Annals of Combinatorics 7 (2003) 155-169
0218-0006/03/020155-15
DOI 10.1007/s00026-003-0179-x

Annals of Combinatorics

X-Trees and Weighted Quartet Systems

Andreas W.M. Dress[1*] and Péter L. Erdős[2†]

[1]Forschungsschwerpunkt Mathematisierung-Struktubildungprozesse, University of Bielefeld
P.O. Box 100131, 33501 Bielefeld, Germany
dress@mathematik.uni-bielefeld.de

[2]A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, P.O. Box 127
1364 Hungary
elp@renyi.hu

Received April 17, 2003

AMS Subject Classification: 05C05, 92D15, 92B05

Abstract. In this note, we consider a finite set X and maps W from the set $\mathcal{S}_{2|2}(X)$ of all 2, 2-splits of X into $\mathbb{R}_{\geq 0}$. We show that such a map W is induced, in a canonical way, by a binary X-tree for which a positive length $\ell(e)$ is associated to every inner edge e if and only if (i) exactly two of the three numbers $W(ab|cd), W(ac|bd)$, and $W(ad|cb)$ vanish, for any four distinct elements a, b, c, d in X, (ii) $a \neq d$ and $W(ab|xc) + W(ax|cd) = W(ab|cd)$ holds for all a, b, c, d, x in X with $\#\{a, b, c, x\} = \#\{b, c, d, x\} = 4$ and $W(ab|cx), W(ax|cd) > 0$, and (iii) $W(ab|uv) \geq \min\big(W(ab|uw), W(ab|vw)\big)$ holds for any five distinct elements a, b, u, v, w in X. Possible generalizations regarding arbitrary $\mathbb{R}$-trees and applications regarding tree-reconstruction algorithms are indicated.

Keywords: biological systematics, phylogeny, phylogenetic combinatorics, evolutionary trees, tree reconstruction, X-trees, quartet methods, quartet systems, weighted quartet systems.

1. Introduction

Let X be a finite set of cardinality n, and let $T = (V, E)$ be an X-tree, i.e., a finite tree without vertices of degree 2 whose set of leaves coincides with X. Further,

(i) let $\binom{X}{i}$ denote, for any natural number i, the set of all subsets of X of cardinality i,
(ii) let $\mathcal{S}_{2|2}(X)$ denote the set of all *partial* 2, 2-*splits* of X:

$$\mathcal{S}_{2|2}(X) := \left\{ \Big\{ \{a, b\}, \{c, d\} \Big\} \Big| \{a, b\}, \{c, d\} \in \binom{X}{2};\ \{a, b\} \cap \{c, d\} = \emptyset \right\},$$

* Supported in part by the DFG.

† Supported by the Alexander v. Humboldt Stiftung and by the Hungarian NSF, under contract Nos. T34702, T37846.

(iii) let $E_0 = E_0(T)$ denote the set of *pending* edges of T, i.e., of edges incident with a leaf:

$$E_0 = E_0(T) := \{e \in E \,\big|\, e \cap X \neq \emptyset\},$$

(iv) let $E_1 = E_1(T)$ denote the complementary set of *inner* edges of T:

$$E_1 = E_1(T) := E \setminus E_0,$$

(v) and let

$$\ell\colon E_1 \to \mathbb{R}_{>0}$$

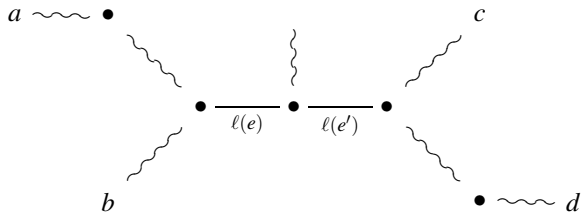
denote an arbitrary, but strictly positive *length function* defined on that set.

For convenience, we will also write $ab|cd$ for the unordered pair $\{\{a,b\},\{c,d\}\}$ of subsets of X of cardinality at most 2, for any $a,b,c,d \in X$ (so that $ab|cd \in \mathcal{S}_{2|2}(X)$ holds if and only if one has $\#\{a,b,c,d\} = 4$).

We are interested in the map $W = W_{T,\ell}$ defined on $\mathcal{S}_{2|2}(X)$ by

$$W\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0},\ ab|cd \mapsto \sum_{e \in E(ab|cd)} \ell(e), \tag{1.1}$$

where the sum runs over the set $E(ab|cd)$ of all edges $e \in E$ that separate the leaves a,b from the leaves c,d. Clearly, the function W measures the total length of the "inner path" of the *quartet tree* $T_{a,b,c,d}$ "spanned" by a,b,c,d in case T contains at least one edge that separates a,b from c,d, and it vanishes otherwise.



The following facts are easily established:

(F1) Given any 4-subset $\{a,b,c,d\}$ of X, at least two of the three numbers $W(ab|cd)$, $W(ac|bd)$, and $W(ad|cb)$ vanish.

(F2) If T is binary, i.e., if all vertices in V outside X have degree 3 or — equivalently — if $\#V = 2n-2$ holds (recall that there is no vertex of degree 2), one has

$$W(ab|cd) + W(ac|bd) + W(ad|cb) > 0 \tag{1.2}$$

for all $\{a,b,c,d\} \in \binom{X}{4}$.

(F3) Given $a,b,c,d,x \in X$ with $\#\{a,b,c,x\} = \#\{b,c,d,x\} = 4$ and

$$W(ab|xc), W(bx|cd) > 0,$$

one has $\#\{a,b,c,d,x\} = 5$ and

$$W(ab|xc) + W(bx|cd) = W(ab|cd). \tag{1.3}$$

(F4) Given any 5-subset $\{a, b, u, v, w\}$ of X, one has

$$W(ab|uw) \geq \min\Big(W(ab|uv), W(ab|vw)\Big), \tag{1.4}$$

i.e., the two smaller ones of the three numbers

$$W(ab|uv), W(ab|uw), W(ab|vw)$$

must coincide or, still in other words, $W(ab|uv) < W(ab|uw)$ implies that $W(ab|uv) = W(ab|vw)$ for all $a, b, u, v, w \in X$ as above.

Our main result is the following:

Theorem 1.1. *A map*

$$W\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$$

is of the form $W_{T,\ell}$ for some finite binary tree T with leave set X and some length function ℓ defined on the set $E_1(T)$ of inner edges of T if and only if W satisfies the conditions (F1) *to* (F4) *above. Moreover, if W satisfies those four conditions, the tree T and the length function $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ with $W = W_{T,\ell}$ are uniquely determined* (*up to canonical isomorphism*) *by W.*

It was established already in 1977 by the psychologists Colonius and Schulze (cf. [5, 6]), the first two papers on quartet analysis that initiated much further work devoted to this topic, cf. [7–39] that, given any subset $\mathcal{Q}$ of $\mathcal{S}_{2|2}(X)$, there exists a binary X-tree $T = (V, E)$ such that the set

$$\mathcal{Q}_T := \big\{ab|cd \in \mathcal{S}_{2|2}(X) \;\big|\; E(ab|cd) \neq \emptyset\big\}$$

of 2|2-splits in $\mathcal{S}_{2|2}(X)$ induced by T coincides with $\mathcal{Q}$ if and only if the following three assertions hold:

(Q1) $\#(\mathcal{Q} \cap \{ab|cd, ac|bd, ad|cb\}) = 1$ holds for all $\{a, b, c, d\} \in \binom{X}{4}$,
(Q2) $ab|cx \in \mathcal{Q}$ and $ax|cd \in \mathcal{Q}$ implies $ab|cd \in \mathcal{Q}$ for all $\{a, b, c, d, x\} \in \binom{X}{5}$,
(Q3) $ab|uv, ab|vw \in \mathcal{Q}$ implies $ab|uw \in \mathcal{Q}$ for all $\{a, b, u, v, w\} \in \binom{X}{5}$,

in which case this tree is uniquely determined by $\mathcal{Q}$.

Thus, the *support*

$$\operatorname{supp}(W) := \big\{ab|cd \in \mathcal{S}_{2|2}(X) \;\big|\; W(ab|cd) \neq 0\big\}$$

of any map $W\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$ that satisfies the conditions (F1) to (F4) above is obviously of the form $\mathcal{Q}_T$ for some unique binary X-tree T. Thus, a proof of the existence part of Theorem 1.1 could easily be based on this observation. In this note however, we want to proceed in a more direct way, not so much to avoid referring to any previous work, but because our direct approach also yields new tree-building strategies.

The paper is organized as follows: In the next section, we will show that the map $W_{T,\ell}\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$ associated with a binary X-tree T and a length function $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ determines T and ℓ up to canonical isomorphism. Then, we will show that a map $W\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$ is of the form $W = W_{T,\ell}$ for some binary X-tree T and length function $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ if and only if W satisfies the conditions (F1) to (F4) above. And finally, we shall discuss various promising directions of future research as well as some simple algorithmic applications of our results in the last section.

2. $W_{T,\ell}$ Determines T and ℓ up to Canonical Isomorphism

Given any two binary X-trees T and T' and maps $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ and $\ell'\colon E_1(T') \to \mathbb{R}_{>0}$, we will show here that $W_{T,\ell} = W_{T',\ell'}$ implies the existence of a unique map $\varphi\colon V(T) \to V(T')$ with $\varphi(x) = x$ for all $x \in X$ and $\{\varphi(u), \varphi(v)\} \in E(T')$ for all $\{u, v\} \in E(T)$, and that this map is necessarily bijective, induces a bijection between $E(T)$ and $E(T')$, and commutes with ℓ and ℓ' (i.e., $\ell(\{u, v\}) = \ell'(\{\varphi(u), \varphi(v)\})$ holds for this map φ and all $\{u, v\} \in E(T)$).

To construct $\varphi(v)$, recall the following facts:

i) Given any finite connected graph $G = (V, E)$, the *standard* graph metric d_G induced on V by G is defined to be the map from $V \times V$ into $\mathbb{N}_0$ that maps each pair $(u, v) \in V \times V$ onto the minimal number $d_G(u, v)$ of edges that constitute a path from u to v in G, i.e., onto the minimum of all $k \in \mathbb{N}_0$ for which vertices $v_0 := u, v_1, \ldots, v_k := v \in V$ exist with $\{v_{i-1}, v_i\} \in E$ for all $i = 1, \ldots, k$.

ii) A finite connected graph $G = (V, E)$ is defined to be a *median* graph if, for all $u, v, w \in V$, there exists a unique vertex $m = \mathrm{med}_G(u, v, w)$ in V with

$$\begin{aligned} d_G(u, v) &= d_G(u, m) + d_G(m, v), \\ d_G(u, w) &= d_G(u, m) + d_G(m, w), \end{aligned}$$

and

$$d_G(v, w) = d_G(v, m) + d_G(m, w),$$

in which case $\mathrm{med}_G(u, v, w) = \mathrm{med}_G(v, u, w) = \mathrm{med}_G(u, w, v)$ and $\mathrm{med}_G(u, u, w) = u$ hold for all $u, v, w \in V$ (cf. [1]).

iii) Any X-tree $T = (V, E)$ is a median graph and every vertex v in V is of the form $v = \mathrm{med}_T(a, b, c)$ for some appropriate leaves a, b, c in X, and one has $\mathrm{med}_T(a, b, c) \in V - X$ for some $a, b, c \in X$ if and only if $\#\{a, b, c\} = 3$ holds.

iv) Given a X-tree $T = (V, E)$, a length function $\ell\colon E_1(T) \to \mathbb{R}_{>0}$, and four distinct leaves $a, b, c, d \in X$, one has $W_{T,\ell}(ab|cd) > 0$ if and only if one has

$$\mathrm{med}_T(a, b, c) = \mathrm{med}_T(a, b, d) \neq \mathrm{med}_T(a, c, d) = \mathrm{med}_T(b, c, d),$$

in which case $E(ab|cd)$ consists exactly of the set of edges $e \in E_1(T)$ on the unique path from $\mathrm{med}_T(a, b, c) = \mathrm{med}_T(a, b, d)$ to $\mathrm{med}_T(a, c, d) = \mathrm{med}_T(b, c, d)$ and $W_{T,\ell}(ab|cd)$ is exactly the length of that path relative to ℓ.

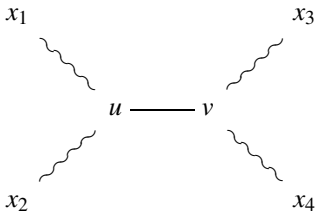
v) If, moreover, T is binary, one has

$$\mathrm{med}_T(a_1, a_2, a_3) = \mathrm{med}_T(b_1, a_2, a_3)$$

for four distinct elements $a_1, a_2, a_3, b_1 \in X$ if and only if one has $W_{T,\ell}(a_1 b_1|a_2 a_3) > 0$, and one has $\mathrm{med}_T(a_1, a_2, a_3) = \mathrm{med}_T(b_1, b_2, b_3)$ for some $a_1, a_2, a_3, b_1, b_2, b_3$ in X with $\#\{a_1, a_2, a_3\} = 3$ if and only if there exists a permutation π of the index set $\{1, 2, 3\}$ with either $a_i = b_{\pi(i)}$ or $\#\{a_1, a_2, a_3, b_{\pi(i)}\} = 4$ and $W_{T,\ell}(a_i b_{\pi(i)}|a_j a_k) > 0$ for all i, j, k in $\{1, 2, 3\}$ with $\{1, 2, 3\} = \{i, j, k\}$ in which case we must also have $\#\{b_1, b_2, b_3\} = 3$ as well as either $b_i = a_{\pi^{-1}(i)}$ or $\#\{b_1, b_2, b_3, a_{\pi^{-1}(i)}\} = 4$ and $W_{T,\ell}(b_i a_{\pi^{-1}(i)}|b_j b_k) > 0$ for all $i, j, k \in \{1, 2, 3\}$ with $\{1, 2, 3\} = \{i, j, k\}$.

In particular, we can decide whether we have $\mathrm{med}_T(a_1, a_2, a_3) = \mathrm{med}_T(b_1, b_2, b_3)$ for some $a_1, a_2, a_3, b_1, b_2, b_3$ in X with $\#\{a_1, a_2, a_3\} = 3$ from exclusively analysing the support of $W_{T,\ell}$.

vi) One can decide whether two distinct vertices u and v in T form an edge by studying medians: Indeed, given any two distinct vertices $u, v \in V$, one can choose elements $x_1, x_2, x_3, x_4 \in X$, not necessarily distinct, as indicated in the figure below:



i.e., with

$$u = \mathrm{med}_T(x_1, x_2, x_3) = \mathrm{med}_T(x_1, x_2, x_4)$$

and

$$v = \mathrm{med}_T(x_1, x_3, x_4) = \mathrm{med}_T(x_2, x_3, x_4),$$

and one has $\{u, v\} \in E(T)$ if and only if

$$\mathrm{med}_T(x_1, x_3, y) \in \big\{\mathrm{med}_T(x_1, x_2, y), \mathrm{med}_T(x_3, x_4, y), u, v\big\}$$

holds for all $y \in X$.

These well-known and easily established facts allow us to define the required map $\varphi\colon V(T) \to V(T')$: For every $x \in X$, we put $\varphi(x) := x$, and for every $v \in V(T) - X$, we choose $a_1, a_2, a_3 \in X$ with $v = \mathrm{med}_T(a_1, a_2, a_3)$ and put

$$\varphi(v) := \mathrm{med}_{T'}(a_1, a_2, a_3).$$

This is clearly well defined in view of Assertion v) above, we have $\varphi(x) = x$ for every $x \in X$ simply by definition, and we have

$$\varphi(v) = \mathrm{med}_{T'}(a_1, a_2, a_3)$$

for *all* $v \in V$ and $a_1, a_2, a_3 \in X$ with $v = \mathrm{med}_T(a_1, a_2, a_3)$ — even in case $v \in X$ because this implies that at least two of the three elements a_1, a_2, a_3 must coincide with v which in turn implies that

$$\mathrm{med}_{T'}(a_1, a_2, a_3) = v = \varphi(v)$$

must hold also in this case. Further, we have $\{\varphi(u), \varphi(v)\} \in E(T')$ for all $\{u, v\} \in E(T)$: Indeed, if $\{u, v\} \in E(T)$ holds, we can choose $x_1, x_2, x_3, x_4 \in X$ as described in Assertion vi) above and, applying φ, we get

$$\begin{aligned}\varphi(u) &= \mathrm{med}_{T'}(x_1, x_2, x_3) = \mathrm{med}_{T'}(x_1, x_2, x_4),\\ \varphi(v) &= \mathrm{med}_{T'}(x_1, x_3, x_4) = \mathrm{med}_{T'}(x_2, x_3, x_4),\end{aligned}$$

as well as

$$\begin{aligned}\mathrm{med}_{T'}(x_2,x_3,y) &= \varphi(\mathrm{med}_T(x_2,x_3,y))\\ &\in \big\{\varphi(\mathrm{med}_T(x_1,x_2,y)),\varphi(\mathrm{med}_T(x_2,x_3,y)),\varphi(u),\varphi(v)\big\}\\ &= \big\{\mathrm{med}_{T'}(x_1,x_2,y),\mathrm{med}_{T'}(x_2,x_3,y),\varphi(u),\varphi(v)\big\}\end{aligned}$$

for all $y \in X$. Hence,

$$\{\varphi(u),\varphi(v)\} \in E(T'),$$

as claimed.

It is also easy to see that any map $\varphi\colon V(T) \to V(T')$ with $\varphi(x) = x$ for all $x \in X$ and $\{\varphi(u),\varphi(v)\} \in E(T')$ for all $\{u,v\} \in E(T)$ is necessarily bijective and induces a bijection between $E(T)$ and $E(T')$ and, hence, also one between $E_1(T)$ and $E_1(T')$: Indeed, the image $\varphi(V(T))$ of $V(T)$ must contain all vertices on all paths between any two leaves in T', and the image $\big\{\{\varphi(u),\varphi(v)\}\,\big|\,\{u,v\} \in E(T)\big\}$ of $E(T)$ must contain all edges on all of those paths. Thus, the map $\varphi\colon V(T) \to V(T')$ as well as the induced map from $E(T)$ into $E(T')$ must be surjective and, hence, bijective because one has $\#V(T) = \#V(T') = 2n-2$ and $\#E(T) = \#E(T') = \#V(T) - 1 = 2n-3$ in view of the fact that both, T and T', were assumed to be binary X-trees.

Finally, we have necessarily

$$\ell(\{u,v\}) = \ell'(\{\varphi(u),\varphi(v)\})$$

for any edge $\{u,v\} \in E_1$ because, as above, we can choose $x_1,x_2,x_3,x_4 \in X$ with $u = \mathrm{med}_T(x_1,x_2,x_3) = \mathrm{med}_T(x_2,x_2,x_3)$ and $v = \mathrm{med}_T(x_2,x_3,x_4) = \mathrm{med}_T(x_1,x_3,x_4)$. Hence,

$$\ell(\{u,v\}) = W_{T,\ell}(x_1x_2\big|x_3x_4) = W_{T',\ell'}(x_1x_2\big|x_3x_4) = \ell'(\{\varphi(u),\varphi(v)\}),$$

as claimed.

It remains to observe that φ is uniquely determined by T and T': However, as observed already above, any map $\psi\colon V(T) \to V(T')$ with $\psi(x) = x$ for all $x \in X$ and $\{\psi(u),\psi(v)\} \in E(T')$ for all $\{u,v\} \in E(T)$ is necessarily bijective and induces a bijection from $E(T)$ onto $E(T')$. Thus, $d_T(x,y) = d_{T'}(x,y)$, and hence,

$$\psi(\mathrm{med}_T(x,y,z)) = \mathrm{med}_{T'}(x,y,z) = \varphi(\mathrm{med}_T(x,y,z))$$

must hold for all $x,y,z \in X$ implying that also $\psi(v) = \varphi(v)$ must hold for all $v \in V$.

3. Deriving T and ℓ from W

In this section, we will assume throughout that W is a map from $\mathcal{S}_{2|2}(X)$ into $\mathbb{R}_{\geq 0}$ that satisfies the conditions (F1) to (F4) stated above, and we want to show that a binary X-tree T and a map $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ with $W = W_{T,\ell}$ then necessarily exist.

To simplify notations, we will say that $W(ab|x|cd)$ holds for some elements a, b, c, d, x in X if and only if the four elements a, b, x, c and the four elements b, x, c, d are distinct and one has $W(ab|xc), W(bx|cd) > 0$. We will begin by collecting some technicalities regarding this quinternary relation. Note first that $W(ab|x|cd)$ implies $\#\{a, b, x, c, d\} = 5$ and

$$W(ab|cd) = W(ab|xc) + W(bx|cd) > W(ab|xc), W(bx|cd) > 0$$

in view of (F3). Hence,

$$W(ab|xc) = W(ab|xd) > 0,\ W(ax|cd) = W(bx|cd) > 0 \tag{3.1}$$

in view of (F4). This proves the implication "**(i)** $\Rightarrow$ **(ii)**" in

Lemma 3.1. *For all a, b, c, d, x in X, the following assertions are equivalent*:

- **(i)** $W(ab|x|cd)$ *holds, i.e., one has* $\#\{a, b, x, c\} = \#\{b, x, c, d\} = 4$ *and* $W(ab|xc)$, $W(bx|cd) > 0$.
- **(ii)** $\#\{a, b, x, c, d\} = 5$, $W(ab|cd) = W(ab|xc) + W(bx|cd)$, $W(ab|xd) = W(ab|xc) > 0$, *and* $W(ax|cd) = W(bx|cd) > 0$.
- **(iii)** $\#\{a, b, c, d\} = \#\{a, b, d, x\} = 4$ *and* $W(ab|cd) > W(ab|xd) > 0$.
- **(iv)** $\#\{a, b, x, c, d\} = 5$, $W(ab|cd) > 0$, $W(ab|xc) = W(ab|xd)$, *furthermore* $W(xa|dc) = W(xb|dc)$.

In particular, given any 5-subset $\{a, b, x, c, d\}$ of X, one has

$$W(ab|x|cd) \Leftrightarrow W(ba|x|cd) \Leftrightarrow W(cd|x|ab) \Leftrightarrow \cdots.$$

Proof. It is obvious that **(ii)** $\Rightarrow$ **(iii)** and **(ii)** $\Rightarrow$ **(iv)** hold.

(iii) $\Rightarrow$ **(i)**: Clearly, we must have $c \neq x$ and, hence, $\#\{a, b, x, c, d\} = 5$. If $W(bx|dc) > 0$ would not hold, we would either have $W(bc|dx) > 0$ and therefore $W(ab|c|dx)$ implying

$$W(ab|cd) > W(ab|xd) = W(ab|cd) + W(bc|xd) > W(ab|cd),$$

an obvious contradiction, or we would have $W(bd|cx) > 0$ and, hence, also $W(ab|d|cx)$ in contradiction to $W(ab|dc) \neq W(ab|dx)$. Thus, $W(ab|x|dc)$, or equivalently, $W(ab|x|cd)$ must hold, as claimed.

(iv) $\Rightarrow$ **(i)**: We must have $W(ab|xc) > 0$ because, otherwise, we would have either $W(xa|bc) > 0$ and therefore $W(xa|b|cd)$, or $W(xb|ac) > 0$ and therefore $W(xb|a|cd)$, both assertions being in contradiction to our assumption $W(xa|cd) = W(xb|cd)$. By symmetry (exchanging a, b with c, d), we must also have $W(bx|cd) > 0$ implying that $W(ab|x|cd) > 0$ must hold indeed. ∎

Corollary 3.2. *If $W(ab|cd) > 0$ and $W(ab|cd) \geq W(ax|cd), W(bx|cd)$ hold for any five distinct elements $a, b, c, d, x \in X$, one has*

$$W(ab|xc) > 0.$$

Proof. Otherwise, we could assume without loss of generality that $W(xa|bc) > 0$ holds which, together with $W(ab|cd) > 0$, would imply $W(xa|b|cd)$, and hence,

$$W(xa|cd) = W(xa|bc) + W(ab|cd) > W(ab|cd),$$

a contradiction. ∎

Corollary 3.3. *If $W(ab|xy), W(ab|yz) > 0$ holds for any five distinct elements $a, b, x, y, z \in X$, one has*

$$W(ax'|y'z') = W(bx'|y'z')$$

for all $x', y', z' \in X$ with $\{x, y, z\} = \{x', y', z'\}$.

Proof. Our assumptions imply $W(ab|xz) \geq \min\{W(ab|xy), W(ab|yz)\} > 0$. Thus, symmetry (relative to x, y, z) allows us to assume, without loss of generality, that $W(bx|yz) > 0$ holds. Together with $W(ab|xy) > 0$, this implies $W(ab|x|yz)$, and hence,

$$W(ax|yz) = W(bx|yz) > 0,$$

which in turn implies that

$$W(ax'|y'z') = W(bx'|y'z')$$

holds for all x', y', z' with $\{x', y', z'\} = \{x, y, z\}$ because both terms vanish in case $x' \neq x$, and both terms coincide with $W(ax|yz) = W(bx|yz)$ in case $x' = x$. ∎

Corollary 3.4. *If*

$$0 < W(ab|xy) \leq W(ab|xz), W(ab|yz)$$

holds for five distinct elements a, b, x, y, z in X, one has either $W(ab|x|yz)$ or $W(ab|y|xz)$ and, hence, in any case

$$W(ab|xz) = W(ab|xy) + W(ay|xz) = W(ab|xy) + W(by|xz) \tag{3.2}$$

as well as

$$W(ab|yz) = W(ab|xy) + W(ax|yz) = W(ab|xy) + W(bx|yz). \tag{3.3}$$

Proof. Clearly, both $W(ab|x|yz)$ and $W(ab|y|xz)$ imply (3.2) and (3.3). Thus, it is enough to show that either $W(bx|yz) > 0$ or $W(by|xz) > 0$ must hold. Yet, otherwise we would have $W(bz|xy) > 0$ implying that $W(ab|z|xy)$ would hold in contradiction to $W(ab|xy) \leq W(ab|xz)$. ∎

Next, we define

$$\underline{W}(ab|**) := \min\left\{ W(ab|xy) \,\middle|\, \{x,y\} \in \binom{X \setminus \{a,b\}}{2} \right\}$$

for any two distinct elements $a, b \in X$.

Note that in case the map W is of the form $W_{T,\ell}$ for some binary X-tree T and some length function $\ell\colon E_1(T) \to \mathbb{R}_{>0}$, we have $\underline{W}(ab|**) > 0$ for any two distinct vertices a and b if and only if the vertices a and b form a *cherry* in T, i.e., the two unique vertices u, v in V with $\{a, u\}, \{b, v\} \in E$ coincide.

Corollary 3.5. *If*

$$W(a_0b_0|c_0d_0) = \max\left\{W(ab|cd)\,\middle|\,ab|cd \in \mathcal{S}_{2|2}(X)\right\}$$

holds for some $a_0b_0|c_0d_0 \in \mathcal{S}_{2|2}(X)$*, one has* $\underline{W}(a_0b_0|**) > 0$ *as well as* $W(a_0x|yz) = W(b_0x|yz)$ *for all* $\{x, y, z\} \in \binom{X\setminus\{a_0,b_0\}}{3}$.

Proof. Corollary 3.2 implies that $W(a_0b_0|xc_0) > 0$ must hold for all x in $X - \{a_0, b_0, c_0\}$ which in turn implies that $W(a_0b_0|xy) > 0$ holds for all $x, y \in X - \{a_0, b_0\}$ with $x \neq y$, in view of (F4) and, therefore, also

$$W(a_0x|yz) = W(b_0x|yz)$$

for all $\{x, y, z\} \in \binom{X\setminus\{a_0,b_0\}}{3}$ in view of Corollary 3.3. ∎

Corollary 3.6. *If* $0 < W(ab|xy) = \underline{W}(ab|**)$ *holds for four distinct elements* $a, b, x, y \in X$*, one has*

$$W(ab|xz) = W(ab|xy) + W(ay|xz)$$

as well as

$$W(ab|yz) = W(ab|xy) + W(ax|yz)$$

for all $z \in (X \setminus \{a, b, x, y\})$.

Proof. This follows directly from Corollary 3.4. ∎

Next, we define

$$\overline{W}_b(a*|cd) := \max\left\{W(az|cd)\,\middle|\,z \in X \setminus \{a, b, c, d\}\right\}$$

for any four distinct elements $a, b, c, d \in X$. The following result will be crucial for our proof of Theorem 1.1:

Lemma 3.7. *If* $\underline{W}(ab|**) > 0$ *holds for two distinct elements* $a, b \in X$*, one has*

$$W(ab|cd) = \underline{W}(ab|**) + \overline{W}_b(a*|cd) \tag{3.4}$$

for any two distinct elements $c, d \in X \setminus \{a, b\}$*. In particular, a map* W *from* $\mathcal{S}_{2|2}(X)$ *into* $\mathbb{R}_{\geq 0}$ *that satisfies the conditions* (F1) *to* (F4) *is completely determined, for any two distinct elements* $a, b \in X$ *with* $\underline{W}(ab|**) > 0$*, by its values on* $\mathcal{S}_{2|2}(X \setminus a) \cup \mathcal{S}_{2|2}(X \setminus b)$ *and the value of* $\underline{W}(ab|**)$.

Proof. In case $W(ab|cd) = \underline{W}(ab|**)$, we have to show that $W(az|cd) = 0$ holds for all $z \in X \setminus \{a, b, c, d\}$ which follows from the fact that $W(az|cd) > 0$ for some $z \in X \setminus \{a, b, c, d\}$ would imply $W(ba|z|cd)$ in view of $W(ba|zc) > 0$ and $W(az|cd) > 0$ in contradiction to $W(ab|cd) = \underline{W}(ab|**) \leq W(ab|zc)$.

Otherwise, we have $W(ab|cd) > \underline{W}(ab|**)$ and we can use (F4) to find some $z \in X \setminus \{a, b, c, d\}$ with $W(ab|zc) = \underline{W}(ab|**)$ and, therefore, $W(ba|z|cd)$ in view of $W(ab|cd) > W(ab|zc) > 0$ and Lemma 3.1, **(iii)** ⇒ **(i)** ⇒ **(ii)** and, thus,

$$\begin{aligned} W(ab|cd) &= W(ab|cz) + W(az|cd) = \underline{W}(ab|**) + W(az|cd) \\ &\leq \underline{W}(ab|**) + \overline{W}_b(a*|cd). \end{aligned}$$

It remains to show that

$$W(az'|cd) \leq W(az|cd)$$

holds for all $z' \in X \setminus \{a, b, c, d\}$. Otherwise, however, we would have $W(az'|cd) > W(az|cd) > 0$ for some $z' \in X \setminus \{a, b, c, d, z\}$ and, hence, $W(az'|z|cd)$ by Lemma 3.1, **(iii)** $\Rightarrow$ **(i)** $\Rightarrow$ **(ii)** which in turn would imply $W(ba|z'|zc)$ in view of $W(az'|zc) > 0$ and $W(ba|z'z) \geq \underline{W}(ab|**) > 0$, and, hence, $W(ab|z'c) < W(ab|zc)$ in contradiction to $W(ab|zc) = \underline{W}(ab|**) \leq W(ab|z'c)$. ∎

We now turn to the remaining part of the proof of Theorem 1.1. We already showed in the previous section that there can be at most one pair T, ℓ with $W = W_{T,\ell}$. So, it remains to show that such an X-tree T and a length function ℓ indeed exist.

To this end, we will use induction relative to the cardinality n of X. Clearly, Theorem 1.1 holds in case $n = 4$. Indeed, if the elements in X are labelled a, b, c, d so that $W(ab|cd) > 0$ and, hence, $W(ac|bd) = W(ad|bc) = 0$ holds, the tree

$$\begin{aligned} T &= T_{ab|cd} \\ &:= \Big(\{a, b, c, d, u_{ab}, u_{cd}\}, \Big\{\{a, u_{ab}\}, \{b, u_{ab}\}, \{c, u_{cd}\}, \{d, u_{cd}\}, \{u_{ab}, u_{cd}\}\Big\}\Big) \end{aligned}$$

with exactly four leaves a, b, c, d and two additional vertices named u_{ab}, u_{cd} of degree 3, u_{ab} adjacent to a, b, and u_{cd}, u_{cd} adjacent to c, d, and u_{ab}, together with the map

$$\ell\colon \big\{\{u_{ab}, u_{cd}\}\big\} \to \mathbb{R}_{>0}, \quad \{u_{ab}, u_{cd}\} \mapsto W(ab|cd)$$

is obviously the unique required pair T, ℓ with $W = W_{T,\ell}$.

To perform induction, we now assume $n > 4$ and choose $a_0b_0\big|c_0d_0 \in \mathcal{S}_{2|2}(X)$ with

$$W(a_0b_0\big|c_0d_0) \geq W(ab\big|cd) \tag{3.5}$$

for all $ab\big|cd \in \mathcal{S}_{2|2}(X)$.

In view Corollary 3.5, this implies that $\underline{W}(a_0b_0|**) > 0$ as well as

$$W(a_0x\big|yz) = W(b_0x\big|yz) \tag{3.6}$$

for any three distinct elements $\{x, y, z\}$ in $X - \{a_0, b_0\}$.

Next, using our inductive hypothesis, we choose a binary $(X \setminus \{a_0\})$-tree T_1 and a length function $\ell_1\colon E_1(T_1) \to \mathbb{R}_{>0}$ with

$$W_{T_1,\ell_1} = W\big|_{\mathcal{S}_{2,2}(X-\{a_0\})}$$

and note that, in view of (3.6), we have also

$$W_{T_2,\ell_2} = W\big|_{\mathcal{S}_{2,2}(X-\{b_0\})},$$

for the binary $(X - \{b_0\})$-tree T_2 and the length function $\ell_2\colon E_1(T_2) \to \mathbb{R}_{>0}$ derived by renaming the vertex a_0 in T_1 by b_0.

Let u_0 denote the unique vertex in $V(T_1)$ with $\{u_0, b_0\} \in E(T_1)$ (and, hence, with $\{u_0, a_0\} \in E(T_2)$). It is clear that u_0 is not a leaf in either T_1 or T_2. Now, choose

some further element w_0 not in any set previously considered and define $T = (V, E)$ and $\ell\colon E_1(T) \to \mathbb{R}_{>0}$ as follows:

$$V := V(T_1) \cup \{a_0, w_0\},$$

$$E := \big\{\{a_0, w_0\}, \{b_0, w_0\}\{u_0, w_0\}\big\} \cup E(T_1) \setminus \big\{\{b_0, u_0\}\big\}.$$

Note that

$$E_1(T) = E_1(T_1) \cup \big\{\{u_0, w_0\}\big\}$$

holds. Put

$$\ell(e) = \ell_1(e)$$

for all $e \in E_1(T_1)$, and

$$\ell(\{u_0, w_0\}) := \underline{W}(a_0 b_0 | * *). \tag{3.7}$$

One has to show that $W = W_{(T,\ell)}$ holds. However, both maps coincide on $\mathcal{S}_{2|2}(X \setminus a_0) \cup \mathcal{S}_{2|2}(X \setminus b_0)$ in view of our construction, and we have also $\underline{W}_{(T,\ell)}(a_0 b_0 | * *) = \ell(\{u_0, w_0\}) = \underline{W}(a_0 b_0 | * *)$. Thus, our claim follows from Lemma 3.7. ∎

The observations leading to this proof immediately suggest various algorithms to construct the tree and to determine the length function: First one has to determine a suitable labelling $X = \{a_1, a_2, \ldots, a_n\}$ of the elements in X and then, in a second run, one builds the tree in a recursive fashion.

4. Discussion

The crucial observation used above that a map $W\colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$ which satisfies the conditions (F1)–(F4) and certain inequalities is uniquely determined by its restriction to a certain subset of $\mathcal{S}_{2|2}(X)$, raises the question for which other collections of inequalities and corresponding subsets of $\mathcal{S}_{2|2}(X)$ this might hold. E.g., one can generalize the observation above and show that, given any four distinct elements a_1, a_2, a_3, a_4 in X with

$$0 < W(a_1 a_2 | a_3 a_4) \leq W(a_1' a_2' | a_3' a_4')$$

for all $\{a_1', a_2', a_3', a_4'\} \in \binom{X}{4}$ with $W(a_1' a_2' | a_3' a_4') > 0$ and

$$\#(\{a_1, a_2, a_3, a_4\} \cap \{a_1', a_2', a_3', a_4'\}) = 3,$$

the map W is uniquely determined by its restriction to all 4-subsets $\{x_1, x_2, x_3, x_4\}$ of X for which $\{x_1, x_2, x_3, x_4\}$ is either contained in

$$A_1 := \{a_1, a_2, a_3\} \cup \Big\{a \in X \setminus \{a_1, a_2, a_3\} \Big| W(a_1 a | a_2 a_3) > 0\Big\},$$

or in

$$A_2 := \{a_1, a_2, a_3\} \cup \Big\{a \in X \setminus \{a_1, a_2, a_3\} \Big| W(a a_2 | a_1 a_3) > 0\Big\},$$

or in

$$A_3 := \{a_1, a_3, a_4\} \cup \Big\{a \in X \setminus \{a_1, a_3, a_4\} \Big| W(a_1 a_4 | a a_3) > 0\Big\},$$

or, finally, in

$$A_4 := \{a_1, a_3, a_4\} \cup \left\{ a \in X \setminus \{a_1, a_3, a_4\} \middle| W(a_1a_3|aa_4) > 0 \right\}.$$

Using this observation, the required X-tree T and length function ℓ with $W = W_{T,\ell}$ can also be constructed as follows: One first chooses two distinct elements a_1, a_2 in X for which some subset $\{x, y\} \in \binom{X \setminus \{a_1, a_2\}}{2}$ with $W(a_1a_2|xy) > 0$ exists, then one chooses two distinct elements a_3, a_4 in $X \setminus \{a_1, a_2\}$ with

$$W(a_1a_2|a_3a_4) = \min \left\{ W(a_1a_2|xy) \middle| \{x, y\} \in \binom{X \setminus \{a_1, a_2\}}{2}, W(a_1a_2|xy) > 0 \right\},$$

and observes that $W(a_1a_2|a_3a_4) \leq W(a'_1a'_2|a'_3a'_4)$ must hold for all $\{a'_1, a'_2, a'_3, a'_4\} \in \binom{X}{4}$ with $W(a'_1a'_2|a'_3a'_4) > 0$ and $\#(\{a_1, a_2, a_3, a_4\} \cap \{a'_1, a'_2, a'_3, a'_4\}) = 3$, then one constructs the subsets A_1, A_2, A_3, A_4 as above and, noting that $a_4 \notin A_1 \cup A_2$ and $a_2 \notin A_3 \cup A_4$ hold, and then one uses the induction hypothesis to find, for each $i \in \{1, 2, 3, 4\}$, an A_i-tree T_i together with a length function ℓ_i such that $W_{T_i,\ell_i} = W|_{\mathcal{S}_{2|2}(A_i)}$ holds. Finally, one "fuses" these four "small" trees in an appropriate (and absolutely canonical) way into one big supertree T and one uses the length function $\ell_1, \ell_2, \ell_3, \ell_4$ to define a length function ℓ for T for which one finally observes that $W = W_{T,\ell}$ must hold by referring to the above generalization of Corollary 3.5.

More generally, one may as well start with any arbitrary labelling

$$X = \{a_1, a_2, \ldots, a_n\}$$

of the elements in X and use the above analysis to construct recursively, starting with the tree $T_3 := (\{a_1, a_2, a_3, v\}, \{\{a_i, v\} | i = 1, 2, 3\})$, a sequence of trees $T^{(i)}$ with leave set $X_i := \{a_1, \ldots, a_i\}$ and length function ℓ_i defined on $E_1(T_i)$ for $i = 4, \ldots, n$ such that

$$W\Big|_{\mathcal{S}_{2|2}(X_i)} = W_{T_i,\ell_i}$$

holds for all $i = 4, \ldots, n$.

Indeed, comparing W-values, one can — for each $i = 4, \ldots, n$ — identify that edge $e_i = \{u_i, v_i\}$ in $T^{(i-1)}$ to which the new pending edge with leaf a_i has to be attached. The tree $T^{(i)}$ then results from $T^{(i-1)}$ by eliminating the edge e_i and adding a new internal vertex w_i as well as three new edges $\{u_i, w_i\}$, $\{w_i, v_i\}$, $\{w_i, a_i\}$, and the length function ℓ_i can then also be defined easily on the (one or two) new internal edges while keeping the value of ℓ_{i-1} on all internal edges of $T^{(i)}$ that are also internal edges of $T^{(i-1)}$.

While, given a map W that satisfies the conditions (F1) to (F4), the outcome of any such recursive construction does, of course, not depend on the labelling of X, the algorithmic procedure will selective only use certain W-values (depending strongly on the chosen labelling) and can thus be applied to any map W from $\mathcal{S}_{2|2}(X)$ into $\mathbb{R}_{\geq 0}$ whether or not (F1) to (F4) are satisfied. And it will always produce a weighted X-tree depending on that map W and the input labelling.

In a forthcoming paper, we will discuss various ideas on how to make a sensible choice of the input labelling in case one starts with a map W that satisfies the conditions (F1) to (F4) only approximately, and present some related experimental results.

Our result also suggests to study arbitrary subsets $\mathcal{X}$ of $\mathcal{S}_{2|2}(X)$ and maps $W_0 \colon \mathcal{X} \to \mathbb{R}_{\geq 0}$ and ask for necessary and/or sufficient conditions on $\mathcal{X}$ and W_0 that imply that there exists at least (or at most) one extension $W = \mathcal{S}_{2|2}(X) \to \mathbb{R}_{>0}$ of W_0 that satisfies the conditions ($\mathcal{X}$) as well as perhaps certain inequalities, or for algorithms that decide extendability and/or construct such an extension if it exists. The results by Boecker and others (cf. [2–4]) suggest that deciding unique extendability might, at least in certain cases, be considerably simpler than just deciding extendability.

Another question that arises naturally in this context is how, given any map $W \colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$, one can find a map $W' \colon \mathcal{S}_{2|2}(X) \to \mathbb{R}_{\geq 0}$ that satisfies the conditions (F1)–(F4) and approximates W as closely as possible (relative to some predefined measure of "closeness"). While prescribing the support of W' (i.e., the topology of the X-tree in question), least square approximations should be easy, a linear-programming approach (similar to that pursued by Weyer-Menkhoff [40], see also [24]) in the case of unweighted X-trees where only the support of W' is of interest) would be welcome whenever any a priori assumptions about that support cannot be provided.

References

1. H.-J. Bandelt and A. Dress, Reconstructing the shape of a tree from observed dissimilarity data, Adv. Appl. Math. **7** (1986) 309–343.
2. S. Böcker, From subtrees to supertrees, Ph.D. Thesis, Universität Bielefeld, 1999, pp. 1–100.
3. S. Böcker, A.W.M. Dress, and M.A. Steel, Patching up X-trees, Ann. Combin. **3** (1999) 1–12.
4. S. Böcker, D. Bryant, A.W.M. Dress, and M.A. Steel, Algorithmic aspects of tree amalgamation, J. Algorithm **37** (2000) 522–537.
5. H. Colonius and H.H. Schultze, Trees constructed from empirical relations, Braunschweiger Berichte aus dem Institut fuer Psychologie **1** (1977).
6. H. Colonius and H.H. Schultze, Tree structure for proximity data, British J. Math. Statist. Psych. **34** (1981) 167–180.
7. J.H. Badger and P. Kearney, Picking fruit from the tree of life, In: Proc. 16th ACM Symp. Appl. Comput., Las Vegas, March 11–14, 2001, pp. 61–67.
8. A. Ben-Dor, B. Chor, D. Graur, R. Ophir, and D. Pelleg, Constructing phylogenies from quartets: elucidation of Eutherian superordinal relationships, J. Comput. Biol. **5** (3) (1998) 377–390.
9. V. Berry, T. Jiang, P. Kearney, M. Li, and T. Wareham, Quartet cleaning: improved algorithms and simulations, In: Algorithms — ESA'99, 7th European Symposium on Algorithms Prague, Chezh Rep. Lect. Notes Comput. Sci., Vol. 1643, 1999, pp. 313–324.
10. V. Berry and O. Gascuel, Inferring evolutionary trees with strong combinatorial evidence, Theoret. Comput. Sci. **240** (2000) 271–298.
11. V. Berry, D. Bryant, T. Jiang, P. Kearney, M. Li, T. Wareham, and H. Zhang, A practical algorithm for recovering the best supported edges of an evolutionary tree (extended abstract), In: ACM Symp. on Discrete Algorithms SODA2000, 2000, pp. 287–296.
12. O. Bininda-Emonds, S.G. Brady, J. Kim, and M.J. Sanderson, Scaling of accuracy in extremely large phylogenetic trees, In: 6th Pacific Symp. on Biocomputing, 2001, pp. 547–558.
13. D.J. Bryant and M.A. Steel, Extension operations on sets of leaf-labelled trees, Adv. Appl. Math. **16** (1995) 425–453.

14. D. Bryant and M. Steel, Fast algorithms for constructing optimal trees from quartets, In: Proc. Tenth Annual ACM-SIAM Symposium on Discrete Algorithms, Baltimore, Maryland, 1999, pp. 147–155.
15. P. Buneman, The recovery of trees from measures of dissimilarity, In: Mathematics in the Archaeological and Historical Sciences, F.R. Hodson, D.G. Kendall, and P. Tautu, Eds., Edinburgh University Press, Edinburgh, 1971, pp. 387–395.
16. B. Chor, Form quartets to phylogenetic trees, In: SOFSEM'98: Theory and Practice of Informatics, B. Rovan, Ed., Lecture Notes in Computer Science, Vol. 1521, Springer-Verlag, 1998, pp. 36–53.
17. M. Csűrös and M-Y. Kao, Provable and accurate recovery of evolutionary trees through harmonic greedy triplets, SIAM J. Comput. **31** (2001) 306–322.
18. M. Csűrös, Fast recovery of evolutionary trees with thousands of nodes, J. Comput. Biol. **9** (2002) 277–297.
19. M.C.H. Dekker, Reconstruction methods for derivation trees, Master's Thesis, Vrije Universiteit, Amsterdam, 1986.
20. A. Dress, M. Hendy, K. Huber, and V. Moulton, Enumerating the vertices of the Buneman graph, Preprints Forschungsschwerpunkt Mathematisierung/Strukturbildungsprozesse, **117**, 1997.
21. P.L. Erdős, M.A. Steel, L.A. Székely, and T. Warnow, Local quartet splits of a binary tree infer all quartet splits via one dyadic inference rule, Comput. Artificial Intelligence **16** (2) (1997) 217–227.
22. P.L. Erdős, M.A. Steel, L.A. Székely, and T. Warnow, Inferring big trees from short quartets, In: Automata, Languages and Programming 24th International Colloquium, ICALP'97, Bologna, Italy, July 7–11, 1997, P. Degano, R. Gorrieri, A. Marchetti-Spaccamela, Eds., Lecture Notes in Computer Science, Vol. 1256, 1997, pp. 827–837.
23. J. Gramm and R. Niedermeier, Minimum quartet inconsistency is fixed parameter tractable, In: Combinatorial Pattern Matching, CPM2001, A. Amir and G.M. Landau Eds., Israel, Jerusalem, LNCS 2089, 2001, pp. 241–256.
24. S. Grünewald, The quartet joining algorithm, manuscript, Bielefeld, 2002.
25. D. Huson, S. Nettles, L. Parida, T. Warnow, and S. Yooseph, The disk-covering method for tree reconstruction, In: Proceedings of "Algorithms and Experiments," ALEX'98, Trento, Italy, 1998, pp. 62–75.
26. D. Huson, S. Nettles, K. Rice, T. Warnow, and S. Yooseph, Hybrid tree reconstruction methods, ACM J. Exp. Alg. **4** (1998) Article 5.
27. D.H. Huson, S.M. Nettles, and T.J. Warnow, Disk-covering, a fast-converging method for phylogenetic tree reconstruction, J. Comput. Biol. **6** (3/4) (1999) 369–386.
28. T. Jiang, P. Kearney, and M. Li, Orchestrating quartets: approximation and data correction, FOCS'98 Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science, 1998, pp. 416–425.
29. T. Jiang, P. Kearney, and M. Li, A polynomial time approximation scheme for inferring evolutionary trees from quartet topologies and its application, SIAM J. Comput. **30** (2000) 1942–1961.
30. P.E. Kearney, The ordinal quartet method (extended abstract), In: RECOMB'98, New York, 1998, pp. 125–133.
31. J. Kim, large-scale phylogenies and measuring the performance of phylogenetic estimators, Syst. Biol. **47** (1998) 43–60.
32. J. Lagergren, Combining polynomial running time and fast convergence for the disk-covering method, J. Comput. System Sci. **65** (2002) 481–493.

33. L. Nakhleh, U. Roshan, K.St. John, J. Sun, and T. Warnow, Designing fast converging phylogenetic methods, In: Bioinformatics, Oxford University Press, ISMB'01 **17** (90001), 2001, S190–S198.
34. V. Ranwez and O. Gascuel, Quartet based phylogenetic inference: improvements and limits, Mol. Biol. Evol. **18** (6) (2001) 1103–1116.
35. K. Strimmer and A. von Haeseler, Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies, Mol. Biol. Evol. **13** (1996) 964–969.
36. K. Strimmer, N. Goldman, and A. von Haeseler, Bayesian probabilities and quartet puzzling, Mol. Biol. Evol. **14** (1997) 210–211.
37. M.A. Steel, L.A. Székely, and P.L. Erdős, The number of nucleotide sites needed to accurately reconstruct large evolutionary trees, DIMACS Technical Report 1996–19.
38. G.D. Vedova and H.T. Wareham, Optimal algorithms for local vertex quartet cleaning, Bioinformatics **18** (2002) 1297–1304.
39. T. Warnow, B.M.E. Moret, and K.St. John, Absolute convergence: true trees from short sequences, In: ACM Symp. on Discrete Algorithms SODA'01, 2001, pp. 186–195.
40. J. Weyer-Menkhoff, Phylogenetic Combinatorics, Ph.D. Thesis, Bielefeld, 2003.