PARITY SPLITS BY TRIPLE POINT DISTANCES IN X-TREES

JÖRGEN BACKELIN AND SVANTE LINUSSON

ABSTRACT. At the conference [3] Andreas Dress defined parity split maps defined by triple point distance and asked for a characterisation of such maps coming from binary phylogenetic X-trees. This article gives an answer to that question. The characterisation for X-trees can be easily described as follows: If all restrictions of a split map to sets of five or fewer elements is a parity split map for an X-tree, then so is the entire map.

To ensure that the parity split map comes from an X-tree which is binary and phylogenetic, we add two more technical conditions also based on studying at most five points at a time.

1. Introduction

Given a finite non-zero set X, recall that an X-tree $\mathcal{T}$ is a pair (T, ϕ), where $T = (V, E)$ is a tree, and $\phi : X \longrightarrow V$ a map such that every vertex in V of degree ≤ 2 is an element in $\phi(X)$. An X-tree $\mathcal{T} = (T, \phi)$ is a *binary phylogenetic X-tree*, if in addition ϕ is injective, $\phi(X)$ consists of the leaves in T, and each vertex in $V \setminus \phi(X)$ has degree 3.

Note, that for any $S = \{x, y\} \in \binom{X}{2}$ (the two-subsets of X) there is a unique path $[x, y]$ with end points x and y, and that for any three-set $S = \{x, y, z\} \in \binom{V}{3}$ there is a unique 'triple point' or 'median' 3_{xyz}, such that $[x, y] \cap [x, z] \cap [y, z] = \{3_{xyz}\}$[1]. Putting e.g. $[x, x] = \{x\}$, the definition works also for some or all of x, y, and z equal. (E.g., $3_{xxz} = x$, as is easily seen.) Finally, let $\mathcal{S}(X) = \big\{\{A, B\} : A \cup B = X \wedge A \cap B = \emptyset\big\}$ (the partitions of X into exactly two parts).

As usual, we get a metric, dist, on T (i.e., formally, on V), and an induced metric on X, by putting $\mathrm{dist}(x, y) = \mathrm{dist}_T(x, y)$ equal to the length of $[x, y]$ (i.e., the numbers of edges therein).

Andreas Dress [3] made the following observation and put the following question for binary phylogenetic X-trees (based on his joint work with Mike Steel, to appear as [2, Remark 3]):

If we define $\Pi_T : \binom{X}{2} \longrightarrow \mathcal{S}(X)$ by demanding that the X elements z and w are in the same part of $\Pi_T(x, y)$ iff $\mathrm{dist}_T(x, 3_{xyz}) \equiv \mathrm{dist}_T(x, 3_{xyw})$

Date: 24th May 2005.

[1] We cannot denote the median simply xyz, as some readers may be used to, since then we would get ambiguities for edges like $a3_{abc}$

(mod 2), *then* Π_T *determines* T *(up to isomorphism). Characterise the resulting split maps* Π_* *in intrinsic* $\mathcal{S}(X)$ *terms!*

1.1. **Results.** It turns out that the parity splits are not in natural manners inherited to binary phylogenetic X'-trees for $X' \subset X$. The reason is, that in forming the binary phylogenetic tree corresponding to X', some edges in the original tree may be merged, changing path length parities. It is however possible to give an easy to state characterisation for possible parity split maps for X-trees, which we do in Theorem 4.1. It is formulated as a four point condition and a five point condition, which the restrictions of the split map to any set of four and five points from X have to satisfy. To answer the original question we add two extra conditions in Theorem 5.2 and Theorem 5.1, which ensures that an X-tree is binary and phylogenetic, respectively. It is noteworthy, that also for the classical problem of splits a more natural condition is given for X-trees, and then special conditions are added for binary and phylogenetic trees; cf. e.g. [4].

1.2. **Extension to parity trees.** It helps our reasoning and makes some proofs easier to formulate if we enlarge the class of trees studied to *parity weighted* binary phylogenetic X-trees. Loosely spoken, we add parities, i.e., elements of $\mathbb{Z}_2$ (thought of as "even" and "odd"), to the edges, and assign edge parity sums as path parities or 'parity distances'; see Section 2.2. (50% of the authors also consider this the most natural setting of the problem.) Two such trees are considered equivalent, if they form the same X-tree when all even edges are contracted. In this setting we thus give a characterisation of equivalence classes.

As Andreas Dress has kindly pointed out to us, there seems also to be a close relationship between our 'parity metrics' (or '$\mathbb{Z}_2$ metrics') on one hand, and the 'symbolic datings' as defined in [1] on the other. It would be very interesting to see this explored in detail.

1.3. **Outline of paper.** We start by making the definitions rigorous in Section 2, in particular formulating the Boolean algebra that is essential for our proof. In Section 3, we analyse the situations for $|X| \leq 5$ fully, including proving the necessity of our conditions. In Section 4, we state and prove the characterisation of parity split maps for X-trees, and finally in Section 5 we give the extra conditions to give a tentative solution of Dress's original problem.

Remark 1.1. Andreas Dress's original question – and our investigation – was not directly motivated by phylogenetic applications, but by the intrinsic value in studying reconstructions of trees in more general settings.

Notwithstanding, it is worth to note that parity distances indeed are of notable interest in one instance of bioinformatics application, namely, in linkage analysis. When analysing whether or not specific alleles at different loci in the same chromosome are more or less likely to be joined or separated by meiotic recombination, what matters is the parity of the number of crossovers occurring between the two loci. In actual analysis, multiple loci situations are often considered, and the situation is said to be fairly complex. The crossover probabilities need neither be independent, nor equally distributed. (If you want to find out more about this, you may read e.g. [5, Chapters 11 and 17]; and especially Sections 11.3 and 11.4. If not, you may feel assured by the fact that no further reference to this will be made in this article; so you may just forget it.)

2. Preliminaries

Let $\mathbb{Z}$ be the set of integers, and $\mathbb{Z}_2 = \mathrm{GF}(2)$ be the field with two elements. (As you recall, thus formally $\mathbb{Z}_2 = \{\{[0],[1]\}$, where $[n]$ denotes the residue class of n modulo 2; but as usual, we let 0 and 1 stand for the residue classes of even and odd numbers, respectively.) We may also identify the elements in $\mathbb{Z}_2$ as *parities*: $0 = [0] =$ 'even', $1 = [1] =$ 'odd'. Finally, we may think of them as *truth values*: $0 =$ 'false', $1 =$ 'true'. In this case, as usual the algebraic operations $+$ and $\cdot$ are interpreted as the logical connectives 'exclusive or' and 'and', respectively.

For any set A, a function $\lambda : A \times A \longrightarrow \mathbb{Z}_2$ is a $\mathbb{Z}_2$ *metric* (on A), if $\lambda(a,b) + \lambda(b,c) = \lambda(a,c)$ for all $a,b,c \in A$. Recalling that 'plus' and 'minus' coincide on $\mathbb{Z}_2$, we find that then $\lambda(a,b) + \lambda(b,c) + \lambda(a,c) = 0$; whence in particular $\lambda(a,a) = \lambda(a,a) + \lambda(a,a) + \lambda(a,a) = 0$, and similarly $\lambda(a,b) = \lambda(b,a)$.

$|M|$ denotes the number of elements in a set M.

Throughout this article, X is always a non-empty, finite set, and all trees are unrooted, undirected, and finite. We use $a,b,c,\dots$ et cetera for elements of X (or $\phi(X)$), and $\alpha,\beta,\dots$ for vertices of the tree.

Mostly, we try to follow the terminology in [4], as regards X-trees. However, informally, we abuse notation freely; speaking of T instead of (T,ϕ) or instead of the vertex set of T, and speaking of a instead of $\phi(a)$. (We may think of X as a set of labels for some of the vertices in T.)

2.1. **Split maps.** Let $\mathcal{S}(X) = \big\{\{A,B\} : A \cup B = X \wedge A \cap B = \emptyset\big\}$ be the set of *splits* of X. A split $\big\{\{a_1,\dots,a_r\},\{b_1,\dots,b_s\}\big\}$ is often denoted $\{a_1,\dots,a_r\}|\{b_1,\dots,b_s\}$ or $a_1\cdots a_r|b_1\cdots b_s$. A *split map* (for X) is a map π from $\binom{X}{2}$ (the set of 2-subsets of X) to $\mathcal{S}(X)$. Given a split map π, $[ab : c_1\cdots c_r|d_1\cdots d_s]$ is the statement that $c_1, \dots, c_r$ and $d_1, \dots, d_s$ belong to different parts of the split $\pi(\{a,b\})$; and we let $(ab : cd)$ be the $\mathbb{Z}_2$ representation of the truth value of the statement $[ab : c|d]$; i.e., $(ab : cd) = 1$ if $[ab : c|d]$ is true, but $(ab : cd) = 0$

else. By convention, we put $(aa : cd) = 0$; this corresponds to defining $\pi(\{a\}) = \{X, \emptyset\}$ for all $a \in X$.

A split map π is *realisable* if $\pi = \Pi_T$ for some X-tree T, in which case we call it a *parity split map*.

For any split map we obviously have some symmetry conditions, and a 'triplet condition' depending on the fact that there are but two parts in each split:

$$(ab : cd) = (ab : dc) = (ba : cd) \tag{1}$$

$$(ab : cd) + (ab : ce) + (ab : de) = 0 \tag{2}$$

(The reader easily may check that these relations hold whether or not some of $a, \ldots, e$ coincide.)

If π is a split map for X, and $X' \subset X$, then the *restriction* $\pi|_{X'}$ of π to X' is defined in the natural manner. I.e., if $a, b \in X'$ and $\pi(a, b) = \{A, B\}$, then $\pi|_{X'}(a, b) = \{A \cap X', B \cap X'\}$.

2.2. **Generalised X-trees with parity weights.** Formally, let a *generalised* X-tree $\mathcal{T}$ be a pair (T, ϕ), where $T = (V, E)$ is a tree, and $\phi : X \longrightarrow V$ is any map. $\mathcal{T}$ is called an *X-tree*, if moreover each element in V is the triple point of three (different or not) elements in $\phi(X)$; or equivalently if every vertex in V of degree ≤ 2 is an element in $\phi(X)$. An X-tree $\mathcal{T} = (T, \phi)$ is a *binary phylogenetic X-tree*, if in addition ϕ is injective, $\phi(X)$ consists of the leaves in T, and each vertex in $V \setminus \phi(X)$ has degree 3. A *parity weight function* ρ on the (generalised) X-tree $T = (V, E)$ is a map $\rho : E \longrightarrow \mathbb{Z}_2$. An X-tree together with a parity weight function is called a *parity weighted* X-tree. An edge with weight 0 (1) is called *even* (*odd*, respectively). For any $\alpha, \beta \in V$, we define the *parity distance* between α and β as the sums of the parities of the edges on the path $[\alpha, \beta]$; i.e., formally,

$$\mathrm{pdist}_T(\alpha, \beta) = \sum_{e \in E : e \subseteq [\alpha, \beta]} \rho(e)\,.$$

Clearly, $\mathrm{pdist} = \mathrm{pdist}_T$ is a $\mathbb{Z}_2$ metric. Any X-tree may be considered as a parity weighted X-tree with *the trivial weight function*, which is constantly 1 (i.e., such that all edges are odd). In this case, 'parity distance' indeed is the same as 'the parity of the (ordinary) distance': $\mathrm{pdist}_T(\alpha, \beta) = [\mathrm{dist}_T(\alpha, \beta)]$. If we do not explicitly denote an X-tree as a parity weighted X-tree, we always assume it to be trivially parity weighted.

We may generalise the definition of binary phylogenetic tree split maps as follows. Given any parity weighted generalised X-tree $T = (V, E, \rho)$, define a split map $\Pi_T : \binom{X}{2} \longrightarrow \mathcal{S}(X)$ in the following manner. For any $\{a, b\} \in \binom{X}{2}$, pick one vertex, say a. Consider the split $\big\{\{x \in X : \mathrm{pdist}_T(a, 3_{abx}) = 0\},\ \{x \in X : \mathrm{pdist}_T(a, 3_{abx})\} = 1.$ Since

pdist_T is a $\mathbb{Z}_2$ metric, picking b instead of a yields the same split; and we define $\Pi_T(a,b)$ as this split. We call Π_T a (general) *parity split map*.

It is easy to see that the restriction of the metric to X may be recovered from the parity split map by

$$\mathrm{pdist}_T(a,b) = (ab : ab)_T\,.$$

A similar construction may be made for every split map. However, the parity split maps for parity weighted X-trees thus yield $\mathbb{Z}_2$ metrics; but not conversely. The precise connections are discussed in remarks 2.4 and 3.2.

Two generalised parity weighted X-trees T_1 and T_2 are *equvivalent* if $\Pi_{T_1} = \Pi_{T_2}$, in which case we write $T_1 \sim T_2$, or $T_1 \sim_X T_2$ in case of ambiguity. There are infinitely many elements in each equivalence class; but (as Theorem 4.1 will show) there is exactly one of minimal size, and that one is the unique element with trivial weight function. Contracting all even edges of any parity weighted X-tree in this class will yield this minimal element. The set of possible parity split maps is thus the same for X-trees (i.e., X-trees with trivial parity weight functions) as for generalised parity weighted X-trees.

For each generalised parity weighted X-tree T, there is at least one parity weighted binary phylogenetic X-tree T' with $T \sim T'$. We may construct T' from the minimal parity weighted X-tree in the equivalence class of T, in the following manner. First, we add an even edge between $\phi(x)$ and the rest of the tree, as long as there is some $x \in X$ with either $|\phi^{-1}(\phi(x))| \geq 2$ or $\deg \phi(x) \geq 2$; which results in an X-tree with X identifiable as its sets of leaves. Then, we split each vertex of degree ≥ 4 into two vertices, with an even edge between them, until all internal vertices have degree 3. In general, T' is not unique, however.

Note, that if $\mathcal{T} = (T, \phi)$ is a generalised parity weighted X-tree and $X' \subset X$, then $\mathcal{T}' = (T, \phi|_{X'})$ is a generalised X'-tree, and

$$\Pi_{\mathcal{T}'} = \Pi_{\mathcal{T}}|_{X'}\,. \tag{3}$$

However, $\mathcal{T}'$ need not be binary, phylogenetic, minimal, or even non-generalised, just because $\mathcal{T}$ is.

2.3. **The four point condition.** If T is a parity weighted binary phylogenetic X-tree with parity distance pdist, and $a, b, c, d \in X$, then c and d are in the same part of the split $\Pi_T(a,b)$ if and only if $\mathrm{pdist}(3_{abc}, 3_{abd}) = 0$. In other words, and using our identification of both parities and truth values with $\mathbb{Z}_2$ elements,

$$\mathrm{pdist}(3_{abc}, 3_{abd}) = (ab : cd)\,.$$

Moreover, if 3_{abc} and 3_{abd} are different, then so are the remaining two triple points formed from this quartet, and in fact then $\{3_{cda}, 3_{cdb}\} = \{3_{abc}, 3_{abd}\}$. Hence, it quickly follows that

$$(ab : cd) = (cd : ab), \quad \text{for all } a, b, c, d \in X. \tag{4}$$

Note that we also include the possibility of some of a, b, c, d being equal. If e.g. $a = b$, then $(ab : cd) = (cd : ab) = 0$; and if $a = c$, then $(ab : cd) = (cd : ab) = \mathrm{pdist}(a, 3_{abc})$.

This motivates the following definition:

Definition 2.1. A split map π over X is said to satisfy the **four point condition** if (4) holds.

We have some immediate 'algebraic' consequences of (4). To begin with, for any $a, b, c, d \in X$ (different or not) and any X split map π, by (1) and (2) we have $(ab : cd) + (ac : bd) + (ad : bc) = (ab : cb) + (ab : db) + (ac : bd) + (ad : bc) = \big((ab : bc) + (ad : bc)\big) + \big((ab : bd) + (ac : bd)\big) = (bd : bc) + (bc : bd)$; whence, if π fulfils (4), then

$$(ab : cd) + (ac : bd) + (ad : bc) = 0\,. \tag{5}$$

In other words, either none or exactly two of these split predicates have the value 1; yielding exactly four possibilities.

Let us classify these possibilities by their combinatorial sense. In fact, if T is a binary phylogenetic X-tree and a, b, c, d are *different* elements of X, there are at most two different triple points defined by the 3-subsets of $Y := \{a, b, c, d\}$; and if there are different such points, then the path between them splits Y into two 2-sets (in the classical sense, i.e. not involving parity split). The combined predicate "The path splits Y as $ab|cd$, and is of odd weight" is true if and only if $(ac : bd)(ad : bc) = 1$, and thus is discernible from the induced parity split map on Y. (Recall that in the truth values interpretation, this product is a 'logical and'.) However, if the path has even weight, we cannot from this deduce how Y is split. If instead at least two of the elements are *equal*, say $c = d$, we get similar but fewer case divisions: then in fact $3_{abc} = 3_{abd}$ and $3_{acd} = c = 3_{bcd}$, and the path $[3_{abc}, c]$ may be odd or even; in the first case ensuredly splitting Y as $ab|c$. This motivates the following definitions.

Definition 2.2. The *quartet parity* of $(a, b, c, d) \in X^4$ (with respect to the split map π) is $\mathrm{Par}(a,b,c,d) = \mathrm{Par}(a,b,c,d)_\pi = \mathrm{Odd}(ab\!-\!\!-cd) + \mathrm{Odd}(ac\!-\!\!-bd) + \mathrm{Odd}(ad\!-\!\!-bc)$,
where $\mathrm{Odd}(ab\!-\!\!-cd) = (ac : bd)(ad : bc)$, and correspondingly. See Figure 1. If π indeed satisfies the four point condition, then $\mathrm{Par}(a,b,c,d)$ is preserved under each permutation of the quadruple.

As you see, the definitions are quite 'algebraic', but encode rather concrete conditions in the case of an (ordinary) X-tree. In fact, for π satisfying (4), the split predicate values may be retrieved by

$$(ab : cd) = \mathrm{Odd}(ab\!-\!\!-cd) + \mathrm{Par}(a,b,c,d)\,. \tag{6}$$

You should also note, that by (5) $\mathrm{Par}(a,b,c,d)$ is odd, if and only if any two of the four vertices split the other two; or in other words that

$$\mathrm{Par}(a,b,c,d) = 1 + (1 + (ab : cd))(1 + (ac : bd))(1 + (ad : bc))\,. \tag{7}$$

Lemma 2.3. *If* $\mathrm{Odd}(ab \,—\, cd) = 1$ *with respect to a split map* Π_T *realised by an* X*-tree* T*, then* $3_{abc} = 3_{abd}$*,* $3_{acd} = 3_{bcd}$*,* $\mathrm{pdist}(3_{abc}, 3_{acd}) = 1$*, and* 3_{abc} *preceeds* 3_{acd} *on the unique path* $[a, c]$ *from* a *to* c*.*

Proof. Since $(ac : bd) = (ad : bc) = 1$, indeed 3_{abc} and 3_{abd} are at odd distances from 3_{acd} and 3_{bcd}, and thus different from these; whence indeed $3_{abc} = 3_{abd}$ and $3_{acd} = 3_{bcd}$, since there are at most two different triple points given by the quartet $\{a, b, c, d\}$. In particular, 3_{acd} cannot precede 3_{abc} in $[a, c]$; for, if it did, then we would have $3_{bcd} = 3_{dbc} = 3_{abc}$. □

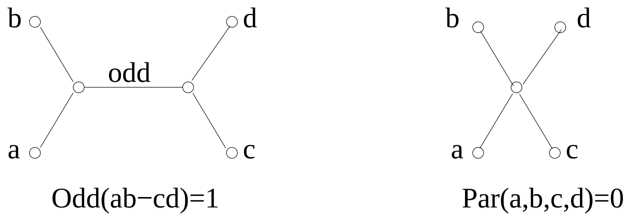

FIGURE 1. Examples when $\mathrm{Odd}(ab — cd) = 1$ and $\mathrm{Odd}(ab — cd) = \mathrm{Par}(a,b,c,d) = 0$ respectively.

Remark 2.4. The restriction of the four point condition to triplets,

$$(ab : ac) = (ac : ab), \text{for all } a, b, c \in X, \tag{8}$$

the *three point condition*, is in fact sufficient to ensure that $\mathrm{pdist}(a, b) := (ab : ab)$ is a $\mathbb{Z}_2$ metric on X. For, if (8) holds, then by (2) and (1) $\mathrm{pdist}(a, b) + \mathrm{pdist}(b, c) = \big((ab : ac) + (ab : bc)\big) + \big((bc : ba) + (bc : ca)\big) = \big((ba : bc) + (bc : ba)\big) + \big((ac : ab) + (ac : cb)\big) = 0 + (ac : ac) = \mathrm{pdist}(a, c)$.

With somewhat more work, we find that in fact the four point condition guarantees that λ is a $\mathbb{Z}_2$ metric on the set X^3 of 'formal triple points', where λ is defined by

$$\lambda\big((a, b, c),\ (d, e, f)\big) := (ab : cf) + (af : be) + (ef : ad)\,;$$

and that if indeed $\pi = \Pi_T$ for any parity weighted X-tree T, then $\mathrm{pdist}(3_{abc}, 3_{def}) = \lambda\big((a, b, c),\ (d, e, f)\big)$. Thus, if several parity weighted X-trees induce the same split maps, they also induce unambiguous parity distances between corresponding pairs of triple points.

2.4. **Analysis of joints.** In this subsection, we have to distinguish $\phi(a)$ from a (et cetera) explicitly.

One important special case of restrictions is when indeed $T = (V, E)$ is a parity weighted X-tree, and $X' = X \setminus \{x\} \neq \emptyset$ for some $x \in X$. In this case, if $\phi(x) = \phi(y)$ for some $y \in X'$ or the degree of $\phi(x)$ is at least 2, then put $T' := T$ and $\xi := \phi(x)$. Else, $\phi(x) \notin \phi(X')$ and is a leaf in T, whence there is a unique edge $\phi(x)\alpha$, say, in T. Then, put $T' = (V \setminus \{\phi(x)\}, E \setminus \{\phi(x)\alpha\})$ and $\xi := \alpha$. In either case, T'

is a generalised parity weighted X'-tree with a specified vertex ξ, and we call ξ the *joint* or *vertex of attachment* of x to T'. Intuitively the problem of finding how to join x to a parity weighted X'-tree by means of the parity split map may be divided into the problem of inserting ξ, the joint of x, properly, and the problem of attaching x to ξ.

Since T is an X-tree, each element in V is a triple point of three (different or equal) elements in $\phi(X)$. Thus there are $a, b \in X'$ and $c \in X$ (not necessarily different), such that $\xi = 3_{\phi(a)\phi(b)\phi(c)}$, and that if $\xi \neq \phi(x)$ then at most one of these may be $\phi(x)$ (since $3_{\phi(x)\phi(x)\alpha} = \phi(x)$). Furthermore, it is easily seen that we can always assume $c = x$ and hence $\xi = 3_{\phi(a)\phi(b)\phi(x)}$.

Next, consider the X parity split map $\pi := \Pi_T$, and its restriction $\pi' := \pi|_{X'}$. π contains the information that π' does, and the following kinds of 'extra information':

(A) For each 2-set $\{d, e\} \subseteq X'$, the information whether or not $[de : d|x]$;
(B) For each $d \in X'$, the restriction of the split $\pi(d, x)$ to X'; and
(C) For each $d \in X'$, the information whether or not $[dx : a|x]$, where $a \in X'$ is as above.

Indeed, the information in π' and (A) is enough to determine all the $\pi(d, e)$ such that $d, e \in X'$; since π' fixes each such split, except as whether x is in one part or in the other, but this is decided by (A). Similarly, the information in (C) completes the information in (B), obtaining the $\pi(d, x)$.

Lemma 2.5. *(A) is entirely decided by* (T', ξ)*; and consequently, so is (B). Moreover (C) in fact contains but one bit of extra information, namely the bit* $\mathrm{pdist}(\phi(x), \xi)$.

Proof. Note, that for $d, e \in X'$, the path $[\phi(d), \phi(e)]$ is entirely contained in T', whence $3_{\phi(d)\phi(e)\phi(x)} = 3_{\phi(d)\phi(e)\xi}$, whence $(de : dx) = \mathrm{pdist}(\phi(d), 3_{\phi(d)\phi(e)\phi(x)}) = \mathrm{pdist}(\phi(d), 3_{\phi(d)\phi(e)\xi})$, which indeed is determined by (T', ξ).

Thus, for $d, e, f \in X'$, $(dx : ef) = (dx : de) + (dx : df) = (de : dx) + (df : dx)$ is determined.

Finally, by (4) $(dx : ax) = (ax : dx)$ for each $d \in X'$; whence (C) is determined by the split $\pi(a, x)$, and more precisely by (B) for a, together with the information whether or not b and x are in the same part of the split $\pi(a, x)$. However, $(ax : bx) = \mathrm{pdist}(\xi, \phi(x))$; and the lemma is proved. □

3. Systems with at most five points

Let us briefly investigate the possible parity weighted binary phylogenetic X-trees $T = (V, E)$, with parity weight function ρ, for small X. (The $|X| \leq 2$ cases are trivial.)

If $|X| = 3$ and $X = \{a, b, c\}$, then (as is well known) V is the 4-set $\{a, b, c, 3_{abc}\}$; E is $\{a3_{abc}, b3_{abc}, c3_{abc}\}$, and as we saw above, e.g. $\rho(a, 3_{abc}) = \mathrm{pdist}(a, 3_{abc}) = (ab : ac)$. Thus, T may be fully reconstructed from Π_T.

If $|X| = 4$ and $X = \{a, b, c, d\}$, then V is a 6-set; E consists of one 'twig' to each leaf, and of one 'internal edge'; the parities of the twigs are determined as in the 3-set case; and the parity of the internal edge is $\mathrm{Par}(a{,}b{,}c{,}d)$. T is fully recoverable if and only if $\mathrm{Par}(a{,}b{,}c{,}d) = 1$. If the quartet parity is even, then T is one of three equivalent parity weighted binary phylogenetic X-trees, depending on how the 'central point' in the minimal X-tree in this equivalent class is split up into two different triple points.

If $|X| = 5$, then we get two internal edges, and three principally different ways of weighting them; cf. Figure 2 below.

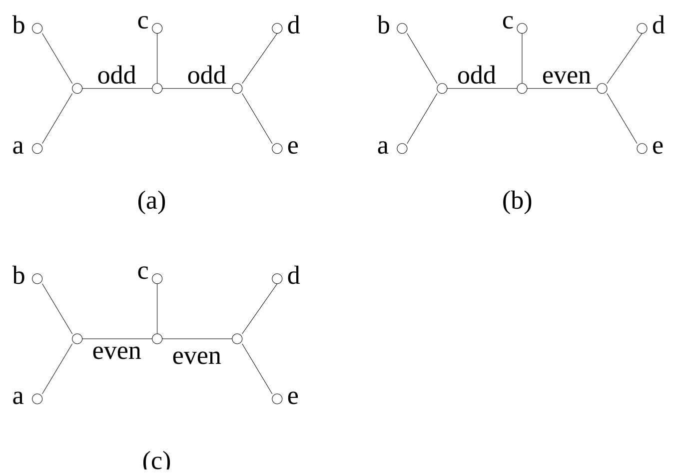

FIGURE 2. The three types of subtrees possible for five leaves.

A glance at the figure (or an obscure formal proof, which we omit) makes it clear, that at least one of the five quartet parities of different X elements must be even. After similary checking the few cases of pentuples in X with some entries equal, and by means of (3), we find that for any X and any generalised parity weighted binary phylogenetic X-tree we have

$$(9) \quad \mathrm{Par}(a{,}b{,}c{,}d)\,\mathrm{Par}(a{,}b{,}c{,}e)\,\mathrm{Par}(a{,}b{,}d{,}e)\,\mathrm{Par}(a{,}c{,}d{,}e)\,\mathrm{Par}(b{,}c{,}d{,}e) = 0$$

(As an exercise, you may prove that in fact (9) follows from (4), in the case where at least two of the elements a, b, c, d, e are equal.)

Definition 3.1. A split map π over X is said to satisfy the **five point condition** if (9) holds (for any $(a, b, c, d, e) \in X^5$).

The reader might suspect that we will continue to define k-point conditions for every k. Have no fear: the essence of Theorem 4.1 is that the four and five point conditions suffice.

Remark 3.2. We have seen that the four point condition implies the three point condition, which implies π inducing a metric. However, the converse implications do not hold, and nor does the four point condition imply the five point one; as the following examples show.

The parity map π for $X = \{a, b, c\}$, given by $\pi(a, b) = \pi(a, c) = \{X, \emptyset\}$, but $[bc : a|bc]$ (i.e., $\pi(b, c) = \big\{\{a\},\ \{b, c\}\big\}$), makes $(ab : ab) = (ac : ac) = (bc : bc) = 0$ and thus induces a parity metric; but $(bc : ba) = 1 \neq 0 = (ba : bc)$.

The parity map π for $X = \{a, b, c, d\}$, given by $[ab : abd|c]$, $[ac : acd|b]$, $[ad : abc|d]$, $[bc : bcd|a]$, $[bd : abc|d]$, and $[cd : abc|d]$, fulfils the three point condition; but not the general four point condition, since $(ab : cd) = 1 \neq 0 = (cd : ab)$.

The parity map π for $X = \{a, b, c, d, e\}$, given by $[ab : abe|cd]$, $[ac : ade|bc]$, $[ad : bde|ac]$, $[bc : ace|bd]$, $[bd : bce|ad]$, $[cd : cde|ab]$, $[ae : abc|de]$, $[be : abd|ce]$, $[ce : bcde|a]$, and $[de : acde|b]$, fulfils the four point condition; but not the five point condition, since all five quartet parities are odd.

We now may sum up what we have proved until now as follows:

Lemma 3.3. *Any realisable split map π for X fulfils the four point and the five point conditions. Conversely, if π fulfils the four point condition, and $|X| \leq 3$, then π is realisable.*

Next, we extend the 'converse' result of that lemma a bit, which will form the base of our inductive proof of Theorem 4.1.

Lemma 3.4. *Assume $4 \leq |X| < \infty$, $x \in X$, π a split map on X, fulfilling (4), and that for each set of four* distinct *elements $a, b, c, d \in X \setminus \{x\}$, $\mathrm{Par}(a,b,c,d) = 0$. Then π is realisable; and its minimal realisation is unique and has at most 2 interior edges.*

Proof. Put $X' := X \setminus \{x\}$, $\pi' := \pi|_{X'}$, and $n := |X'| \geq 3$. (By Remark 2.4, pdist_π is a metric.)

Let s be an element not in X, and let T' be the star with vertex set $X' \cup \{s\}$ and edge set $\{ys : y \in X'\}$). For any $a \in X'$, pick two distinct $b, c \in X' \setminus \{a\}$, and let $\rho(as) := (ab : ac)$. This defines a parity weight on the X'-tree T'; and we claim that its parity split map $\Pi_{T'} = \pi'$:

First of all, if d is a fourth element in X', then

$$(ac : ab) = (ab : ac) = \rho(as)\,,$$

and

$$(ab : ad) = (ab : ac) + (ab : cd) = \rho(as) + \mathrm{Par}(a,b,c,d) = \rho(as)\,,$$

by (4) and (2); i.e., $\rho(as)$ is independent of the choices of b and c. Next, $(ab:ab)_{T'} = \mathrm{pdist}_{T'}(a,s)+\mathrm{pdist}_{T'}(s,b) = (ab:ac)+(ab:bc) = (ab:ab)$ indeed, by (1); and, for $b \neq c$,

$$(ab:ac)_{T'} = \mathrm{pdist}_{T'}(a,s) = (ab:ac)\,.$$

Finally, for a, b, c, and d all distinct,

$$(ab:cd)_{T'} = 0 = (ab:cd)\,.$$

Thus inded the claim is proved.

Next, consider any minimal parity weighted X-tree T, such that $T^* := T|_{X'} \sim T'$. Let ξ be the joint of x with $T|_{X'}$. Then T^* is one of the following:

(A) T^* is a star, with $\xi = s$ (1 possibility);
(B) $\xi s \in E_{T^*}$; and for precisely one of the $a \in X'$, either $\xi = a$, or $\xi a \in E_{T^*}$, depending on whether $\mathrm{dist}_{T'}(a,s)$ is 1 or 0 (n possibilities); and
(C) There is a partition $X' = X_1 \cup X_2$, with both $|X_i| \geq 2$, and a split of s into two vertices s_1 and s_2, such that X_i is attached to s_i, and that both $\xi s_i \in E_{T^*}$ ($2^{n-1}-n-1$ possibilities).

In each of these cases, there are two possibilities for T, depending on $\mathrm{dist}_T(x,\xi)$. Thus, there are 2^n different such minimal parity weighted X-trees T; each of them with a different parity split map fulfilling (4) and extending π'. In order to prove that indeed π is one of these parity split maps, it is enough to prove that there are at most 2^n split maps that extend π' and fulfil (4).

Fix an $a \in X'$, and an order $b_1, \ldots, b_{n-1}$ on the elements in $X' \setminus \{a\}$; and put $b_n := x$. Now, let ψ be any split map on X, such that $\psi|_{X'} = \pi'$. Give ψ the *signature* $\big((ab_i:ax)\big)_{i=1}^{n}$. There are clearly not more than 2^n possible signatures; and we claim that ψ is completely determined by its signature (and π'). For, for $i \leq n-1$, $\psi(a,b_i)|_{X'} = \pi'(a,b_i)$, and the i'th entry of the signature determines which of the two parts of $\psi(a,b_i)$ (namely, the part containing a, or the other part), that contains x. This indeed fixes the split $\psi(a,b_i)$. Since $(ax:ab_i) = (ab_i:ax)$ by (4), and $(ax:ax)$ is the last entry of the signature, it is determined exactly which of the b_i ($i = 1,\ldots,n$) that are in the part of $\psi(a,x)$ that contains a; i,e, $\psi(a,b_n) = \psi(a,x)$ is fixed, too. Finally, for any $c,d,e \in X$ (different or not), $(cd:ae)_\psi = (ae:cd)_\psi$, which fixes $\psi(c,d)$. Thus the last claim, and consequently the lemma, is proved. □

Corollary 3.5. If a split map $\pi : \binom{X}{2} \longrightarrow \mathcal{S}(X)$ fulfils (4) and (9), and $|X| \leq 5$, then indeed X is realisable, and minimal realisations are isomorphic.

Proof. For $|X| \leq 3$, we are through by Lemma 3.3. For $|X| = 4$, the extra condition in the lemma trivially is fulfilled for any $x \in X$, since $X \setminus \{x\}$ does not contain four different elements. Finally, for

$|X| = 5$, by (9) we may write $X = \{a, b, c, d, e\}$ in such a manner that Par(a,b,c,d) $= 0$; whence taking $x := e$ in the lemma, indeed $X \setminus \{x\} = \{a, b, c, d\}$ fulfils the extra condition of the lemma. $\square$

Finally, we need a couple of more technical lemmata. Note that both the formulations and the 'algebraic' proofs of the first two work, even if $Y = \{a, b, c, d, e\}$ actually has cardinality smaller than 5. If you prefer 'purer' divide-and-conquer combinatorial proofs, then just note that the split map restriction to Y is realisable, by Corollary 3.5; and classify the possible realisations by means of Figure 2 (*mutatis mutandis*, if Y is not a 5-set).

Lemma 3.6. *If π is a split map on X that fulfils (4) and (9), and $a, b, c, d, e \in X$, then either none, three or four of the five quartet parities, defined by taking four of a, b, c, d, e, are odd.*

Proof. Either $(xy : zu) = 0$ for all distinct letters $x, y, z, u \in Y$, in which case all five quartet parities are even; or e.g. $(ab : cd) = 1$. In the latter case, Par(a,b,c,d) $= 1$; and by (2) exactly one of $(ab : ce)$ and $(ab : de)$, and exactly one of $(ae : cd)$ and $(be : cd)$, are odd, yielding at least two more odd quartet parities. All five cannot be odd because of (9). $\square$

Lemma 3.7. *Assume $a, b, c, d, e \in X$ and π a split map satisfying the four and five point conditions (4) and (9). If* Odd(ab — cd) $=$ Odd(bc — de) $= 1$*, then* Odd(ab — ce) $=$ Odd(ac — de) $= 1$*, too, but* Par(a,b,d,e) $= 0$.

Proof. By the assumptions, $(ab : cd) = (bc : de) = 0$, but $(ac : bd) = (ad : bc) = (bd : ce) = (be : cd) = 1$. Thus, by (1) and (2) $(cd : ae) = (cd : ab) + (cd : be) = 0+1 = 1$, and similarly $(ae : bc) = 1$; whence by (7) Par(a,b,c,d) $=$ Par(a,b,c,e) $=$ Par(a,c,d,e) $=$ Par(b,c,d,e) $= 1$; whence by (9) Par(a,b,d,e) $= 0$, i.e., $(ab : de) = (ad : be) = (ae : bd) = 0$. Thus, $(ab : ce) = (ab : de) + (ab : cd) + (ab : de) = 0 + 0 = 0$ and $(ac : be) = (ab : ce) + (ae : bc) = 0 + 1 = 1$ yields Odd(ab— ce) $=$ Par(a,b,c,e) $= 1$; and similarly, $(bc : de) = 0$ and Odd(bc—de) $= (ad : ce) = 1$. $\square$

Lemma 3.8. *If π is a split map on X that fulfils (4) and (9), $x \in X$, and* Par(a,b,c,x) $= 0$ *for all* distinct $a, b, c \in X' := X \setminus \{x\}$*, then* Par($a$,$b$,$c$,$d$) $= 0$ *for all distinct $a, b, c, d \in X'$.*

Proof. Given distinct $a, b, c, d \in X'$, at least four of the quartet parities in $\{a, b, c, d, x\}$ are zero, whence so is the fifth by Lemma 3.6. $\square$

4. Proof of main result

This section is devoted to proving our main theorem, which is the following.

Theorem 4.1. *A split map π over a set X is realisable, i.e. $\pi = \Pi_T$, for some parity weighted X-tree T, if and only if π satisfies the four point condition (4) and the five point condition (9). Moreover, up to isomorphisms there is exactly one such X-tree T with all edges of odd parity.*

It is clear from Section 3 that the conditions are necessary. In order to prove the sufficiency and the uniqueness statements, we shall use induction over $|X|$. In the proofs, we will almost entirely employ only trivially weighted X-trees, i.e., with only odd edges. Proving the existence and uniqueness for such trees proves the theorem.

For $|X| \leq 5$, all claims in the theorem hold by Corollary 3.5. Let π be a split map X, $|X| \geq 6$, satisfying the four and five point conditions, and make the natural assumptions of induction.

Take $x \in X$, by induction we know that there exists a unique (up to isomorphism) X_x-tree T_x for which $\pi|_{X_x}$ is the parity split map. We now want to show that there is a unique way of 'extending' T_x, to an X-tree T with π as parity split map.

By the discussion in Section 2.2, we must first determine the position of the joint of x (possibly splitting one T_x vertex into three in the process), and then determine whether x is mapped to the joint, or to a new vertex adjacent to it. (The second part is almost trivial though.)

We first define a direction for some edges in T_x, and show that this is well-defined in Lemma 4.3. The intuition is that the edges are pointing towards the joint of x. Some edges may not have a direction.

Definition 4.2. Let DT_x be T_x together with directions on *some* of its edges, determined as follows: If $\mathrm{Odd}(ax — bc) = 1$ then all edges on the paths $[b, 3_{abc}]$ and $[c, 3_{abc}]$ in T_x will be directed towards 3_{abc}.

Lemma 4.3. *DT_x is well defined, i.e., no edge can receive contradicting directions.*

Proof. Assume for contradiction that the edge $e = \{\beta_0, \beta_1\}$ in T' is given two directions; i.e., that there exists $a, b, c \in X'$ such that e is directed towards β_0 by $\mathrm{Odd}(ax — bc) = 1$, with e on the path from b to 3_{abc}, and there exists $u, v, w \in X'$ such that e is directed towards β_1 by $\mathrm{Odd}(ux — vw) = 1$, with e on the path from v to 3_{uvw}. Let C_1 and C_2 be the two connected components of $T' \backslash \{e\}$, with $a, c, v \in C_1$ and $b, u, w \in C_2$.

First, note that for any $z \in C_2 \cap X$ we have $\mathrm{Odd}(ax — zc) = 1$. For, since $\mathrm{pdist}(\beta_0, \beta_1) = 1$, there is exactly one $\beta \in e$, such that $\mathrm{pdist}(3_{abc}, \beta) = 1$; and since $\beta \in [3_{abc}, z]$, there is a $y \in X$ such that $\beta = 3_{ayb} = 3_{ayz} = 3_{cyb} = 3_{cyz}$. In particular, $\mathrm{Odd}(ac — by) = \mathrm{Odd}(ac — yz) = 1$; and $\mathrm{Odd}(xa — cb) = 1$ by assumption. Thus, by Lemma 3.7 twice, $\mathrm{Odd}(xa — cb) = \mathrm{Odd}(ac — by) = 1$ yields $\mathrm{Odd}(xa — cy) = 1 = \mathrm{Odd}(ac — yz)$ yields $\mathrm{Odd}(xa — cz) = 1$ indeed.

In particular, $\mathrm{Odd}(ax — uc) = \mathrm{Odd}(ax — wc) = 1$; and the latter gives $\mathrm{Odd}(cx — aw) = 0$. However, and by interchanging the rôles of (a, b, c) and (u, v, w), we find that $\mathrm{Odd}(ux — aw) = 1 = \mathrm{Odd}(cu — xa)$, again by means of Lemma 3.7 yielding $\mathrm{Odd}(cx — aw) = 1$, and the sought contradiction. □

Remark 4.4. In fact all edges on the same connected component C as (e.g.) b on $T \backslash 3_{abc}$ will be directed towards 3_{abc} in DT_x.

Proof. The statement is trivially true if $b = 3_{abc}$. Otherwise, let $d \in X_x$ be any point such that $\mathrm{dist}(3_{abc}, 3_{abd}) = 1$ and $3_{abd} \in [3_{abc}, b]$. Then $\mathrm{Odd}(ax — bc) = \mathrm{Odd}(ac — bd) = 1$, which by Lemma 3.7 gives $\mathrm{Odd}(ax — cd) = 1$. Hence also all edges from d to 3_{acd} are directed that way. Now, every element $e \in C$ such that $3_{abe} \neq 3_{abd}$ satisfies $3_{ade} = 3_{abd}$, which similarly implies $\mathrm{Odd}(ax — ce) = 1$. □

As the reader may have suspected, there is no guarantee for having *any* directed edges in DT_x. However, if no edge has received a direction, then we have $\mathrm{Par}(a,b,c,x) = 0$ for all distinct $a, b, c \in X_x$. In Lemma 3.8 we proved that this implies $\mathrm{Par}(a,b,c,d) = 0$ for all distinct $a, b, c, d \in X_x$, whence indeed π is uniquely realisable by Lemma 3.4.

Thus, in the sequel we may assume that there is at least one directed edge in DT_x, whence by Remark 4.4 there is also a directed edge $e = \{\beta, \upsilon\}$ pointing away from a leaf β in DT_x. Assume the leaf β is labeled $Y \subseteq X$.

Claim 4.5. *$\pi(x, y_1) = \pi(x, y_2)$, for $y_1, y_2 \in Y$.*

Proof. First note that $(xy_1 : cd) + (xy_2 : cd) = (y_1y_2 : cd)$ by (2). If $c, d \in X_x$ this is zero by $\pi|_{X_x} = \Pi|_{T_x}$ and thus $(xy_1 : cd) = (xy_2 : cd)$ as wanted. Again by (2) we get $(y_1y_2 : cx) = (y_1y_2 : dx)$ for all $c, d \in X_x$. We must show it is zero.

By the direction of the edge e and by Remark 4.4, there exists $a, b \in X_x \backslash Y$ such that $\mathrm{Odd}(ax — by) = 1$ for all $y \in Y$. In particular, $(ax : by_1) = (ax : by_2) = 0$. This gives $(y_1y_2 : ax) = (ax : y_1y_2) = 0$ as needed. The claim is proved. □

We may thus assume that $Y = \{y\}$.

Recall that by definition the joint of the label y with the X_x-tree T_x is the node υ. Let $X_{x,y}$ denote $X \backslash \{x, y\}$ and let T_{xy} denote the $X_{x,y}$-tree with $\pi_{X_{x,y}} = \Pi_{T_{xy}}$ which we know exists and is unique (up to isomorphism) by induction. There are two possibilities:

- **I** υ has degree at least four in T_x or is labeled and is hence present also in T_{xy}, i.e., T_{xy} is a subtree of T_x.
- **II** υ is not labeled and has degree at most 3 in T_x. Hence υ is not present in T_{xy} and its two neighbours φ' and φ'' in T_x are merged to a single node φ in T_{xy}.

We will call the process of adjusting an X_x-tree T_x to a subset the **adaptation** of T_x to the subset. Note that since we sometimes have to remove and identify vertices the adaptation will in general not be a subtree.

By induction we know that we can extend T_{xy} with x in a unique way to a X_y-tree T_y. We now want to put back β, with label y to T_y and form a X-tree T. Let ξ be the joint of x in T_y. In this case the distance between x and ξ can be zero or one, i.e., x might be mapped to ξ or to a leaf adjacent to ξ, but we do not need to distinguish these cases below. By definition T_{xy} is the adaptation of T_y to $X_{x,y}$, and we have again two different possibilities.

- **a** ξ has degree at least four in T_y or has labels (other than possibly x), and ξ is hence present also in T_{xy}, i.e., T_{xy} is a subtree of T_y.
- **b** ξ has no other labels and has degree at most 3 in T_y. Hence ξ is not present in T_{xy}, and its two neighbours γ' and γ'' in T_y are merged to a single node γ in T_{xy}.

We will now for each of the four cases Ia, Ib, IIa and IIb (notation as above) describe how to build the tree T. The tree T must first of all be such that its adaptation to X_x is T_x and its adaptation to X_y is T_y. In each case we will first describe the possible candidates T with this property and then check that there is a unique (up to isomorphism) among them with $\Pi_T = \pi$.

It is easily seen that for any such candidate T, $\Pi_T(c, d)$ is completely determined for all $c, d \in X_{x,y}$ by the adaptations to T_x and T_y. In fact, there is only one bit of information not determined. If for some $c, d \in X_{x,y}$ we know $(cx : dy)_T$., then for any $u \in X_x, v \in X_{x,y}$ we have by (2) that $(ux : vy)_T = (cx : vy)_T + (cu : vy)_{T_x} = (cx : dy)_T + (cx : vd)_{T_y} + (cu : vy)_{T_x}$. For $v = x$ we use similar expansions, and the split map is thus fully determined.

If ξ (or γ) and υ (or φ) are sufficiently far apart in T_{xy} the construction of T is easy, but if they are close more care has to be taken. Each of the four main cases below will be further divided into two subcases depending on this distance. The details will be similar, especially when the distance is 'large', and we will write them out completely only for case Ia.

Case Ia) Let $[\upsilon, \xi]_{T_y} = (\upsilon = \delta_0, \delta_1, \dots, \delta_k = \xi)$.
(i) If $k \geq 2$ we extend T_y to T by joining y to υ as it is joined in T_{xy}. This is clearly the only candidate tree. We may then choose $a, b, c \in X \backslash \{x, y\}$ such that $3_{by\xi} = \delta_0$, $3_{ay\xi} = \delta_1$ and $3_{cy\xi} = \delta_2$ in T. We then have $\mathrm{Odd}(ac — by)_T = \mathrm{Odd}(ac — by)_\pi = 1$ and $\mathrm{Odd}(ab — cx)_T = \mathrm{Odd}(ab — cx)_\pi = 1$, which by Lemma 3.7 yields $\mathrm{Odd}(ax — by)_T = \mathrm{Odd}(ax — by)_\pi = 1$ and hence $(ux : vy)_T = (ux : vy)_\pi$, for all $u, v \in X \backslash \{x, y\}$. We thus have $\Pi_T = \pi$.

(ii) If $k = 1$ there are exactly two ways of adding y depending on the values of $(ux : vy)_\pi$. Let C_υ and C_ξ be the two connected components if the edge $\{\upsilon, \xi\}$ is removed from T_y, with $\upsilon \in C_\upsilon, \xi \in C_\xi$. The first way to add y back in order to form the tree T, is the obvious one of letting the joint of y be υ. This works if $(bx : ay)_\pi = 1$, for $b \in C_\upsilon, a \in C_\xi$. If on the other hand $(bx : ay)_\pi = 0$, for $b \in C_\upsilon, a \in C_\xi$, we form the X-tree T' by splitting υ into two nodes υ', υ'', and splitting ξ into ξ', ξ''. They will form a path $\upsilon', \xi'', \upsilon'', \xi'$, and υ' (ξ') will take the place of υ (ξ) in C_υ (C_ξ), whereas υ'' will be the joint of y and ξ'' the joint of x. The trees T and T' will have the same split functions, except that y will be placed in a different part of $\pi(c, x)$, for all $c \in X_x$. See Figure 3.

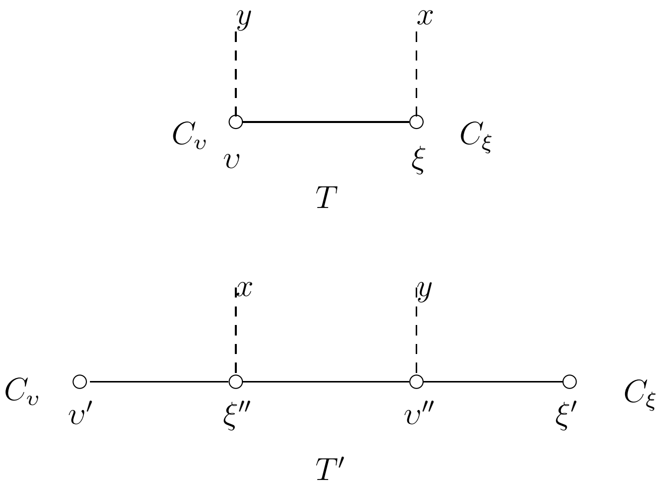

FIGURE 3. Construction of T and T' in Case Ia(ii)

Note that $k = 0$ is not a possibility. If $\upsilon = \xi$, then x and y must belong to the same part in all splits $\pi(u, v), u, v \in X_{x,y}$. Thus $(xy : uv) = 0$ for all $u, v \in X_{x,y}$, which contradicts the fact that the edge e was given a direction by $\mathrm{Odd}(ax — by) = 1$.

Case Ib) We again let k be the distance between υ and ξ in T_y. When $k \geq 2$ we may argue as in Ia(i) to show that $\pi = \Pi_T$ for a unique T.

If $k = 1$, this means that $\upsilon = \gamma'$ (or γ''). As in Case Ia(ii) we can now construct two candidate X-trees T and T' with γ' and γ'' as the joint of y. The split map for T and T' will differ only on the values of $(ux : vy)_\pi$ and the adaptations of both T and T' is T_x and T_y. Hence exactly one of $\pi = \Pi_T$ and $\pi = \Pi_{T'}$ will be true.

Case IIa) Here we will let k be the distance between υ and ξ in T_x. Then this case is symmetric to case Ib).

Case IIb)

(i) Assume $\gamma \neq \varphi$ in T_{xy}. In this case, T will be constructed in the obvious manner from T_x by splitting γ, as T_y is constructed from T_{xy}. That $\Pi_T = \pi$ is shown as in Ia(i).

(ii) Assume $\gamma = \varphi$ in T_{xy}. Here we need to examine the structure of T_x and T_y carefully. Let C_1 be the component containing γ' and C_2 the component containing γ'' in $T_y \backslash \{\xi\}$. Similarly let C_3, C_4 be the connected components containing φ', φ'' when removing v from T_x. First we shall show that that there cannot be $a, b, c, d \in X_{x,y}$ such that $a \in C_1 \cap C_3, b \in C_1 \cap C_4, c \in C_2 \cap C_3, d \in C_2 \cap C_4$. Assume for contradiction that such a, b, c, d exist. Some of them could be labels of γ in T_{xy}, but the others must by necessity be in different components of $T_{xy} \backslash \{\gamma\}$. In any case $3_{abx} = \gamma', 3_{cdx} = \gamma''$ in T_y and $3_{acy} = \varphi', 3_{bdy} = \varphi''$ in T_x. This means that we have the following values for π:
$(ab : cx) = 0, (ac : bx) = 1, (ax : bc) = 1,$
$(ab : dx) = 0, (ad : bx) = 1, (ax : bd) = 1$ and
$(ab : cy) = 1, (ac : by) = 0, (ay : bc) = 1,$
$(ab : dy) = 1, (ad : by) = 1, (ay : bd) = 0.$

This implies directly using (2), where $\epsilon = 0$ or 1.

$(ab : cx) = 0$	$(ac : bx) = 1$	$(ax : bc) = 1$
$(ab : cy) = 1$	$(ac : by) = 0$	$(ay : bc) = 1$
$(ab : xy) = 1$	$(ax : by) = \epsilon$	$(ay : bx) = 1 - \epsilon$
$(ac : xy) = 1$	$(ax : cy) = 1 - \epsilon$	$(ay : cx) = \epsilon$
$(bc : xy) = 0$	$(bx : cy) = \epsilon$	$(by : cx) = \epsilon$

By the five point condition at least one row in this table must be all zeros and this must thus be the last row, so $\epsilon = 0$. However, we immediately get

$(ab : dx) = 0$	$(ad : bx) = 1$	$(ax : bd) = 1$
$(ab : dy) = 1$	$(ad : by) = 1$	$(ay : bd) = 0$
$(ab : xy) = 1$	$(ax : by) = 0$	$(ay : bx) = 1$
$(ad : xy) = 0$	$(ax : dy) = 1$	$(ay : dx) = 1$
$(bd : xy) = 1$	$(bx : dy) = 0$	$(by : dx) = 1$

,

where the last two rows follows from repeated use of (2). This contradicts the five point condition (9).

Thus, without loss of generality we may assume that C_2 and C_4 contain no common labels from $X_{x,y}$. We may view C_1 as a subtree of T_{xy}, if we rename γ' as γ (keeping the labels of γ' though). Similarly C_2, C_3, C_4 may viewed as subtrees. Let now $T(i, j)$ be the subtree of T_{xy} restricted to $C_i \cap C_j$, where the labels of γ form the intersection of the labels from C_i and C_j. It is clear that the only way to construct T from T_y respecting π is with a path $\gamma'', \xi, \gamma'(= \varphi'), v, \varphi''$ where x and y are joined to ξ and v respectively. The tree $T(2, 3)$ is attached to the path by identifying γ and γ''. Similarly $T(1, 3)$ is attached to γ' and $T(1, 4)$ to φ''. Thus, $\text{dist}(v, \xi) = 2$, and we may use the same argument as in Ia) to show that $\Pi_T = \pi$.

The last possibility would be that C_1 and C_4 have the same labels and so have C_2 and C_3. This implies that x and y would belong to the same part in all splits $\pi(u, v), u, v \in X_{x,y}$. Thus $(xy : uv) =$

$0, u, v \in X_{x,y}$, which contradicts that the edge e was given a direction by $\mathrm{Odd}(ax — by) = 1$.

Thus indeed in each case we have constructed a unique T, except as regards whether we should put $\phi(x) = \xi$ or we should add a new leaf $\phi(x)$ with an edge $\phi(x)\xi$. However, this is decided by the truth value of $\mathrm{pdist}(x, \xi)$, which indeed is uniquely determined, as proved in Lemma 2.5.

5. Characterisation of Binary phylogenetic X-trees

Theorem 4.1 gives exact conditions for when a split map π is equal to Π_T for some X-tree T. Below we give two conditions which together with the four and five point conditions give necessary and sufficient conditions to determine which split maps come from binary phylogenetic X-trees. We challenge the reader to find other and perhaps better answers to the question of Dress.

Theorem 5.1. *Let (T, ϕ) be an X-tree with parity split map Π_T. Then T is phylogenetic if and only if*

$$\text{(10)} \quad \textit{for all distinct } a, b, c \in X, \textit{ such that } (ac : bc) = 0 \quad \textit{there exists } x \in X, \textit{ such that } \mathrm{Odd}(ab — cx) = 1.$$

Proof. The conditions $a, b, c \in X, a, b \neq c$ and $(ac : bc) = 0$ imply that $\mathrm{dist}(c, 3_{abc})$ is even. If T is phylogenetic, then the distance cannot be zero, whence there exists an $x \in X$ with $\mathrm{dist}(3_{abc}, 3_{acx}) = 1$. This proves the "only if" direction. For the converse, assume that T is not phylogenetic. This means that there exists some $c \in X$ that is not mapped to a leaf, i.e., $c \in [a, b]$ for some other elements $a, b \in X$. This violates (10), which proves the theorem. □

Theorem 5.2. *Let (T, ϕ) be a phylogenetic X-tree with parity split map Π_T. Then T is binary if and only if*

$$\text{(11)} \quad \textit{for all distinct } a, b, c, d \in X, \textit{ such that } \mathrm{Par}(a,b,c,d) = 0 \quad \textit{there exists } x \in X, \textit{ such that}$$
$$\mathrm{Par}(a,b,c,x) = \mathrm{Par}(a,b,x,d) = \mathrm{Par}(a,x,c,d) = \mathrm{Par}(x,b,c,d) = 1.$$

Proof. First assume that T is binary. $\mathrm{Par}(a,b,c,d) = 0$ for distinct $a, b, c, d \in X$ means that the distance between the triple points must be even and at least two. Without loss of generality, we may assume that $3_{abc} = 3_{abd} \neq 3_{acd} = 3_{bcd}$. Thus there is some $x \in X$ such that $3_{acx} \in [3_{abc}, 3_{acd}]$ and with $\mathrm{dist}(3_{acx}, 3_{abc})$ odd. We are now in case (a) of Figure 2, and the "only if" direction of the theorem follows.

For the converse, assume that T is not binary. Since T is phylogenetic there must be a vertex β of degree 4 or more. Then we can take $a, b, c, d \in X$ in different parts of $T \backslash \beta$. Thus $3_{abc} = 3_{abd} = 3_{acd} = 3_{bcd}$, and no x can exist as in (11). □

Corollary 5.3. A split map π over a set X satisfies $\pi = \Pi_T$ for some binary phylogenetic X-tree T, if and only if π satisfies the four point condition (4), the five point condition (9), and (10) and (11). Moreover this tree is unique among binary phylogenetic X-trees.

Acknowledgements. We would like to thank Andreas Dress for introducing us to this problem, and for helpful criticism; and Maria Grunewald for helping us to a better understanding of the relation between parity questions and linkage analysis.

References

[1] Sebastian Böcker and Andreas Dress: *Recovering symbolically dated, rooted trees from symbolic ultrametrics*, Advances in Mathematics **138**, 105–125 (1998)

[2] Andreas Dress and Mike Steel: *Mapping edge sets to splits in trees: The path index and parsimony*, preprint 2005.

[3] Andreas Dress, Talk at the **PCA04** conference in Uppsala, Sweden, July 5–9, 2004

[4] Charles Semple and Michael Steel: *Phylogenetics*, Oxford University Press, 2003 (ISBN 0 19 850942 1)

[5] Tom Strachan and Andrew P. Read: *Human molecule genetics*, second edition, BIOS Scientific Publishers Ltd, 1999 (ISBN 1 85996 202 5)

Jörgen Backelin, Department of Mathematics, Stockholms universitet, SE-106 91 Stockholm, Sweden
E-mail address: `joeb@math.su.se`

Svante Linusson, Department of Mathematics, Linköpings universitet, SE-581 83 Linköping, Sweden
E-mail address: `linusson@mai.liu.se`