The Tree of Life and Other Affine Buildings

Andreas Dress, Werner Terhalle

In this note, we discuss some mathematics which has proven to be of use in the analysis of molecular evolution – and, actually, was discovered in this context (cf. [D]).

According to evolutionary theory, the spectrum of present-day species (or biomolecules) arose from their common ancestors according to a well-defined scheme of bi-(or multi-)furcation steps. The task of phylogenetic analysis as defined by E. Haeckel is to unravel that scheme by comparing systematically all data available regarding present and extinct species. This task has been simplified enormously in recent years through the availability of molecular sequence data, first used for that purpose by W. Fitch and E. Margoliash in their landmark paper from 1967 dealing with Cytochrome C sequences [FM]. The basic idea in that field is that species (or molecules) which appear to be closely related should have diverged more recently than species which appear to be less closely related.

A standard formalization is to measure relatedness by a metric defined on the set of species (or molecules) in question. The task then is to construct an ($\mathbb{R}$-)tree which represents the metric (and hence the bifurcation scheme) as closely as possible. Below, we discuss necessary and sufficient conditions for the existence of such a tree that represents the metric *exactly*, as well as some constructions which lead to that tree if those conditions are fulfilled, and to more or also less treelike structures if not. Remarkably, the theory we developed in this context allowed also to view affine buildings (which in the rank 1 case are $\mathbb{R}$-trees) from a new perspective.

Here are some basic definitions and results:

Definition 1: *Given a non-empty set E, an integer $m \geq 2$, and a map*

$$v : E^m \to \{-\infty\} \cup \mathbb{R},$$

the pair (E, v) is called a valuated matroid of rank *m if the following properties hold:*

(VM0) for every $e \in E$, there exist some $e_2, \dots, e_m \in E$ such that

$$v(e, e_2, \dots, e_m) \neq -\infty,$$

(VM1) v is totally symmetric,

(VM2) for $e_1, \dots, e_m \in E$ with $\#\{e_1, \dots, e_m\} < m$, one has

$$v(e_1, \dots, e_m) = -\infty,$$

(VM3) for all $e_1, \ldots, e_m, f_1, \ldots, f_m \in E$*, one has*

$$\begin{aligned} & v(e_1, \ldots, e_m) + v(f_1, \ldots, f_m) \\ \leq \; & \max_{1 \leq i \leq m} \{v(f_1, e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_m) + v(e_i, f_2, \ldots, f_m)\}. \end{aligned}$$

Condition (VM3) is also called the VALUATED EXCHANGE PROPERTY.

If $\{b_1, \ldots, b_m\} \subseteq E$ *satisfies* $v(b_1, \ldots, b_m) \neq -\infty$*, then* $\{b_1, \ldots, b_m\}$ *is called a* BASE *of the valuated matroid* (E, v).

Note that (VM3) implies the bases exchange property of ordinary matroids for the set $B_{(E,v)}$ of bases of (E, v).

Here is a "generic" example:
Let K be a field with a non-archimedean valuation $w : K \to \{-\infty\} \cup \mathbb{R}$, that is a map satisfying the conditions

$$w(x) = \infty \iff x = 0,$$

$$w(x \cdot y) = w(x) + w(y),$$

and

$$w(x + y) \leq \max\{w(x), w(y)\}$$

for all $x, y \in K$; then – in view of the GRASSMANN-PLÜCKER identity

$$\begin{aligned} & \det(e_1, \ldots, e_m) \cdot \det(f_1, \ldots, f_m) \\ = \; & \sum_{i=1}^{m} \det(e_1, \ldots, e_{i-1}, f_1, e_{i+1}, \ldots, e_m) \cdot \det(e_i, f_2, \ldots, f_m) \end{aligned}$$

$(e_1, \ldots, e_m, f_1, \ldots, f_m \in K^m)$ – the pair $(K^m \setminus \{0\}, w \circ \det)$ is a valuated matroid of rank m.

DEFINITION 2: *Given a valuated matroid* (E, v) *of rank* m*, we put*

$$T_{(E,v)} := \{p : E \to \mathbb{R} \mid \forall e \in E : p(e) = \max_{e_2, \ldots, e_m \in E} \{v(e, e_2, \ldots, e_m) - \sum_{i=2}^{m} p(e_i)\}\}.$$

$T_{(E,v)}$ *is also called the* tight span *of* (E, v) *or its* T-CONSTRUCTION.

The following proposition details this set of maps:

PROPOSITION 1: *Let* $H := \{(t_1, \ldots, t_m) \in \mathbb{R}^m \mid \sum_{i=1}^{m} t_i = 0\}$*. Then, for every base* $\{b_1, \ldots, b_m\} \in B_{(E,v)}$ *of a valuated matroid* (E, v) *of rank* m*, the map* $\Phi_{b_1, \ldots, b_m} : H \to \mathbb{R}^E$ *which maps each* $(t_1, \ldots, t_m) \in H$ *to the map*

$$E \to \mathbb{R} : e \mapsto \max_{1 \leq i \leq m} \{v(e, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_m) + t_i\} - \tfrac{m-1}{m} v(b_1, \ldots, b_m)$$

is an injective map into $T_{(E,v)}$.

Furthermore, one has

$$T_{(E,v)} = \bigcup_{\{b_1,\ldots,b_m\}\in B_{(E,v)}} \Phi_{b_1,\ldots,b_m}(H).$$

Thus, $T_{(E,v)}$ is a union of (images of) affine hyperplanes of dimension $m-1$, called the *apartments* in $T_{(E,v)}$.

These apartments intersect as follows:

Proposition 2:

1) *Given two bases $B, B' \subseteq E$ of a valuated matroid (E,v) of rank m, with suitable orderings of their elements as $B = \{b_1,\ldots,b_m\}$ and $B' = \{b'_1,\ldots,b'_m\}$, resp., one has*

$$\Phi_{b_1,\ldots,b_m}(H) \cap \Phi_{b'_1,\ldots,b'_m}(H) = \bigcap_{i=0}^{m} \Phi_{b_1,\ldots,b_i,b'_{i+1},\ldots,b'_m}(H).$$

2) *Given a base $\{b_1,\ldots,b_m\} \in B_{(E,v)}$, an element $b_0 \in E \setminus \{b_1,\ldots,b_m\}$, and a subset $I \subseteq \{1,\ldots,m\}$ so that $\{b_0,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_m\}$ is a base if and only if $i \in I$, then one has*

$$\begin{aligned} &\Phi^{-1}(\Phi_{b_1,\ldots,b_m}(H) \cap \Phi_{b_0,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_m}(H)) \\ = \ &\{(t_1,\ldots,t_m) \in H \ \mid \ t_i + v(b_0,b_1,\ldots,b_{i-1},b_{i+1},\ldots,b_m) = \\ &\qquad \max_{j\in I}\{t_j + v(b_0,b_1,\ldots,b_{j-1},b_{j+1},\ldots,b_m)\}\} \end{aligned}$$

for every $i \in I$.

We return to our generic example mentioned above, that is, to the valuated matroid $(E := K^m \setminus \{0\}, v = w \circ \det)$, with K a field with a non-archimedean valuation $w : K \twoheadrightarrow \{-\infty\} \cup \mathbb{Z}$. By $\Gamma_{\mathbb{Z}}$, we denote the group of all affine maps from H to itself consisting of a translation by an integer vector and a permutation of coordinates, that is,

$$\begin{aligned} \Gamma_{\mathbb{Z}} := \{\gamma : H \to H \ \big| &(t_1,\ldots,t_m) \mapsto (t_{\sigma(1)} + a_1,\ldots,t_{\sigma(m)} + a_m) \text{ for some} \\ &(a_1,\ldots,a_m) \in H \cap \mathbb{Z}^m \text{ and some } \sigma \in S_m\}. \end{aligned}$$

Every subset

$$\begin{aligned} C = \{\Phi_{b_1,\ldots,b_m} \circ \gamma(t_1,\ldots,t_m) \ \mid \ &(t_1,\ldots,t_m) \in H \text{ with} \\ &t_1 \le t_2 \le \ldots \le t_m \le t_1 + 1\} \end{aligned}$$

with $\{b_1,\ldots,b_m\} \in B_{(E,v)}$ some base and $\gamma \in \Gamma_{\mathbb{Z}}$ is called a chamber of $T_{(E,v)}$; in case $\{b_1,\ldots,b_m\}$ is the canonical base of the vector space K^m and γ equals id_H, the resulting chamber C_0 is called the fundamental chamber, while the

apartment $A_0 = \Phi_{b_1,\ldots,b_m}(H)$ for the canonical base is called the FUNDAMENTAL APARTMENT.
If a map p in $T_{(E,v)}$ satisfies $p(e) \equiv \frac{i}{m} \mod 1$ for some $i \in \{0,\ldots,m-1\}$ and every $e \in E$, then p is called a VERTEX of $T_{(E,v)}$ (OF TYPE i).
It is easy to see that the general linear group $GL_m(K)$ acts transitively on the set of vertices of $T_{(E,v)}$ via its group action defined on $T_{(E,v)}$ by

$$\begin{array}{rcl} GL_m(K) \times T_{(E,v)} & \to & T_{(E,v)} : \\ (X,p) & \mapsto & (E \to \mathbb{R} : e \mapsto p(X^{-1}e) + \frac{1}{m} w \circ \det(X)) \end{array}$$

$(X \in GL_m(K), p \in T_{(E,v)}, e \in E)$.
This action induces a transitive action of the group $SL_m(K)$ on the set of apartments as well as on the set of chambers of $T_{(E,v)}$; since the stabilizers of these actions give rise to a BN-pair in the sense of building theory, one has

THEOREM 1: *For the valuated matroid* $(E = K^m \setminus \{0\}, v = w \circ \det)$ *with* K *a field with a non-archimedean valuation* $w : K \twoheadrightarrow \{-\infty\} \cup \mathbb{Z}$*, the* T*-construction* $T_{(E,v)}$ *is a geometrical realization of the affine building defined for the group* $GL_m(K)$.

Now, we come back to the general case of an arbitrary valuated matroid (E, v) of rank m.

LEMMA 1: *For every* $p \in T_{(E,v)}$*, the map*

$$\begin{array}{l} d_p : E \times E \to \mathbb{R} \\ \qquad (e,f) \mapsto e^{\sup\{v(e,f,e_3,\ldots,e_m) - p(e) - p(f) - \sum_{i=3}^{m} p(e_i) \mid e_3,\ldots,e_m \in E\}} \end{array}$$

(with $e^{-\infty} := 0$*) is a (pseudo-ultra-)metric on* E*.*
In addition, for any two maps p *and* q *in* $T_{(E,v)}$*, the metrics* d_p *and* d_q *are topologically equivalent.*

DEFINITION 3: *A valuated matroid* (E, v) *of rank* m *is called* COMPLETE *if, for some (or equivalently: for every)* $p \in T_{(E,v)}$*, the metric space* (E, d_p) *is complete.*

Up to "projective equivalence" and identifying "parallel elements" (we refer to [DT1] for details), one has

THEOREM 2: *Every valuated matroid has an (essentially unique) completion.*

(In fact, one only has to complete (E, d_p) to a metric space $(\hat{E}, \hat{d})$ and then to define $\hat{v} : \hat{E}^m \to \{-\infty\} \cup \mathbb{R}$ as the continuous extension of v.)

Concerning the T-construction, one has the following result:

THEOREM 3: *Let* $(\hat{E}, \hat{v})$ *be a completion of the valuated matroid* (E, v) *with* $\hat{E} \supseteq E$*. Then the restriction map from* $T_{(\hat{E},\hat{v})} \subseteq \mathbb{R}^{\hat{E}}$ *to* $\mathbb{R}^E$*, mapping every* $p \in T_{(\hat{E},\hat{v})}$ *to* $p\big|_E$*, is a bijection into* $T_{(E,v)}$*.*

From now on, we assume for simplicity (E, v) to be a complete valuated matroid of rank m.

DEFINITION 4: *An* END *of* $T_{(E,v)}$ *is a map* ε *from* $T_{(E,v)}$ *to* $\mathbb{R}$ *satisfying*

(E1) for every base $\{b_1, \ldots, b_m\}$, there exist some $r \in \{1, \ldots, m\}$, some affine map $\gamma : H \to H$ with a coordinate permutation as linear component, and some $c \in \mathbb{R}$ such that, for every $(t_1, \ldots, t_m) \in H$, the equation

$$\varepsilon \circ \Phi_{b_1,\ldots,b_m} \circ \gamma(t_1, \ldots, t_m) = \max_{1 \le i \le r} t_i + c$$

holds;

(E2) there exist some base $\{b_1, \ldots, b_m\}$ and some $c \in \mathbb{R}$ such that, for every $(t_1, \ldots, t_m) \in H$,

$$\varepsilon \circ \Phi_{b_1,\ldots,b_m}(t_1, \ldots, t_m) = t_1 + c.$$

The set of all ends of $T_{(E,v)}$ will be denoted by $\mathcal{E}_{T_{(E,v)}}$.

With this definition, one has

PROPOSITION 3: *For every $e \in E$, the map*

$$\begin{aligned} \varepsilon_e : T_{(E,v)} &\to \mathbb{R} \\ p &\mapsto p(e) \end{aligned}$$

is an end of $T_{(E,v)}$.
And, for every $\varepsilon \in \mathcal{E}_{T_{(E,v)}}$, there exist some $e \in E$ and some $c \in \mathbb{R}$ such that $\varepsilon = \varepsilon_e + c$.

And one has

THEOREM 4: *If one defines a map w from the set $\mathcal{E}^m_{T_{(E,v)}}$ of m-tupels of ends of (E, v) to $\{-\infty\} \cup \mathbb{R}$ by*

$$w(\varepsilon_1, \ldots, \varepsilon_m) := \inf_{p \in T_{(E,v)}} \sum_{i=1}^{m} \varepsilon_i(p)$$

for $\varepsilon_1, \ldots, \varepsilon_m \in \mathcal{E}_{T_{(E,v)}}$, then one has

$$w(\varepsilon_{e_1}, \ldots, \varepsilon_{e_m}) = v(e_1, \ldots, e_m)$$

for all $e_1, \ldots, e_m \in E$. That is, $(\mathcal{E}_{T_{(E,v)}}, w)$ is a complete valuated matroid of rank m which – up to "parallel elements" – is isomorphic to (E, v).

We now restrict ourselves to the case that the rank m equals 2. Here, $T_{(E,v)}$ is a *path-infinite* $\mathbb{R}$-tree, that is an $\mathbb{R}$-tree being the union of isometric images of the real line – namely the apartments from above: for any two $p, q \in T_{(E,v)}$, there exists some base $\{b_1, b_2\}$ such that $p, q \in \Phi_{b_1,b_2}(H)$, say $p = \Phi_{b_1,b_2}((s, -s))$ and $q = \Phi_{b_1,b_2}((t, -t))$ for some $s, t \in \mathbb{R}$; then putting $d(p, q) := |s - t|$ leads to a (well-defined) metric on $T_{(E,v)}$ having the desired property.
And the ends of $T_{(E,v)}$ in our sense correspond to its ends in the way ends are defined for $\mathbb{R}$-trees, that is, they correspond to (equivalence classes of) isometric embeddings of real halflines into $T_{(E,v)}$.

An example which we found particularly intriguing is the following one: Let E denote the set of subsets of $\mathbb{R}$ which are bounded from above, and for $e, f \in E$, let $v(e,f) := \sup(e \triangle f)$ be the supremum of their symmetric difference. Then it is easy to see that (E, v) is a valuated matroid of rank 2. The corresponding $\mathbb{R}$-tree has the particular property that omitting any point leads to the same "number" of connected components, and this number equals $\#\mathfrak{P}(\mathbb{R})$, the cardinality of the powerset of $\mathbb{R}$.

Now, it is well-known that, for the metric d of an $\mathbb{R}$-tree T, the so-called *four-point condition*

$$d(x,y) + d(z,w) \le \max \left\{ \begin{array}{c} d(x,z) + d(y,w), \\ d(x,w) + d(y,z) \end{array} \right\}$$

holds for all $x, y, z, w \in T$. But this four-point condition is literally the exchange property (VM3) in the rank 2 case! Of course, one has $d(x,x) = 0$ instead of $d(x,x) = -\infty$ (cf. (VM2)).

This observation led us to the definition of matroidal trees:

Definition 5: *A* matroidal tree *or, for short,* matree*, is a pair (X, u) consisting of a non-empty set X together with a map $u : X \times X \to \{-\infty\} \cup \mathbb{R}$ satisfying the following three conditions:*

(MT0) for every $x \in X$, there exists some $y \in X$ with $u(x,y) \neq -\infty$,

(MT1) u is symmetric,

(MT2) for all $x_1, x_2, y_1, y_2 \in X$, one has

$$u(x_1,x_2) + u(y_1,y_2) \le \max \left\{ \begin{array}{c} u(y_1,x_2) + u(x_1,y_2), \\ u(y_1,x_1) + u(x_2,y_2) \end{array} \right\}$$

(and no restriction on the diagonal corresponding to (VM2)).

Note that, for every matree (X, u), the restriction $u \;\Big|_{\{x \in X \mid u(x,x)=0\}^2}$ is a (pseudo)metric.

Now, let's have a look at the set

$$\begin{aligned} H_{(X,u)} :=\{f : X \to \{-\infty\} \cup \mathbb{R} \mid f(x) + u(y,z) \le \\ \max \left\{ \begin{array}{c} f(y) + u(x,z), \\ f(z) + u(x,y) \end{array} \right\} \text{ for all } x, y, z \in X, f \not\equiv -\infty\}, \end{aligned}$$

the set of all *one-point extensions* of a matree (X, u) (containing at least all maps

$$\begin{aligned} h_a : X &\to \{-\infty\} \cup \mathbb{R} : \\ x &\mapsto u(a,x) \end{aligned}$$

for $a \in X$).

If one wants to make a new matree $(H_{(X,u)}, w)$ from this set, and one wants the map $w : H_{(X,u)} \times H_{(X,u)} \to \{-\infty\} \cup \mathbb{R}$ to satisfy $w(h_x, h_y) = u(x,y)$ for all $x, y \in X$

(in order to have an "homomorphism" $X \to H_{(X,u)} : x \mapsto h_x$), and, slightly more general, $w(f, h_x) = f(x)$ for every $f \in H_{(X,u)}$ and every $x \in X$, then w necessarily has to satify

$$w(f,g) + u(x,y) \le \max \left\{ \begin{array}{c} f(x) + g(y), \\ f(y) + g(x) \end{array} \right\}$$

for all $f, g \in H_{(X,u)}$ and all $x, y \in X$.
And, indeed, one has

Theorem 5: *If, for a matree (X, u) and for $H_{(X,u)}$ as above, one defines*

$$\begin{aligned} w := w_{(X,u)} : H_{(X,u)} \times H_{(X,u)} &\to \{-\infty\} \cup \mathbb{R} \\ (f,g) &\mapsto \inf_{x,y \in X} \left\{ \max \left\{ \begin{array}{c} f(x) + g(y), \\ f(y) + g(x) \end{array} \right\} - u(x,y) \right\} \end{aligned}$$

(with the convention $(-\infty) - (-\infty) := +\infty$), then $(H_{(X,u)}, w)$ is again a matree. In addition, for every $f \in H_{(X,u)}$ and every $x \in X$, one has

$$w(f, h_x) = f(x)$$

– in particular, one has $w(h_x, h_y) = u(x, y)$ for all $x, y \in X$.

The matree $(H_{(X,u)}, w)$ can be seen as a "hull" of (X, u), as one has

Theorem 6: *If, for $F \in H_{(H_{(X,u)}, w_{(X,u)})}$, one defines*

$$\begin{aligned} \varphi(F) : X &\to \{-\infty\} \cup \mathbb{R} \\ x &\mapsto F(h_x), \end{aligned}$$

and, for $f \in H_{(X,u)}$,

$$\begin{aligned} \psi(f) : H_{(X,u)} &\to \{-\infty\} \cup \mathbb{R} \\ g &\mapsto w_{(X,u)}(f,g), \end{aligned}$$

then φ is a bijective map from $H_{(H_{(X,u)}, w_{(X,u)})}$ to $H_{(X,u)}$, and ψ is a bijective map in the other direction; both maps are inverse to each other; and for all $f, g \in H_{(X,u)}$, one has

$$w_{(H_{(X,u)}, w_{(X,u)})}(\psi(f), \psi(g)) = w(f,g).$$

Thus, $(H_{(H_{(X,u)}, w_{(X,u)})}, w_{(H_{(X,u)}, w_{(X,u)})})$ and $(H_{(X,u)}, w_{(X,u)})$ are canonically isomorphic matrees.

In addition, one has

Theorem 7: *$H_{(X,u)}$ is the smallest set of maps $X \to \{-\infty\} \cup \mathbb{R}$ that a) contains $\{h_x \mid x \in X\}$ and b) is closed under addition of constants, under suprema, and under limites.*
More precisely: for every $f \in H_{(X,u)}$, one of the following three possibilities hold:

(i) there exist some $x \in X$ and some $c \in \mathbb{R}$ such that $f = h_x + c$,

(ii) there exist some $x, y \in X$ and some $b, c \in \mathbb{R}$ such that

$$f = \max\{h_x + d, h_y + c\},$$

(iii) there exist sequences $(x_n)_{n\in\mathbb{N}}$ in X and $(c_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ such that

$$f = \lim_{n\to\infty}(h_{x_n} + c_n).$$

Essential for the study of matrees is the following

FUNDAMENTAL LEMMA: *Let (X, u) be a matree; for $x, y \in X$ with $u(x, y) \neq -\infty$, put*

$$s_{x,y} := \frac{1}{2}(u(y,y) - u(x,y)) \in \{-\infty\} \cup \mathbb{R},$$
$$s^{x,y} := \frac{1}{2}(u(x,y) - u(x,x)) \in \mathbb{R} \cup \{+\infty\},$$

and $I(x,y) := [s_{x,y}, s^{x,y}] \cap \mathbb{R}$; for $t \in \mathbb{R}$, define $h^t \in H_{(X,u)}$ by

$$h^t := \max\{h_x + t, h_y - t\} - \frac{1}{2}xy.$$

Then the map $I(x,y) \to H_{(X,u)} : t \mapsto h^t$ is a surjective isometry onto the set

$$\{f \in H_{(X,u)} \mid w(h_x, h_y) = w(f, h_x) + w(f, h_y) \text{ and } w(f,f) = 0\}$$

– with isometry meaning that $w(h^s, h^t) = |s - t|$ holds for all $s, t \in I(x,y)$.

COROLLARY: *The set*

$$\{f \in H_{(X,u)} \mid w(f,f) = 0\}$$

is connected; hence – since the restriction of w to it is a metric satisfying the four-point condition – it is an $\mathbb{R}$-tree relative to the restriction of w (cf. [D]).

We want to close this section by a short discussion on the relationship between $H_{(E,v)}$ and $T_{(E,v)}$ for a valuated matroid (E, v) of rank 2.
For this, let

$$T'_{(E,v)} := \{p : E \to \mathbb{R} \mid p(e) = \sup_{f\in E}\{v(e,f) - p(f)\} \text{ for every } e \in E\}$$

– note the "sup" instead of "max" as for $T_{(E,v)}$; and define the canonical metric d on $T'_{(E,v)}$ by $d(p,q) := \sup_{e\in E}|p(e) - q(e)|$. It is easy to see that $T'_{(E,v)}$ is the completion of $T_{(E,v)}$ relative to this metric.
One should remark that $T'_{(E,v)}$ is the set of all minimal elements in the polytope

$$P_{(E,v)} := \{p : E \to \mathbb{R} \mid p(e) + p(f) \geq v(e,f) \text{ for all } e, f \in E\}$$

relative to the order $p \leq q :\iff p(e) \leq q(e)$ for every $e \in E$.

Coming back to the comparison of $H_{(E,v)}$ with $T'_{(E,v)}$, the following holds: The maps $p \in T'_{(E,v)}$ are exactly those maps in $H_{(E,v)}$ satisfying

$$w_{(E,v)}(p,p) = 0,$$

and one has

$$d = w_{(X,u)} \Big|_{\{p \in H_{(E,v)} | w_{(E,v)}(p,p)=0\}^2} .$$

Slightly more general, one has

$$\{p \in H_{(E,v)} \,\big|\, w_{(E,v)}(p,p) \neq -\infty\} = \{p + c \,\big|\, p \in T'_{(E,v)}, c \in \mathbb{R}\}.$$

And the maps $p \in H_{(E,v)}$ satisfying $w_{(E,v)}(p,p) = -\infty$ correspond to the ends of the $\mathbb{R}$-tree $T_{(E,v)}$.

Based on these considerations, an algorithm for analyzing distance data and for constructing phylogenetic trees if those data fit exactly into trees and *phylogenetic networks* based on the T-construction if the data do not fit into a tree has been developed jointly with D. Huson and others which is available via `http://bibiserv.techfak.uni-bielefeld.de/splits/` where also further references can be found.

References

[BD1] H.-J. Bandelt, A. W. M. Dress: *Reconstructing the Shape of a Tree from Observed Dissimilarity Data*, Advances in Applied Mathematics 7 (1986), 309-343.

[BD2] H.-J. Bandelt, A. W. M. Dress: *A canonical decomposition theory for metrics on a finite set*, Advances in Mathematics 92 (1992), 47-105.

[BD3] H.-J. Bandelt, A. W. M. Dress: *Split Decomposition: A New and Useful Approach to Phylogenetic Analysis of Distance Data*, Molecular Phylogenetics and Evolution 1, No. 3 (1992), 242-252.

[D] A. W. M. Dress: *Trees, Tight Extensions of Metric Spaces, and the Cohomological Dimension of Certain Groups: A Note on Combinatorial Properties of Metric Spaces*, Advances in Mathematics 53 (1984), 321-402.

[DDH] A. W. M. Dress, J. Dopazo, A. v. Haeseler: *Split decomposition: a new technique to analyse viral evolution*, PNAS 90 (1993), 10320-10324.

[DEW1] A. W. M. Dress, M. Eigen, R. Winkler-Oswatitsch: *Statistical geometry in sequence space: A method of quantitative comparative sequence analysis*, Proc. Natl. Acad. Sci. USA 85 (1988), 5913-5917.

[DEW2] A. W. M. Dress, M. Eigen, R. Winkler-Oswatitsch: *How old is the genetic code?*, Science, 244 (1989), 673-679.

[DMT] A. W. M. Dress, V. L. Moulton, W. F. Terhalle: *T-Theory — An Overview*, European Journal of Combinatorics 17 (1996) 161-175.

[DT1] A. W. M. Dress, W. F. Terhalle: *A Combinatorial Approach to $\wp$-adic Geometry, Part I: The Process of Completion*, Geometriae Dedicata 46 (1993), 127-148.

[DT2] A. W. M. Dress, W. F. Terhalle: *The Real Tree*, Advances in Mathematics 120 (1996) 283-301.

[DW] A. W. M. Dress, R. Wetzel: *The Human Organism – a Place to Thrive for the Immuno-Deficiency Virus*, in: New Approaches in Classification and Data Analysis, ed. E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, B. Burtschy, Springer-Verlag, 1994, 636-643.

[FM] W. M. Fitch, E. Margoliash: *The Construction of Phylogenetic Trees — A Generally Applicable Method Utilizing Estimates of the Mutation Distance Obtained from Cytochrome c Sequences*, Science 155 (1967), 279-284.

[T1] W. F. Terhalle: *Ein kombinatorischer Zugang zu $\wp$-adischer Geometrie: Bewertete Matroide, Bäume und Gebäude*, Dissertation, Universität Bielefeld, 1992.

[T2] W. F. Terhalle: *Coordinatizing $\mathbb{R}$-Trees in Terms of Universal c-Trees*, Annals of Combinatorics 1 (1997), 183-196.

[T3] W. F. Terhalle: *$\mathbb{R}$-Trees and Symmetric Differences of Sets*, European Journal of Combinatorics 18 (1997), 825-833.

[T4] W. F. Terhalle: *Matroidal Trees. A Unifying Theory of Treelike Spaces and Their Ends*, Habilitationsschrift, Universität Bielefeld, submitted.

[WWW1] The www page THE TREE OF LIFE (http://phylogeny.arizona.edu/tree/phylogeny.html) provides uptodate information regarding our present knowledge of (or believe in) the detailed branching structure of species evolution.

[WWW2] The www page SPLITSTREE 2 (http://bibiserv.techfak.uni-bielefeld.de/splits) allows to use or to download the SplitsTree algorithm via the net and to view what that algorithm does to an appropriately specified data input.

[WWW2] The www page COMPUTATIONAL MOLECULAR EVOLUTION (http://dexter.gnets.ncsu.edu/lab/moleevol.html) provides links to (almost) all www pages relevant in this field, from data banks to algorithms (incl. SplitsTree 2) to journals.

Andreas Dress, Werner Terhalle
FSP Mathematisierung
Universität Bielefeld
Postfach 10 01 31
D-33501 Bielefeld
dress@mathematik.uni-bielefeld.de
terhalle@mathematik.uni-bielefeld.de