



Order Distances and Split Systems

Vincent Moulton¹ · Andreas Spillner² 

Received: 9 September 2020 / Accepted: 1 September 2021 / Published online: 20 September 2021
© The Author(s) 2021

Abstract

Given a pairwise distance D on the elements in a finite set X , the *order distance* $\Delta(D)$ on X is defined by first associating a total preorder \leq_x on X to each $x \in X$ based on D , and then quantifying the pairwise disagreement between these total preorders. The order distance can be useful in relational analyses because using $\Delta(D)$ instead of D may make such analyses less sensitive to small variations in D . Relatively little is known about properties of $\Delta(D)$ for general distances D . Indeed, nearly all previous work has focused on understanding the order distance of a *treelike distance*, that is, a distance that arises as the shortest path distances in a tree with non-negative edge weights and X mapped into its vertex set. In this paper we study the order distance $\Delta(D)$ for distances D that can be decomposed into sums of simpler distances called split-distances. Such distances D generalize treelike distances, and have applications in areas such as classification theory and phylogenetics.

Keywords Total preorder · Order distance · Treelike distance · Kalmanson distance · Circular split system · Flat split system

1 Introduction

In areas such as phylogenetics, it is common to take a distance between elements in some set X of interest as a first step towards representing the relationships between elements in X using some discrete structure such as a tree [1, Ch. 7]. However, small inaccuracies in measuring the distance can lead to large changes in the final representation. One approach to deal with this problem is to quantify the difference between the orderings on X induced by the distance between elements in X and build the final representation based on these differences rather than from the distance directly. This can make the whole process less sensitive to small variations in the measured distance [2].

✉ Andreas Spillner
andreas.spillner@hs-merseburg.de

Vincent Moulton
v.moulton@uea.ac.uk

¹ School of Computing Sciences, University of East Anglia, NR4 7TJ, Norwich, UK

² Merseburg University of Applied Sciences, 06217, Merseburg, Germany

A formal way to define this process is as follows. Suppose that D is a distance on a finite, non-empty set X , i.e. a symmetric map $D : X \times X \rightarrow \mathbb{R}$ with $D(x, x) = 0$ and $D(x, y) \geq 0$ for all $x, y \in X$. To any element $x \in X$ we associate the binary relation \preceq_x on X by putting $y \preceq_x z$ if $D(x, y) \leq D(x, z)$ for all $y, z \in X$. In addition, we use $y \prec_x z$ to express $D(x, y) < D(x, z)$. Note that \preceq_x is transitive and total, a type of binary relation that is sometimes called a *total preorder* or a *weak order* (see e.g. the discussion in [3, p. 630]). To measure how much \preceq_x and $\preceq_{x'}$ disagree for $x, x' \in X$, we penalize every pair of elements $y, z \in X$ by 2 if they are ordered oppositely in \preceq_x and $\preceq_{x'}$ and by 1 if they are neither ordered oppositely nor equally in \preceq_x and $\preceq_{x'}$. More formally, the penalization term $\delta_{y,z}(x, x')$ is defined as

- 2 if $(y \prec_x z \text{ and } z \prec_{x'} y)$ or $(y \prec_{x'} z \text{ and } z \prec_x y)$,
- 1 if $(y \prec_x z, z \not\prec_{x'} y \text{ and } y \not\prec_{x'} z)$ or $(y \prec_{x'} z, z \not\prec_x y \text{ and } y \not\prec_x z)$,
- 0 else.

The order distance $\Delta(D)$ on X is defined by setting

$$\Delta(D)(x, x') = \frac{1}{2} \cdot \sum_{y \in X} \sum_{z \in X} \delta_{y,z}(x, x').$$

This way of measuring the disagreement between total preorders was proposed, for example, in [4] and later in [3].

Most previous results on order distances have concerned properties of $\Delta(D)$ for D a *treelike* distance (see e.g. [2, 5–7]). These are precisely those distances D on X that arise through a map $\varphi : X \rightarrow V$ of X into the vertex set V of a tree $T = (V, E)$ with non-negative edge weighting $\ell : E \rightarrow \mathbb{R}_{\geq 0}$ and then putting $D(x, y)$ to be the length of the shortest path from $\varphi(x)$ to $\varphi(y)$ in T . The pair (T, φ) is called an *X-tree* and, to ensure that, up to isomorphism, the tree T together with the maps φ and ℓ is uniquely determined by the treelike distance D , it is usually assumed that the image $\varphi(X)$ contains all vertices of T with degree at most two and that ℓ is strictly positive (see e.g. [1, Sec. 7.1]). We denote the treelike distance D arising from an edge-weighted X -tree (T, φ, ℓ) by $D_{(T, \varphi, \ell)}$ and say that the edge-weighted X -tree *generates* D . In Fig. 1 we illustrate the main result

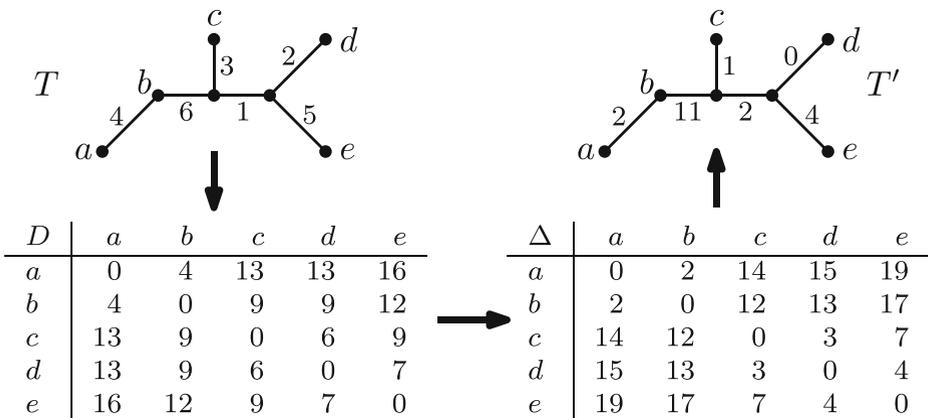


Fig. 1 An edge-weighted X -tree (T, φ, ℓ) with $X = \{a, b, c, d, e\}$ and the distance $D = D_{(T, \varphi, \ell)}$. The associated order distance $\Delta = \Delta(D)$ is generated by an edge-weighted X -tree (T', φ', ℓ') obtained by adjusting the weights of the edges in T

of Bonnot et al. in [2]: The order distance $\Delta(D)$ of a treelike distance $D = D_{(T, \varphi, \ell)}$ on X is always treelike and the edge-weighted X -tree (T', φ', ℓ') with $\Delta(D) = D_{(T', \varphi', \ell')}$ is obtained by adjusting the edge-weights of T . This result is important since it shows that the order distance $\Delta = \Delta(D)$ of an unknown treelike distance D on X allows to gain information about the structure of the X -tree that generates D .

In this paper, we aim to better understand to what extent Bonnot et al.’s result can be generalized to discrete structures that include X -trees as a special case and are commonly used in phylogenetics. To describe them, we call an unordered pair of non-empty subsets A and B of a finite set X with $A \cup B = X$ and $A \cap B = \emptyset$ a *split* of X . We denote such a pair by $A \mid B$ and, since it is an unordered pair, $B \mid A$ refers to same split of X . Moreover, we denote by $D_{A \mid B}$ the *split-distance* on X obtained by taking $D_{A \mid B}(x, y) = 1$ if $|A \cap \{x, y\}| = |B \cap \{x, y\}| = 1$ and $D_{A \mid B}(x, y) = 0$ otherwise for all $x, y \in X$. A set \mathcal{S} of splits of X is called a *split system* on X and an ordered pair (\mathcal{S}, ω) consisting of a split system \mathcal{S} on X and a non-negative weighting $\omega : \mathcal{S} \rightarrow \mathbb{R}_{\geq 0}$ is called a *weighted split system* on X . In addition, we call

$$D_{(\mathcal{S}, \omega)} = \sum_{S \in \mathcal{S}} \omega(S) \cdot D_S$$

the distance *generated* by the weighted split system (\mathcal{S}, ω) . Distances D on X that are generated by a weighted split system on X are called ℓ_1 -distances and always satisfy the triangle inequality (see e.g. [8, Ch. 4]). Every treelike distance is an ℓ_1 -distances (see e.g. [1, Sec. 7.4]). Bonnot et al.’s result, phrased in the language of split systems, suggests to look for split systems \mathcal{S} on X with the property that for every non-negative weighting ω of the splits in \mathcal{S} , there exists a non-negative weighting ω' of the splits in \mathcal{S} such that

$$\Delta(D_{(\mathcal{S}, \omega)}) = D_{(\mathcal{S}, \omega')}.$$

We call such split systems *orderly* and will mainly investigate these here using the language of splits and distances to establish some of their basic properties. We suspect, however, that further insights into the structure of orderly split systems will require a better understanding of the finite space of those collections $\{\preceq_x : x \in X\}$ of total preorders that arise from a split system on X through the distances generated by weighting the split system.

The rest of the paper is structured as follows. In Section 2 we introduce the *preorder split system* \mathcal{S}_D of a distance D , a concept which was implicitly considered in [9] and that is key to many of our arguments. In addition, we formally introduce some classes of split systems that commonly arise in phylogenetics and for which we will study in the remainder of this paper to what extend they are orderly. In Section 3 we show that so-called *circular* split systems [10] are orderly in case they have maximum size (Theorem 1). We also briefly consider the consequences of this result for efficiently computing the order distance for distances that are generated by circular split systems. In Section 4 we then explore orderly split systems within the class of so-called *flat* split systems [11, 12]. In particular, we show that within the class of maximum sized flat split systems, the orderly split systems are precisely those that are circular (Theorem 3). We conclude in Section 5 with some possible directions for future work. The appendix contains the proofs of some technical lemmas and auxiliary results presented in Sections 2 and 4.

2 Preorders and Split Systems

In this section we present some concepts related to split systems that will be used later. For the rest of this paper X will denote a finite non-empty set with $|X| = n$ and $\binom{X}{2}$ denotes the set of all 2-element subsets of X . Moreover, by $\mathcal{S}(X)$ we denote the set consisting of all possible splits of X . It will be convenient to use the notation $A \mid B$ introduced for splits of X also for the unordered pair $\emptyset \mid X$, even though this pair is not considered a split of X .

Given a distance D on X and its associated collection $\{\preceq_x : x \in X\}$ of total preorders, we define, for all $u, v \in X$ with $u \neq v$, the set

$$X_{u,v} = \{x \in X : u \prec_x v\}.$$

The *preorder split system* \mathcal{S}_D associated to D is the set of splits of X of the form $S_{u,v} = X_{u,v} \mid X - X_{u,v}$ for $u, v \in X$ with $u \neq v$ and $X_{u,v} \neq \emptyset, X$. \mathcal{S}_D is closely related to the so-called *midpath phylogeny* introduced in [9]. The relevance of the preorder split system for the work presented here comes from the fact that, as established in [2], the order distance $\Delta(D)$ associated to D can always be written as

$$\Delta(D) = \sum_{S \in \mathcal{S}_D} \omega_D(S) \cdot D_S, \tag{1}$$

where $\omega_D(S)$ equals the number of $(u, v) \in X \times X$ with $u \neq v$ such that $S_{u,v} = S$. Note that this implies that, for all distances D , the order distance $\Delta(D)$ is an ℓ_1 -distance, even if D itself is not an ℓ_1 -distance.

In the next lemma, whose proof can be found in Appendix A, we present a useful upper bound on the size of \mathcal{S}_D . Note that, as a direct consequence of the definition of \mathcal{S}_D , it follows that \mathcal{S}_D contains at most $n(n - 1)$ splits. In view of $|\mathcal{S}(X)| = 2^{n-1} - 1$, we immediately see that for $n \leq 5$ this upper bound on $|\mathcal{S}_D|$ is not tight. Using a computer program, the smallest n for which we found a distance D on a set X with n elements and $|\mathcal{S}_D| = n(n - 1)$ is $n = 16$.

Lemma 1 *Let D be a distance on a set X with $n \geq 1$ elements. Then we have $|\mathcal{S}_D| \leq n(n - 1)$ and, for all sufficiently large $n \in \mathbb{N}$, there exists a distance D on X such that equality holds.*

In view of (1) a necessary condition for a split system \mathcal{S} on X to be orderly is that, for all distances D generated by weighting \mathcal{S} , the preorder split system \mathcal{S}_D belongs to the same class of split systems as \mathcal{S} . In the remainder of this section we formally define the classes of split systems we consider in this paper.

Two splits $A_1 \mid B_1$ and $A_2 \mid B_2$ of X are *compatible* if at least one of the intersections $A_1 \cap A_2, A_1 \cap B_2, B_1 \cap A_2$, and $B_1 \cap B_2$ is empty. In addition, we also call a split system \mathcal{S} on X compatible if the splits in \mathcal{S} are pairwise compatible. A split system \mathcal{S} on X is compatible if and only if there exists an X -tree (T, φ) such that the splits in \mathcal{S} are precisely those that arise by removing an edge e of T and then considering the two subsets of X mapped by φ into the resulting connected components of $T - e$. Such an X -tree will be said to *represent* the compatible split system and it is unique up to isomorphism (see e.g. [1, Sec. 3.1]). Since an X -tree has at most $2n - 3$ edges (see e.g. [1, Sec. 2.1]) it follows that a compatible split system on X contains at most $2n - 3$ splits. In Fig. 2(a) we give an example of a compatible split system \mathcal{S} and Fig. 2(b) displays the unique X -tree that represents \mathcal{S} .

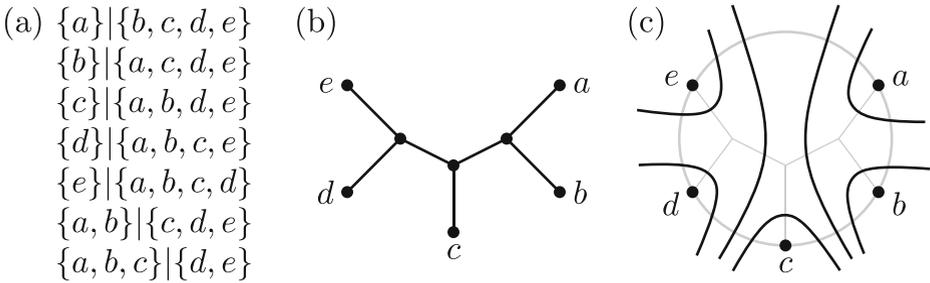


Fig. 2 (a) A split system \mathcal{S} consisting of 7 pairwise compatible splits of $X = \{a, b, c, d, e\}$. (b) The X -tree representing the split system \mathcal{S} . (c) The compatible split system \mathcal{S} is circular (but not maximum circular) and fits on the permutation $\pi = (a, b, c, d, e)$. The maximum circular split system \mathcal{S}_π contains 10 splits

The following lemma, whose proof can also be found in Appendix A, establishes that there is a 6-point condition that characterizes when the preorder split system \mathcal{S}_D of a distance D on X is compatible. To state the lemma, we denote the *restriction* of a distance D on X to a subset $M \subseteq X$ by $D|_M$. In [9] an example of a distance D on a 6-element set X is provided such that $\mathcal{S}_{D|_M}$ is compatible for all 5-element subsets $M \subseteq X$, but \mathcal{S}_D is not compatible. Hence, compatibility of \mathcal{S}_D cannot be characterized by a k -point condition for $k < 6$.

Lemma 2 *Let D be a distance on a set X with $n \geq 6$ elements. The preorder split system $\mathcal{S}_D \subseteq \mathcal{S}(X)$ is compatible if and only if for every subset $M \subseteq X$ with $|M| = 6$ the preorder split system $\mathcal{S}_{D|_M} \subseteq \mathcal{S}(M)$ is compatible.*

Next we consider a class of split systems that has applications in the analysis of biological data where it underlies certain networks which generalize X -trees (see e.g. [13]). A split system \mathcal{S} on X is *circular* if there exists a permutation $\pi = (x_1, x_2, \dots, x_n)$ of the elements in X such that, for any split $S \in \mathcal{S}$, there exist $1 \leq i \leq j < n$ with

$$S = \{x_i, x_{i+1}, \dots, x_j\} | X - \{x_i, x_{i+1}, \dots, x_j\},$$

in which case we will say that \mathcal{S} (and also each single split $S \in \mathcal{S}$) *fits* on π . We denote the circular split system consisting of all splits that fit on a permutation π of X by \mathcal{S}_π and remark that π is, up to cyclic shifting and reversal, uniquely determined by \mathcal{S}_π . It follows immediately from the definition that a circular split system \mathcal{S} on X contains at most $\binom{n}{2}$ splits and if \mathcal{S} contains precisely $\binom{n}{2}$ splits it is called *maximum circular*. Circular split systems naturally appear in the context of the so-called *split decomposition* of a distance (see [10, Sec. 3], where they are introduced). As illustrated in Fig. 2(c), every compatible split system is circular, but not vice versa [10]. In Section 3 we will further study distances for which the preorder split system is circular.

Again motivated by applications in phylogenetics (see e.g. [14]), a generalization of circular split systems was introduced in [11]. To define this class of split systems we first associate to any permutation $\pi = (x_1, x_2, \dots, x_n)$ of X , and any $k \in \{1, 2, \dots, n - 1\}$ the permutation

$$\pi(k) = (x_1, \dots, x_{k-1}, x_{k+1}, x_k, x_{k+2}, \dots, x_n),$$

that is, the permutation obtained by swapping the elements at positions k and $k + 1$ in π . We denote the set of the two elements that are swapped by $sw(\pi, k) = \{x_k, x_{k+1}\}$ and define the split

$$S(\pi, k) = \{x_1, \dots, x_k\} \mid \{x_{k+1}, \dots, x_n\}.$$

Then, putting $m = \binom{n}{2}$, we consider pairs (π, κ) consisting of a permutation π of X and a sequence $\kappa = (k_1, \dots, k_m) \in \{1, 2, \dots, n - 1\}^m$. With each such pair (π, κ) we associate the sequence $\pi_0, \pi_1, \dots, \pi_m$ of permutations of X defined by putting $\pi_0 = \pi$ and $\pi_i = \pi_{i-1}(k_i)$, $1 \leq i \leq m$. The pair (π, κ) is called *allowable* if $sw(\pi_{i-1}, k_i) \neq sw(\pi_{j-1}, k_j)$ holds for all $1 \leq i < j \leq m$. Such a sequence of permutations is commonly called in the literature a *simple allowable sequence* and it can be graphically represented by a so-called *wiring diagram* [15]. In the wiring diagram the elements in X are represented by x -monotone curves (the wires) with each pair of curves crossing precisely once (representing the swaps, see Fig. 3 for an example). A split system $\mathcal{S} \subseteq \mathcal{S}(X)$ is called *flat* if there exists an allowable pair (π, κ) with

$$\mathcal{S} \subseteq \mathcal{S}_{(\pi, \kappa)} = \{S(\pi_{i-1}, k_i) : 1 \leq i \leq m\}.$$

As for circular split systems, it follows immediately from the definition that a flat split system on X contains at most $\binom{n}{2}$ splits and if \mathcal{S} contains precisely $\binom{n}{2}$ splits it is called *maximum flat*. As illustrated in Fig. 3, every circular split system is flat, but not vice versa [12]. Flat split systems will be considered in Section 4 where we will use the following characterization of maximum circular split systems (see Appendix A for a proof):

Lemma 3 *A maximum flat split system $\mathcal{S} \subseteq \mathcal{S}(X)$ that contains the split $\{x\} \mid X - \{x\}$ for all $x \in X$ is maximum circular.*

We conclude this section with a class of split systems that provides a useful conceptual framework for all classes of split systems mentioned so far. A split system \mathcal{S} on X is *linearly independent* if the set of split distances $\{D_S : S \in \mathcal{S}\}$ arising from \mathcal{S} is linearly independent when viewed as a finite subset of the vector space of all symmetric bivariate maps $D : X \times X \rightarrow \mathbb{R}$ with $D(x, x) = 0$ for all $x \in X$. Linearly independent split systems were introduced in [11], where it is also shown that every flat split system (referred to in [11] as *pseudo-affine* split system) is linearly independent, but not vice versa. Again it follows

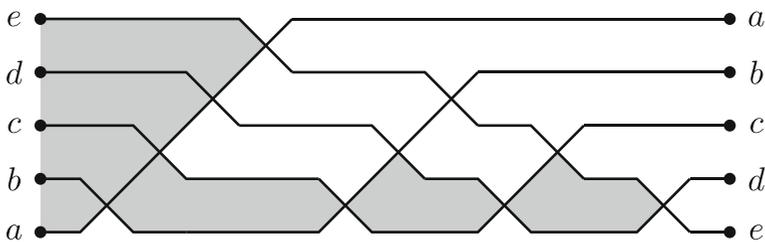


Fig. 3 A wiring diagram representing the simple allowable sequence of permutations of $X = \{a, b, c, d, e\}$ obtained from the allowable pair (π, κ) with $\pi = (a, b, c, d, e)$ and $\kappa = (1, 2, 3, 4, 1, 2, 3, 1, 2, 1)$. The face immediately to the left of each crossing yields a split $A \mid B$ of X with A and B corresponding to those elements in X whose wires are below and above that face, respectively. The maximum circular split system \mathcal{S}_π from Fig. 2(c) equals the maximum flat split system $\mathcal{S}_{(\pi, \kappa)}$ and the faces corresponding to splits in the compatible split system \mathcal{S} from Fig. 2(a) are shaded gray

immediately from the definition that a linearly independent split system on X contains at most $\binom{n}{2}$ splits and if \mathcal{S} contains precisely $\binom{n}{2}$ splits it is called *maximum* linearly independent. The next proposition (see Appendix A for a proof) gives a characterization of maximum flat split systems among all maximum linearly independent split systems that will be used in Section 4. A split system \mathcal{S} on X satisfies the *pairwise separation property* if for any two distinct elements $x, y \in X$, there exist disjoint subsets A and B of $X - \{x, y\}$ with $A \cup B = X - \{x, y\}$ such that $A \cup \{x, y\} \mid B, A \cup \{x\} \mid B \cup \{y\}, A \cup \{y\} \mid B \cup \{x\}$ and $A \mid B \cup \{x, y\}$ are all contained in $\mathcal{S} \cup \{\emptyset \mid X\}$.

Proposition 1 *A maximum linearly independent split system \mathcal{S} on a set X with $n \geq 2$ elements is maximum flat if and only if it satisfies the pairwise separation property.*

3 Circular Split Systems

The key result of this section (Lemma 4) establishes that the preorder split system of a distance generated by a weighted circular split system is always circular. This will provide a useful starting point for applying the data analysis process mentioned in the introduction to the computation of networks such as those described [13]. Using the notation introduced in the introduction, we define a distance D on X to be *circular* if there exists a circular split system \mathcal{S} on X together with a non-negative weighting ω of \mathcal{S} such that $D = D_{(\mathcal{S}, \omega)}$. It follows from split decomposition theory [10] that if such a pair (\mathcal{S}, ω) exists then it is unique up to the addition or removal of splits with weight 0.

Circular distances are also known as *Kalmanson distances* [16] and in the proof of Lemma 4 below we will use the following characterization of these distances given in [17] (see also [18]): A distance D on X is circular if and only if there exists a permutation $\pi = (x_1, x_2, \dots, x_n)$ of the elements in X such that

$$\max(D(x_i, x_j) + D(x_k, x_l), D(x_i, x_l) + D(x_j, x_k)) \leq D(x_i, x_k) + D(x_j, x_l) \quad (2)$$

holds for all $1 \leq i < j < k < l \leq n$. In particular, if $D = D_{(\mathcal{S}, \omega)}$ for a weighted circular split system (\mathcal{S}, ω) such that \mathcal{S} fits on a certain permutation π of the elements in X then D satisfies condition (2) for this π .

Lemma 4 *Let (\mathcal{S}, ω) be a circular split system on X with non-negative weighting ω such that \mathcal{S} fits on the permutation $\pi = (x_1, x_2, \dots, x_n)$ of the elements in X . Then, for $D = D_{(\mathcal{S}, \omega)}$, the preorder split system \mathcal{S}_D is circular and fits on π .*

Proof If $\mathcal{S}_D = \emptyset$ then \mathcal{S}_D is circular and it fits on π . So assume that $\mathcal{S}_D \neq \emptyset$ and consider any split $S \in \mathcal{S}_D$. By the definition of \mathcal{S}_D there exist $u, v \in X$ with $u \neq v$ and $S = X_{u,v} \mid X - X_{u,v}$. By the definition of the set $X_{u,v}$, we must have $u \in X_{u,v}$ and $v \in X - X_{u,v}$. Assume for a contradiction that neither $X_{u,v}$ nor $X - X_{u,v}$ form an interval of consecutive elements in π . This implies that there exist $u' \in X_{u,v}$ with $u' \neq u$ and $v' \in X - X_{u,v}$ with $v' \neq v$ such that, after possibly shifting and/or reversing π , the restriction of π to $\{u, v, u', v'\}$ is (u, v', u', v) . Then, in view of the definition of $X_{u,v}$, we must have

$$\begin{aligned} D(u, u') &< D(v, u') \text{ and} \\ D(v, v') &\leq D(u, v'). \end{aligned}$$

This implies

$$D(u, u') + D(v, v') < D(v, u') + D(u, v'),$$

contradicting condition (2). □

As a consequence of Lemma 4 we obtain the main result of this section:

Theorem 1 *Every maximum circular split system is orderly.*

Proof Let \mathcal{S} be a maximum circular split system together with a non-negative weighting ω and put $D = D_{(\mathcal{S}, \omega)}$. By Eq. 1 we have

$$\Delta(D) = \sum_{S \in \mathcal{S}_D} \omega_D(S) \cdot D_S.$$

By Lemma 4, \mathcal{S}_D fits onto the unique permutation π of the elements in X with $\mathcal{S} = \mathcal{S}_\pi$. Hence, we have $\mathcal{S}_D \subseteq \mathcal{S}$ and, therefore,

$$\Delta(D) = \sum_{S \in \mathcal{S}} \omega'(S) \cdot D_S,$$

where $\omega'(S) = \omega_D(S)$ if $S \in \mathcal{S}_D$ and $\omega'(S) = 0$ otherwise. Since ω was chosen arbitrarily, it follows that \mathcal{S} is orderly. □

The corollary below follows from Theorem 1 in view of the definition of an orderly split system and the fact that every circular distance is generated by some weighted maximum circular split system.

Corollary 1 *The order distance $\Delta(D)$ of a circular distance D is always circular.*

Note that in Theorem 1 we assume that the circular split system is maximum. To illustrate that we cannot remove this assumption, consider, for example, the non-maximum circular split system

$$\mathcal{S} = \{\{b\} \mid \{a, c, d\}, \{a, b\} \mid \{c, d\}, \{a, d\} \mid \{b, c\}\}$$

on $X = \{a, b, c, d\}$ and the weighting ω that assigns weight 1 to every split in \mathcal{S} . This yields the order distance $\Delta = \Delta(D)$ associated to $D = D_{(\mathcal{S}, \omega)}$ with

$$\begin{aligned} \Delta(a, b) &= \Delta(a, c) = \Delta(b, c) = 8, \\ \Delta(a, d) &= \Delta(c, d) = 4 \text{ and } \Delta(b, d) = 10. \end{aligned}$$

which is generated as $\Delta = D_{(\mathcal{S}', \omega')}$ by the weighted circular split system $\mathcal{S}' \supseteq \mathcal{S}$ with

$$\begin{aligned} \omega'(\{a, b\} \mid \{c, d\}) &= \omega'(\{a, d\} \mid \{b, c\}) = 3, \\ \omega'(\{a\} \mid \{b, c, d\}) &= \omega'(\{c\} \mid \{a, b, d\}) = 1 \text{ and } \omega'(\{b\} \mid \{a, c, d\}) = 4. \end{aligned}$$

So, in general, if D is generated by a non-maximum circular split system \mathcal{S} on X the order distance associated to D may be generated only by a proper superset of \mathcal{S} . We will explore this phenomenon further in the next section when we consider so-called closed split systems.

In the remainder of this section, we briefly look into some consequences of Lemma 4 for computing order distances. Applying an algorithm described in [19] for computing the disagreement between two total preorders to each pair of elements in X , $\Delta(D)$ can be computed in $\mathcal{O}(n^3 \log n)$ time for general distances D on X . For the special case of a treelike distance D a running time in $\mathcal{O}(n^2 \log n)$ can be achieved using the approach developed

in [9] for computing the midpath phylogeny. In the following theorem we show that this running time can also be achieved for circular distances D . Clearly, no algorithm for computing $\Delta(D)$ can run faster than the size of the output, that is, we have a lower bound of $\Omega(n^2)$.

Theorem 2 *The order distance $\Delta(D)$ of a circular distance D on a set X with n elements can be computed in $\mathcal{O}(n^2 \log n)$ time.*

Proof The first step in the computation of $\Delta(D)$ from D is to obtain a permutation $\pi = (x_1, x_2, \dots, x_n)$ of the elements in X such that, $D = D_{(\mathcal{S}, \omega)}$ for a suitable non-negative weighting ω of the maximum circular split system $\mathcal{S} = \mathcal{S}_\pi$. In view of the assumption that D is circular such a permutation must exist and it can be computed (along with a suitable weighting ω that we do not use here) in $\mathcal{O}(n^2)$ time with the algorithm presented in [18].

In view of Lemma 4, we have $\mathcal{S}_D \subseteq \mathcal{S}_\pi$. More specifically, every split $S = A | B \in \mathcal{S}_D$ must fit on π and, therefore, be such that the elements in either A or B form an interval $I(S) = \{x_i, x_{i+1}, \dots, x_j\}$, $1 \leq i \leq j < n$, of consecutive elements in π . We count, for all $1 \leq i \leq j < n$, the number $p(i, j)$ of ordered pairs $(u, v) \in X \times X$ with $u \neq v$, $S_{u,v} \in \mathcal{S}_D$ and $I(S_{u,v}) = \{x_i, x_{i+1}, \dots, x_j\}$. Since D satisfies the triangle inequality, it follows from the definition of the sets $X_{u,v}$ and $X_{v,u}$ that $S_{u,v} = S_{v,u} = \emptyset | X$ for any two distinct $u, v \in X$ with $D(u, v) = 0$. Thus, it suffices to consider those $(u, v) \in X \times X$ with $D(u, v) > 0$ and, again in view of the definition of the set $X_{u,v}$, we then have $S_{u,v} \in \mathcal{S}_D$. For each of these pairs (u, v) we use binary search to compute the indices i and j with $I(S_{u,v}) = \{x_i, x_{i+1}, \dots, x_j\}$ in $\mathcal{O}(\log n)$ time. This yields all the required numbers $p(i, j)$, $1 \leq i \leq j < n$, in $\mathcal{O}(n^2 \log n)$ time. Then, in view of Eq. 1, we have $\Delta(D) = D_{(\mathcal{S}_\pi, \omega')}$, where we put $\omega'(S) = p(i, j)$ for the split $S \in \mathcal{S}_\pi$ with $I(S) = \{x_i, x_{i+1}, \dots, x_j\}$, $1 \leq i \leq j < n$. The distance $D_{(\mathcal{S}_\pi, \omega')}$ can be computed from π and ω' in $\mathcal{O}(n^2)$ time (see e.g. [20] where this and related computational problems on split systems are discussed). \square

4 Flat Split Systems

The main result of this section (Theorem 3) establishes that maximum flat split systems that are orderly are necessarily circular. This result implies that, in contrast to circular split systems, the data analysis process mentioned in the introduction cannot be used directly to compute networks based on flat split systems such as those described in [14].

We begin by presenting some properties of linearly independent split systems that may be of independent interest in future work but are mainly provided here to subsequently be applied to the special case of flat split systems. The next lemma provides a useful link between the combinatorial structure of a linearly independent split system and the order distances that it generates. We call two splits $A_1 | B_1$ and $A_2 | B_2$ of X *incompatible*, for short, if they are not compatible. Moreover, we call a split system \mathcal{S} on a set X with $n \geq 4$ elements *closed* if for any two incompatible splits $A_1 | B_1$ and $A_2 | B_2$ in \mathcal{S} at least one of the following holds:

- (a) \mathcal{S} also contains the splits $A_1 \cap A_2 | X - (A_1 \cap A_2)$, $B_1 \cap A_2 | X - (B_1 \cap A_2)$, $A_1 \cap B_2 | X - (A_1 \cap B_2)$ and $B_1 \cap B_2 | X - (B_1 \cap B_2)$.
- (b) $|A_1 \cap A_2| \cdot |B_1 \cap B_2| = |A_1 \cap B_2| \cdot |B_1 \cap A_2|$ and \mathcal{S} also contains the split $(A_1 \cap A_2) \cup (B_1 \cap B_2) | (A_1 \cap B_2) \cup (B_1 \cap A_2)$.
- (c) $|A_1 \cap A_2| \cdot |B_1 \cap B_2| > |A_1 \cap B_2| \cdot |B_1 \cap A_2|$ and \mathcal{S} also contains the splits $(A_1 \cap A_2) \cup (B_1 \cap B_2) | (A_1 \cap B_2) \cup (B_1 \cap A_2)$, $A_1 \cap A_2 | X - (A_1 \cap A_2)$ and $B_1 \cap B_2 | X - (B_1 \cap B_2)$.

- (d) $|A_1 \cap A_2| \cdot |B_1 \cap B_2| < |A_1 \cap B_2| \cdot |B_1 \cap A_2|$ and \mathcal{S} also contains the splits $(A_1 \cap A_2) \cup (B_1 \cap B_2) \mid (A_1 \cap B_2) \cup (B_1 \cap A_2)$, $B_1 \cap A_2 \mid X - (B_1 \cap A_2)$ and $A_1 \cap B_2 \mid X - (A_1 \cap B_2)$.

Lemma 5 *Let X be a set with $n \geq 4$ elements and \mathcal{S} a linearly independent split system on X . If \mathcal{S} is orderly then \mathcal{S} is closed.*

The somewhat technical proof of Lemma 5 can be found in Appendix A, where it is also shown that Lemma 5 together with Theorem 1 yields the following characterization of orderly split systems amongst all maximum linearly independent split systems on sets with 5 elements.

Proposition 2 *A maximum linearly independent split system \mathcal{S} on a set X with 5 elements is orderly if and only if it is circular.*

We remark that maximum circular split systems on sets with 4 elements cannot be characterized as in Proposition 2. It would be interesting to know if Proposition 2 can be extended to all values of $n \geq 5$, but we have been unable to find a proof or counter-example. If we restrict to those maximum linearly independent split systems that are flat, however, we obtain the following characterization.

Theorem 3 *Let \mathcal{S} be a maximum linearly independent split system on a set X with $n \geq 5$ elements. Then the following properties are equivalent:*

- (i) \mathcal{S} is maximum circular.
- (ii) \mathcal{S} is maximum flat and orderly.
- (iii) \mathcal{S} is closed and satisfies the pairwise separation property.

Proof (i) \Rightarrow (ii): Assume that \mathcal{S} is a maximum circular split system. Then \mathcal{S} is maximum flat and, in view of Theorem 1, orderly.

(ii) \Rightarrow (iii): Assume that \mathcal{S} is a maximum flat split system that is orderly. Then, by Proposition 1, \mathcal{S} satisfies the pairwise separation property and, by Lemma 5, \mathcal{S} is closed.

(iii) \Rightarrow (i): Assume that \mathcal{S} is closed and satisfies the pairwise separation property. Then, by Proposition 1, \mathcal{S} is maximum flat. Consider any $x \in X$. We claim that the split $\{x\} \mid X - \{x\}$ is contained in \mathcal{S} .

Consider first the case that there exists some $y \in X - \{x\}$ such that one of the subsets A_y and B_y of X , that must exist for x and y according to the pairwise separation property, are empty. Assume without loss of generality that $A_y = \emptyset$. Then, in view of the pairwise separation property, the split $A_y \cup \{x\} \mid B_y \cup \{y\} = \{x\} \mid X - \{x\}$ is contained in \mathcal{S} , as claimed.

It remains to consider the case that for all $y \in X - \{x\}$ the subsets A_y and B_y of X are both non-empty. Then, in view of the pairwise separation property, \mathcal{S} contains the two incompatible splits $S_y = A_y \cup \{x\} \mid B_y \cup \{y\}$ and $S'_y = A_y \cup \{y\} \mid B_y \cup \{x\}$. Since $n \geq 5$, we have

$$\begin{aligned} & |(A_y \cup \{x\}) \cap (A_y \cup \{y\})| \cdot |(B_y \cup \{x\}) \cap (B_y \cup \{y\})| = |A_y| \cdot |B_y| \\ & > |\{x\}| \cdot |\{y\}| = |(A_y \cup \{x\}) \cap (B_y \cup \{x\})| \cdot |(B_y \cup \{y\}) \cap (A_y \cup \{y\})|. \end{aligned}$$

Thus, condition (a) or (c) from the definition of closedness must hold for the two splits S_y and S'_y , for all $y \in X - \{x\}$.

If there exists some $y \in X - \{x\}$ such that condition (a) holds for the splits S_y and S'_y then \mathcal{S} contains the split

$$(A_y \cup \{x\}) \cap (B_y \cup \{x\}) \mid X - ((A_y \cup \{x\}) \cap (B_y \cup \{x\})) = \{x\} \mid X - \{x\},$$

as claimed. So, assume that for all $y \in X - \{x\}$ condition (a) does not hold for the splits S_y and S'_y . Then there must exist two distinct $y_1, y_2 \in X - \{x\}$ such that condition (c) holds for the two splits S_{y_i} and $S'_{y_i}, i \in \{1, 2\}$. Thus, \mathcal{S} contains the split of X into the subsets

$$\begin{aligned} ((A_{y_i} \cup \{x\}) \cap (A_{y_i} \cup \{y_i\})) \cup ((B_{y_i} \cup \{x\}) \cap (B_{y_i} \cup \{y_i\})) &= A_{y_i} \cup B_{y_i} \text{ and} \\ ((A_{y_i} \cup \{x\}) \cap (B_{y_i} \cup \{x\})) \cup ((B_{y_i} \cup \{y_i\}) \cap (A_{y_i} \cup \{y_i\})) &= \{x, y_i\}. \end{aligned}$$

It follows that \mathcal{S} contains the two incompatible splits $\{x, y_1\} \mid X - \{x, y_1\}$ and $\{x, y_2\} \mid X - \{x, y_2\}$. As above, since $n \geq 5$, only condition (a) or (c) can hold for these two splits. In both cases \mathcal{S} contains the split

$$\{x, y_1\} \cap \{x, y_2\} \mid X - (\{x, y_1\} \cap \{x, y_2\}) = \{x\} \mid X - \{x\},$$

establishing the claim.

Since x was chosen arbitrarily, the maximum flat split system \mathcal{S} must contain the split $\{x\} \mid X - \{x\}$ for all $x \in X$. Therefore, in view of Lemma 3, \mathcal{S} is maximum circular. \square

5 Concluding Remarks

In this paper, we have started to explore which classes of split systems allow to gain information about them only knowing the order distances $\Delta(D)$ that indirectly arise through the distances D that they generate. We expect that further progress on this will require a deeper understanding of the precise relationship between the space of collections $\{\preceq_x : x \in X\}$ of total preorders that can arise from a particular class of split systems and their associated preorder split systems. A specific question that remains open is whether or not the generalization of Proposition 2 to $n \geq 6$ is true? Computational experiments that we have performed on a large number of randomly generated maximum linearly independent split systems seem to indicate that, at least for $n = 6$, if counterexamples exist they are very rare.

In another direction, note that an ℓ_1 -distance D on X , as defined in the introduction, can usually be generated by many different maximum linearly independent split systems on X . Therefore, for any ℓ_1 -distance D , one might hope to always find *some* split system that would generate both D and $\Delta(D)$. However, by an exhaustive search through the 34 isomorphism classes of maximum linearly independent split systems on a set X with $n = 5$ elements with a computer program we found the following: For every maximum linearly independent split system \mathcal{S} on X that is not maximum flat, there exists a non-negative weighting ω such that for the distance $D = D_{(\mathcal{S}, \omega)}$ and its associated order distance $\Delta(D)$ there is *no* maximum linearly independent split system that generates both D and $\Delta(D)$. It would be interesting to explore whether or not this is the case in general.

Finally, in future work it could also be worth studying variants of the order distance that are obtained by employing any of the other distance measures on total preorders considered in [3]. In particular, it would be interesting to know for which of them the resulting order distance is guaranteed to be an ℓ_1 -distance and, if such variants exist, which classes of split systems are orderly for them.

Appendix A

Proof of Lemma 1 It remains to show that the upper bound $n(n - 1)$ is tight for all sufficiently large n . To this end, consider a distance D on X such that, for all $\{u, v\} \in \binom{X}{2}$, the value $D(u, v) = D(v, u)$ is selected independently and uniformly at random from the set $\{1, 2\}$. We now argue that, for sufficiently large n , the probability that $|\mathcal{S}_D| = n(n - 1)$ is strictly greater than 0.

By construction, D satisfies the triangle inequality, and $D(u, v) > 0$ for all $\{u, v\} \in \binom{X}{2}$. This implies, in view of the definition of the sets $X_{u,v}$ and $X_{v,u}$, that $u \in X_{u,v}$, $v \in X_{v,u}$, $S_{u,v} \in \mathcal{S}_D$ and $S_{v,u} \in \mathcal{S}_D$. In order to have $|\mathcal{S}_D| = n(n - 1)$, for any two distinct $\{u, v\}, \{a, b\} \in \binom{X}{2}$, the splits $S_{u,v}, S_{v,u}, S_{a,b}$ and $S_{b,a}$ must be pairwise distinct.

Again in view of the definition of the sets $X_{u,v}$ and $X_{v,u}$, $S_{u,v} = S_{v,u}$ can only hold if $D(u, x) \neq D(v, x)$ for all $x \in X$. By the construction of D , the probability of this is at most $\left(\frac{1}{2}\right)^{n-2}$. Similarly, $S_{u,v} = S_{a,b}$ implies that $X_{u,v} = X_{a,b}$ (renaming the involved elements of X , if necessary). This can only hold if, for all $x \in X - \{a, b, u, v\}$, we have $D(u, x) < D(v, x)$ whenever $D(a, x) < D(b, x)$ and vice versa. By the construction of D , the probability of this is at most $\left(\frac{3}{4}\right)^{n-4}$. Applying an analogous analysis for any two of the four splits $S_{u,v}, S_{v,u}, S_{a,b}$ and $S_{b,a}$, it follows that the probability that at least two of these splits coincide is bounded by $d \cdot c^n$ for some constants $0 < d$ and $0 < c < 1$. This implies that the probability of $|\mathcal{S}_D| < n(n - 1)$ is at most

$$\binom{n}{2} \cdot \left(\binom{n}{2} - 1 \right) \cdot d \cdot c^n,$$

which is strictly less than 1 for sufficiently large n , as required. □

Proof of Lemma 2 First note that if two splits $S_1 = A_1 | B_1$ and $S_2 = A_2 | B_2$ of X are such that at least one of the intersections $A_1 \cap A_2, A_1 \cap B_2, B_1 \cap A_2$ and $B_1 \cap B_2$ is empty then also the restrictions $A_1 \cap M | B_1 \cap M$ and $A_2 \cap M | B_2 \cap M$ of these splits to a subset $M \subseteq X$ satisfy this property, assuming that these restrictions both form bipartitions of M into two non-empty subsets at all. Thus, if \mathcal{S}_D is compatible then $\mathcal{S}_{D|M}$ is also compatible for every 6-element subset $M \subseteq X$ in view of the fact that every split in $\mathcal{S}_{D|M}$ must be of the form $A \cap M | B \cap M$ for some split $A | B \in \mathcal{S}_D$.

It remains to establish that if \mathcal{S}_D is not compatible then there exists some 6-element subset $M \subseteq X$ such that $\mathcal{S}_{D|M}$ is not compatible. So assume that there exist $u, v, u', v' \in X$ with $u \neq v$ and $u' \neq v'$ such that the splits $S_{u,v} = X_{u,v} | X - X_{u,v}$ and $S_{u',v'} = X_{u',v'} | X - X_{u',v'}$ are incompatible. Then there must exist four distinct elements $a, b, c, d \in X$ with $a \in X_{u,v} \cap X_{u',v'}$, $b \in X_{u,v} \cap (X - X_{u',v'})$, $c \in (X - X_{u,v}) \cap X_{u',v'}$ and $d \in (X - X_{u,v}) \cap (X - X_{u',v'})$. Moreover, in view of the definition of the sets $X_{u,v}$ and $X_{u',v'}$, we have $u \in X_{u,v}, v \in X - X_{u,v}, u' \in X_{u',v'}$ and $v' \in X - X_{u',v'}$. Therefore, we can choose the elements a, b, c, d in such a way that $|\{a, b, c, d\} \cup \{u, u', v, v'\}| = 6$. Hence $M = \{a, b, c, d\} \cup \{u, u', v, v'\}$ is a 6-element subset such that $\mathcal{S}_{D|M}$ is not compatible. □

Proof of Lemma 3 As remarked in [21], the lemma can be proven using graph-theoretical concepts from [22]. In the following we provide a direct proof for completeness. Consider an allowable pair (π, κ) such that the maximum flat split system $\mathcal{S} = \mathcal{S}_{(\pi, \kappa)}$ contains all

splits of the form $\{x\} \mid X - \{x\}$. We represent the simple allowable sequence of permutations of X associated with (π, κ) by a wiring diagram \mathcal{W} . Since any two wires cross precisely once, there must exist in \mathcal{W} , for all $x \in X$, a unique face that is bounded to the right such that only the wire associated with x is either below or above that face. We form the sequence $F = (f_1, f_2, \dots, f_n)$ of these faces obtained by ordering them as they occur from left to right in \mathcal{W} , first those at the bottom of \mathcal{W} and then those at the top. The sequence F then yields a permutation $\pi^* = (x_1, x_2, \dots, x_n)$ of the elements in X in view of the fact that each f_i is associated with a unique element $x_i \in X$. This is illustrated in Fig. 4.

We claim that every split $S = A \mid B \in \mathcal{S}$ fits on π^* . To show this, consider the unique face f in \mathcal{W} that is bounded to the right and has all wires corresponding to elements in A below it and all wires corresponding to elements in B above it (renaming these sets, if necessary). Consider the rightmost point q on the boundary of the face f . In q the wires of two distinct elements $u, v \in X$ cross. Let $i, j \in \{1, 2, \dots, n\}$ be such that $u = x_i$ and $v = x_j$. Assume without loss of generality that $i < j$, that is, u comes before v in π^* . Then, in view of the fact that any two wires in \mathcal{W} cross precisely once, we must have $S = \{x_i, x_{i+1}, \dots, x_{j-1}\} \mid X - \{x_i, x_{i+1}, \dots, x_{j-1}\}$, as required. \square

To prove Proposition 1, we first show that for a maximum linearly independent split system \mathcal{S} on X that satisfies the pairwise separation property, linear independence is preserved in the restriction of \mathcal{S} to $X - \{y\}$, where the *restriction* of a split system \mathcal{S} on X to a subset $Y \subseteq X$ is the split system

$$\mathcal{S}|_Y = \{A \cap Y \mid B \cap Y : A \mid B \in \mathcal{S}, A \cap Y \neq \emptyset, B \cap Y \neq \emptyset\}.$$

Lemma 6 *Let \mathcal{S} be a maximum linearly independent split system on a set X with $n \geq 3$ elements that satisfies the pairwise separation property. Then, for any $y \in X$, the restriction of \mathcal{S} to $X - \{y\}$ is a maximum linearly independent split system that satisfies the pairwise separation property.*

Proof Fix any $y \in X$. For every $c \in X - \{y\}$ let A_c and B_c denote the two subsets of $X - \{y, c\}$ that must exist for the pair $\{y, c\}$ according to the pairwise separation property. Consider the set

$$\mathcal{S}_c = \{A_c \mid X - (A_c \cup \{y\}), B_c \mid X - (B_c \cup \{y\})\}$$

for each $c \in X - \{y\}$ and put

$$\mathcal{S}_{\leftrightarrow} = \bigcup_{c \in X - \{y\}} \mathcal{S}_c.$$

The set $\mathcal{S}_{\leftrightarrow}$ contains splits of $X - \{y\}$ and, possibly, also the unordered pair $\emptyset \mid X - \{y\}$.

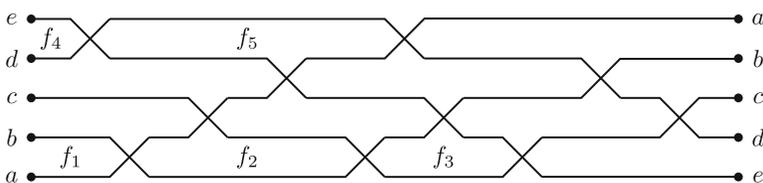


Fig. 4 A wiring diagram displaying the maximum flat split system $\mathcal{S} = \mathcal{S}_{(\pi, \kappa)}$ on $X = \{a, b, c, d, e\}$ with $\pi = (a, b, c, d, e)$ and $\kappa = (4, 1, 2, 3, 1, 4, 2, 1, 3, 2)$. \mathcal{S} contains the split $\{x\} \mid X - \{x\}$ for all $x \in X$, each corresponding to a unique face that is bounded to the right and has precisely one wire either above or below it. The sequence $(f_1, f_2, f_3, f_4, f_5)$ of these faces yields the permutation $\pi^* = (a, b, c, e, d)$ of the elements in X and all splits in \mathcal{S} fit on π^*

We claim that $|\mathcal{S}_{\leftrightarrow}| \geq n - 1$. To see this, consider the graph G_{\leftrightarrow} with vertex set $\mathcal{S}_{\leftrightarrow}$ in which there is an edge between two distinct $S, S' \in \mathcal{S}_{\leftrightarrow}$ if there exists $c \in X - \{y\}$ such that $\{S, S'\} = \mathcal{S}_c$. Thus, by construction, G_{\leftrightarrow} has precisely $n - 1$ edges, each uniquely associated with an element $c \in X - \{y\}$. Moreover, the two splits $A | B$ and $A' | B'$ that are connected by the edge associated with element $c \in X - \{y\}$ are such that $A' = A - \{c\}$ and $B' = B \cup \{c\}$ (after naming the subsets involved in a suitable way). In this sense, every edge of G_{\leftrightarrow} corresponds to the move of a single element from $X - \{y\}$ from one subset in an unordered pair to the other subset in that pair. But this implies that, if G_{\leftrightarrow} contains a cycle this cycle must contain all edges of G_{\leftrightarrow} , since otherwise it is impossible to return to any particular split along that cycle. Hence, G_{\leftrightarrow} must have at least $n - 1$ vertices, implying that $|\mathcal{S}_{\leftrightarrow}| \geq n - 1$, as claimed.

Next note that $|\mathcal{S}_{\leftrightarrow}| \geq n - 1$ implies that, when restricting \mathcal{S} to $X - \{y\}$, we obtain at most $\binom{n}{2} - (n - 1)$ splits of $X - \{y\}$ in view of the fact that, for every $S' \in \mathcal{S}_{\leftrightarrow}$, there are, according to the pairwise separation property, at least two distinct splits in \mathcal{S} that restrict to S' . On the other hand, in view of the fact that the square matrix consisting of the $\binom{n}{2}$ column vectors formed by the distances D_S for $S \in \mathcal{S}$ has full rank, the restriction of this matrix to the rows associated to the pairs of distinct elements $x, x' \in X - \{y\}$ must have rank $\binom{n}{2} - (n - 1)$. But this implies that $\mathcal{S}|_{X-\{y\}}$ must contain at least $\binom{n}{2} - (n - 1)$ splits. Thus $\mathcal{S}|_{X-\{y\}}$ contains precisely $\binom{n}{2} - (n - 1) = \binom{n - 1}{2}$ splits, implying that $\mathcal{S}|_{X-\{y\}}$ is a maximum linearly independent split system on $X - \{y\}$. That $\mathcal{S}|_{X-\{y\}}$ also satisfies the pairwise separation property follows immediately from the definition of this property. \square

To prove Proposition 1 we shall also use [21, Theorem 15] which states that a maximum linearly independent split system \mathcal{S} on a set X with $n \geq 2$ elements is a maximum flat split system if and only if, for every 4-element subset $Y \subseteq X$, the restriction $\mathcal{S}|_Y$ contains precisely 6 splits.

Proof of Proposition 1 In [11] it was remarked that every maximum flat split system satisfies the pairwise separation property. In the following we briefly explain why this is the case. Let \mathcal{S} be a maximum flat split system. Then, by definition, there exists an allowable pair (π, κ) with $\mathcal{S} = \mathcal{S}_{(\pi, \kappa)}$. We represent the simple allowable sequence of permutations of X associated with (π, κ) by a wiring diagram \mathcal{W} . To show that \mathcal{S} satisfies the pairwise separation property, consider two distinct elements $x, y \in X$. By the definition of a simple allowable sequence, there exists precisely one point q where the wires associated to x and y cross in \mathcal{W} . Putting A to be the set of those elements in X whose wires are below q and, similarly, B to be the set of those elements in X whose wires are above q , we obtain the two subsets of $X - \{x, y\}$ required by the pairwise separation property. This is illustrated in Fig. 5.

Next assume that \mathcal{S} is a maximum linearly independent split system that satisfies the pairwise separation property. Then, for $n \in \{2, 3, 4\}$, the fact that \mathcal{S} consists of precisely $\binom{n}{2}$ splits immediately implies that \mathcal{S} is a maximum flat split system by [21, Theorem 15]. So, assume that $n \geq 5$. Then, every 4-element subset $Y \subseteq X$ can be obtained by removing from X the elements in $X - Y$ one by one. Hence, the restriction $\mathcal{S}|_Y$ is the last element of a sequence of restrictions, each to a subset with one element less, and to each such

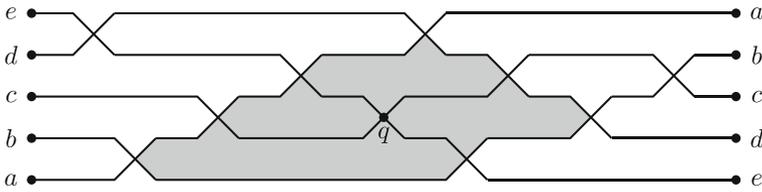


Fig. 5 A wiring diagram displaying the maximum flat split system $\mathcal{S}_{(\pi, \kappa)}$ on $X = \{a, b, c, d, e\}$ with $\pi = (a, b, c, d, e)$ and $\kappa = (4, 1, 2, 3, 2, 4, 1, 3, 2, 3)$. For the elements c and e we obtain the subsets $A = \{b\}$ and $B = \{a, d\}$ required by the pairwise separation property. The elements in A correspond to those wires that are below the point q where the wires of c and e cross, and the elements in B correspond to those wires that are above q . The faces corresponding to the splits $A \cup \{c, e\} \mid B, A \cup \{c\} \mid B \cup \{e\}, A \cup \{e\} \mid B \cup \{c\}$ and $A \mid B \cup \{c, e\}$ are shaded gray

restriction Lemma 6 applies. Therefore, $\mathcal{S}|_Y$ must be a linearly independent split system on Y containing $\binom{4}{2} = 6$ splits. But this implies, again by [21, Theorem 15], that \mathcal{S} is maximum flat. □

Proof of Lemma 5 If \mathcal{S} is compatible then it is trivially closed. So assume that \mathcal{S} is orderly and contains two incompatible splits $S_1 = A_1 \mid B_1$ and $S_2 = A_2 \mid B_2$. Let ω be the weighting of \mathcal{S} with $\omega(S_1) = \omega(S_2) = 1$ and $\omega(S) = 0$ for all other $S \in \mathcal{S}$. We consider the order distance $\Delta = \Delta(D)$ with $D = D_{(\mathcal{S}, \omega)}$. Then, putting $n_1 = |A_1 \cap A_2|, n_2 = |B_1 \cap A_2|, n_3 = |A_1 \cap B_2|$ and $n_4 = |B_1 \cap B_2|$, we have, for all $x_1 \in A_1 \cap A_2, x_2 \in B_1 \cap A_2, x_3 \in A_1 \cap B_2$ and $x_4 \in B_1 \cap B_2$,

$$\begin{aligned} \Delta(x_1, x_2) &= n_1n_4 + 2(n_1n_2 + n_3n_4) + n_2n_3, \\ \Delta(x_1, x_3) &= n_1n_4 + 2(n_1n_3 + n_2n_4) + n_2n_3, \\ \Delta(x_1, x_4) &= 2(n_1n_4 + n_1n_2 + n_3n_4 + n_1n_3 + n_2n_4), \\ \Delta(x_2, x_3) &= 2(n_2n_3 + n_1n_2 + n_3n_4 + n_1n_3 + n_2n_4), \\ \Delta(x_2, x_4) &= n_1n_4 + 2(n_1n_3 + n_2n_4) + n_2n_3, \\ \Delta(x_3, x_4) &= n_1n_4 + 2(n_1n_2 + n_3n_4) + n_2n_3 \end{aligned}$$

and, for all other $u, v \in X$, we have $\Delta(u, v) = 0$. Since \mathcal{S} is orderly, there must exist some non-negative weighting ω' of \mathcal{S} with

$$\Delta = \sum_{S \in \mathcal{S}} \omega'(S) \cdot D_S.$$

Put $\mathcal{S}' = \{S \in \mathcal{S} : \omega'(S) > 0\}$ and $\mathcal{C} = \{A_1 \cap A_2, B_1 \cap A_2, A_1 \cap B_2, B_1 \cap B_2\}$. In view of $\Delta(u, v) = 0$ for all $u, v \in X$ with $\{u, v\}$ a subset of one of the sets in \mathcal{C} , we must have

$$\begin{aligned} \mathcal{S}' &\subseteq \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\} = \mathcal{S}^* \text{ with} \\ S_1 &= A_1 \cap A_2 \mid X - (A_1 \cap A_2), \quad S_2 = B_1 \cap A_2 \mid X - (B_1 \cap A_2), \\ S_3 &= A_1 \cap B_2 \mid X - (A_1 \cap B_2), \quad S_4 = B_1 \cap B_2 \mid X - (B_1 \cap B_2), \\ S_5 &= A_1 \mid B_1, \quad S_6 = A_2 \mid B_2, \quad S_7 = (A_1 \cap A_2) \cup (B_1 \cap B_2) \mid (A_1 \cap B_2) \cup (B_1 \cap A_2) \end{aligned}$$

The columns in the following matrix represent the split distances associated with the splits in \mathcal{S}^* , with each row corresponding to an unordered pair of distinct sets in \mathcal{C} .

	S_1	S_2	S_3	S_4	S_5	S_6	S_7
$\{A_1 \cap A_2, A_1 \cap B_2\}$	1	0	1	0	0	1	1
$\{A_1 \cap A_2, B_1 \cap A_2\}$	1	1	0	0	1	0	1
$\{A_1 \cap A_2, B_1 \cap B_2\}$	1	0	0	1	1	1	0
$\{A_1 \cap B_2, B_1 \cap A_2\}$	0	1	1	0	1	1	0
$\{A_1 \cap B_2, B_1 \cap B_2\}$	0	0	1	1	1	0	1
$\{B_1 \cap A_2, B_1 \cap B_2\}$	0	1	0	1	0	1	1

Removing any column from this matrix yields a matrix of rank 6. This implies that the split system \mathcal{S}^* is not linearly independent but every 6-element subset of \mathcal{S}^* is. Thus, the space of solutions $(\omega_1, \omega_2, \dots, \omega_7) \in \mathbb{R}^7$ of the equation

$$\Delta = \sum_{i=1}^7 \omega_i \cdot D_{S_i}$$

is 1-dimensional. More specifically, these solutions have the form

$$(n_1n_4 - \alpha, n_2n_3 - \alpha, n_2n_3 - \alpha, n_1n_4 - \alpha, 2(n_1n_2 + n_3n_4) + \alpha, 2(n_2n_4 + n_1n_3) + \alpha, \alpha)$$

with $\alpha \in \mathbb{R}$. Only the solutions for $\alpha = 0$ and $\alpha = \min(n_1n_4, n_2n_3)$, however, yield Δ as a sum of linearly independent split distances. This implies that \mathcal{S}' must correspond to one of these solutions and, thus, consist of those 6 splits that receive a positive weight in that solution. But this implies that \mathcal{S} satisfies at least one of the conditions (a)-(d) in the definition of a closed split system, as required. □

Proof of Proposition 2 Let \mathcal{S} be a maximum linearly independent split system on a set X with 5 elements. Then, in view of Theorem 1, if \mathcal{S} is circular it is orderly.

It remains to show that if \mathcal{S} is orderly then it is circular. So, assume that \mathcal{S} is orderly. Then, in view of Lemma 5, \mathcal{S} is closed. Moreover, since \mathcal{S} is a maximum linearly independent split system, we have $|\mathcal{S}| = 10$. Thus, as any compatible split system on X contains at most 7 splits, \mathcal{S} must contain two incompatible splits S_1 and S_2 . Relabeling the elements in X , if necessary, we assume without loss of generality that $S_1 = \{a, b\} \mid \{c, d, e\}$ and $S_2 = \{b, c\} \mid \{a, d, e\}$. Then, since \mathcal{S} is closed, it must also contain the splits $\{b\} \mid \{a, c, d, e\}$ and $S_3 = \{d, e\} \mid \{a, b, c\}$. Moreover, in view of $|\mathcal{S}| = 10$, \mathcal{S} must contain an additional split $S_4 = \{x, y\} \mid X - \{x, y\}$ for a 2-element subset $\{x, y\} \subseteq X$ with $\{x, y\} \not\subseteq \{\{a, b\}, \{b, c\}, \{d, e\}\}$. We consider three cases.

Case 1: $S_4 = \{a, c\} \mid \{b, d, e\}$. Then S_1 and S_4 are incompatible. Thus, since \mathcal{S} is closed, we have $\{a\} \mid \{b, c, d, e\} \in \mathcal{S}$. Similarly, S_2 and S_4 are incompatible and, therefore, we have $\{c\} \mid \{a, b, d, e\} \in \mathcal{S}$. Put

$$\mathcal{S}^* = \{S_1, S_2, S_3, S_4\} \cup \{\{x\} \mid X - \{x\} : x \in \{a, b, c\}\}.$$

For all weightings ω of \mathcal{S}^* we have $D_{(\mathcal{S}^*, \omega)}(d, e) = 0$ and $D_{(\mathcal{S}^*, \omega)}(d, x) = D_{(\mathcal{S}^*, \omega)}(e, x)$ for all $x \in \{a, b, c\}$. Thus, the matrix whose columns are the split distances for the 7 splits in \mathcal{S}^* has rank at most 6, implying that $\mathcal{S}^* \subseteq \mathcal{S}$ is not linearly independent, a contradiction.

Case 2: $S_4 = \{b, e\} \mid \{a, c, d\}$. (Note that the case $S_4 = \{b, d\} \mid \{a, c, e\}$ is symmetric.) Then S_3 and S_4 are incompatible. Therefore, since \mathcal{S} is closed, $\{a, c\} \mid \{b, d, e\} \in \mathcal{S}$. From this we obtain a contradiction as in Case 1.

Case 3: $S_4 = \{c, d\} \mid \{a, b, e\}$. (Note that the cases $S_4 = \{c, e\} \mid \{a, b, d\}$, $S_4 = \{a, d\} \mid \{b, c, e\}$ and $S_4 = \{a, e\} \mid \{b, c, d\}$ are symmetric.) Then, using again that \mathcal{S} is closed, the fact that S_2 and S_4 are incompatible implies that the splits $\{c\} \mid \{a, b, d, e\}$ and

$S_5 = \{a, e\} \mid \{b, c, d\}$ are contained in \mathcal{S} . Similarly, since S_3 and S_4 are incompatible, we have $\{d\} \mid \{a, b, c, e\} \in \mathcal{S}$. Since S_1 and S_5 are incompatible, we have $\{a\} \mid \{b, c, d, e\} \in \mathcal{S}$. And since S_3 and S_5 are incompatible, we have $\{e\} \mid \{a, b, c, d\} \in \mathcal{S}$. Thus, $\mathcal{S} = \mathcal{S}_\pi$ for the permutation $\pi = (a, b, c, d, e)$ of the elements in X , implying that \mathcal{S} is circular, as required. \square

Acknowledgements The authors would like to thank Barbara Holland and Katharina Huber for interesting discussions on the ordering of treelike distances which led them to consider the connection between order distances and split systems. The authors would also like to thank two anonymous reviewers for their very helpful comments on earlier versions of this manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability statement Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Semple, C., Steel, M.: *Phylogenetics*. Oxford University Press, Oxford (2003)
2. Bonnot, F., Guénoche, A., Perrier, X.: Properties of an order distance associated to a tree distance. In: Diday, E. (ed.) *Ordinal and Symbolic Data Analysis*, pp. 252–261. Springer, Berlin (1996)
3. Fagin, R., Kumar, R., Mahdian, M., Sivakumar, D., Vee, E.: Comparing partial rankings. *SIAM J. Discret. Math.* **20**, 628–648 (2006). <https://doi.org/10.1137/05063088X>
4. Giakoumakis, V., Monjardet, B.: Coefficients d'accord entre deux préordres totaux. *Stat. Anal. Données* **12**, 46–99 (1987)
5. Guénoche, A.: Order distance associated with a hierarchy. *J. Classif.* **14**, 101–115 (1997). <https://doi.org/10.1007/s003579900005>
6. Guénoche, A.: Order distances in tree reconstruction. In: Mirkin, B. (ed.) *Mathematical Hierarchies and Biology*, pp. 171–182. American Mathematical Society, Providence (1997)
7. Guénoche, A.: Ordinal properties of tree distances. *Discret. Math.* **192**, 103–117 (1998). [https://doi.org/10.1016/S0012-365X\(98\)00068-5](https://doi.org/10.1016/S0012-365X(98)00068-5)
8. Deza, M., Laurent, M.: *Geometry of Cuts and Metrics*. Springer, Berlin (1997)
9. Kearney, P.: A six-point condition for ordinal matrices. *J. Comput. Biol.* **4**, 143–156 (1997). <https://doi.org/10.1089/cmb.1997.4.143>
10. Bandelt, H.-J., Dress, A.: A canonical decomposition theory for metrics on a finite set. *Adv. Math.* **92**, 47–105 (1992). [https://doi.org/10.1016/0001-8708\(92\)90061-O](https://doi.org/10.1016/0001-8708(92)90061-O)
11. Bryant, D., Dress, A.: Linearly independent split systems. *Eur. J. Combi.* **28**, 1814–1831 (2007). <https://doi.org/10.1016/j.ejc.2006.04.007>
12. Spillner, A., Nguyen, B., Moulton, V.: Constructing and drawing regular planar split networks. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **9**, 395–407 (2012). <https://doi.org/10.1109/TCBB.2011.115>
13. Bryant, D., Moulton, V.: NeighborNet: An agglomerative method for the construction of phylogenetic networks. *Mol. Biol. Evol.* **21**, 255–265 (2004). <https://doi.org/10.1093/molbev/msh018>
14. Balvočiūtė, M., Spillner, A., Moulton, V.: FlatNJ: A novel network-based approach to visualize evolutionary and biogeographical relationships. *Syst. Biol.* **63**, 383–396 (2014). <https://doi.org/10.1093/sysbio/syu001>

15. Felsner, S.: Geometric Graphs and Arrangements. Vieweg, Wiesbaden (2004)
16. Kalmanson, K.: Edgeconvex circuits and the travelling salesman problem. *Can. J. Math.* **27**, 1000–1010 (1975). <https://doi.org/10.4153/CJM-1975-104-6>
17. Chepoi, V., Fichet, B.: A note on circular decomposable metrics. *Geom. Dedicata.* **69**, 237–240 (1998). <https://doi.org/10.1023/A:1004907919611>
18. Christopher, G., Farach, M., Trick, M.: The structure of circular decomposable metrics. In: Proc. 4th Annual European Symposium on Algorithms. LNCS, pp. 486–500. Springer, Berlin (1996). https://doi.org/10.1007/3-540-61680-2_77
19. Bansal, M., Fernández-Baca, D.: Computing distances between partial rankings. *Inf. Process. Lett.* **109**, 238–241 (2009). <https://doi.org/10.1016/j.ipl.2008.10.010>
20. Moulton, V., Spillner, A.: Optimal algorithms for computing edge weights in planar split networks. *J. Appl. Math. Comput.* **39**, 1–13 (2012). <https://doi.org/10.1007/s12190-011-0506-z>
21. Balvočiūtė, M., Bryant, D., Spillner, A.: When can splits be drawn in the plane? *SIAM J. Discret. Math.* **31**, 839–856 (2017). <https://doi.org/10.1137/15M1040852>
22. Dress, A., Huson, D.: Constructing splits graphs. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **1**, 109–115 (2004). <https://doi.org/10.1109/TCBB.2004.27>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.