j,k\}$. Note that

$$M^{x,y} = \mathrm{M}_{\pi_x,\pi_y} = \begin{pmatrix} 1\,0\,1\,0 \\ 0\,0\,1\,0 \\ 1\,0\,0\,1 \\ 1\,1\,0\,1 \end{pmatrix},$$

$$M^{x,z} = \mathrm{M}_{\pi_x,\pi_z} = \begin{pmatrix} 0\,1\,1 \\ 0\,0\,1 \\ 1\,0\,0 \\ 1\,1\,1 \end{pmatrix} \quad \text{and} \quad M^{y,z} = \mathrm{M}_{\pi_y,\pi_z} = \begin{pmatrix} 1\,0\,1 \\ 1\,1\,1 \\ 0\,1\,1 \\ 1\,1\,0 \end{pmatrix}.$$

1.2 Intersection graph and Buneman graph

A graph G is a pair of sets $\big(\mathrm{V}(G), \mathrm{E}(G)\big)$, where $\mathrm{V}(G)$ is the vertex set of G and $\mathrm{E}(G) \subseteq \binom{\mathrm{V}(G)}{2}$ is the edge set of G.

Let $\mathcal{C} = \{\{\varphi_1, \ldots, \varphi_n\}\}$ be a character system on a finite set X. The *intersection graph* of $\mathcal{C}$, denoted by $\mathrm{int}(\mathcal{C})$, is the graph with $\{(i,A):\ i \in [n], A \in \varphi_i\}$ as its vertex set and with $\big\{\{(i,A),(j,B)\} \in \binom{\mathrm{V}(\mathrm{int}(\mathcal{C}))}{2}:\ i \neq j \text{ and } A \cap B \neq \emptyset\big\}$ as its edge set. The intersection graph $\mathrm{int}(\mathcal{C})$ has a *restricted chordal completion* provided there exists a set $E' \subseteq \big\{\{(i,A),(j,B)\} \in \binom{\mathrm{V}(\mathrm{int}(\mathcal{C}))}{2}:\ i \neq j\big\}$ such that $(\mathrm{V}(\mathrm{int}(\mathcal{C})), \mathrm{E}(\mathrm{int}(\mathcal{C})) \cup E')$ is a chordal graph. The *Buneman graph* [Bun71] of a character system $\mathcal{C}$ on X is the graph $\mathcal{B}(\mathcal{C})$ satisfying

$$\mathrm{V}\left(\mathcal{B}(\mathcal{C})\right) = \{\alpha \in (2^X)^{\mathcal{C}}:\ \alpha(\varphi) \in \varphi, \alpha(\varphi) \cap \alpha(\psi) \neq \emptyset \text{ for all } \varphi, \psi \in \mathcal{C}\}$$

and

$$\mathrm{E}\left(\mathcal{B}(\mathcal{C})\right) = \big\{\{\alpha,\beta\}:\ \|\{\{\varphi \in \mathcal{C}:\ \alpha(\varphi) \neq \beta(\varphi)\}\}\| = 1\big\}.$$

Remark 3 Let X be a finite set and let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$ be a partition system on X. From the intersection array $\mathrm{M}_{\mathcal{C}} = \mathrm{M}_{\pi_1,\pi_2}$ of $\mathcal{C}$, we can construct a $(0,1)$ matrix

$$A = \begin{pmatrix} 0 & \mathrm{M}_{\mathcal{C}} \\ \mathrm{M}_{\mathcal{C}}^{\top} & 0 \end{pmatrix}, \tag{1}$$

and then form a graph G whose adjacency matrix is A. Up to isomorphism, it is evident that the intersection graph $\mathrm{int}(\mathcal{C})$ of $\mathcal{C}$ is just G and that the Buneman graph $\mathcal{B}(\mathcal{C})$ of $\mathcal{C}$ is the line graph $\mathrm{L}(G)$ of G. We summarize these observations in Fig. 2.

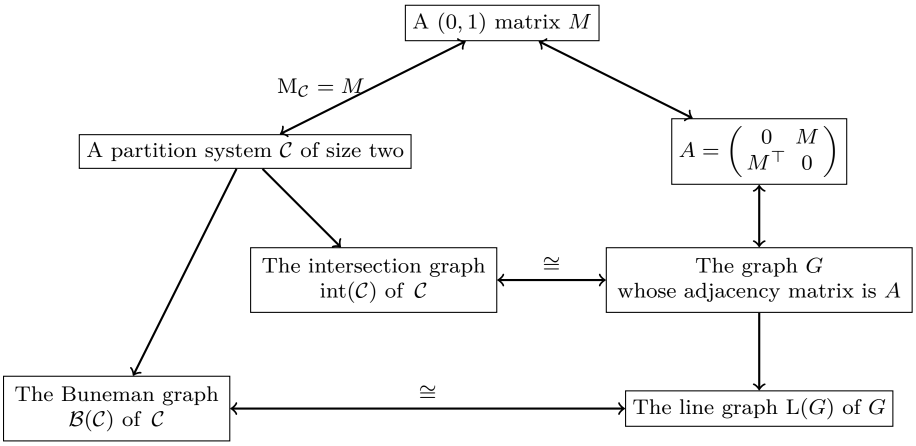

Fig. 2: Various presentations of a size-2 partition system.

Example 2 Let $X = \{a, b, c, d, e, f, g, h, i, j\}$ and let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$ be a partition system on X where $\pi_1 = a|bcd|ef|ghij$ and $\pi_2 = ab|c|deg|f|h|i|j$. The intersection array of $\mathcal{C}$ is

$$\mathrm{M}_{\mathcal{C}} = \mathrm{M}_{\pi_1,\pi_2} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 5 & 6 & 7 & 8 & 9 & 10 & 11 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \end{array}. \tag{2}$$

Let A be the $(0,1)$ matrix specified by Eq. (1) and let G be the graph whose adjacency matrix is A. On the left of Fig. 3 we display the graph G, while on its right we depict the line graph $\mathrm{L}(G)$. As predicted in Remark 3, G is isomorphic to the intersection graph of $\mathcal{C}$ and $\mathrm{L}(G)$ is isomorphic to the Buneman graph $\mathcal{B}(\mathcal{C})$ of $\mathcal{C}$.

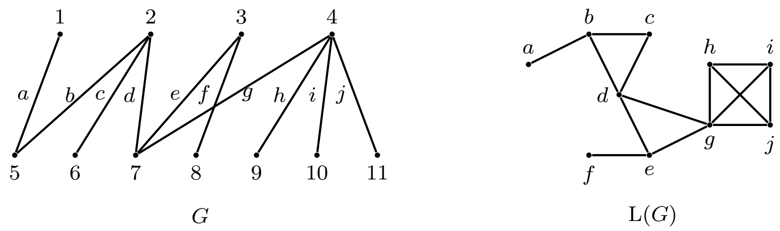

Fig. 3: A graph and a line graph generated from (2).

A *forest* is an acyclic graph and a *tree* is a connected forest. Fix a graph G. For any $E \subseteq \mathrm{E}(G)$, denote by $G - E$ the graph whose vertex set is $\mathrm{V}(G)$ and whose edge set is $\mathrm{E}(G) \setminus E$. A vertex of G with degree at most one is called a *leaf* of G. A set $V' \subseteq \mathrm{V}(G)$ is *convex* in G provided it contains all vertices lying in geodesic paths in G joining elements of V'. For any $V' \subseteq \mathrm{V}(G)$, the *convex hull* of V' in G, denoted by $\mathrm{Conv}_G(V')$, is the smalles convex set which contains V'.

Let X be a finite set. An *X-labeled graph* $\mathcal{G}$ is a tuple (G, ℓ_G) such that G is a graph and ℓ_G is a map from X to $\mathrm{V}(G)$. When G is a tree, we call the X-labeled graph (G, ℓ_G) an *X-labeled tree*. An X-labeled tree (T, ℓ_T) is called an *X-tree* if the image set of ℓ_T contains all the vertices of T with degree at most two. A *phylogenetic* X-tree is an X-tree (T, ℓ_T) in which ℓ_T is a bijection from X to the set of leaves of T. A *fully labeled X-tree* is an X-tree (T, ℓ_T) such that the image of ℓ_T is $\mathrm{V}(T)$. For an X-labeled graph $\mathcal{G} = (G, \ell_G)$, we say that $\mathcal{G}$ *displays* a character φ on X if the convex hulls $\mathrm{Conv}_G\big(\ell_G(A)\big)$ and $\mathrm{Conv}_G\big(\ell_G(B)\big)$ are disjoint for every two states A and B of φ. We say that a character system on X can be displayed on an X-labeled graph $\mathcal{G}$ if $\mathcal{G}$ displays every character in the character system. A character system on X is said to have a *perfect phylogeny* [SS03, Ste16] if there exists a phylogenetic X-tree displaying it. A character system having a perfect phylogeny is also called *compatible* [DS92, EM77, HMSW14, HT11, Ste92, SVFB13]. Let us mention that for any character system $\mathcal{C}$ on X, the following are equivalent:

- There is an X-labeled tree displaying $\mathcal{C}$;
- There is an X-tree displaying $\mathcal{C}$;
- There is a perfect phylogeny for $\mathcal{C}$.

The *perfect phylogeny problem* is to determine if a given character system is compatible. For any positive integer r, the *r-state perfect phylogeny problem* [LGS11] is to decide the existence of a perfect phylogeny for a given character system whose members are r-state characters on a common finite set.

The general perfect phylogeny problem is NP-complete [Ste92]. So is the problem of determining if a given perfect phylogeny is unique for a given character system [HS13]. For any fixed parameter r, the r-state perfect phylogeny problem is polynomial time solvable [AFB94, BFW92, MWW94].

Theorem 1 *[EM77, Theorem] A character system of size two is compatible if and only if its intersection graph is acyclic.*

Theorem 2 *[DHK+12, Theorem 3.3] A split system is compatible if and only every two-element subset of it is compatible.*

Here is a common generalization of Theorems 1 and 2.

Theorem 3 *[Bun74, Theorem 2.1, Theorem 2.7][Ste16, Theorem 4.2.2] A character system is compatible if and only if its intersection graph has a restricted chordal completion. Moreover, given a restricted chordal completion* *for the intersection graph of a character system, a perfect phylogeny of the character system can be constructed in linear time [HM91].*

Let X be a finite set and let A and B be two subsets of X. We say that A and B *crosses* each other, denoted by $A \ddagger B$, if the four sets $A \cap B, A \setminus B, B \setminus A, X \setminus (A \cup B)$ are all nonempty. Let π and ρ be two partitions of X. We call $A \in \pi$ a *dependent state* of π with *witness* ρ whenever we can find two different states $B, C \in \rho$ such that both $A \ddagger B$ and $A \ddagger C$ happen. For any partition system $\mathcal{C}$ on X, we call a state A of $\pi \in \mathcal{C}$ a *dependent state* of π with witness $\mathcal{C}$ provided it is a dependent state of π with some $\rho \in \mathcal{C}$ as a witness.

Theorem 4 *[SFB12, Theorem 7] Let X be a finite set and let $\mathcal{C}$ be a family of 3-state partitions on X. Then there is a perfect phylogeny for $\mathcal{C}$ if and only if every element of $\mathcal{C}$ has at most one dependent state with $\mathcal{C}$ as a witness.*

For any positive integer r, let $\theta(r)$ be the maximum integer k such that there exists a partition system $\mathcal{C}$ consisting of some partitions with at most r states on a set such that $\mathcal{C}$ is not compatible but all subsystems of $\mathcal{C}$ of size at most $k-1$ are compatible. Note that Theorem 2 says nothing but $\theta(2) = 2$.

Theorem 5 *(i) ([Mea83], [LGS11, Theorem 7.1]) It holds $\theta(r) \geq r$ for every $r \geq 2$.*
(ii) ([LGS11, Theorem 2.4]) $\theta(3) = 3$.
(iii) ([HT11, Section 4]) $\theta(4) \geq 5$.
(iv) ([SVFB13, Corollary 2]) It holds $\theta(r) \geq \lfloor \frac{r}{2} \rfloor \times \lceil \frac{r}{2} \rceil + 1$ for every $r \geq 2$.

One may be tempted to conjecture that $\theta(r)$ is finite for all r. This is recently refuted by Van Iersel, Jones and Kelk [IJK19].

Theorem 6 *[IJK19, Theorem 20] $\theta(8) = +\infty$.*

1.4 Sparse properties

The leading role of our study is the sparse properties of character systems as defined below (Definition 1). The purpose of this work is to use this concept as a microscope to observe various slices at different scales of a sparse $(0,1)$ array and to elaborate how this concept can give birth to a rich line of combinatorial problems and some nice answers.

Definition 1 (Sparse character system) Let X be a finite set and let $\mathcal{C} = \{\{\varphi_1, \ldots, \varphi_n\}\}$ be a character system on X. Take $k \in [n]$. We say that $\mathcal{C}$ has the property Q_k provided

$$|\varphi_{i_1}[Y] \wedge \cdots \wedge \varphi_{i_k}[Y]|_{\mathrm{s}} \leq \sum_{j=1}^{k} \left(|\varphi_{i_j}[Y]|_{\mathrm{s}} - 1 \right) + 1 \qquad (3)$$

holds for every $\{i_1, \ldots, i_k\} \in \binom{[n]}{k}$ and every $Y \in 2^X \setminus \{\emptyset\}$. We say that $\mathcal{C}$ has the property $\overline{\mathrm{Q}}_k$ provided

$$|\varphi_{i_1} \wedge \cdots \wedge \varphi_{i_k}|_{\mathrm{s}} = \sum_{j=1}^{k} \left(\|\varphi_{i_j}\| - 1\right) + 1$$

holds for every $\{i_1, \ldots, i_k\} \in \binom{[n]}{k}$.

It is obvious that every character system has the property Q_1. For a 3-state partition system $\mathcal{C}$, we can check that $\mathcal{C}$ has the property Q_2 if each partition in $\mathcal{C}$ has at most one dependent state with witness $\mathcal{C}$, and hence, in view of Theorem 4, when $\mathcal{C}$ is compatible.

Definition 2 (Sparse tensor) Let M be an n-dimensional $(0,1)$ array of size $a_1 \times \cdots \times a_n$. We say that M has the *sparse property* Q_n if

$$\sum_{(\alpha_1,\ldots,\alpha_n) \in I_1 \times \cdots \times I_n} M(\alpha_1, \ldots, \alpha_n) \leq \sum_{j=1}^{n} (|I_j| - 1) + 1 \tag{4}$$

holds for all nonempty sets $I_1 \subseteq [a_1], \ldots, I_n \subseteq [a_n]$, and we say that M has the *sparse property* $\overline{\mathrm{Q}}_n$ if equality holds in (4) when $I_1 = [a_1], \ldots, I_n = [a_n]$. Let $k \in [n]$. We say that M has the *sparse property* Q_k if every k-dimensional projection of M has the sparse property Q_k, and we say that M has the *sparse property* $\overline{\mathrm{Q}}_k$ if every k-dimensional projection of M has the sparse property $\overline{\mathrm{Q}}_k$.

Let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system. Since we mainly concern ourselves with the sparse property which is irrelevant with how we order the elements in the multiset $\mathcal{C}$, we call $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ the intersection array of $\mathcal{C}$ rather than the intersection array of the sequence $\pi_1, \ldots, \pi_n$, and we even write $\mathrm{M}_{\mathcal{C}}$ for $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ when no confusion can arise. We often follow the same convention when we are focusing on a property of some construction from the sequence $\pi_1, \ldots, \pi_n$ which is not affected by any change of their ordering, that is, we abuse the notation by putting $\mathcal{C}$ as the index of that construction.

Remark 4 A sparse graph is a graph in which the number of edges is a small multiple of the number of vertices [Spi17]. Analogously, we need to control from above the number of 1-entries in a $(0,1)$ array so that we can call it "sparse". Regarding the way of measuring sparsity as in Definitions 1 and 2, we like to point out that

$$\Omega \doteq \sum_{j=1}^{n} (|I_j| - 1) + 1$$

is the dimension of the linear space of all functions f on the disjoint union of $I_j, j \in [n]$, such that

$$\sum_{i \in I_1} f(i) = \cdots = \sum_{i \in I_n} f(i).$$

For each set I, let $\mathcal{S}^I$ be the probability simplex on I, namely $\mathcal{S}^I$ consists of all nonnegative functions on I that sum to 1. Let $\mathcal{S} \doteq \mathcal{S}^{I_1} \times \cdots \times \mathcal{S}^{I_n} \subseteq \mathbb{R}^{I_1 \cup \cdots \cup I_n}$ where we view $I_1, \ldots, I_n$ as disjoint sets. By Carathéodory's Theorem, each element in $\mathcal{S}$ is a convex combination of at most Ω integral points from $\mathcal{S}$. When all $|I_j|$ are equal, this is related to a fact on fuzzy partitions [BH79, Theorem 1]. By coincidence, Ω appears in some other context, say [AHJ19, Theorem 1.7], when all $|I_j|$ are equal.

Remark 5 Let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system and let $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ be the intersection array of $\pi_1, \ldots, \pi_n$. It is straightforward that

$$|\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} = \sum_{(\alpha_1,\ldots,\alpha_n) \in [\|\pi_1\| \times \cdots \times [\|\pi_n\|]} \mathrm{M}_{\pi_1,\ldots,\pi_n}(\alpha_1, \ldots, \alpha_n).$$

For each $k \in [n]$, we read from the above equation that $\mathcal{C}$ has the property $\overline{\mathrm{Q}}_k$ if and only if $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ has the property $\overline{\mathrm{Q}}_k$.

Theorem 7 *Let n and k be integers such that $k \in [n]$, let X be a finite set, and let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system on X. The following are equivalent.*

(i) The partition system $\mathcal{C}$ has the sparse property Q_k.
(ii) The intersection array $\mathrm{M}_{\pi_1,\ldots,\pi_n}$ has the sparse property Q_k.

Owing to Remark 5 and Theorem 7, we see that Definition 2 is consistent with Definition 1 and so is really about the meet semilattice generated by the corresponding partition system.

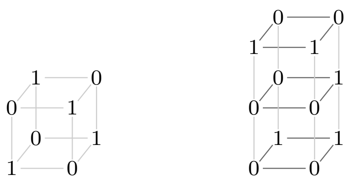

Fig. 4: Two 3-dimensional $(0, 1)$ arrays.

Example 3 Let $X = [6]$ and let $\mathcal{C} = \{\{123|456, 245|136, 5|24|136\}\}$. The intersection array $\mathrm{M}_{\mathcal{C}}$ of $\mathcal{C}$ is displayed on the right of Fig. 4. One can see that both $\mathcal{C}$ and $\mathrm{M}_{\mathcal{C}}$ satisfy Q_3 but do not have the property Q_2, which coincides with the claim in Theorem 7.

Example 4 For every $(0, 1)$ matrix M, it is not difficult to prove that the matrix

$$\begin{pmatrix} 0 & M \\ M^\top & 0 \end{pmatrix}$$

is the adjacency matrix of a tree if and only if M has the properties Q_2 and $\overline{\mathrm{Q}}_2$. Those readers who do not bother to do the above exercise can simply check Example 2 to convince himself/herself the validity of our statement.

Theorem 8 *Let M be a 2-dimensional $(0,1)$ array of size $a \times b$. The following are equivalent.*

(*i*) *The array M has the sparse property Q_2.*
(*ii*) *For all sets $J' \subseteq [a]$ and $J \subseteq [b]$ with $|J'| = |J| > 0$, it holds*

$$\sum_{(\alpha,\beta)\in J'\times J} M(\alpha,\beta) \le 2|J| - 1. \tag{5}$$

Remark 6 Note that Theorem 8 does not hold for arrays of dimension $n \ge 3$. Let $X = [6]$, and let $\pi_1 = 12|34|56, \pi_2 = 135|246$ and $\pi_3 = 145|236$ be three elements of $\mathrm{P}(X)$. It is easy to check that

$$\sum_{(\alpha_1,\alpha_2,\alpha_3)\in I_1\times I_2\times I_3} \mathrm{M}_{\pi_1,\pi_2,\pi_3}(\alpha_1,\alpha_2,\alpha_3) \le 3(|I| - 1) + 1$$

is valid for every $I_1 \subseteq [3]$, $I_2 \subseteq [2]$ and $I_3 \subseteq [2]$ such that $|I_1| = |I_2| = |I_3| > 0$. Since

$$\sum_{(\alpha_1,\alpha_2,\alpha_3)\in[3]\times[2]\times[2]} \mathrm{M}_{\pi_1,\pi_2,\pi_3}(\alpha_1,\alpha_2,\alpha_3) = 6 > 5 = (3-1)+(2-1)+(2-1)+1,$$

the partition system $\{\{\pi_1, \pi_2, \pi_3\}\}$ does not have the property Q_3.

Algorithm 1 Check the sparse property Q_2 for a given matrix

Input: a $(0,1)$ matrix M
1: **while** $M \ne M'$ **do**
2: $\quad M \doteq M'$
3: $\quad M' =$ the matrix obtained from M by deleting those columns and rows of M with at most one nonzero entry
4: **end while**
Output YES! (M satisfies Q_2) if M' vanishes; NO! (M does not satisfy Q_2) otherwise.

Algorithm 1 is a recognition algorithm for the Q_2 property. The correctness of this algorithm can be directly verified. A forest is characterized by the property that it can be reduced to the empty graph by repeatedly removing leaves. As we will report in Remark 7 (1), this characterization of forests is essentially the same with the characterization of Q_2 in terms of Algorithm 1. It worths noting that, even though Theorem 8 and Algorithm 1 suggest that matrices with the Q_2 property are very tree-like, we should not think that we have understood everything about them, not mentioning those general $(0,1)$ arrays having the Q_2 property. See a result about the Q_2 property in 2 of Section 3.

For each character system $\mathcal{C}$, the properties Q_k and $\overline{\mathrm{Q}}_k$ are properties about its k-dimensional margins, namely they are local properties about k-wise relationships for elements of $\mathcal{C}$. We reveal some relationships between properties of different margins of a common high-dimensional data set in the next result.

Theorem 9 *Let M be a $(0,1)$ n-tensor and let k be an integer with $3 \leq k \leq n$.*

(i) *If M has the sparse property Q_{k-1}, then M has the sparse property Q_k.*
(ii) *If M satisfies Q_k and if the one dimensional projection M^n of M to $\{n\}$ has at most one 1-entry, then the $n-1$ dimensional projection $M^{1,\ldots,n-1}$ of M to $[n-1]$ has the property Q_{k-1}.*
(iii) *If M has $n+1$ 1-entries, it can have the sparse property Q_n but fail to have the sparse property Q_{n-1}.*
(iv) *If M has the sparse properties Q_{n-1} and $\overline{\mathrm{Q}}_n$, then M has the sparse properties Q_ℓ for all $\ell \in [n]$.*
(v) *If M has the sparse properties Q_{n-1} and $\overline{\mathrm{Q}}_n$, then M satisfies $\overline{\mathrm{Q}}_\ell$ for all $\ell \in [n]$.*

Example 5 Let $X = [4]$ and let $\mathcal{C} = \{\{\pi_1, \pi_2, \pi_3\}\}$ be a partition system on X, where $\pi_1 = 12|34$, $\pi_2 = 13|24$, $\pi_3 = 14|23$. The intersection array $\mathrm{M}_{\mathcal{C}}$ of $\mathcal{C}$ can be drawn as the left of Fig. 4. It is evident that $\mathcal{C}$ satisfies Q_3 but does not satisfy Q_2. Note that this construction is different from the one used in the proof of Theorem 9 (iii) in Section 2.3.

Example 6 Let $X = [5]$ and let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$ be a partition system on X, where $\pi_1 = 12|34|5$, $\pi_2 = 13|24|5$. The intersection array of $\mathcal{C}$ is

$$\mathrm{M}_{\pi_1,\pi_2} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $\mathcal{C}$ satisfies Q_1 and $\overline{\mathrm{Q}}_2$ but not Q_2. Compare this with Theorem 9 (iv).

Theorem 10 *For any partition system $\mathcal{C}$ of size $n \geq 3$ and satisfying the property $\overline{\mathrm{Q}}_n$, exactly one of the following holds:*

- *The partition system $\mathcal{C}$ has the properties $\overline{\mathrm{Q}}_k$ and Q_k for all $k \in [n]$;*
- *The partition system $\mathcal{C}$ has the property Q_k for no $k \in \{2, \ldots, n-1\}$.*

1.5 Sparse completion and compatible partition system

Theorems 3 and 4 characterize compatible character systems via the concepts of restricted chordal completition and dependent states, respectively. The compatibility problem for a collection of phylogenetic trees, which is very close to the compatibility problem for character system as addressed here, has been widely studied [BL06, FBV18, GKL15]. Let us mention that, via the concept of reduced display graph, Fernández-Baca and Vakati [FBV18] produced characterizations of compatibility of trees based on triangulations and tree decompositions. In the same vein, we will see that perfect phylogeny is characterized by some sparse properties.

Definition 3 (Sparse completion) Let M and M' be n-dimensional $(0,1)$ arrays of equal size. The array M' is called a *sparse completion* of M provided $M' - M$ is a $(0,1)$ array and that M' satisfies

$$\begin{cases} \overline{\mathrm{Q}}_1, & \text{if } n = 1; \\ \mathrm{Q}_2 \text{ and } \overline{\mathrm{Q}}_n, & \text{if } n \geq 2. \end{cases} \tag{6}$$

Let $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ be a partition system on X and $\mathcal{C}' = \{\{\pi'_1, \dots, \pi'_n\}\}$ be a partition system on $X' \supseteq X$. We call $\mathcal{C}'$ a *sparse completion* of $\mathcal{C}$ if $\mathrm{M}_{\mathcal{C}'}$ has the property specified in (6) and that $\pi'_i[X] = \pi_i$ for all $i \in [n]$ – consequently, $\mathrm{M}_{\mathcal{C}'}$ is a sparse completion of $\mathrm{M}_{\mathcal{C}}$.

The next result links perfect phylogeny with the sparse properties specified in Definitions 1 and 3. Observe the resemblence between Theorem 3 and Theorem 11 (ii). Also note that Theorem 11 strengthens the claims about the Q_2 property given after Definition 1.

Theorem 11 *Let X be a finite set, and let $\mathcal{C}$ be a partition system on X of size n.*

(i) If $n = 2$, then $\mathcal{C}$ is compatible if and only if $\mathcal{C}$ has the sparse property Q_2.
(ii) The partition system $\mathcal{C}$ is compatible if and only if $\mathcal{C}$ has a sparse completion.

Example 7 Let $X = \{x, y, z, p, q\}$ and let $\mathcal{C} = \{\{xy|q|zp, xq|y|zp, x|yz|pq\}\}$. Note that $\mathcal{C}$ has the sparse property Q_2, but there is no sparse completion for it. Indeed, $\mathcal{C}$ does not have a perfect phylogeny, but it can be displayed on the circular split network [DHK+12, HRS10] as depicted in Fig. 5.

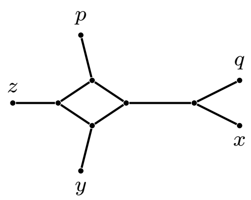

Fig. 5: A labeled circular split system.

Remark 7 Let the character system $\mathcal{C}$ and the graph $G \cong \mathrm{int}(\mathcal{C})$ be as given in Remark 3.

(1) By Theorem 1 and Theorem 11, G is a forest if and only if $\mathcal{C}$ is compatible if and only if $\mathrm{M}_{\mathcal{C}}$ has the sparse property Q_2 and if and only if $\mathrm{M}_{\mathcal{C}}$ has a sparse completion. Especially, checking if a $(0,1)$ matrix $M = \mathrm{M}_{\mathcal{C}}$ satisfies Q_2 amounts to checking whether or not the graph G is acyclic. Note that the adjacency matrix of G is specified by Eq. (1).

(2) Remembering Example 4 and Definition 3, we can further find that G is a tree if and only if $\mathrm{M}_{\mathcal{C}}$ has the sparse properties Q_2 and $\overline{\mathrm{Q}}_2$ and if and only if $\mathcal{C}$ is a sparse completion of itself.

Example 8 We have discussed how to display a partition system on X with an X-labeled tree, namely a tree with a vertex-labeling map. Instead of labeling vertices, we can also label edges. In Fig. 3 we have encountered an edge-labeled tree G.

We think of the edges of G as "sticks" glued together by vertices. If S is a set of vertices of G, we can think of removing the glue at all vertices in S, breaking the tree into pieces (we will see more than one pieces unless all vertices in S are pendant), thereby inducing a partition of the edges of G. We call this the partition induced by S. When the splitting set S is $\{5,7\}$, we can see the partition $\pi_1 = a|bcd|ef|ghij$; When the splitting set S is $\{2,3\}$ instead, the induced partition becomes $\pi_2 = ab|c|deg|f|h|i|j$.

Remark 8 Let $\mathcal{C} = \{\{\varphi_1, \varphi_2\}\}$ be a partition system on a finite set X. For every $x \in X$, there exists a unique edge $e_x = \{(1, A), (2, B)\}$ of $\mathrm{int}(\mathcal{C})$ such that $x \in A \cap B$. Let ℓ_T be the labeling map sending every element x of X to the corresponding unique edge e_x. Note that $\big(\mathrm{int}(\mathcal{C}), \ell_T\big)$ is an edge-labeled graph displaying $\mathcal{C}$. Moreover, from Theorem 1 we derive that $\mathcal{C}$ can be displayed on an edge-labeled tree if and only if $\mathcal{C}$ is compatible.

For any fixed r, can we tell if an n-dimensional $(0,1)$ array of size $\overbrace{r \times \cdots \times r}^{n}$ possess a sparse completion by examining its local projections? Based on those results on perfect phylogeny as seen in Section 1.3, we find that the answer depends on the size of r.

Theorem 12 *(i) Let M be a $(0,1)$ n-tensor of size $\overbrace{2 \times \cdots \times 2}^{n}$. If M^{i_1,i_2} has a sparse completion for every $\{i_1, i_2\} \in \binom{[n]}{2}$, then M has a sparse completion.*

(ii) Let M be a $(0,1)$ n-tensor of size $\overbrace{3 \times \cdots \times 3}^{n}$. If M^{i_1,i_2,i_3} has a sparse completion for every $\{i_1, i_2, i_3\} \in \binom{[n]}{3}$, then M has a sparse completion.

(iii) Let r be a positive integer with $4 \le r \le 7$ and let $d = \lfloor \frac{r}{2} \rfloor \times \lceil \frac{r}{2} \rceil$. There exists a positive integer n and a $(0,1)$ n-tensor M of size $\overbrace{r \times \cdots \times r}^{n}$ such that M does not have a sparse completion but $M^{i_1,\ldots,i_d}$ has a sparse completion for every $\{i_1, \ldots, i_d\} \in \binom{[n]}{d}$.

(iv) For any positive integers d and r with $r \ge 8$, there exists a positive integer n and a $(0,1)$ n-tensor M of size $\overbrace{r \times \cdots \times r}^{n}$ such that M does not have a sparse completion but $M^{i_1,\ldots,i_d}$ has a sparse completion for every $\{i_1, \ldots, i_d\} \in \binom{[n]}{d}$.

1.6 Minimum displaying tree

To represent the possible evolutionary processes underlying a character system, several constructions have been proposed in the literature, say Buneman graphs [Bun71], relation graphs [HM02], quasi-median graphs [BHM02, HMS04], and minimum average-distance clique trees of restricted chordal completions [XGG15].

Algorithm 2 From a character system to a labeled graph (CG)

Input: A character system $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ on X.

1: Let Y be the set of states of the character $\pi_1 \wedge \cdots \wedge \pi_n$.

2: For each $i \in [n]$, let T_i be a forest satisfying that $\mathrm{V}(T_i) = Y$ and that a subset Y' of Y is a connected component of T_i if and only if $\bigcup_{A \in Y'} A$ is a state of the character $\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n$.

3: Let $\bar{T}$ be the graph such that $\mathrm{V}(\bar{T}) = Y$ and $\mathrm{E}(\bar{T}) = \cup_{i \in [n]} \mathrm{E}(T_i)$.

4: Add the minimum number of edges to get a connected graph T such that $\mathrm{V}(T) = \mathrm{V}(\bar{T})$ and $\mathrm{E}(T) \supseteq \mathrm{E}(\bar{T})$.

Output: The graph T and the surjective map ℓ_T from X to $\mathrm{V}(T)$ sending every element in a state A of $\pi_1 \wedge \cdots \wedge \pi_n$ to the vertex A of T.

When $\mathcal{C}$ is a partition system on X, there is a canonical labeling map $\phi_{\mathcal{C}}$ from X to $\mathrm{V}\left(\mathcal{B}(\mathcal{C})\right)$ that sends $x \in X$ to $\phi_{\mathcal{C}}(x) = \alpha \in (2^X)^{\mathcal{C}}$ such that $\alpha(\varphi)$ is the part of φ that contains x for all $\varphi \in \mathcal{C}$ [DHK+12, p. 55]. Along this line of research, we further suggest the CG construction in Algorithm 2, which is a variant of the intersection graph construction of character systems.

Let X be a finite set and let $\mathcal{C}$ be a character system on X. We call an X-tree (T, ℓ_T) a *minimum* X-tree displaying $\mathcal{C}$ if $|\mathrm{V}(T)| \leq |\mathrm{V}(T')|$ holds for all X-trees $(T', \ell_{T'})$ which display $\mathcal{C}$. By the principle of Ockham's razor, minimum displaying trees for a given character system will be preferred as they explain the observed data with as few evolutionary events as possible. The usefulness of the CG construction in reconstructing trees can be seen from the subsequent result.

Theorem 13 *Let X be a finite set, and let $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ be a partition system on X.*

(i) If $n = 2$ and $\mathcal{C}$ is compatible, then an X-tree (T, ℓ_T) is a minimum X-tree displaying $\mathcal{C}$ if and only if (T, ℓ_T) is a possible output of Algorithm 2 applied to $\mathcal{C}$.

(ii) If $\mathcal{C}$ is a sparse completion of $\mathcal{C}$ itself, then an X-tree (T, ℓ_T) is a minimum X-tree displaying $\mathcal{C}$ if and only if (T, ℓ_T) is a possible output of CG *applied to $\mathcal{C}$.*

(iii) If $\mathcal{C}$ is a sparse completion of $\mathcal{C}$ itself, then the Buneman graph $\mathcal{B}(\mathcal{C})$ is a connected graph and the set of all possible outputs of CG *applied to $\mathcal{C}$ consists of $(T, \phi_{\mathcal{C}})$ where T runs through the set of all spanning trees of $\mathcal{B}(\mathcal{C})$.*

In view of the CG construction, all the outputs of Algorithm 2 carry an onto labeling map and so all the minimum X-trees mentioned in Theorem 13 are fully labeled trees.

Example 9 Let $X = [7]$ and let (T, ℓ_T) be the X-labeled tree as shown on the right of Fig. 6. Note that $\mathrm{E}(T)$ is divided into three parts, the set of horizontal (green) edges E_{green}, the set of vertical (blue) edges E_{blue}, and the set of remaining (red) edges E_{red}. Let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$, where $\pi_1 = 124|5|6|3|7$ and $\pi_2 = 1|23|45|67$ are partitions of X. Note that (T, ℓ_T) is a minimum displaying tree for $\mathcal{C}$ and the intersection array of $\mathcal{C}$ is shown on the left of Fig. 6. If we apply Algorithm 2 to $\mathcal{C}$, we see that $T_1 = (\mathrm{V}(T), E_{\text{green}}), T_2 = (\mathrm{V}(T), E_{\text{blue}})$, and $\bar{T} = (\mathrm{V}(T), E_{\text{green}} \cup E_{\text{blue}})$, and so (T, ℓ_T) is a possible output of Algorithm 2. Note that the only edge in E_{red} is outside of $\mathrm{E}(\mathcal{B}(\mathcal{C}))$ and so T is not a subgraph of $\mathcal{B}(\mathcal{C})$.

Fig. 6: A 2-dimensional $(0, 1)$ array and one of its displaying trees.

Example 10 Let $X = [9]$, and let $\pi_1 = 47|5|8|12369, \pi_2 = 45789|3|26|1, \pi_3 = 123|45|6789$ be partitions of X. Then the partition system $\mathcal{C} = \{\{\pi_1, \pi_2, \pi_3\}\}$ has the sparse properties Q_2 and $\overline{\mathrm{Q}}_3$. On the left of Fig. 7 we give the intersection array $\mathrm{M}_{\mathcal{C}}$ while in the middle of Fig. 7 you see its Buneman graph $\mathcal{B}(\mathcal{C})$. The X-labeled tree on the right of Fig. 7 is a spanning tree of $\mathcal{B}(\mathcal{C})$ and is a possible output of Algorithm 2 applied on $\mathcal{C}$.

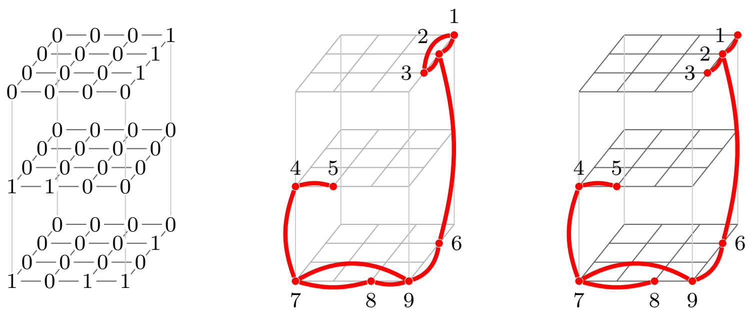

Fig. 7: A $(0, 1)$ array, its Buneman graph and one of its displaying trees.

Example 11 Let the character system $\mathcal{C}$ and graph G be what are presented in Example 2. By Theorem 13 (iii), the set of spanning trees of $\mathrm{L}(G)$, which is demonstrated on the right of Fig. 3, coincides with the set of minimum displaying trees of $\mathcal{C}$.

Remark 9 Let G be a tree and let M be the adjacency matrix of G. From Algorithm 1 (or directly from the definition of Q_2, or from Example 4), we can see that M satisfies Q_2. Whenever $|\,\mathrm{V}(G)| \geq 2$, we can obtain a partition system of size two as in Example 1. According to Theorem 11 (i), this partition system is compatible and so Theorem 13 (i) tells us how to obtain its minimum displaying trees. We are wondering what is the relationship between the topology of these displaying trees and the original tree G.

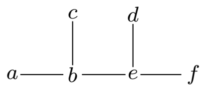

Fig. 8: A tree.

Example 12 Let G be the tree in Fig. 8 and let M be its adjacency matrix. Let $X = \{uv \in \mathrm{V}(G) \times \mathrm{V}(G) :\ \{u, v\} \in \mathrm{E}(G)\}$ and let $\mathcal{C} = \{\{\,\pi_1, \pi_2\,\}\}$ be a partition system on X where

$$\begin{cases} \pi_1 = \{ab\}|\{ba, bc, be\}|\{cb\}|\{de\}|\{eb, ed, ef\}|\{fe\}, \\ \pi_2 = \{ba\}|\{ab, cb, eb\}|\{bc\}|\{ed\}|\{be, de, fe\}|\{ef\}. \end{cases}$$

One can check that $\mathrm{M}_{\mathcal{C}} = M$, and that $\mathcal{C}$ can be displayed on the X-tree in Fig. 9, which is obtained by applying Algorithm 2 to $\mathcal{C}$.

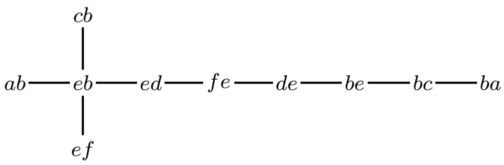

Fig. 9: A displaying tree of the partition system generated by the tree in Fig. 8.

Note that a character $\varphi \in \mathrm{C}(X)$ can be displayed on a phylogenetic X-tree if and only if there is a phylogenetic X-tree displaying the partition $\varphi \vee \mathrm{I}_X$ of X. For each character system $\mathcal{C}$, we call each sparse completion of $\{\{\,\varphi \vee \mathrm{I}_X :\ \varphi \in \mathcal{C}\,\}\}$ a sparse completion $\mathcal{C}$. Combining Theorem 11 (ii) and Theorem 13 (ii), on the condition that we know of a sparse completion of a character system, we can construct a displaying tree of the character system in a running-time which is a polynomial function of both the size of the sparse completion and the size of the ground set of the sparse completion. We mention that Xu, Gysel and Gusfield [XGG15] proposed another method to get compact bushy displaying trees for a compatible character system.

The rest of this paper is organized as follows. We first prove Theorems 7 and 8 in Section 2.1 and then study the perfect phylogeny problem for partition systems of size two in Section 2.2. Thanks to these preparations, we can verify Theorems 9 and 10 in Section 2.3, establish Theorems 11 and 12 in Section 2.4, and finally demonstrate Theorem 13 in Section 2.5. To close the paper, we present some research problems in Section 3.

2 Proofs

2.1 Sparse properties reformulated

Let X be a finite set, let n be a positive integer, and let $\pi_1, \ldots, \pi_n$ be a family of partitions of X. We define $\mathscr{M}_{\pi_1,\ldots,\pi_n}$ to be the map from $2^{[\|\pi_1\|]} \times \cdots \times 2^{[\|\pi_n\|]}$ to $2^{[\|\pi_1\|]\times\cdots\times[\|\pi_n\|]}$ such that

$$\begin{aligned}&\mathscr{M}_{\pi_1,\ldots,\pi_n}(I_1,\ldots,I_n)\\&=\{(\alpha_1,\ldots,\alpha_n)\in I_1\times\cdots\times I_n:\ \pi_1^{-1}(\alpha_1)\cap\cdots\cap\pi_n^{-1}(\alpha_n)\neq\emptyset\}\\&=\{(\alpha_1,\ldots,\alpha_n)\in I_1\times\cdots\times I_n:\ \mathrm{M}_{\pi_1,\ldots,\pi_n}(\alpha_1,\ldots,\alpha_n)=1\}.\end{aligned} \tag{7}$$

Note that

$$\sum_{(\alpha_1,\ldots,\alpha_n)\in I_1\times\cdots\times I_n}\mathrm{M}_{\pi_1,\ldots,\pi_n}(\alpha_1,\ldots,\alpha_n)=|\mathscr{M}_{\pi_1,\ldots,\pi_n}(I_1,\ldots,I_n)| \tag{8}$$

for all $(I_1, \ldots, I_n) \in 2^{[\|\pi_1\|]} \times \cdots \times 2^{[\|\pi_n\|]}$, and that

$$|\pi_1\wedge\cdots\wedge\pi_n|_{\mathrm{s}}=|\mathscr{M}_{\pi_1,\ldots,\pi_n}([\|\pi_1\|],\ldots,[\|\pi_n\|])|. \tag{9}$$

Proof (Proof of Theorem 7) For every $\{i_1, \ldots, i_k\} \in \binom{[n]}{k}$, denote by $f_{i_1,\ldots,i_k}$ the map from $\mathscr{M}_{\pi_{i_1},\cdots,\pi_{i_k}}([\|\pi_{i_1}\|], \ldots, [\|\pi_{i_k}\|])$ to the set of states of the partition $\pi_{i_1} \wedge \cdots \wedge \pi_{i_k}$ such that

$$f_{i_1,\ldots,i_k}(\alpha_1,\ldots,\alpha_k)=\pi_{i_1}^{-1}(\alpha_1)\cap\cdots\cap\pi_{i_k}^{-1}(\alpha_k).$$

Clearly, $f_{i_1,\ldots,i_k}$ is bijective.

We start by showing that (i) implies (ii). Take $\{i_1, \ldots, i_k\} \in \binom{[n]}{k}$ and nonempty sets $I_{i_1} \subseteq [\|\pi_{i_1}\|], \ldots, I_{i_k} \subseteq [\|\pi_{i_k}\|]$. Let

$$Y\doteq\bigcup_{(\alpha_1,\ldots,\alpha_k)\in\mathscr{M}_{\pi_{i_1},\cdots,\pi_{i_k}}(I_{i_1},\ldots,I_{i_k})}f_{i_1,\ldots,i_k}(\alpha_1,\ldots,\alpha_k). \tag{10}$$

If $Y=\emptyset$, then $|\mathscr{M}_{\pi_{i_1},\ldots,\pi_{i_k}}(I_{i_1},\ldots,I_{i_k})| = 0 \le \sum_{j=1}^{k}\left(|I_{i_j}|-1\right)+1$; If $Y\neq\emptyset$, then

$$\begin{aligned} |\mathscr{M}_{\pi_{i_1},\ldots,\pi_{i_k}}(I_{i_1},\ldots,I_{i_k})| &= |\pi_{i_1}[Y]\wedge\cdots\wedge\pi_{i_k}[Y]|_{\mathrm{s}} && \text{By Eq. (10)}\\ &\le \textstyle\sum_{j=1}^{k}\left(\left|\pi_{i_j}[Y]\right|_{\mathrm{s}}-1\right)+1 && \text{By (3)}\\ &\le \textstyle\sum_{j=1}^{k}\left(|I_{i_j}|-1\right)+1. && \text{By Eq. (10)} \end{aligned}$$

In both cases, Eq. (8) enables us deduce (ii).

We are now going to prove (i) by assuming the truth of (ii). Pick $\{i_1,\ldots,i_k\}\in\binom{[n]}{k}$ and $Y\in 2^X\setminus\{\emptyset\}$. Let $S=\{A\in\pi_{i_1}\wedge\cdots\wedge\pi_{i_k}:\ A\cap Y\neq\emptyset\}$, and let

$$I_{i_j} = \bigcup_{(i_1,\ldots,i_k)\in f^{-1}_{i_1,\ldots,i_k}(S)} \{i_j\} \tag{11}$$

for each $j\in[k]$. It is easy to verify that

$$\begin{aligned} &|\pi_{i_1}[Y]\wedge\cdots\wedge\pi_{i_k}[Y]|_{\mathrm{s}}\\ \le\ & |\mathscr{M}_{\pi_{i_1},\ldots,\pi_{i_k}}(I_{i_1},\ldots,I_{i_k})| && \text{By Eq. (11)}\\ =\ & \textstyle\sum_{(\alpha_1,\ldots,\alpha_k)\in I_{i_1}\times\cdots\times I_{i_k}} \mathrm{M}_{\pi_{i_1},\ldots,\pi_{i_k}}(\alpha_1,\ldots,\alpha_k) && \text{By Eq. (8)}\\ \le\ & \textstyle\sum_{j=1}^{k}\left(|I_{i_j}|-1\right)+1 && \text{By (4)}\\ =\ & \textstyle\sum_{j=1}^{k}\left(\left|\pi_{i_j}[Y]\right|_{\mathrm{s}}-1\right)+1, && \text{By Eq. (11)} \end{aligned}$$

finishing the proof. □

Proof (Proof of Theorem 8) It is trivial that (i) implies (ii). Now we turn to the other direction. Take nonempty sets $I\subseteq[a]$ and $J\subseteq[b]$. Because of symmetry, we may assume that $|I|=i\ge j=|J|$. Enumerate the elements of I as $\alpha_1,\ldots,\alpha_i$ such that, for every $k\in[i-1]$, it holds $\sum_{\beta\in J}M(\alpha_k,\beta)\le\sum_{\beta\in J}M(\alpha_{k+1},\beta)$. This along with (5) for $J'=\{\alpha_{i-j+1},\ldots,\alpha_i\}$ implies

$$\sum_{\beta\in J} M(\alpha_k,\beta)\le 1 \tag{12}$$

for all $k\in[i-j+1]$. Now we can obtain

$$\begin{aligned} &\textstyle\sum_{(\alpha,\beta)\in I\times J} M(\alpha,\beta)\\ =\ & \textstyle\sum_{k=1}^{i-j}\sum_{\beta\in J} M(\alpha_k,\beta)+\sum_{(\alpha,\beta)\in\{\alpha_{i-j+1},\ldots,\alpha_i\}\times J} M(\alpha,\beta)\\ \le\ & (i-j)+(2j-1) && \text{By (12) and (5)}\\ =\ & i+j-1, \end{aligned}$$

which shows that M has the sparse property Q_2. □

Lemma 1 *Let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system on a finite set X. If $\mathcal{C}$ can be displayed on an X-labeled tree (T, ℓ_T), then $|\mathrm{V}(T)| \geq |\pi_1 \wedge \cdots \wedge \pi_n|_s$. Moreover, if $|\mathrm{V}(T)| = |\pi_1 \wedge \cdots \wedge \pi_n|_s$, then the map sending $v \in \mathrm{V}(T)$ to $\ell_T^{-1}(v) \in \pi_1 \wedge \cdots \wedge \pi_n$ is a bijection from $\mathrm{V}(T)$ to the set of states of $\pi_1 \wedge \cdots \wedge \pi_n$.*

Proof Let x and y be two elements of X which are contained in different states of $\pi_1 \wedge \cdots \wedge \pi_n$. It is immediate that they fall into different states of π_i for some $i \in [n]$. This implies that $\ell_T(x) \neq \ell_T(y)$, which completes the proof. □

Lemma 2 *Let X be a finite set, let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$ be a partition system on X, and let (T, ℓ_T) be an output of the algorithm* CG *applied to $\mathcal{C}$. Assume that $\mathcal{C}$ has the sparse property* Q_2*. Then T is a tree and the X-labeled tree (T, ℓ_T) is a minimum X-tree displaying $\mathcal{C}$. Moreover, (T, ℓ_T) is a fully labeled X-tree with $|\mathrm{V}(T)| = |\pi_1 \wedge \pi_2|_s$.*

Proof Write $\pi_1 = A_1|\cdots|A_a$ and $\pi_2 = B_1|\cdots|B_b$, where $a \doteq \|\pi_1\|$ and $b \doteq \|\pi_2\|$. Let T_1, T_2 and $\bar{T}$ be the graphs appeared in the algorithm CG applied to $\mathcal{C}$.

We begin by proving that T is a tree. By virtue of the rule of Algorithm 2, it is sufficient to show that $\bar{T}$ is a forest. For every $i \in [2]$ and every connected component A of T_i, we label all edges of $\bar{T}$ with endpoints in A by A. It is immediate that every edge of $\bar{T}$ has a unique label. If $\bar{T}$ has a cycle P, we can assume that the labels along P are read in the following cyclic order:

$$A_{i_1}, \ldots, A_{i_1}, B_{j_1}, \ldots, B_{j_1}, A_{i_2}, \ldots, A_{i_2}, \ldots, B_{j_m}, \ldots, B_{j_m}. \tag{13}$$

Let us suppose that the cycle P is chosen so that its parameter m is as small as possible. This implies that $I \doteq \{i_1, \ldots, i_m\} \in \binom{[a]}{m}$ and $J \doteq \{j_1, \ldots, j_m\} \in \binom{[b]}{m}$. If $m \leq 1$, we will find two distinct vertices of $\bar{T}$ which are incident to edges of labels A_{i_1} and B_{j_1} and hence must be both $A_{i_1} \cap B_{j_1}$, which is absurd. Note that every two consecutive elements in the circular sequence $A_{i_1} B_{j_1} A_{i_2} \cdots B_{j_m}$ have a nonempty intersection. Accordingly, from $m \geq 2$ we can deduce

$$|\mathscr{M}_{\pi_1,\pi_2}(I, J)| \geq 2m. \tag{14}$$

We now proceed to find that

$$\begin{aligned} \textstyle\sum_{(\alpha,\beta)\in I\times J} \mathrm{M}_{\pi_1,\pi_2}(\alpha, \beta) &= |\mathscr{M}_{\pi_1,\pi_2}(I, J)| && \text{By Eq. (8)} \\ &\geq 2m && \text{By (14)} \\ &> 2m - 1 \\ &= |I| + |J| - 1. \end{aligned}$$

By Theorem 7, this is a contradiction with the assumption that $\mathcal{C}$ has the sparse property Q_2. Consequently, $\bar{T}$ is a forest.

Note that $|\mathrm{V}(T)| = |\pi_1 \wedge \pi_2|_\mathrm{s}$. In view of Lemma 1, we find that (T, ℓ_T) is a minimum X-tree which displays $\mathcal{C}$, finishing the proof. □

Lemma 3 *Let X be a finite set and let $\mathcal{C}$ be a partition system on X of size two. Then $\mathcal{C}$ has the sparse property* Q_2 *if and only if $\mathcal{C}$ is compatible.*

Proof The forward direction is already proved in Lemma 2. We now turn to the backward direction. Let $\mathcal{C} = \{\{\pi_1, \pi_2\}\}$ and assume that $\mathcal{C}$ can be displayed on an X-tree (T, ℓ_T). Pick $Y \in 2^X \setminus \{\emptyset\}$ and assume that $|\pi_1[Y]|_s = a$ and $|\pi_2[Y]|_s = b$. This means that we can find $E_1 \in \binom{\mathrm{E}(T)}{a-1}$ and $E_2 \in \binom{\mathrm{E}(T)}{b-1}$ such that elements in different states of $\pi_i[Y]$ are mapped by ℓ_T to different connected components of $T - E_i$ for $i \in [2]$. Accordingly, we see that elements in different states of $\pi_1[Y] \wedge \pi_2[Y]$ are mapped by ℓ_T to different connected components of $T - (E_1 \cup E_2)$. This shows that $|\pi_1[Y] \wedge \pi_2[Y]|_s \leq |E_1 \cup E_2| + 1 \leq a + b - 1$, as wanted. □

2.3 Relations among sparse properties

For any nonempty set A, let K_A denote the complete graph on the vertex set A, namely $\mathrm{V}(\mathrm{K}_A) = A$ and $\mathrm{E}(\mathrm{K}_A) = \binom{A}{2}$. The *Cartesian product* of two graphs G and H is the graph $G \Box H$ with $\mathrm{V}(G \Box H)$ being

$$\mathrm{V}(G) \times \mathrm{V}(H) = \{(u, v) : \ u \in \mathrm{V}(G), v \in \mathrm{V}(H)\}$$

and $\mathrm{E}(G \Box H)$ being the union of

$$\big\{\{(u, v), (u, v')\} : \ u \in \mathrm{V}(G), \{v, v'\} \in \mathrm{E}(H)\big\}$$

and

$$\big\{\{(u, v), (u', v)\} : \ \{u, u'\} \in \mathrm{E}(G), v \in \mathrm{V}(H)\big\}.$$

This definition of Cartesian product naturally extends to a product of any number of graphs. We call any Cartesian product of a family of complete graphs a *Hamming graph*. Given positive integers $m_1, \ldots, m_n$, the Hamming graph $\mathrm{K}_{[m_1]} \Box \cdots \Box \mathrm{K}_{[m_n]}$ will be denoted by $\mathrm{H}_{m_1,\ldots,m_n}$. For every edge $e = \{(u_1, \ldots, u_n), (v_1, \ldots, v_n)\}$ of $\mathrm{H}_{m_1,\ldots,m_n}$, there exists a unique $i \in [n]$ such that $u_i \neq v_i$, which is said to be the *type* of the edge e in $\mathrm{H}_{m_1,\ldots,m_n}$. Since each Buneman graph is a vertex induced subgraph of a corresponding Hamming graph, we will also talk about the type of an edge in a Buneman graph.

Lemma 4 *Let $m_1, \ldots, m_n$ be positive integers, let P be a closed walk in the Hamming graph $\mathrm{H}_{m_1,\ldots,m_n}$, and let V and E stand for the set of vertices and edges passed in the walk P, respectively. For each $i \in [n]$, let $J_i = \{u_i : (u_1, \ldots, u_n) \in V \subseteq [m_1] \times \cdots \times [m_n]\}$, and let $t = |\{i \in [n] : \ |J_i| > 1\}|$. Then it holds $|E| \geq \sum_{i=1}^{n}(|J_i| - 1) + t$. In particular, when P is a cycle of length ℓ in $\mathrm{H}_{m_1,\ldots,m_n}$, we have $\ell \geq \sum_{i=1}^{n}(|J_i| - 1) + t$.*

Proof We divide the set E into n parts $E_1, \ldots, E_n$ such that E_i is the set of edges of type i for all $i \in [n]$. For each $i \in [n]$, it is not hard to see that

$$\begin{cases} |E_i| = 0, & \text{if } |J_i| = 1; \\ |E_i| \geq |J_i|, & \text{if } |J_i| > 1. \end{cases}$$

This indicates that $|E| \geq \sum_{i=1}^{n}(|J_i| - 1) + t$. □

Lemma 5 *Let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system having the property Q_n. Then, for any cycle of the Buneman graph $\mathcal{B}(\mathcal{C})$, its edges are of the same type and its vertices form a clique of size at least 3.*

Proof Write $\|\pi_i\| = m_i$ for each $i \in [n]$. Note that $\mathcal{B}(\mathcal{C})$ is the subgraph of $\mathrm{H}_{m_1,\ldots,m_n}$ induced by $\mathscr{M}_{\pi_1,\ldots,\pi_n}([m_1], \ldots, [m_n])$. Let P be a cycle in $\mathcal{B}(\mathcal{C})$, let V be the set of vertices of P, let $J_i = \{u_i : (u_1, \ldots, u_n) \in V \subseteq [m_1]\times\cdots\times[m_n]\}$ for each $i \in [n]$, and let $t = |\{i \in [n] : |J_i| > 1\}| \geq 1$. Lemma 4 indicates that

$$|V| \geq \sum_{i=1}^{n}(|J_i| - 1) + t. \tag{15}$$

Since $\mathcal{C}$ has the property Q_n, we obtain that $|V| \leq |\mathscr{M}_{\pi_1,\ldots,\pi_n}(J_1, \ldots, J_n)| \leq \sum_{i=1}^{n}(|J_i| - 1) + 1$. This together with (15) forces $t = 1$, which implies that the edges of P are of the same type and that V is a clique in $\mathcal{B}(\mathcal{C})$. □

Let X be a finite set and let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be a partition system on X. Write

$$\mathrm{S}(\mathcal{C}) \doteq \mathscr{M}_{\pi_1,\ldots,\pi_n}([\|\pi_1\|] \times \cdots \times [\|\pi_n\|]).$$

It follows from Eq. (9) that

$$|\,\mathrm{S}(\mathcal{C})| = |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}}\,. \tag{16}$$

For each $i \in [n]$, let

$$\mathrm{S}_i(\mathcal{C}) \doteq \mathscr{M}_{\pi_1,\ldots,\pi_{i-1},\pi_{i+1},\ldots,\pi_n}([\|\pi_1\|] \times \cdots \times [\|\pi_{i-1}\|] \times [\|\pi_{i+1}\|] \times \cdots \times [\|\pi_n\|]),$$

let $\mathrm{f}_i^{\mathcal{C}}$ be the map from $\mathrm{S}(\mathcal{C})$ to $[\|\pi_i\|]$ such that

$$\mathrm{f}_i^{\mathcal{C}}\left((\alpha_1, \ldots, \alpha_n)\right) \doteq \min\{\bar{\alpha}_i : (\alpha_1, \ldots, \alpha_{i-1}, \bar{\alpha}_i, \alpha_{i+1}, \ldots, \alpha_n) \in \mathrm{S}(\mathcal{C})\},$$

and let

$$\bar{\mathrm{S}}_i(\mathcal{C}) \doteq \{(\alpha_1, \ldots, \alpha_n) \in \mathrm{S}(\mathcal{C}) : \alpha_i > \mathrm{f}_i^{\mathcal{C}}\left((\alpha_1, \ldots, \alpha_n)\right)\}.$$

Pick $i \in [n]$. It holds

$$|\,\mathrm{S}_i(\mathcal{C})| = |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} \tag{17}$$

and

$$|\bar{\mathrm{S}}_i(\mathcal{C})| = |\,\mathrm{S}(\mathcal{C})| - |\,\mathrm{S}_i(\mathcal{C})|. \tag{18}$$

Therefore,

$$\begin{aligned} & |\bar{\mathrm{S}}_i(\mathcal{C})| \\ = & |\,\mathrm{S}(\mathcal{C})| - |\,\mathrm{S}_i(\mathcal{C})| && \text{By Eq. (18)} \\ = & |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}}\,. && \text{By Eqs. (16) and (17)} \end{aligned} \tag{19}$$

For every $\{k,j\} \in \binom{[n]}{2}$, let $\mathrm{g}_j^k(\mathcal{C})$ be the map from $\bar{\mathrm{S}}_j(\mathcal{C})$ to $\binom{\mathrm{S}_k(\mathcal{C})}{2}$ such that

$$\begin{aligned} &\mathrm{g}_j^k(\mathcal{C})\big((\alpha_1,\ldots,\alpha_n)\big) \\ \doteq &\begin{cases} \{\alpha_{\hat{k}},(\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_{j-1},\mathrm{f}_j^{\mathcal{C}}(\alpha_1,\ldots,\alpha_n),\alpha_{j+1},\ldots,\alpha_n)\}, & \text{if } k<j, \\ \{\alpha_{\hat{k}},(\alpha_1,\ldots,\alpha_{j-1},\mathrm{f}_j^{\mathcal{C}}(\alpha_1,\ldots,\alpha_n),\alpha_{j+1},\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_n)\}, & \text{if } k>j, \end{cases} \end{aligned} \tag{20}$$

where $\alpha_{\hat{k}} \doteq (\alpha_1,\ldots,\alpha_{k-1},\alpha_{k+1},\ldots,\alpha_n)$. For each $k \in [n]$, let $\mathrm{G}_k(\mathcal{C})$ be the graph with

$$\begin{cases} \mathrm{V}\left(\mathrm{G}_k(\mathcal{C})\right) & \doteq \mathrm{S}_k(\mathcal{C}), \\ \mathrm{E}\left(\mathrm{G}_k(\mathcal{C})\right) & \doteq \bigcup_{j\in[n]\setminus\{k\}} \{\mathrm{g}_j^k(\mathcal{C})\big((\alpha_1,\ldots,\alpha_n)\big) :\ (\alpha_1,\ldots,\alpha_n) \in \bar{\mathrm{S}}_j(\mathcal{C})\}. \end{cases} \tag{21}$$

Lemma 6 *Let n be an integer with $n \geq 3$ and let $\mathcal{C} = \{\{\pi_1,\ldots,\pi_n\}\}$ be a partition system on a finite set X. Assume that $\mathcal{C}$ satisfies Q_{n-1}. Then the following hold.*

(i) For every $\{k,j\} \in \binom{[n]}{2}$, the map $\mathrm{g}_j^k(\mathcal{C})$ is injective.
(ii) It holds $\sum_{j\in[n]\setminus\{k\}} |\bar{\mathrm{S}}_j(\mathcal{C})| = |\,\mathrm{E}(\mathrm{G}_k(\mathcal{C}))|$ for all $k \in [n]$.
(iii) For every $k \in [n]$, the graph $\mathrm{G}_k(\mathcal{C})$ is acyclic.

Proof (i) Without loss of generality, we may assume that $j = n-1$ and $k = n$. For the sake of contradiction, we may suppose that $\alpha = (\alpha_1,\ldots,\alpha_n)$ and $\alpha' = (\alpha_1,\ldots,\alpha_{n-1},\alpha'_n)$ are two different elements of $\bar{\mathrm{S}}_{n-1}(\mathcal{C})$ such that $\mathrm{g}_{n-1}^n(\mathcal{C})(\alpha) = \mathrm{g}_{n-1}^n(\mathcal{C})(\alpha')$. This indicates that

$$\mathrm{f}_{n-1}^{\mathcal{C}}(\alpha) = \mathrm{f}_{n-1}^{\mathcal{C}}(\alpha'). \tag{22}$$

We can see that

$$\begin{aligned} & 4 \\ = & \,|\,\mathrm{g}_{n-1}^1(\mathcal{C})(\alpha) \cup \mathrm{g}_{n-1}^1(\mathcal{C})(\alpha')| && \text{By Eq. (20)} \\ \leq & \,|\,\mathscr{M}_{\pi_2,\ldots,\pi_n}(\{\alpha_2\},\ldots,\{\alpha_{n-2}\},\{\alpha_{n-1},\mathrm{f}_{n-1}^{\mathcal{C}}(\alpha)\},\{\alpha_n,\alpha'_n\})| && \text{By Eq. (22)} \\ = & \sum_{\substack{\beta_{n-1}\in\{\alpha_{n-1},\mathrm{f}_{n-1}^{\mathcal{C}}(\alpha)\} \\ \beta_n\in\{\alpha_n,\alpha'_n\}}} \mathrm{M}_{\pi_2,\ldots,\pi_n}(\alpha_2,\ldots,\alpha_{n-2},\beta_{n-1},\beta_n). && \text{By Eq. (8)} \end{aligned} \tag{23}$$

As $\mathcal{C}$ satisfies Q_{n-1}, the right hand side of (23) is not bigger than $(1-1)+\cdots+(1-1)+(2-1)+(2-1)+1 = 3$, which is absurd.

(ii) By Eqs. (20) and (21), $\mathrm{E}(\mathrm{G}_k(\mathcal{C}))$ is the disjoint union of

$$\{\,\mathrm{g}_j^k(\mathcal{C})\big((\alpha_1,\ldots,\alpha_n)\big) :\ (\alpha_1,\ldots,\alpha_n) \in \bar{\mathrm{S}}_j(\mathcal{C})\},\ j \in [n]\setminus\{k\}.$$

The result is now ensured by Lemma 6 (i).

(iii) For each $k \in [n]$, let $\mathcal{C}_k = \{\{\pi_1,\ldots,\pi_{k-1},\pi_{k+1},\ldots,\pi_n\}\}$ having the property Q_{n-1}. Note that $\mathrm{G}_k(\mathcal{C})$ is the subgraph of $\mathcal{B}(\mathcal{C}_k)$ such that the edges of the same type in $\mathrm{G}_k(\mathcal{C})$ form a disjoint union of some star graphs. This combined with Lemma 5 implies that $\mathrm{G}_k(\mathcal{C})$ is acyclic. □

Lemma 7 *Let X be a finite set and let $\mathcal{C} = \{\{\,\pi_1, \dots, \pi_n\,\}\}$ be a partition system of size $n \geq 3$ having the sparse property Q_{n-1}. Then there exists $i \in [n]$ such that*

$$|\pi_1 \wedge \cdots \wedge \pi_n|_s \leq |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_s + |\pi_i|_s - 1. \tag{24}$$

Proof We suppose, to the contrary, that (24) does not hold. This along with Eq. (19) says that

$$|\bar{\mathrm{S}}_i(\mathcal{C})| \geq |\pi_i|_{\mathrm{s}} \tag{25}$$

for all $i \in [n]$. We are now able to derive

$$\begin{aligned} |\,\mathrm{V}(\mathrm{G}_n(\mathcal{C}))| &= |\,\mathrm{S}_n(\mathcal{C})| && \text{By Eq. (21)} \\ &= |\pi_1 \wedge \cdots \wedge \pi_{n-1}|_{\mathrm{s}} && \text{By Eq. (17)} \\ &\leq \textstyle\sum_{i=1}^{n-1}(|\pi_i|_{\mathrm{s}} - 1) + 1 && \mathcal{C} \text{ satisfies } \mathrm{Q}_{n-1} \\ &< \textstyle\sum_{i=1}^{n-1}|\pi_i|_{\mathrm{s}} && \text{By } n \geq 3 \\ &\leq \textstyle\sum_{i=1}^{n-1}|\bar{\mathrm{S}}_i(\mathcal{C})| && \text{By (25)} \\ &= |\,\mathrm{E}(\mathrm{G}_n(\mathcal{C}))|, && \text{By Lemma 6 (ii)} \end{aligned}$$

which yields a contradiction with Lemma 6 (iii). □

Lemma 8 *Let $\mathcal{C}$ be a partition system of size $n \geq 3$. If $\mathcal{C}$ has the property Q_{n-1}, then $\mathcal{C}$ satisfies Q_n.*

Proof Write $\mathcal{C} = \{\{\,\pi_1, \dots, \pi_n\,\}\}$. By the definition of Q_{n-1} and Q_n, we only need to derive

$$|\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} \leq \sum_{k=1}^{n}(|\pi_k|_{\mathrm{s}} - 1) + 1. \tag{26}$$

Since $\mathcal{C}$ has the sparse property Q_{n-1}, it holds

$$|\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} \leq \sum_{k \in [n]\setminus\{i\}} (|\pi_k|_{\mathrm{s}} - 1) + 1, \tag{27}$$

and, as a result of Lemma 7, there exists $i \in [n]$ such that (24) holds. The combination of (27) and (24) leads to (26), as wanted. □

Lemma 9 *Let X be a finite set and let $\mathcal{C} = \{\{\,\pi_1, \dots, \pi_n\,\}\}$ be a partition system of size $n \geq 2$ having the sparse property*

$$\begin{cases} \mathrm{Q}_2 & \text{and} \quad \overline{\mathrm{Q}}_2, \text{ if } n = 2; \\ \mathrm{Q}_{n-1} & \text{and} \quad \overline{\mathrm{Q}}_n, \text{ if } n \geq 3. \end{cases} \tag{28}$$

Then

$$\|\pi_i\| = |\pi_i|_s \tag{29}$$

for all $i \in [n]$.

Proof Observe that Eq. (29) will be a consequence of the assumption that $\mathcal{C}$ satisfies both Q_n and $\overline{\mathrm{Q}}_n$. Therefore, we only need to consider the case of $n \geq 3$. But we know that $\mathcal{C}$ satisfies Q_{n-1} and so Lemma 8 tells us that it also satisfies Q_n. This is the proof! □

Lemma 10 *Let X be a finite set and let $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ be a partition system of size $n \geq 2$ which satisfies the sparse property* (28) *Then $\mathcal{C}$ has the sparse property $\overline{\mathrm{Q}}_{n-1}$.*

Proof Because $\mathcal{C}$ fulfils $\overline{\mathrm{Q}}_n$, Lemma 9 reduces our task to establishing

$$|\pi_j|_{\mathrm{s}} - 1 = |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - |\pi_1 \wedge \cdots \wedge \pi_{j-1} \wedge \pi_{j+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} \tag{30}$$

for every $j \in [n]$. When $n = 2$, Eq. (29) says that Eq. (30) is nothing but $\overline{\mathrm{Q}}_2$. Henceforth, we may assume $n \geq 3$ hereafter. From Lemma 7, we derive that there exists $i \in [n]$ such that (24) holds. We may assume, without losing generality, that $i = n$ in (24), and obtain

$$\begin{aligned}
&\textstyle\sum_{j=1}^{n}(|\pi_j|_{\mathrm{s}} - 1) + 1 & \\
&= \textstyle\sum_{j=1}^{n}(\|\pi_j\| - 1) + 1 & \text{By Eq. (29)} \\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} & \mathcal{C} \text{ satisfies } \overline{\mathrm{Q}}_n \\
&\leq |\pi_1 \wedge \cdots \wedge \pi_{n-1}|_{\mathrm{s}} + |\pi_n|_{\mathrm{s}} - 1 & \text{By assumption} \\
&\leq \textstyle\sum_{j=1}^{n}(|\pi_j|_{\mathrm{s}} - 1) + 1. & \mathcal{C} \text{ satisfies } \mathrm{Q}_{n-1}
\end{aligned}$$

This implies

$$|\pi_1 \wedge \cdots \wedge \pi_{n-1}|_{\mathrm{s}} = \sum_{j=1}^{n-1}(|\pi_j|_{\mathrm{s}} - 1) + 1. \tag{31}$$

Letting $\rho_j = \pi_1 \wedge \cdots \wedge \pi_{j-1} \wedge \pi_{j+1} \wedge \cdots \wedge \pi_n$, we can now find that

$$\begin{aligned}
&\textstyle\sum_{j=1}^{n-1}(|\pi_j|_{\mathrm{s}} - 1) & \\
&= |\rho_n|_{\mathrm{s}} - 1 & \text{By Eq. (31)} \\
&= |\,\mathrm{S}_n(\mathcal{C})| - 1 & \text{By Eq. (17)} \\
&= |\,\mathrm{V}\left(\mathrm{G}_n(\mathcal{C})\right)| - 1 & \text{By Eq. (21)} \\
&\geq |\,\mathrm{E}\left(\mathrm{G}_n(\mathcal{C})\right)| & \text{By Lemma 6 (iii)} \\
&= \textstyle\sum_{j=1}^{n-1} |\bar{\mathrm{S}}_j(\mathcal{C})| & \text{By Lemma 6 (ii)} \\
&= \textstyle\sum_{j=1}^{n-1} \left(|\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - |\rho_j|_{\mathrm{s}}\right) & \text{By Eq. (19)} \\
&= \textstyle\sum_{j=1}^{n-1} \left(\sum_{i=1}^{n}(|\pi_i|_{\mathrm{s}} - 1) + 1 - |\rho_j|_{\mathrm{s}}\right) & \mathcal{C} \text{ satisfies } \overline{\mathrm{Q}}_n \\
&\geq \textstyle\sum_{j=1}^{n-1} \left(\sum_{i=1}^{n}(|\pi_i|_{\mathrm{s}} - 1) + 1 - \sum_{i \in [n]\setminus\{j\}}(|\pi_i|_{\mathrm{s}} - 1) - 1\right) & \mathcal{C} \text{ satisfies } \mathrm{Q}_{n-1} \\
&= \textstyle\sum_{j=1}^{n-1}(|\pi_j|_{\mathrm{s}} - 1), &
\end{aligned}$$

forcing the second inequality to hold as equality, namely

$$|\rho_j|_{\mathrm{s}} = \sum_{i \in [n]\setminus\{j\}} (|\pi_i|_{\mathrm{s}} - 1) + 1$$

for every $j \in [n-1]$. This along with Eq. (31) proves Eq. (30). □

Lemma 11 *Let n be an integer with $n \geq 2$, let X be a finite set, let $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ be a partition system on X. Let (T, ℓ_T) be an output of the algorithm* CG *applied to $\mathcal{C}$ and let $T_1, \dots, T_n$ and $\bar{T}$ be the graphs appeared when executing* CG *on the input $\mathcal{C}$. Assume that $\mathcal{C}$ possesses the sparse property* (28)*. Then the following hold.*

(*i*) (T, ℓ_T) *is a fully labeled* X*-tree with* $|\operatorname{V}(T)| = |\pi_1 \wedge \cdots \wedge \pi_n|_s$*.*
(*ii*) $\operatorname{E}(T_1)|\cdots|\operatorname{E}(T_n)$ *is a partition of* $\operatorname{E}(T)$*, namely* $\bar{T} = T$*.*
(*iii*) (T, ℓ_T) *is a minimum* X*-tree displaying* $\mathcal{C}$*.*
(*iv*) *The partition system* $\mathcal{C}$ *is compatible.*

Proof Let Y be the set of states of the partition $\pi_1 \wedge \cdots \wedge \pi_n$. Note that the trees $T_1, \ldots, T_n$ and T all have Y as their vertex sets. Regard the vertex $v \in Y$ as $(\alpha_1, \ldots, \alpha_n)$ if $v = \pi_1^{-1}(\alpha_1) \cap \cdots \cap \pi_n^{-1}(\alpha_n)$. We can see that $\bar{T}$ is a subgraph of the Hamming graph $\operatorname{H}_{\|\pi_1\|,\ldots,\|\pi_n\|}$ and that every edge of T_i is an edge of type i in $\operatorname{H}_{\|\pi_1\|,\ldots,\|\pi_n\|}$.

(i) It suffices to prove that $\bar{T}$ is acyclic. Because $\mathcal{C}$ possesses the sparse property (28), it follows from Lemma 8 that $\mathcal{C}$ satisfies Q_n. In light of the rule of CG, we see that $\bar{T}$ is the subgraph of $\mathcal{B}(\mathcal{C})$ and that the set of edges of type i in $\bar{T}$ is T_i, which is a forest. This along with Lemma 5 says that $\bar{T}$ is acyclic, proving (i).

(ii) It is trivial that $\operatorname{E}(T_i) \cap \operatorname{E}(T_j) = \emptyset$ for all $\{i, j\} \in \binom{[n]}{2}$. It remains to verify that $|\operatorname{E}(T)| = \sum_{i=1}^{n} |\operatorname{E}(T_i)|$. For each $i \in [n]$, let c_i represent the number of connected components of T_i. The rule of Algorithm 2 asserts

$$c_i = |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}}\,. \tag{32}$$

Consequently, for every $i \in [n]$ it holds

$$\begin{aligned}
&|\operatorname{E}(T_i)| \\
&= |\operatorname{V}(T_i)| - c_i && T_i \text{ is acyclic}\\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - c_i && \text{By Lemma 11 (i)}\\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} && \text{By Eq. (32)}\\
&= \textstyle\sum_{j=1}^{n}(\|\pi_j\| - 1) + 1 - |\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} && \mathcal{C} \text{ satisfies } \overline{\operatorname{Q}}_n\\
&= \textstyle\sum_{j=1}^{n}(\|\pi_j\| - 1) + 1 - \sum_{j \in [n]\setminus\{i\}}(\|\pi_j\| - 1) + 1 && \text{By Lemma 10}\\
&= \|\pi_i\| - 1.
\end{aligned} \tag{33}$$

This in turn gives

$$\begin{aligned}
|\operatorname{E}(T)| &= |\operatorname{V}(T)| - 1 && T \text{ is a tree}\\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - 1 && \text{By Lemma 11 (i)}\\
&= \textstyle\sum_{i=1}^{n}(\|\pi_i\| - 1) && \mathcal{C} \text{ satisfies } \overline{\operatorname{Q}}_n\\
&= \textstyle\sum_{i=1}^{n} |\operatorname{E}(T_i)|, && \text{By Eq. (33)}
\end{aligned}$$

as desired.

(iii) By Lemma 1 and Lemma 11 (i), we only need to prove that (T, ℓ_T) displays $\mathcal{C}$. For each $i \in [n]$, let ρ_i be the partition of X such that every part of ρ_i is the union of a connected components of the forest $T - \operatorname{E}(T_i)$. For all $i \in [n]$, it is clear that (T, ℓ_T) displays ρ_i and that

$$|\rho_i|_{\mathrm{s}} = |\operatorname{E}(T_i)| + 1. \tag{34}$$

To prove that (T, ℓ_T) displays $\mathcal{C}$, we are going to verify that $\rho_i = \pi_i$ for all $i \in [n]$. For each $i \in [n]$, we have

$$\begin{cases} |\rho_i|_{\mathrm{s}} = \|\pi_i\|, & \text{By Eqs. (33) and (34)} \\ \|\rho_i\| = |\rho_i|_{\mathrm{s}}\,, & \text{By the definition of } \rho_i \\ \|\pi_i\| = |\pi_i|_{\mathrm{s}}\,. & \text{By Lemma 9} \end{cases}$$

Consequently, our task is to show that, in the partition lattice on X, it holds $\rho_i \preceq \pi_i$ for all $i \in [n]$.

To this end, pick $i \in [n]$ and let $u = (\alpha_1, \ldots, \alpha_n)$ and $v = (\beta_1, \ldots, \beta_n)$ be two elements in Y such that u and v are contained in different states of π_i. This asserts $\alpha_i \neq \beta_i$. Let P be the unique path in T from u to v. Note that P contains an edge in $\mathrm{E}(T_i)$, according to the definition of $\mathrm{E}(T_i)$ in Algorithm 2. This along with Lemma 11 (ii) says that u and v fall into different connected components of $T - \mathrm{E}(T_i)$, namely they belong to different states of ρ_i, as wanted.

(iv) It is direct from Lemma 11 (iii). □

Proof (Proof of Theorem 9) (i) Let $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ be the partition system for which $M = \mathrm{M}_{\mathcal{C}}$ holds. Pick $\{i_1, \ldots, i_k\} \in \binom{[n]}{k}$ and let $\mathcal{C}' = \{\{\pi_{i_1}, \ldots, \pi_{i_k}\}\}$. Clearly, $\mathcal{C}'$ is a partition system of size k having the property Q_{k-1}. Applying Lemma 8 to $\mathcal{C}'$, we see that $\mathcal{C}'$ has the property Q_k. This together with Theorem 7 gives us

$$\sum_{(\alpha_1,\ldots,\alpha_k)\in I_{i_1}\times\cdots\times I_{i_k}} \mathrm{M}_{\pi_{i_1},\ldots,\pi_{i_k}}(\alpha_1, \ldots, \alpha_k) \leq \sum_{j=1}^{k}(|I_{i_j}| - 1) + 1$$

for every nonempty sets $I_{i_1} \subseteq [\|\pi_{i_1}\|], \ldots, I_{i_k} \subseteq [\|\pi_{i_k}\|]$, proving that M has the property Q_k.

(ii) If M is an n-tensor of all zeros, $M^{1,\ldots,n-1}$ surely satisfies Q_{k-1}. We can thus assume that $M = \mathrm{M}_{\mathcal{C}}$ where $\mathcal{C} = \{\{\pi_1, \ldots, \pi_n\}\}$ is a partition system with $|\pi_n|_{\mathrm{s}} = 1$.

Pick $\{i_1, \ldots, i_{k-1}\} \in \binom{[n-1]}{k-1}$, let $t \in [\|\pi_n\|]$ such that $\pi_n^{-1}(t) \neq \emptyset$, and let I_{i_j} be a nonempty subset of $[\|\pi_{i_j}\|]$ for every $j \in [k-1]$. We have

$$\begin{aligned} &\textstyle\sum_{(\alpha_1,\ldots,\alpha_{k-1})\in I_{i_1}\times\cdots\times I_{i_{k-1}}} M_{\pi_{i_1},\ldots,\pi_{i_{k-1}}}(\alpha_1, \ldots, \alpha_{k-1}) \\ &= \textstyle\sum_{(\alpha_1,\ldots,\alpha_k)\in I_{i_1}\times\cdots\times I_{i_{k-1}}\times\{t\}} M_{\pi_{i_1},\ldots,\pi_{i_{k-1}},\pi_n}(\alpha_1, \ldots, \alpha_k) && |\pi_n|_{\mathrm{s}} = 1 \\ &\leq \textstyle\sum_{j=1}^{k-1}(|I_{i_j}| - 1) + (|\{t\}| - 1) + 1 && \mathcal{C} \text{ satisfies } \mathrm{Q}_k \\ &= \textstyle\sum_{j=1}^{k-1}(|I_{i_j}| - 1) + 1. \end{aligned}$$

This says that $M^{1,\ldots,n-1}$ has the property Q_{k-1}.

(iii) Let $X = [n+1]$, let $A_1 = \{1\}$, let $A_i = \{i, n+1\}$ for all $i \in [n] \setminus [1]$, let $\pi_i = A_i|(X \setminus A_i)$ for all $i \in [n]$ and let $M = \mathrm{M}_{\pi_1,\ldots,\pi_n}$. It is easy to check that $|\mathscr{M}_{\pi_2,\ldots,\pi_n}([2] \times \cdots \times [2])| = |X| = n + 1 > n$. This together with Eq. (8) implies that M does not have the sparse property Q_{n-1}. A case by case analysis shows that M does have the sparse property Q_n.

(iv) Let $\mathcal{C}$ be a partition system such that $M = \mathrm{M}_{\mathcal{C}}$. It is trivial that $\mathcal{C}$ has the sparse property Q_1. Taking into account Lemma 3 and Lemma 11 (iv), we find that $\mathcal{C}$ has the sparse property Q_2. Now, an application of Theorem 9 (i) yields the claim that $\mathcal{C}$ has the sparse properties Q_ℓ for all $\ell \in [n]$.

(v) According to Lemma 9, $\mathcal{C}$ has the property $\overline{\mathrm{Q}}_1$. We then derive from Lemma 10 that $\mathcal{C}$ satisfies $\overline{\mathrm{Q}}_{n-1}$. If $n = 3$, the claim already follows. If $n \geq 4$, by Theorem 9 (iv), $\mathcal{C}$ has the properties Q_ℓ for all $\ell \in [n]$. Henceforth, every $n-1$ partitions of $\mathcal{C}$ form a partition system with properties Q_{n-2} and $\overline{\mathrm{Q}}_{n-1}$. Invoking Lemma 10 to these partition systems shows that $\mathcal{C}$ has the sparse property $\overline{\mathrm{Q}}_{n-2}$. Continuing this process, we will eventually get to the fact that $\mathcal{C}$ has the properties $\overline{\mathrm{Q}}_n, \ldots, \overline{\mathrm{Q}}_2$, completing the proof. □

Proof (Proof of Theorem 10) This is immediate from Theorem 9 (i), (iv) and (v). □

2.4 Compatible partition system

Lemma 12 *Let X be a finite set and let $\mathcal{C}$ be a partition system on X. If $\mathcal{C}$ is compatible, then $\mathcal{C}$ has a sparse completion.*

Proof Write $\mathcal{C} = \{\{\, \pi_1, \ldots, \pi_n \,\}\}$. When $n = 1$, let ρ be a partition of some set X' with $X' \supseteq X$ such that every part of ρ is nonempty and $\rho[X] = \pi_1$. Clearly, the array M_ρ is an all-ones vector, hence satisfying the property $\overline{\mathrm{Q}}_1$, and it holds $\mathrm{M}_{\pi_1} \leq \mathrm{M}_\rho$. This says that $\{\{\, \rho \,\}\}$ is a sparse completion of $\mathcal{C}$.

We now assume that $n \geq 2$. We know that $\mathcal{C}$ can be displayed on some X-labeled tree (T, ℓ_T). For each $i \in [n]$, there exists $E_i \subseteq \mathrm{E}(T)$ such that two elements of X are in different states of π_i if and only if they are mapped by ℓ_T to different connected components of the forest $T - E_i$. We may assume, without losing generality, that ℓ_T is injective, that $E_1, \ldots, E_n$ are pairwise disjoint, and that

$$|E_i| = \|\pi_i\| - 1. \tag{35}$$

Let X' be the set with $X' \supseteq X$ such that there exists a bijection ℓ'_T from X' to $\mathrm{V}(T)$ sending every element $x \in X \subseteq X'$ to $\ell'_T(x) = \ell_T(x)$, namely, $X' = \ell'^{-1}_T\big(\mathrm{V}(T)\big)$. For each $i \in [n]$, let π'_i be the partition of X' such that each part of π'_i is mapped by ℓ'_T to a connected component of the forest $T - E_i$. Denote by $\mathcal{C}'$ the family $\{\{\, \pi'_1, \ldots, \pi'_n \,\}\}$. It remains to prove that $\mathcal{C}'$ is a sparse completion of $\mathcal{C}$. For all $i \in [n]$, it is immediate that (T, ℓ'_T) displays π'_i and that $\pi'_i[X] = \pi_i$. It follows from Lemma 3 that $\mathcal{C}'$ has the property Q_2. Meanwhile, we have

$$\begin{aligned}
|\pi'_1 \wedge \cdots \wedge \pi'_n|_{\mathrm{s}} &= \textstyle\sum_{i=1}^n |E_i| + 1 && E_1, \ldots, E_n \text{ are pairwise disjoint} \\
&= \textstyle\sum_{i=1}^n (\|\pi_i\| - 1) + 1 && \text{By Eq. (35)} \\
&= \textstyle\sum_{i=1}^n (\|\pi'_i\| - 1) + 1,
\end{aligned}$$

which implies that $\mathcal{C}'$ has the property $\overline{\mathrm{Q}}_n$, completing the proof. □

Proof (Proof of Theorem 11) (i) This is already proved in Lemma 3.

(ii) The forward direction is verified in Lemma 12. We now aim to show the other direction.

If $n = 1$, there is nothing to prove. If $n = 2$, we know that $\mathcal{C}$ has the property Q_2 and so, in view of Theorem 11 (i), $\mathcal{C}$ is compatible.

Finally, let us consider the case of $n \geq 3$. Let $\mathcal{C}'$ be the sparse completion of $\mathcal{C}$ so that $\mathcal{C}'$ is a partition system on some finite set $X' \supseteq X$ and that $\mathcal{C}'$ fulfils Q_2 and $\overline{\mathrm{Q}}_n$. It follows from Theorem 9 (i) that $\mathcal{C}'$ has the property Q_{n-1}. Applying Lemma 11 (iv) to $\mathcal{C}'$, we see that there exists an X'-tree (T, ℓ'_T) displaying $\mathcal{C}'$. Let ℓ_T be the restriction of ℓ'_T to X. Then the X-labeled tree (T, ℓ_T) displays $\mathcal{C}$, showing that $\mathcal{C}$ is compatible. □

Proof (Proof of Theorem 12) According to Theorem 11 (ii), we can derive claim (i) and claim (ii) from Theorem 2 and Theorem 5 (ii), respectively.

The proof of the remaining claims are similar. In view of Theorem 11 (ii), claim (iii) is a result of Theorem 5 (iv), and claim (iv) comes from Theorem 6. We only present below a complete proof for claim (iii).

By Theorem 5 (iv), there exists a partition system $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ such that the following are valid:

(a) $|\pi_i|_s \leq r$ for all $i \in [n]$;
(b) $\mathcal{C}$ is not compatible;
(c) $\{\{\pi_{i_1}, \dots, \pi_{i_d}\}\}$ is compatible for all $\{i_1, \dots, i_d\} \in \binom{[n]}{d}$.

Let $\mathcal{C}' = \{\{\pi'_1, \dots, \pi'_n\}\}$ be the partition system such that for each $i \in [n]$, $\|\pi'_i\| = r$ and π'_i has the same states with π_i. It is straightforward to check the following:

(b′) $\mathcal{C}'$ is not compatible;
(c′) $\{\{\pi'_{i_1}, \dots, \pi'_{i_d}\}\}$ is compatible for all $\{i_1, \dots, i_d\} \in \binom{[n]}{d}$.

Let $M = \mathrm{M}_{\mathcal{C}'}$ be a $(0, 1)$ n-tensor of size $\overbrace{r \times \cdots \times r}^{n}$. Theorem 11 (ii) says that M does not have a sparse completion but $M^{i_1,\dots,i_d}$ has a sparse completion for every $\{i_1, \dots, i_d\} \in \binom{[n]}{d}$, finishing the proof. □

2.5 Tree reconstruction

Lemma 13 *Let n be an integer with $n \geq 2$ and let $\mathcal{C} = \{\{\pi_1, \dots, \pi_n\}\}$ be a compatible partition system on a finite set X. Suppose that $\mathcal{C}$ has the sparse property $\overline{\mathrm{Q}}_n$ and $\mathcal{C}$ can be displayed on a fully labeled X-tree (T, ℓ_T) with $|\mathrm{V}(T)| = |\pi_1 \wedge \cdots \wedge \pi_n|_s$. Then there exists a partition $E_1|\cdots|E_n$ of $\mathrm{E}(T)$ such that, for all $i \in [n]$, states of π_i coincide with the sets of labels assigned by ℓ_T to a common connected component of $T - E_i$, and so*

$$|E_i| = |\pi_i|_s - 1. \tag{36}$$

Proof In light of Lemma 1, we assume that $\mathrm{V}(T)$ is the set of states of the partition $\pi_1 \wedge \cdots \wedge \pi_n$, and that ℓ_T is the map sending every vertex x to the unique state A of $\pi_1 \wedge \cdots \wedge \pi_n$ which contains x. We can deduce from Lemma 3 that $\mathcal{C}$ satisfies Q_2, and so Theorem 9 (i) and Lemma 9 allow us to obtain Eq. (29). For each $i \in [n]$, as π_i is displayed on T by the surjective map ℓ_T, there are exactly $|\pi_i|_\mathrm{s} - 1$ edges of T whose two endpoints fall into different parts of π_i and we use E_i to designate this set of $|\pi_i|_\mathrm{s} - 1$ edges. For each $i \in [n]$, it is trivial that Eq. (36) holds and the partition of $\mathrm{V}(T)$ into the connected components of $T - E_i$, when viewed as a partition of X in the natural way, is the partition π_i.

It remains to show that $\{E_1, \ldots, E_n\}$ is a partition of $\mathrm{E}(T)$. For each edge $\{A, B\}$ of T, since A and B must be different states of $\pi_1 \wedge \cdots \wedge \pi_n$, there exists $i \in [n]$ such that A and B are different states of π_i and so $\{A, B\} \in E_i$. This proves $\bigcup_{i \in [n]} E_i = \mathrm{E}(T)$. We further notice that

$$
\begin{aligned}
|\textstyle\bigcup_{i\in[n]} E_i| &= |\,\mathrm{E}(T)| \\
&= |\,\mathrm{V}(T)| - 1 && T \text{ is a tree} \\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_\mathrm{s} - 1 \\
&= \textstyle\sum_{i=1}^{n} (\|\pi_i\| - 1) && \mathcal{C} \text{ satisfies } \overline{\mathrm{Q}}_n \\
&= \textstyle\sum_{i=1}^{n} (|\pi_i|_\mathrm{s} - 1) && \text{By Eq. (29)} \\
&= \textstyle\sum_{i=1}^{n} |E_i|, && \text{By Eq. (36)}
\end{aligned}
$$

which then confirms that $E_1 | \cdots | E_n$ is a partition of $\mathrm{E}(T)$. □

Proof (Proof of Theorem 13) (i) By Theorem 11 (i), $\mathcal{C}$ has the property Q_2. We thus get the backward implication from Lemma 2.

We proceed to prove the forward direction. Let (T, ℓ_T) be a minimum X-tree displaying $\mathcal{C}$. It follows from Lemmas 1 and 2 that ℓ_T is surjective. For each $i \in [2]$, let E_i be the set of edges of T whose endpoints are mapped by ℓ_T^{-1} to different states of π_i, and let $T_i = T - E_{3-i}$. Let $\bar{T}$ be the graph such that $\mathrm{V}(\bar{T}) = \mathrm{V}(T)$ and $\mathrm{E}(\bar{T}) = \mathrm{E}(T_1) \cup \mathrm{E}(T_2) \subseteq \mathrm{E}(T)$. Because T is a tree, T is a connected graph obtained from $\bar{T}$ by adding a minimum number of edges. To prove that (T, ℓ_T) is a possible output of Algorithm 2 applied to $\mathcal{C}$, we only need to show that T_1 and T_2 both can appear as an outpout forest in the course of executing Algorithm 2 on the input $\mathcal{C}$. By symmetry, it suffices to consider the case of $i = 1$.

To this end, it remains to show that, for every two vertices u and v of T, u and v are in the same component of T_1 if and only if $\ell_T^{-1}(u)$ and $\ell_T^{-1}(v)$ are in the same state of π_2. We first assume that $\{u, v\} \in \mathrm{E}(T_1) = \mathrm{E}(T) \setminus E_2$. The definition of E_2 leads to the conclusion that $\ell_T^{-1}(u)$ and $\ell_T^{-1}(v)$ are in the same state of π_2. This is the "only if" direction. We now turn to the "if" direction. Let P be the path connecting u and v in T. Since (T, ℓ_T) displays π_2 and $\ell_T^{-1}(u)$ and $\ell_T^{-1}(v)$ are from the same state of π_2, we see that $\ell_T^{-1}(w)$ are from the same state of π_2 for all vertex w in P. This says that none of the edges along P are in E_2. Therefore u and v are from the same component of T_1, as wanted.

(ii) The case of $n = 1$ is trivial. When $n = 2$, Theorem 11 (ii) says that the result follows from Theorem 13 (i). In the rest of the proof we assume $n \geq 3$. The "if" direction is given by Lemma 11 (iii). It remains to show the other direction.

Let (T, ℓ_T) be a minimum X-tree displaying $\mathcal{C}$. It follows from Lemma 1 and Lemma 11 (i) that $|\,\mathrm{V}(T)| = |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}}$ and that (T, ℓ_T) is a fully labeled X-tree. This allows us to apply Lemma 13 to $\mathcal{C}$ and (T, ℓ_T). Let $E_1|\cdots|E_n$ be a partition of $\mathrm{E}(T)$ as claimed by Lemma 13. For each $i \in [n]$, let $T_i = (\mathrm{V}(T), E_i)$ and let ρ_i be the partition of X without empty parts such that two elements x and y of X are mapped by ℓ_T to the vertices in the same connected component of T_i if and only if x and y are contained in the same state of ρ_i. Based on the procedures of CG, it suffices to prove that ρ_i has the same states with the partition $\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n$ for all $i \in [n]$.

By Theorem 9 (i) and Lemma 9 we get Eq. (29). By Theorem 9 (i) and Theorem 9 (v), we see that $\mathcal{C}$ has the property $\overline{\mathrm{Q}}_{n-1}$. For every $i \in [n]$, observe that the number of connected components of T_i equals $|\rho_i|_{\mathrm{s}}$, namely

$$|\rho_i|_{\mathrm{s}} = |\,\mathrm{V}(T)| - |E_i| = |\,\mathrm{E}(T)| - |E_i| + 1. \tag{37}$$

Therefore, for each $i \in [n]$ we have

$$\begin{aligned}
&|\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n|_{\mathrm{s}} & \\
=& \textstyle\sum_{j \in [n] \setminus \{i\}} (\|\pi_i\| - 1) + 1 & \mathcal{C} \text{ satisfies } \overline{\mathrm{Q}}_{n-1} \\
=& \textstyle\sum_{j \in [n] \setminus \{i\}} (|\pi_i|_{\mathrm{s}} - 1) + 1 & \text{By Eq. (29)} \\
=& \textstyle\sum_{j \in [n] \setminus \{i\}} |E_i| + 1 & \text{By Eq. (36)} \\
=& |\,\mathrm{E}(T)| - |E_i| + 1 & E_1|\cdots|E_n \text{ is a partition of } \mathrm{E}(T) \\
=& |\rho_i|_{\mathrm{s}}\,. & \text{By Eq. (37)}
\end{aligned}$$

This means that, for the purpose of showing that ρ_i and $\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n$ share the same set of states, we need to demonstrate

$$\rho_i \preceq \pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n \tag{38}$$

for each $i \in [n]$. Take $x, y \in X$ which are in the same state of ρ_i. The definition of ρ_i tells us that $\ell_T(x)$ and $\ell_T(y)$ are contained in the same connected component of T_i. Since $E_1|\cdots|E_n$ is a partition of $\mathrm{E}(T)$, we see that $\ell_T(x)$ and $\ell_T(y)$ are contained in the same connected component of $T - E_j$ for all $j \neq i$. It then follows from Lemma 13 that x and y are in the same state of π_j for all $j \neq i$, which simply says that x and y are in the same state of $\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n$. This proves (38) and so we are done.

(iii) For each vertex α of $\mathcal{B}(\mathcal{C})$, we use $(\alpha_1, \ldots, \alpha_n) \in [\|\pi_1\|] \times \cdots \times [\|\pi_n\|]$ to stand for it if $\alpha(\pi_i) = \pi_i^{-1}(\alpha_i)$ for all $i \in [n]$. Recall from Theorem 11 (ii) that $\mathcal{C}$ is compatible. By the Helly property of subtrees in a tree, this means that there is a natural bijection from $\mathrm{V}\left(\mathcal{B}(\mathcal{C})\right)$ to the set of states of $\pi_1 \wedge \cdots \wedge \pi_n$, mapping every vertex $(\alpha_1, \ldots, \alpha_n)$ to $\pi_1^{-1}(\alpha_1) \cap \cdots \cap \pi_n^{-1}(\alpha_n)$. This enables us to identify $(\alpha_1, \ldots, \alpha_n)$ with $\pi_1^{-1}(\alpha_1) \cap \cdots \cap \pi_n^{-1}(\alpha_n)$. Note that

$$\left|\mathrm{V}\left(\mathcal{B}(\mathcal{C})\right)\right| = |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}}\,. \tag{39}$$

If $n = 1$, the Buneman graph $\mathcal{B}(\mathcal{C})$ is a complete graph. It is easy to see that Algorithm 2 outputs all the spanning trees of $\mathcal{B}(\mathcal{C})$ with the input $\mathcal{C}$.

In the rest of the proof, we assume $n \geq 2$. Since $\mathcal{C}$ has the properties Q_2 and $\overline{\mathrm{Q}}_n$. An application of Theorem 9 (i) yields that $\mathcal{C}$ satisfies property (28). Consequently, Lemma 9 gives Eq. (29). Let G be the underlying graph of an output of CG applied to $\mathcal{C}$. By Theorem 13 (ii), we know that G is a tree. To show that $\mathcal{B}(\mathcal{C})$ is connected, we are going to verify that G is a spanning tree of $\mathcal{B}(\mathcal{C})$. To this end, note that $\mathrm{V}(G)$ is the set of states of $\pi_1 \wedge \cdots \wedge \pi_n$, we only need to prove $\mathrm{E}(G) \subseteq \mathrm{E}\left(\mathcal{B}(\mathcal{C})\right)$. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$ be two adjacent vertices of G. Theorem 9 (i) allows us to apply Lemma 11 (ii) to $\mathcal{C}$, and hence there exists a unique $i \in [n]$ such that $\pi_1^{-1}(\alpha_1) \cap \cdots \cap \pi_n^{-1}(\alpha_n)$ and $\pi_1^{-1}(\beta_1) \cap \cdots \cap \pi_n^{-1}(\beta_n)$ are contained in the same state of $\pi_1 \wedge \cdots \wedge \pi_{i-1} \wedge \pi_{i+1} \wedge \cdots \wedge \pi_n$. In other words, i is the unique index such that $\alpha_i \neq \beta_i$. Therefore $\{\alpha, \beta\}$ is an edge of $\mathcal{B}(\mathcal{C})$, as wanted.

Let T be a spanning tree of $\mathcal{B}(\mathcal{C})$. It remains to prove that $(T, \phi_{\mathcal{C}})$ is a possible output of CG for the input $\mathcal{C}$. By Eq. (39), Lemma 1 and Theorem 13 (ii), it suffices to check that the X-tree $(T, \phi_{\mathcal{C}})$ displays $\mathcal{C}$.

To this end, note that $\mathcal{B}(\mathcal{C})$ is a subgraph of the Hamming graph $\mathrm{H}_{\|\pi_1\|,\ldots,\|\pi_n\|}$. For every edge $\{A, B\}$ of $\mathcal{B}(\mathcal{C})$, there exists a unique $i \in [n]$ such that $\{A, B\}$ is a type-i edge of $\mathrm{H}_{\|\pi_1\|,\ldots,\|\pi_n\|}$. For each $i \in [n]$, let F_i consist of all type-i edges of $\mathcal{B}(\mathcal{C})$. It is obvious that $F_1|\cdots|F_n$ is a partition of $\mathrm{E}\left(\mathcal{B}(\mathcal{C})\right)$. For each $i \in [n]$, let $E_i = \mathrm{E}(T) \cap F_i$. Since $\mathrm{E}(T) \subseteq \mathrm{E}\left(\mathcal{B}(\mathcal{C})\right)$, we see that $E_1|\cdots|E_n$ is a partition of $\mathrm{E}(T)$. We can obtain

$$
\begin{aligned}
\textstyle\sum_{i=1}^n |E_i| &= |\mathrm{E}(T)| && E_1|\cdots|E_n \text{ is a partition of } \mathrm{E}(T) \\
&= |\mathrm{V}\left(\mathcal{B}(\mathcal{C})\right)| - 1 && T \text{ is a spanning tree of } \mathcal{B}(\mathcal{C}) \\
&= |\pi_1 \wedge \cdots \wedge \pi_n|_{\mathrm{s}} - 1 && \text{By Eq. (39)} \\
&= \textstyle\sum_{i=1}^n (\|\pi_i\| - 1). && \mathcal{C} \text{ satisfies } \overline{\mathrm{Q}}_n \\
&= \textstyle\sum_{i=1}^n (|\pi_i|_{\mathrm{s}} - 1). && \text{By Eq. (29)}
\end{aligned}
\tag{40}
$$

For each $i \in [n]$ and $j \in [\|\pi_i\|]$, let $T_i = T - E_i$, let $V_i^{(j)} = \{(\alpha_1, \ldots, \alpha_n) \in \mathrm{V}\left(\mathcal{B}(\mathcal{C})\right) : \alpha_i = j\}$, and let $T_i^{(j)}$ be the subgraph of T induced by $V_i^{(j)}$, namely $\mathrm{V}(T_i^{(j)}) = V_i^{(j)}$ and $\mathrm{E}(T_i^{(j)}) = \mathrm{E}(T) \cap \binom{V_i^{(j)}}{2}$. It is obvious that $V_i^{(1)}|\cdots|V_i^{(\|\pi_i\|)}$ is a partition of $\mathrm{V}(T)$ and that $E_i|\,\mathrm{E}(T_i^{(1)})|\cdots|\,\mathrm{E}(T_i^{(\|\pi_i\|)})$ is a partition of $\mathrm{E}(T)$. for all $i \in [n]$.

Pick $i \in [n]$. It follows that

$$
\begin{aligned}
&|\mathrm{E}(T)| \\
&= |E_i| + \textstyle\sum_{j=1}^{\|\pi_i\|} |\mathrm{E}(T_i^{(j)})| && E_i|\,\mathrm{E}(T_i^{(1)})|\cdots|\,\mathrm{E}(T_i^{(\|\pi_i\|)}) \in \mathrm{P}\left(\mathrm{E}(T)\right) \\
&\leq |E_i| + \textstyle\sum_{j=1}^{\|\pi_i\|} |\mathrm{V}(T_i^{(j)})| - \|\pi_i\| && T_i^{(j)} \text{ is an induced subgraph of } T \\
&= |E_i| - \|\pi_i\| + |\mathrm{V}(T)| && V_i^{(1)}|\cdots|V_i^{(\|\pi_i\|)} \in \mathrm{P}\left(\mathrm{V}(T)\right) \\
&= |E_i| - \|\pi_i\| + 1 + |\mathrm{E}(T)| && T \text{ is a tree} \\
&= |E_i| - |\pi_i|_{\mathrm{s}} + 1 + |\mathrm{E}(T)|. && \text{By Eq. (29)}
\end{aligned}
\tag{41}
$$

This together with Eq. (40) forces that

$$|E_i| = |\pi_i|_{\mathrm{s}} - 1. \tag{42}$$

For each $j \in [\|\pi_i\|]$, as $T_i^{(j)}$ is a forest, the combination of (41) and (42) shows that $|\,\mathrm{V}(T_i^{(j)})| = |\,\mathrm{E}(T_i^{(j)})|+1$, and thus $T_i^{(j)}$ is indeed a tree. In light of Eq. (42) again, we now find that the connected components of T_i are $V_i^{(1)}, \ldots, V_i^{(\|\pi_i\|)}$, and so the X-tree $(T, \phi_{\mathcal{C}})$ displays π_i. This is the proof. □

3 Further problems

1. Theorem 8 is a result about the property Q_2. Is there a counterpart of it for general Q_k? Note that Remark 6 tells us that some direct generalization may not work.
2. Let X and J be two finite sets. For each $j \in J$, let π_j be a partition of X. For any subset K of J, let $\omega(K)$ denote $|\wedge_{k\in K}\pi_k|_{\mathrm{s}}$. When will the margin weight function ω on 2^J be submodular, namely

$$\omega(A \cup B) + \omega(A \cap B) \le \omega(A) + \omega(B) \tag{43}$$

   holds for all $A, B \in 2^J$? We have conjectured that (43) follows from the assumption that $\{\{\,\pi_j :\ j \in J\,\}\}$ satisfies Q_2 and this is verified by Yinfeng Zhu. Can we characterize those functions on 2^J which are margin weight functions of some partition systems? We mention that this problem is motivated by Lemma 7.
3. McMorris, Warnow and Wimer [MWW94] presented an algorithm for the perfect phylogeny problem which runs in $O(r^{n+1}n^{n+1} + mn^2)$ time, where the character system consists of n r-state characters on a size-m set. By virtue of Theorem 11, it will be interesting to find an algorithm to detect if a given $(0,1)$ array M of size $a_1 \times \cdots \times a_n$ possesses a sparse completion in any time better than $O(r^{n+1}n^{n+1} + mn^2)$, where m is the number of 1-entries of M and $r \doteq \max\{a_1, \ldots, a_n\}$. Bryant and Lagergren [BL06] found a $O(mf(n))$ time algorithm for the tree compatibility problem, where m is the total number of leaves and n is the number of trees. Algorithm 1 is an easy polynomial time algorithm to check whether or not a given $(0,1)$ array has the property Q_2. What is the complexity of checking Q_k for an integer k bigger than two? Is there a fast algorithm to check whether or not a given $(0,1)$ array has a sparse completion?
4. For any $(0,1)$ n-tensor M of size $b_1 \times \cdots \times b_n$, let us write $\mathrm{Size}(M)$ for $b_1 + \cdots + b_n$. Given positive integers $a_1, \ldots, a_n$, what is the value of

$$\min_{\mathcal{C}} \max_{\mathrm{M}^0_{\mathcal{C}}} \mathrm{Size}(\mathrm{M}^0_{\mathcal{C}}),$$

   where $\mathcal{C}$ runs through all compatible partition systems $\mathcal{C} = \{\{\,\pi_1, \ldots, \pi_n\,\}\}$ such that $\|\pi_i\| = a_i$ for all $i \in [n]$ and $\mathrm{M}^0_{\mathcal{C}}$ runs through those all-zeros subarrays of $\mathrm{M}_{\mathcal{C}}$?

5. Every n-uniform hypergraph (simplicial complex) on a finite ground set naturally corresponds to a symmetric $(0, 1)$ n-tensor and hence a corresponding character system. There are quite some definitions of sparse/tree-like hypergraphs [DKM16, FGT11, GGS14, Tow18]. What is the relation between the sparse measures of a hypergraph (simplicial complex) and the sparse measures of the corresponding character system?
6. Gambette, Huber and Scholz constructed the 1-nested uprooted phylogenetic networks for a split system by using insights into the structure of the Buneman graph [GHS17]. Theorem 13 shows a connection between Buneman graph construction and the minimum displaying trees of a partition system. Will this allow us to generalize the work of Gambette et al. to partition systems?
7. We may consider the sparsity properties of those partition systems which can be displayed on tree-like graphs, say a circular split network [DP17, DHK+12]; See Example 7. We can also consider the sparsity properties of those k-compatible or weakly compatible partition systems [GKMW12]. We mention that the partition system given in Example 5 is not weakly compatible, while the one in Example 3 is weakly compatible [DHK+12, p. 149]. For a partition system $\mathcal{C}$, one may ask what is the minimum number of X-trees so that each element of $\mathcal{C}$ can be displayed in at least one of the given X-trees. Due to cancer genomics and other applications [DVPRS17, GGP+96, HHM+18], quite some variants of perfect phylogeny have been proposed and they suggest different ways of relaxing the condition on displaying a data set on a tree. It is expected that examining these different tree-like representations of a data set can lead us to discover and compare some more general natural sparsity measures.
8. Let X be a finite point set in a geodesic metric space Y. For each partition π of X, we call it a *Tverberg partition* provided the convex hulls of the parts of π have a common point and we call it an *anti-Tverberg partition* provided the convex hulls of the parts of π are pairwise disjoint. If $\mathcal{C}$ is a family of anti-Tverberg/Tverberg partitions of X, how to understand its sparsity imposed by the geometry of Y? When Y is a metric tree and when the partition system is anti-Tverberg, Theorem 11 is our contribution in this work towards the previous question. If Y is the d-dimensional Euclidean space $\mathbb{R}^d$, Tverberg partitions have been well studied and named due to a famous result of Tverberg [Tve81]. When $Y = \mathbb{R}^d$ and the partition system $\mathcal{C}$ is anti-Tverberg, how to measure the tree-likeness of $\mathcal{C}$?
9. A linear lattice is just a lattice of commuting equivalence relations [KRY09, Section 3.4]. If the elements of a partition system are all from a linear lattice, can we get a counterpart of Theorem 11 in this case?
10. Let X be a finite set, let T be an X-tree, and let $\mathcal{C}_T$ be the set of partitions of X that can be displayed on T. It is easy to see that $\mathcal{C}_T$ ordered by the refinement relation $\preceq$ is a lattice. Can we characterize all those lattices arising from X-trees in this way?

11. Let r be an integer with $4 \leq r \leq 7$ and let M be a $(0,1)$ n-tensor of size $\overbrace{r \times \cdots \times r}^{n}$. Does there exist a positive integer k_r determined solely by r such that M has a sparse completion whenever $M^{i_1,\ldots,i_{k_r}}$ has a sparse completion for all $\{i_1,\ldots,i_{k_r}\} \in \binom{[n]}{k_r}$? Note that any progress on this question will be one step forward from Theorem 12.
12. Let X be a finite set and let $\mathcal{C} = \{\{\varphi_1,\ldots,\varphi_n\}\}$ be a character system on X. Take $k \in [n]$ and let f be a real function defined on all k-tuples of nonnegative integers. We say that $\mathcal{C}$ has the sparse property $\mathrm{Q}_k(f)$ if it holds
$$|\varphi_{i_1}[Y] \wedge \cdots \wedge \varphi_{i_k}[Y]|_{\mathrm{s}} \leq f(|\varphi_{i_1}[Y]|_{\mathrm{s}},\ldots,|\varphi_{i_k}[Y]|_{\mathrm{s}})$$
for every $\{i_1,\ldots,i_k\} \in \binom{[n]}{k}$ and every $Y \in 2^X \setminus \{\emptyset\}$. If $f(n_1,\ldots,n_k) = \sum_{j=1}^{k}(n_j-1)+2$, does the sparse property $\mathrm{Q}_k(f)$ follow from $\mathrm{Q}_{k-1}(f)$ for $k \geq 3$? A set $\tau \subseteq 2^X$ is slim [HMS18, §4] provided $|\bigcup_{S\in\tau'} S| \geq \sum_{S\in\tau'}(|S|-2)+2$ holds for every nonempty set $\tau' \subseteq \tau$. Accordingly, what can we say on the sparse property $\mathrm{Q}_k(f)$ for which $f(n_1,\ldots,n_k) = \sum_{j=1}^{k}(n_j-2)+2$? Can we get a counterpart of Theorems 9 and 12 for general sparse property $\mathrm{Q}_k(f)$ defined by some other function f?
13. In Remark 8 we mention a fact on displaying a partition system of size two on vertex-labeled trees and edge-labeled trees. What is a possible generalization of it for partition systems of general sizes?
14. For a size-n partition system $\mathcal{C}$ on a finite set X, how to find all minimum X-trees displaying $\mathcal{C}$? When $n = 2$, Theorem 13 (i) gives an answer. When $n \geq 3$, Theorem 13 (ii) and (iii) provide an answer under some additional assumptions.
15. Let F be a forest with n vertices and let $C_1,\ldots,C_m$ enumerate all the connected components of F. We may assume $n_i = |C_i|$ for each $i \in [m]$. Write
$$t \doteq |\{T : T \text{ is a tree}, \mathrm{V}(T) = \mathrm{V}(F) \text{ and } \mathrm{E}(T) \supseteq \mathrm{E}(F)\}|.$$
It follows from Kirchhoff's matrix-tree theorem that $t = \frac{1}{m}\lambda_1\lambda_2\cdots\lambda_{m-1}$, where $\lambda_1,\lambda_2,\cdots,\lambda_{m-1}$ are the non-zero eigenvalues of the following Laplacian matrix:
$$\begin{pmatrix} n_1(n-n_1) & -n_1n_2 & \cdots & -n_1n_m \\ -n_2n_1 & n_2(n-n_2) & \cdots & -n_2n_m \\ \vdots & \vdots & & \vdots \\ -n_mn_1 & -n_mn_2 & \cdots & n_m(n-n_m) \end{pmatrix}.$$
The value of t may help us count the number of all possible outputs of Algorithm 2 when the input is a compatible partition system on X of size two, which, by Theorem 13 (i), is the number of all minimum X-trees displaying that given partition system. Can we get more tidy formula for the number of all minimum X-trees displaying a given compatible partition system of size two? To know that whether or not this number is one is

basically the problem of capturing a perfect phylogeny with two partitions [Ste16, §5.1.1].

16. An r-state fuzzy partition of a set X is a multiset of r nonnegative functions over X whose sum is the all-ones function in $\mathbb{R}^X$ [BH79]. How to generalize the discussion on partition systems here to fuzzy partition systems?

References

ABC⁺18. Noga Alon, Manu Basavaraju, L. Sunil Chandran, Rogers Mathew, and Deepak Rajendraprasad. Separation dimension and sparsity. *J. Graph Theory*, 89(1):14–25, 2018. doi:10.1002/jgt.22236.

AFB94. Richard Agarwala and David Fernández-Baca. A polynomial-time algorithm for the perfect phylogeny problem when the number of character states is fixed. *SIAM J. Comput.*, 23(6):1216–1224, 1994. doi:10.1137/S0097539793244587.

AHJ19. Ron Aharoni, Ron Holzman, and Zilin Jiang. Rainbow fractional matchings. *Combinatorica*, Oct 2019. doi:10.1007/s00493-019-4019-y.

Bai17. Rosemary A. Bailey. Relations among partitions. In *Surveys in Combinatorics 2017*, volume 440 of *London Math. Soc. Lecture Note Ser.*, pages 1–86. Cambridge Univ. Press, Cambridge, 2017. doi:10.1017/9781108332699.002.

BFW92. Hans L. Bodlaender, Mike R. Fellows, and Tandy J. Warnow. Two strikes against perfect phylogeny. In *Automata, Languages and Programming (Vienna, 1992)*, volume 623 of *Lecture Notes in Comput. Sci.*, pages 273–283. Springer, Berlin, 1992. doi:10.1007/3-540-55719-9_80.

BH79. James C. Bezdek and J. Douglas Harris. Convex decompositions of fuzzy partitions. *J. Math. Anal. Appl.*, 67(2):490–512, 1979. doi:10.1016/0022-247X(79)90039-8.

BHM02. Hans-Jürgen Bandelt, Katharina T. Huber, and Vincent Moulton. Quasi-median graphs from sets of partitions. *Discrete Appl. Math.*, 122(1):23–35, 2002. doi:10.1016/S0166-218X(01)00353-5.

BL06. David Bryant and Jens Lagergren. Compatibility of unrooted phylogenetic trees in FPT. *Theoret. Comput. Sci.*, 351(3):296–302, 2006. doi:10.1016/j.tcs.2005.10.033.

Bun71. Peter Buneman. The recovery of trees from measures of dissimilarity. In *Mathematics the Archeological and Historical Sciences*, pages 387–395, United Kingdom, 1971. Edinburgh University Press. URL: http://homepages.inf.ed.ac.uk/opb/homepagefiles/phylogeny-scans/manuscripts.pdf.

Bun74. Peter Buneman. A characterisation of rigid circuit graphs. *Discrete Math.*, 9(3):205–212, 1974. doi:10.1016/0012-365X(74)90002-8.

DHK⁺12. Andreas Dress, Katharina T. Huber, Jacobus Koolen, Vincent Moulton, and Andreas Spillner. *Basic Phylogenetic Combinatorics*. Cambridge University Press, Cambridge, 2012. doi:10.1017/CBO9781139019767.

DKM16. Art M. Duval, Caroline J. Klivans, and Jeremy L. Martin. Simplicial and cellular trees. In *Recent Trends in Combinatorics*, volume 159 of *IMA Vol. Math. Appl.*, pages 713–752. Springer, [Cham], 2016. doi:10.1007/978-3-319-24298-9_28.

DP17. Satyan L. Devadoss and Samantha Petti. A space of phylogenetic networks. *SIAM J. Appl. Algebra Geom.*, 1(1):683–705, 2017. doi:10.1137/16M1103129.

DS92. Andreas Dress and Michael Steel. Convex tree realizations of partitions. *Appl. Math. Lett.*, 5(3):3–6, 1992. doi:10.1016/0893-9659(92)90026-6.

DVPRS17. Gianluca Della Vedova, Murray Patterson, Raffaella Rizzi, and Mauricio Soto. Character-based phylogeny construction and its application to tumor evolution. In *Unveiling Dynamics and Complexity*, volume 10307 of *Lecture Notes in Comput. Sci.*, pages 3–13. Springer, Cham, 2017. doi:10.1007/978-3-319-58741-7_1.

EM77. G. F. Estabrook and F. R. McMorris. When are two qualitative taxonomic characters compatible? *J. Math. Biol.*, 4(2):195–200, 1977. doi:10.1007/BF00275985.

FBV18. David Fernández-Baca and Sudheer R. Vakati. On compatibility and incompatibility of collections of unrooted phylogenetic trees. *Discrete Appl. Math.*, 245:42–58, 2018. doi:10.1016/j.dam.2017.05.002.

FGT11. Fedor V. Fomin, Petr A. Golovach, and Dimitrios M. Thilikos. Approximating width parameters of hypergraphs with excluded minors. *SIAM J. Discrete Math.*, 25(3):1331–1348, 2011. doi:10.1137/080743226.

GGP+96. Leslie Ann Goldberg, Paul W. Goldberg, Cynthia A. Phillips, Elizabeth Sweedyk, and Tandy Warnow. Minimizing phylogenetic number to find good evolutionary trees. *Discrete Appl. Math.*, 71(1-3):111–136, 1996. doi:10.1016/S0166-218X(96)00060-1.

GGS14. Georg Gottlob, Gianluigi Greco, and Francesco Scarcello. Treewidth and hypertree width. In *Tractability*, pages 3–38. Cambridge Univ. Press, Cambridge, 2014. URL: https://www.mat.unical.it/~ggreco/files/GottlobGrecoScarcello.pdf.

GHS17. P. Gambette, K. T. Huber, and G. E. Scholz. Uprooted phylogenetic networks. *Bull. Math. Biol.*, 79(9):2022–2048, 2017. doi:10.1007/s11538-017-0318-x.

GKL15. Alexander Grigoriev, Steven Kelk, and Nela Lekić. On low treewidth graphs and supertrees. *J. Graph Algorithms Appl.*, 19(1):325–343, 2015. doi:10.7155/jgaa.00361.

GKMW12. Stefan Grünewald, Jack H. Koolen, Vincent Moulton, and Taoyang Wu. The size of 3-compatible, weakly compatible split systems. *J. Appl. Math. Comput.*, 40(1-2):249–259, 2012. doi:10.1007/s12190-012-0546-z.

HHM+18. Ademir Hujdurović, Edin Husić, Martin Milanič, Romeo Rizzi, and Alexandru I. Tomescu. Perfect phylogenies via branchings in acyclic digraphs and a generalization of Dilworth's theorem. *ACM Trans. Algorithms*, 14(2):Art. 20, 26, 2018. doi:10.1145/3182178.

HM91. W. L. Hsu and T. H. Ma. Substitution decomposition on chordal graphs and applications. In *ISA '91. Algorithms (Taipei, 1991)*, volume 557 of *Lecture Notes in Comput. Sci.*, pages 52–60. Springer, Berlin, 1991. doi:10.1007/3-540-54945-5_49.

HM02. Katharina T. Huber and Vincent Moulton. The relation graph. *Discrete Math.*, 244(1):153–166, 2002. Algebraic and Topological Methods in Graph Theory. doi:10.1016/S0012-365X(01)00080-2.

HMS04. Katharina T. Huber, Vincent Moulton, and Charles Semple. Replacing cliques by stars in quasi-median graphs. *Discrete Appl. Math.*, 143(1-3):194–203, 2004. doi:10.1016/j.dam.2004.03.002.

HMS18. Katharina T. Huber, Vincent Moulton, and Mike Steel. Phylogenetic flexibility via Hall-type inequalities and submodularity. *Bull. Math. Biol.*, pages 1–20, 2018. doi:10.1007/s11538-018-0419-1.

HMSW14. Katharina T. Huber, Vincent Moulton, Charles Semple, and Taoyang Wu. Representing partitions on trees. *SIAM J. Discrete Math.*, 28(3):1152–1172, 2014. doi:10.1137/130906192.

HRS10. Daniel H. Huson, Regula Rupp, and Celine Scornavacca. *Phylogenetic Networks: Concepts, Algorithms and Applications*. Cambridge University Press, Cambridge, 2010. URL: http://www.phylogenetic-networks.org/.

HS13. Michel Habib and Juraj Stacho. Unique perfect phylogeny is intractable. *Theoret. Comput. Sci.*, 476:47–66, 2013. doi:10.1016/j.tcs.2012.12.048.

HT11. Michel Habib and T.-H. To. On a conjecture about compatibility of multistates characters. In *Algorithms in Bioinformatics*, volume 6833 of *Lecture Notes in Comput. Sci.*, pages 116–127. Springer, Heidelberg, 2011. doi:10.1007/978-3-642-23038-7_11.

IJK19. Leo van Iersel, Mark Jones, and Steven Kelk. A third strike against perfect phylogeny. *Systematic Biology*, 68(5):814–827, 2019. doi:10.1093/sysbio/syz009.

JK15. Tibor Jordán and Viktória E. Kaszanitzky. Sparse hypergraphs with applications in combinatorial rigidity. *Discrete Appl. Math.*, 185:93–101, 2015. doi:10.1016/j.dam.2014.12.020.

KRY09. Joseph P. S. Kung, Gian-Carlo Rota, and Catherine H. Yan. *Combinatorics: the Rota way*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2009. doi:10.1017/CBO9780511803895.

LGS11. Fumei Lam, Dan Gusfield, and Srinath Sridhar. Generalizing the splits equivalence theorem and four gamete condition: perfect phylogeny on three-state characters. *SIAM J. Discrete Math.*, 25(3):1144–1175, 2011. doi:10.1137/090776305.

Mea83. C. A. Meacham. Theoretical and computational considerations of the compatibility of qualitative taxonomic characters. In *Numerical Taxonomy*, pages 304–314, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg. doi:10.1007/978-3-642-69024-2_34.

MWW94. F. R. McMorris, Tandy J. Warnow, and Thomas Wimer. Triangulating vertex-colored graphs. *SIAM J. Discrete Math.*, 7(2):296–306, 1994. doi:10.1137/S0895480192229273.

NOdM12. Jaroslav Nešetřil and Patrice Ossona de Mendez. *Sparsity: Graphs, Structures, and Algorithms*, volume 28 of *Algorithms and Combinatorics*. Springer, Heidelberg, 2012. doi:10.1007/978-3-642-27875-4.

Rot97. Gian-Carlo Rota. The many lives of lattice theory. *Notices Amer. Math. Soc.*, 44(11):1440–1445, 1997. URL: http://www.ams.org/notices/199711/comm-rota.pdf.

SFB12. Brad Shutters and David Fernández-Baca. A simple characterization of the minimal obstruction sets for three-state perfect phylogenies. *Appl. Math. Lett.*, 25(9):1226–1229, 2012. doi:10.1016/j.aml.2012.02.060.

Spi17. Daniel A. Spielman. Graphs, vectors, and matrices. *Bull. Amer. Math. Soc. (N.S.)*, 54(1):45–61, 2017. doi:10.1090/bull/1557.

SS03. Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003. URL: https://global.oup.com/academic/product/phylogenetics-9780198509424?cc=cn&lang=en&.

ST11. Ileana Streinu and Louis Theran. Natural realizations of sparsity matroids. *Ars Math. Contemp.*, 4(1):141–151, 2011. URL: https://amc-journal.eu/index.php/amc/article/viewFile/197/126.

Ste92. Michael Steel. The complexity of reconstructing trees from qualitative characters and subtrees. *J. Classification*, 9(1):91–116, 1992. doi:10.1007/BF02618470.

Ste16. Mike Steel. *Phylogeny–Discrete and Random Processes in Evolution*, volume 89 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016. doi:10.1137/1.9781611974485.ch1.

SVFB13. Brad Shutters, Sudheer Vakati, and David Fernández-Baca. Incompatible quartets, triplets, and characters. *Algorithms for Molecular Biology*, 8(11), Apr 2013. 10 pp. doi:10.1186/1748-7188-8-11.

Tar16. A. A. Taranenko. Permanents of multidimensional matrices: properties and applications. *Diskretn. Anal. Issled. Oper.*, 23(4):35–101, 2016. doi:10.1134/S1990478916040141.

Tow18. Henry Towsner. An analytic approach to sparse hypergraphs: hypergraph removal. *Discrete Anal.*, 2018. Paper No. 3, 47 pp. URL: https://discreteanalysisjournal.com/article/3104-an-analytic-approach-to-sparse-hypergraphs-hypergraph-removal.

Tve81. H. Tverberg. A generalization of Radon's theorem. II. *Bull. Austral. Math. Soc.*, 24(3):321–325, 1981. doi:10.1017/S0004972700004858.

Whi89. Walter Whiteley. A matroid on hypergraphs, with applications in scene analysis and geometry. *Discrete Comput. Geom.*, 4(1):75–95, 1989. doi:10.1007/BF02187716.

XGG15. Shou-Jun Xu, Rob Gysel, and Dan Gusfield. Minimum average distance clique trees. *SIAM J. Discrete Math.*, 29(3):1706–1734, 2015. doi:10.1137/15M1021052.