THE CATEGORY OF X-NETS

ANDREAS DRESS

Address:
Department of Combinatorics and Geometry,
CAS-MPG Partner Institute for Computational Biology,
Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences,
320 Yue Yang Road, Shanghai, China, Postcode: 200031,
and Max Planck Institute for Mathematics in the Sciences,
Inselstrasse 22 -26, D 04103 Leipzig, Germany

email: dress@sibs.ac.cn and dress@mis.mpg.de

Short title: X**-Nets**

Abstract. While, in current discussions, networks are most often described in terms of (more or less ornamented) graphs, in this report, we prefer to describe networks in terms of *metric spaces.* The reason for this is that the concepts and results to be presented here using metric spaces as the basic notion are of some use in the context of phylogenetic analysis where the *lengths* of edges customarily used in pictorial representations of results are not only highly informative as they indicate the presumed time span of evolutionary phases under investigation, they often are already quite essential for deriving such results algorithmically by searching for results that provide edge lengths which are "optimally" adapted to the given data. More specifically, the theory of X-nets described here in terms of metric spaces can be used as a natural framework for taxonomic analysis in terms of *phylogenetic networks* in analogy to the framework offered by the theory of X-trees supporting taxonomic analysis in terms of (the much more familiar) phylogenetic trees.

Keywords and Phrases:
X-tree, X-net, X-split, X-split system, Buneman complex, metric space, L_1-space, graph, hypercube, hypercuboid, path metric, isometric subgraph, median graph

Thanks:
I thank the Center for Combinatorics, Nankai University, for supporting this research, the organizers of the Conference on "Networks: From Biology To Theory" for allowing me to present these results, and the Sino-German Centre in Beijing for sponsoring this meeting. A paper including all missing proofs will be published jointly with members of the Center for Combinatorics at Nankai University.

Date: 6/11/2005.

1. **Introduction**

While, in current discussions, networks are most often described in terms of (more or less ornamented) graphs, in this report on recent work done at the Center for Combinatorics at Nankai University, we prefer to describe networks in terms of *metric spaces*.

The reason for this is that the concepts and results to be presented here using metric spaces as the basic notion are of some use in the context of phylogenetic analysis where the *length* of edges customarily used in pictorial representations of results are not only highly informative as they indicate the presumed time span of evolutionary phases under investigation, they often are already quite essential for *deriving* such results algorithmically by searching for results that provide edge lengths which are "optimally" adapted to the given data.

More specifically, the *category of X-nets* described here in terms of metric spaces can be used as a natural framework for taxonomic analysis in terms of *phylogenetic networks* (cf. [**?**, **?**, **?**, **?**, **?**]) in analogy to the framework offered by the *theory of X-trees* supporting taxonomic analysis in terms of (the much more familiar) *phylogenetic trees* (cf. [**?**, **?**]):

$$\begin{tikzcd}
\text{phylogenetic trees} \arrow[r] \arrow[d] & X\text{-trees} \arrow[d] \\
\text{phylogenetic networks} \arrow[r] & X\text{-nets}
\end{tikzcd}$$

In the next section, we will present some basic terminology. In Section 3, we collect some basic results (whose proofs need still to be written up, in publishable form, jointly with members of the Center for Combinatorics at Nankai University). In Section 4, it will be discussed how X-nets are "classified" by $\mathbb{R}$-valued split systems, and in the last section, some relevant examples are presented, extending from phylogenetic trees and networks to nets related to (i) subjectively perceived similarity of colours, (ii) geographic data from a road atlas, and (iii) the structural relatedness of world languages.

2. **Basic Terminology**

In this note, we consider finite metric spaces $\mathbf{M} = (M, D)$, i.e., pairs consisting of a finite set M and a (proper) metric[1] $D : M \times M \to \mathbb{R} : (u, v) \mapsto uv$ defined on M.

Given any two points u, v in M, we define the interval $[u, v]$ spanned by u and v by

$$[u, v] = [u, v]_D := \{w \in M : uv = uw + wv\},$$

we write $u \leq_w v$ for some $w \in M$ whenever $u \in [w, v]$ holds and note that the binary relation "$\leq_u$" is a partial order of M for every $u \in M$, and we define a binary relation "$\|$" on M^2 by putting

$$uu' \| vv' \iff_{\text{def}} u', v \in [u, v'] \text{ and } u, v' \in [u', v]$$

for all pairs $(u, u'), (v, v')$ in M^2.

Next, the L_1-*product* of any two metric spaces $\mathbf{M} = (M, D)$ and $\mathbf{M'} = (M', D')$, denoted by $\mathbf{M} \times \mathbf{M'}$, is defined to be the metric space

$$\mathbf{M} \times \mathbf{M'} := (M \times M', D \oplus D')$$

whose point set is the cartesian product $M \times M'$ of the point sets M and M' of the two given spaces, and whose metric $D \oplus D'$ is defined by putting

$$(D \oplus D')\big((u, u'), (v, v')\big) := D(u, v) + D'(u', v')$$

for all $(u, u'), (v, v') \in M \times M'$ — note that the k-dimensional *standard* L_1-space $\mathbf{L_1(k)} := (\mathbb{R}^k, L_1^{(k)})$ whose metric $L_1^{(k)} : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ is given by

$$L_1^{(k)}(\mathbf{x}, \mathbf{y}) := \|\mathbf{x}, \mathbf{y}\|_1 := \sum_{i=1}^{k} |x_i - y_i|$$

for all $\mathbf{x} := (x_1, \ldots, x_k), \mathbf{y} := (y_1, \ldots, y_k) \in \mathbb{R}^k$ is nothing but the L_1-product of k copies of the space $\mathbf{L_1(1)}$, that is, the real line (endowed with its standard metric).

Further, given any metric space $\mathbf{M} = (M, D : V \times V \to \mathbb{R} : (u, v) \mapsto uv)$,

(1) we define $\mathbf{M}$ to be an (abstract) L_1-*space* if the relation "$\|$" is an equivalence relation on M^2 — implying that the L_1-product of any two abstract L_1-spaces and, therefore, also (as good sense would require) the standard L_1-spaces $\mathbf{L_1(k)}$ are abstract L_1-spaces (as the real line is easily seen to be

---

[1] Recall that a metric or, more precisely, a *proper* metric defined on a set M is a bivariate map $D : M \times M \to \mathbb{R} : (u, v) \mapsto uv$ such that $uv = 0 \iff u = v$ and $uv + vw \geq wu$ — and, therefore, also $uv = vu \geq 0$ — holds for all $u, v, w \in M$. Recall also that, according to J. Isbell (cf. [**?**]), (i) the most appropriate way of defining *The Category of Metric Spaces* denoted by `MET` is to define, for any two metric spaces $\mathbf{M} = (M, D)$ and $\mathbf{M'} = (M', D')$, the set of morphisms from $\mathbf{M}$ into $\mathbf{M'}$ to consist of the set of all *non-expansive* maps from M into M', i.e., all maps $\varphi : M \to M'$ with $D'(\varphi(u), \varphi(v)) \leq D(u, v)$ for all $u, v \in M$, — with composition of morphisms defined in the obvious way, (ii) there is a canonical class of monomorphisms in that category — including all isomorphisms — which are the "isometric embeddings", i.e., the maps $\varphi : M \to M'$ from a metric space $\mathbf{M} = (M, D)$ into a metric space $\mathbf{M'} = (M', D')$ for which $D'(\varphi(u), \varphi(v)) = D(u, v)$ holds for all $u, v \in M$, and that (iii) two metric spaces are called *isometric* if they are isomorphic objects in that category.

one: indeed, one has $uu'\|vv'$ for any two pairs $(u,u'),(v,v')$ of real numbers relative to $L_1^{(1)}$ if and only if either $\#\{u,u'\} = \#\{v,v'\} = 1$ or and $(u,v) = (u',v')$ holds),

(2) we define two elements $u, v \in M$ to be forming a *primitive pair* (in $\mathbf{M}$) if and only if $\#[u,v] = 2$ — or, equivalently, $u \neq v$ and $\{u,v\} = [u,v]$ — holds, and we denote the set of all primitive pairs in $\mathbf{M}$ by $\mathrm{Prim}(\mathbf{M})$, i.e., we put

$$\mathrm{Prim}(\mathbf{M}) := \{\{u,v\} \subseteq M : \#[u,v] = 2\},$$

(3) we define a sequence $a_0, a_1, a_2, \dots, a_k$ of points in M to be
   (i) a *geodesic* sequence (in $\mathbf{M}$) if the identity $a_0a_k = \sum_{i=1}^{k} a_{i-1}a_i$ holds (in which case — even stronger — $a_ia_j = \sum_{\ell=i+1}^{j} a_{\ell-1}a_\ell$ holds for all $i, j \in \{0,1,2,\dots,k\}$ with $i < j$),
   (ii) a *path* (in $\mathbf{M}$) if all pairs $\{a_{i-1}, a_i\}$ $(i = 1,2,\dots,k)$ are primitive pairs in $\mathbf{M}$,
   (iii) and, of course, a *geodesic path* (in $\mathbf{M}$) if it is a geodesic sequence that is, simultaneously, a path in $\mathbf{M}$,

(4) we put

$$\mathbf{M}(u < v) := \{w \in M : u \leq_w v\} \quad (= \{w \in M : u \in [w,v]\})$$

for all $u, v \in M$, we define two elements $u, v \in M$ to be forming a *bipartioning pair* (in $\mathbf{M}$) if $M = \mathbf{M}(u < v) \cup \mathbf{M}(v < u)$ holds — implying that a subset $\{u,v\} \subseteq V$ is a bipartioning pair if and only if $u \neq v$ holds and $\{u,v\}$ is a *gated* subset of $\mathbf{M}$ (i.e., if and only if $\#\{u,v\} = 2$ holds and there exists, for every $x \in M$, some point $y \in \{u,v\}$ — the *gate* of x in $\{u,v\}$ — with $xw = xy + yw$ for each element $w \in \{u,v\}$), and that every bipartioning pair must also be a primitive pair,

(5) we define $\mathbf{M}$ to be *bipartite* if, conversely, every primitive pair in $\mathbf{M}$ is also a bipartioning pair,

(6) and we recall that $\mathbf{M}$ is said to be a *median* space if and only if one has $\#([u,v] \cap [v,w] \cap [w,u]) = 1$ for all $u, v, w \in M$ in which case the single element in $[u,v] \cap [v,w] \cap [w,u]$ is denoted by $\mathrm{med}(u,v,w) = \mathrm{med}_D(u,v,w)$ and dubbed *the* median of u, v, and w — note that the L_1-product of any two median spaces and, therefore, also the standard L_1-spaces $\mathbf{L_1(k)}$ are median spaces (as the real line $\mathbf{L_1(1)}$ is a median space, the median $\mathrm{med}(x,y,z)$ of any three real numbers x, y, z in $\mathbf{L_1(1)}$ being given by $\mathrm{med}(x,y,z) = x + y + z - \max(x,y,z) - \min(x,y,z)$).

Remarks: (R1) Note that the primitive pairs in a finite metric space correspond to the edges in a connected finite graph. More precisely, given any finite metric space $\mathbf{M} = (M, D)$, we can associate to $\mathbf{M}$ the necessarily finite and connected simple graph

$$G_{\mathbf{M}} := (M, \mathrm{Prim}(\mathbf{M}))$$

with vertex set M and edge set $\mathrm{Prim}(\mathbf{M}) \subseteq \binom{M}{2}$ and, conversely, to any finite and connected simple graph $G = (V, E)$ with vertex set V and edge set $E \subseteq \binom{V}{2}$, the finite metric space $\mathbf{M}_G := (V, D_G)$ with point set V whose metric D_G is the standard graph metric on V, i.e., the (well-defined and unique) largest metric D defined on V for which $D(u, v) \leq 1$ holds for every edge $\{u, v\}$ in E.

This yields in particular a canonical one-to-one correspondence between

(i) the isometry classes of finite metric spaces $\mathbf{M}$ for which $D(u, v) = 1$ holds for every primitive pair $\{u, v\}$ in $\mathrm{Prim}(\mathbf{M})$ and
(ii) the isomorphism classes of finite and connected simple graphs.

Recall also that a subgraph $G' = (V', E')$ of a finite and connected simple graph $G = (V, E)$ is called an *isometric* subgraph of G if it is connected and $D_G(u', v') = D'_{G'}(u', v')$ holds for all $u', v' \in V'$.

(R2) More generally, given a *weighted* finite and connected simple graph, i.e., a triple $\mathbf{G} = (V, E; L)$ consisting of a finite and connected simple graph $G = (V, E)$ together with an edge weighting $L : E \rightarrow \mathbb{R}_{>0} := \{\rho \in \mathbb{R} : \rho > 0\}$, we may associate to L the unique largest metric $D = D_L$ defined on V for which $D(u, v) \leq L(\{u, v\})$ holds for every edge $\{u, v\}$ in E, this way setting up a canonical one-to-one correspondence between the isometry classes of finite metric spaces $\mathbf{M} = (M, D)$ and the isomorphism classes of weighted finite and connected simple graphs $\mathbf{G} = (V, E; L)$ for which

$$L(\{u, v\}) < \sum_{i=1}^{k} L(\{u_{i-1}, u_i\})$$

holds for every edge $\{u, v\}$ in E and all sequences $u_0, u, \ldots, u_k$ of vertices in V of length $k > 1$ with $u_0 = u, u_k = v$, and $\{u_{i-1}, u_i\} \in E$ for all $i = 1, \ldots, k$ (or, equivalently, for which $D_L(u, v) < D_L(u, w) + D_L(w, v)$ holds for every edge $\{u, v\}$ in E and every $w \in V - \{u, v\}$). Indeed, if $\mathbf{M} = (M, D)$ is a finite metric space, the edge weighting $L = L_{\mathbf{M}}$ defined on the set $\mathrm{Prim}(\mathbf{M})$ of edges of the associated graph $G_{\mathbf{M}}$ by $L_{\mathbf{M}} : \mathrm{Prim}(\mathbf{M}) \rightarrow \mathbb{R} : \{u, v\} \mapsto D(u, v)$ satisfies the above condition and one has $D = D_{L_{\mathbf{M}}}$ while conversely, if an edge weighting L defined on the set E of edges of a finite and connected simple graph $G = (V, E)$ satisfies this condition, one has $G = G_{(\mathbf{M}_L)}$ for the associated finite metric space $\mathbf{M}_L = \mathbf{M}_{(G,L)} := (V, D_L)$ while L coincides with the corresponding edge weighting $L_{(\mathbf{M}_L)}$.

(R3) Note that, given a finite and connected simple graph $G = (V, E)$, the above condition holds for *every* edge weighting $L : E \rightarrow \mathbb{R}_{>0}$ if and only if G is a tree. The resulting metrics D_L will be called T-metrics, and the resulting metric spaces $\mathbf{M}_L$ will be called T-spaces. Consequently, a finite metric space $\mathbf{M} = (M, D)$ is a T-space if and only if the associated graph $G_{\mathbf{M}} = (M, \mathrm{Prim}(\mathbf{M}))$ is a tree in which case the metric D_L induced by the edge weighting

$$L := L_{\mathbf{M}} : \mathrm{Prim}(\mathbf{M}) \rightarrow \mathbb{R} : \{u, v\} \mapsto D(u, v)$$

coincides with D.

(R4) Note also that a connected finite simple graph G is a bipartite or a median graph if and only if the associated metric space $\mathbf{M}_G$ is a bipartite or a median metric space, respectively, and that every T-space is a bipartite as well as a median metric space.

Finally, we'll need the following definitions:

(7) A finite metric space $\mathbf{H}$ is called a *hypercuboid* if it is isometric to the L_1-product $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_k$ of a finite number of metric spaces all of whose point sets have cardinality 2 — implying that every hypercuboid is a median L_1-space (as any metric space of cardinality 2 is such a space), that any hypercuboid derived from k factors can be embedded isometrically into the k-dimensional standard L_1-space $\mathbf{L_1(k)}$, and that the graph $G_{\mathbf{H}}$ associated to a hypercuboid $\mathbf{H}$ is a hypercube, i.e., it is isomorphic to a graph of the form $(\{0,1\}^k, E^{(k)})$ with $E^{(k)}$ defined by

$$E^{(k)} := \{\{(x_1, \ldots, x_k), (y_1, \ldots, y_k)\} \subseteq \{0,1\}^k : \sum_{i=1}^{k} |x_i - y_i| = 1\},$$

i.e., the graph associated to the subspace of the standard L_1-space $\mathbf{L_1(k)}$ whose point set is $\{0,1\}^k$.

(8) Another, yet equivalent way to describe hypercuboids is to associate, to any weighted set $\mathbf{E}$, i.e., to a pair $\mathbf{E} := (E, L)$ consisting of a finite set E and a map $L : E \rightarrow \mathbb{R}_{>0}$, the metric space $\mathbf{M}^{\mathbf{E}} := (\mathcal{P}(E), D^{\mathbf{E}})$ whose point set is the power set $\mathcal{P}(E)$ of E while its metric $D^{\mathbf{E}}$ is given by the map

$$D^{\mathbf{E}} : \mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathbb{R} : (F, F') \mapsto L_+(F \triangle F')$$

(where, as usual, $F \triangle F'$ denotes the *symmetric difference* $(F - F') \cup (F' - F)$ of the two subsets F and F' of E, and $L_+(F)$ denotes, for a weighted set $\mathbf{E} = (E, L)$ and a subset F of E, the sum $L_+(F) := \sum_{e \in F} L(e)$), and then to define a finite metric space $\mathbf{H}$ to be a hypercuboid if it is isometric to a metric space of that form, i.e., if a weighted set $\mathbf{E}$ as above exists so that $\mathbf{H}$ is isometric to $\mathbf{M}^{\mathbf{E}}$ (as $\mathbf{M}^{\mathbf{E}}$ is apparently isometric to the L_1-product $\Pi_{e \in E}(\{0,1\}, D_e)$ where $D_e : \{0,1\} \times \{0,1\} \rightarrow \mathbb{R}$ is, of course, defined by $D_e(0,1) = D_e(1,0) := L(e)$ and $D_e(0,0) = D_e(1,1) := 0$). So, $\mathbf{M}^{\mathbf{E}} = (\mathcal{P}(E), D^{\mathbf{E}})$ must, in particular, be a median space — and it is indeed also easily verified directly that the median of any three subsets F, F', F'' of E, considered as points in the point set $\mathcal{P}(E)$ of $\mathbf{M}^{\mathbf{E}}$, always exists, and always coincides with the subset $(F \cap F') \cup (F' \cap F'') \cup (F'' \cap F)$, independently of the choice of L.

(9) A *net* $\mathbf{N}$ is a metric space $\mathbf{N} = (N, D)$ with point set N and metric D that can be embedded into a hypercuboid $\mathbf{M}$ so that any two points in $\mathbf{N}$ can be connected by a geodesic path $a_0, a_1, a_2, \ldots, a_k$ in $\mathbf{M}$ all of whose points are points in $\mathbf{N}$.

(10) We define *the category* `NET` *of nets* to be the category whose objects are the nets while the morphisms from one net $\mathbf{N} = (N, D)$ into another net

$\mathbf{N}' = (N', D')$ are defined to be exactly those morphisms from $\mathbf{N}$ into another net $\mathbf{N}'$ in the category `MET` (i.e., those non-expansive maps φ from N into N') that are *additive*, i.e., one has

$$D'(\varphi(u), \varphi(v)) + D'(\varphi(v), \varphi(w)) = D'(\varphi(u), \varphi(w))$$

for all $u, v, w \in V$ with $uv + vw = uw$ (or, equivalently, $\varphi(w) \in [\varphi(u), \varphi(v)]$ holds for all $u, v, w \in N$ with $w \in [u, v]$), and for which

$$\{\varphi(u), \varphi(v)\} \in \mathrm{Prim}(V', D')$$

holds for every primitive pair $\{u, v\}$ in $\mathrm{Prim}(V, D)$ with $\varphi(u) \neq \varphi(v)$. And any such morphism φ is called an *isometric embedding* if φ, considered as a morphism in `MET`, is an isometric embedding, i.e., if $D'(\varphi(u), \varphi(v)) = D(u, v)$ holds for all $u, v \in N$.

(11) Given a finite set X, we define an X-net $\mathcal{N}$ to be a pair $\mathcal{N} := (\mathbf{N}, \psi)$ consisting of a net $\mathbf{N} = (N, D)$ together with a map $\psi : X \rightarrow N$ such that

$$\psi(X) \cap \mathbf{N}(u < v) \cap \mathbf{N}(u' < v') \neq \emptyset$$

holds for all primitive pairs $\{u, v\}, \{u', v'\}$ in $\mathbf{N}$ for which the intersection $\mathbf{N}(u < v) \cap \mathbf{N}(u' < v')$ is non-empty.

(12) And we define *the category* `X—NET` *of X-nets* to be the category whose objects are the X-nets while the morphisms from one X-net $\mathcal{N} = (\mathbf{N}, \psi)$ into another X-net $\mathcal{N}' = (\mathbf{N}', \psi')$ are those morphisms φ from $\mathbf{N}$ into $\mathbf{N}'$ in `NET` for which $\psi' = \varphi \circ \psi$ holds. As above, any such morphism will also be called an *isometric embedding* if it is one, considered as a morphism in `NET` (or, equivalently, in `MET`).

3. **Some Basic Results**

Clearly, the definition of a net given above is rather a "descriptive" or "constructive" than a structural or "intrinsic" definition. However, the other definitions collected above allow us to present the following seven characterizations of nets two of which are "intrinsic":

Theorem 3.1. *Given a finite metric space $\mathbf{M} = (M, D)$, the following assertions all are equivalent:*

(i) $\mathbf{M}$ *is a net,*
(ii) $\mathbf{M}$ *is a bipartite L_1-space,*
(iii) $\mathbf{M}$ *is bipartite and the relation "$\|$" defined — by abuse of notation — on* $\mathrm{Prim}(\mathbf{M})$ *by putting*

$$\{u, u'\} \| \{v, v'\} \iff_{\mathrm{def}} uu' \| vv' \text{ or } uu' \| v'v$$

for all $\{u, u'\}, \{v, v'\} \in \mathrm{Prim}(\mathbf{M})$ *is an equivalence relation on* $\mathrm{Prim}(\mathbf{M})$,
(iv) *the graph* $G_{\mathbf{M}} = (M, \mathrm{Prim}(\mathbf{M}))$ *is an isometric subgraph of a hypercube, and $uu' = vv'$ holds for all u, u', v, v' in M for which $\{u, u'\}, \{v, v'\}$ are parallel edges in that hypercube,*

(v) *there exists a pair* $(\mathbf{E}, \Delta)$ *consisting of a weighted finite set* $\mathbf{E} = (E, L)$ *and a map* $\Delta : M^2 \rightarrow \mathcal{P}(E)$ *with*

$$\mathrm{Prim}(\mathbf{M}) = \{\{u, v\} \subseteq M : \#\Delta(u, v) = 1\}$$

such that

$$\Delta(u, v) = \Delta(u, w) \bigtriangleup \Delta(w, v)$$

and

$$D(u, v) = L_+(\Delta(u, v))$$

holds for all $u, v, w \in M$ *in which case*

(a) *the map*

$$\psi_v : M \rightarrow \mathcal{P}(E) : u \mapsto \Delta(u, v)$$

is an isometry from $\mathbf{M}$ *into the metric space* $\mathbf{M}^{\mathbf{E}}$ *for given any point* $v \in M$,

(b) $\Delta(u, v) = \Delta(v, u)$ *and* "$\Delta(u, v) = \emptyset \Longleftrightarrow u = v$" *holds for all* u, v *in* M,

(c) *and* $uu' \| vv'$ *holds for some* u, u', v, v' *in* M *if and only if*

$$\Delta(u, u') = \Delta(v, v') \text{ and } \Delta(u, u') \cap \Delta(u', v') = \emptyset$$

and, hence, also

$$\Delta(u, v) = \Delta(u, v') \bigtriangleup \Delta(v', v) = \Delta(v', u) \bigtriangleup \Delta(u, u') = \Delta(u', v')$$

as well as

$$\Delta(u, v') = \Delta(u, u') \cup \Delta(u', v') = \Delta(u', u) \cup \Delta(u, v) = \Delta(u', v)$$

holds,

(vi) *there exists some* $k \in \mathbb{N}$ *and an isometric embedding* φ *of* $\mathbf{M}$ *into the standard* k*-dimensional* L_1*-space* $\mathbf{L_1(k)}$ *such that*

$$\mathrm{med}_{\mathbf{L_1(k)}}(\varphi(u), \varphi(v), \varphi(w)) \in \{\varphi(u), \varphi(v)\}$$

holds for all $u, v, w \in M$ *with* $\{u, v\} \in \mathrm{Prim}(\mathbf{M})$,

(vii) *there exists an isometric embedding* φ *of* $\mathbf{M}$ *into some median* L_1*-space* $\mathbf{M}'$ *with such that* $\mathrm{med}_{\mathbf{M}'}(\varphi(u), \varphi(v), \varphi(w)) \in \{\varphi(u), \varphi(v)\}$ *holds for all* $u, v, w \in M$ *with* $\{u, v\}$ *in* $\mathrm{Prim}(\mathbf{M})$.

To establish Theorem **??**, the following more detailed results are required:

Theorem 3.2. *A path* $a_0, a_1, \ldots, a_k$ *in a finite bipartite metric space* $\mathbf{M}$ *is a geodesic path if and only if* $\{a_{i-1}, a_i\} \| \{a_{j-1}, a_j\}$ *implies* $i = j$ *for all* $i, j = 1, \ldots, k$. *Furthermore, if* $a_0, a_1, \ldots, a_k$ *is a geodesic path in* $\mathbf{M}$, *one has* $k' \geq k$ *for any other path* $a'_0, a'_1, \ldots, a'_{k'}$ *of points in* $\mathbf{M}$ *with* $a_0 = a'_0$ *and* $a_k = a'_{k'}$ *while equality* $k = k'$ *holds if and only if the path* $a'_0, a'_1, \ldots, a'_{k'}$ *is also a geodesic path in which case there exists a permutation* π *of the index set* $\{1, \ldots, k\}$ *such that* $a_{i-1}a_i = a'_{\pi(i)-1}a'_{\pi(i)}$ *holds for all* $i = 1, \ldots, k$.

Theorem 3.3. *If* $\mathbf{M} = (M, D)$ *is a finite bipartite metric space for which the binary relation* "$\|$" *defined above on* $\mathrm{Prim}(\mathbf{M})$ *is an equivalence relation on* $\mathrm{Prim}(\mathbf{M})$, *one has* $\{u, v\} \| \{u', v'\}$ *for two primitive pairs* $\{u, v\}, \{u', v'\}$ *in* $\mathrm{Prim}(\mathbf{M})$ *if and only if the two bipartitions* $\{\mathbf{M}(u < v), \mathbf{M}(v < u)\}$ *and* $\{\mathbf{M}(u' < v'), \mathbf{M}(v' < u')\}$ *of the point set* M *of* $\mathbf{M}$ *associated with* $\{u, v\}$ *and* $\{u', v'\}$ *coincide.*

Thus, denoting the set of "$\|$"-equivalence classes in $\mathrm{Prim}(\mathbf{M})$ by $\mathrm{E}(\mathbf{M})$ and associating to any path $a_0, a_1, \dots, a_k$ in $\mathbf{M}$ the set

$$\Delta(a_0, a_1, \dots, a_k) := \{e \in \mathrm{E}(\mathbf{M}) : \#\{i \in \{1, \dots, k\} : \{a_{i-1}, a_i\} \in e\} \equiv 1 \mod 2\}$$

of "$\|$"-equivalence classes e in $\mathrm{E}(\mathbf{M})$ represented by an odd number of pairs of the form $\{a_{i-1}, a_i\}$ $(i = 1, \dots, k)$, the following can be established:

Theorem 3.4. *If $\mathbf{M}$ is a finite bipartite metric space for which the binary relation "$\|$" defined above on* $\mathrm{Prim}(\mathbf{M})$ *is an equivalence relation on* $\mathrm{Prim}(\mathbf{M})$*, a path $a_0, a_1, \dots, a_k$ in $\mathbf{M}$ is a geodesic path in $\mathbf{M}$ if and only if the cardinality of set $\Delta(a_0, a_1, \dots, a_k)$ coincides with k in which case one has $\Delta(a_0, a_1, \dots, a_k) = \Delta(a'_0, a'_1, \dots, a'_{k'})$ for any other path $a'_0, a'_1, \dots, a'_{k'}$ in $\mathbf{M}$ with $a_0 = a'_0$ if and only if $a_k = a'_{k'}$ holds — allowing us to (well-)define the map $\Delta : V^2 \rightarrow \mathcal{P}(\mathrm{E}(\mathbf{M}))$ as described in* Theorem **??** (v) *by letting $\Delta(u, v)$ denote the set $\Delta(a_0, a_1, \dots, a_k)$ for one — or, as well, for all — paths $a_0 := u, a_1, \dots, a_k; = v$ from u to v in $\mathbf{M}$*

Together, these results can be used to establish

Theorem 3.5. (i) *For any two X-nets $\mathcal{N}$ and $\mathcal{N}'$, there exists at most one morphism in* `X—NET` *from $\mathcal{N}$ into $\mathcal{N}'$.*
(ii) *Whenever a (necessarily unique) morphism φ from an X-net $\mathcal{N} = (\mathbf{N}, \psi)$ into an X-net $\mathcal{N}' = (\mathbf{N}', \psi')$ exists, this morphism is an isometric embedding if and only if ψ and ψ' induce the same metric on X, i.e., if and only if*

$$D(\psi(x), \psi(y)) = D'(\psi'(x), \psi'(y))$$

holds for all $x, y \in X$.
(iii) *For every X-net $\mathcal{N}$, there exists an X-net $\mathcal{N}^*$, also called the* injective hull *of $\mathcal{N}$, together with an isometric embedding $\varphi_{\mathcal{N}}$ from $\mathcal{N}$ into $\mathcal{N}^*$ such that, for every isometric embedding φ of $\mathcal{N}$ into another X-net $\mathcal{N}'$, there exists a (necessarily unique) isometric embedding φ' from $\mathcal{N}'$ into $\mathcal{N}^*$ with $\varphi_{\mathcal{N}} = \varphi' \circ \varphi$ (implying, as usual, that both, $\mathcal{N}^*$ and $\varphi_{\mathcal{N}}$ are uniquely determined up to canonical isomorphism by $\mathcal{N}$, and that $\mathcal{N}^*$ is also the injective hull of every X-net $\mathcal{N}'$ for which an isometric embedding from $\mathcal{N}$ into $\mathcal{N}'$ exists).*
Moreover, any morphism φ in `X—NET` *from an X-net $\mathcal{N}_1$ into an X-net $\mathcal{N}_2$ induces a morphism φ^* in* `X—NET` *from the injective hull $\mathcal{N}_1^*$ of $\mathcal{N}_1$ into the injective hull $\mathcal{N}_2^*$ of $\mathcal{N}_2$.*
(iv) *And, given an X-net $\mathcal{N} = (\mathbf{N}, \psi)$, the underlying metric space N of $\mathcal{N}$ is a median metric space if and only if every isometric embedding from $\mathcal{N}$ into any other X-net $\mathcal{N}'$ is an isomorphism if and only if the morphism $\varphi_{\mathcal{N}} : \mathcal{N} \rightarrow \mathcal{N}^*$ is an isomorphism, that is, if and only if $\mathcal{N}$ is its own injective hull (implying that the underlying metric space of the injective hull $\mathcal{N}^*$ of any X-net $\mathcal{N}$ is a median metric space).*

Corollary 3.6. *Every T-space is a net while a pair $(\mathbf{N}, \psi)$ consisting of T-space $\mathbf{N} = (N, D)$ and a map $\psi : X \rightarrow N$ from X into the point set N of $\mathbf{N}$ is an X-net if and only if the tree $G_{\mathbf{N}}$ together with the map ψ from X into its vertex set N is an X-tree, i.e., if and only if every vertex in the graph*

$$G_{\mathbf{N}} = (N, \mathrm{Prim}(\mathbf{N}))$$

of degree less than 3 is contained in the image of ψ.

4. X-nets and Split Systems over X

Now, recall that given any finite set X, one denotes

- by $\mathcal{S}(X)$ the collection
$$\mathcal{S}(X) := \{\{A, B\} : A, B \subseteq X, A \cup B = X, A \cap B = \emptyset\}$$
of all X-*splits*,
- by $S(x)$, for every X-split $S = \{A, B\} \in \mathcal{S}(X)$ and every element $x \in X$, that subset, A or B, in S that contains x,
- by $\mathcal{S}^*(X)$ the collection
$$\mathcal{S}^*(X) := \{\{A, B\} : A, B \subseteq X, A \cup B = X, A \cap B = \emptyset \neq A, B\}$$
of all *bipartitions* of X, or *proper X-splits*,
- and by $\mathcal{S}^*(X|\,\mathbb{R})$ the $\mathbb{R}$-vectorspace consisting of all maps μ from $\mathcal{S}(X)$ into $\mathbb{R}$ with $\mu(\{X, \emptyset\} = 0$, that is, all maps μ from $\mathcal{S}(X)$ into $\mathbb{R}$ whose support
$$\operatorname{supp}(\mu) := \{S \in \mathcal{S}(X) : \mu(S) \neq 0\}$$
is contained in $\mathcal{S}^*(X)$.

Any such map μ will also be called an ($\mathbb{R}$-weighted) split system over X, and it will be called an $\mathbb{R}_{\geq 0}$-weighted split system over X if $\mu(S) \geq 0$ holds for all $S \in \mathcal{S}(X)$.

There is a close connection between X-nets and $\mathbb{R}_{\geq 0}$-weighted split systems over X. To explicate this, note first that, given a finite set X and an X-net $\mathcal{N} = (\mathbf{N}, \psi)$, one can associate, to any primitive pair $\{u, v\} \in \operatorname{Prim}(\mathbf{N})$, the corresponding X-split
$$S_{u,v} = S^{\mathcal{N}}_{u,v} := \{X(u < v), X(v < u)\}$$
whose two parts $X(u < v)$ and $X(v < u)$ are the pre-images (relative to ψ) of the two parts of the split $\{\mathbf{N}(u < v), \mathbf{N}(v < u)\}$ associated to the pair $\{u, v\}$ in $\operatorname{Prim}(\mathbf{N})$, i.e., the two subsets
$$X(u < v) = X^{\mathcal{N}}(u < v) := \{x \in X : \psi(x) \in \mathbf{N}(u < v)\}$$
and
$$X(v < u) = X^{\mathcal{N}}(v < u) := \{x \in X : \psi(x) \in \mathbf{N}(v < u)\}$$
of X. The following is a simple corollary of the definitions and results collected above:

Corollary 4.1. *Given a finite set X, an X-net $\mathcal{N} = (\mathbf{N}, \psi)$, and two primitive pairs $\{u, v\}\{u', v'\} \in \operatorname{Prim}(\mathbf{N})$, one has $\{u, v\} \| \{u', v'\}$ if and only if $S_{u,v} = S_{u',v'}$ holds. In particular, one has $uv = u'v'$ for any two primitive pairs $\{u, v\}, \{u', v'\}$ in $\operatorname{Prim}(\mathbf{N})$ with $S_{u,v} = S_{u',v'}$.*

In consequence, one can associate, to any X-net $\mathcal{N} = (\mathbf{N}, \psi)$, a corresponding $\mathbb{R}_{\geq 0}$-weighted split system $\mu = \mu_{\mathcal{N}}$ over X that maps any split $S \in \mathcal{S}(X)$ of the form $S = S_{u,v}$ for some primitive pair $\{u, v\} \in \operatorname{Prim}(\mathbf{N})$, onto the positive real number uv, and all other splits $S \in \mathcal{S}(X)$ (including the split $\{X, \emptyset\}$) onto 0.

Conversely, given any $\mathbb{R}_{\geq 0}$-weighted split system μ over X, one can associate to μ the X-net $\mathcal{N} = \mathcal{N}_\mu = (\mathbf{N}_\mu, \psi_\mu) = ((N_\mu, D_\mu), \psi_\mu)$ for which $\mu = \mu_{\mathcal{N}}$ holds that is defined as follows: One defines N_μ to consist of all maps $v : X \to \mathcal{P}(\mathrm{supp}(\mu))$ with

$$\binom{\mathrm{supp}(\mu)}{2} = \bigcup_{x \in X} \binom{\mathrm{supp}(\mu) - v(x)}{2} \tag{1}$$

for which

$$v(x) \triangle v(y) = \Delta_\mu(x,y) := \{S \in \mathrm{supp}(\mu) : S(x) \neq S(y)\} \tag{2}$$

or, equivalently,

$$v(x) = v(y) \triangle \Delta_\mu(x,y)$$

holds for all $x, y \in X$ — Condition (**??**) just requiring that there exists, for any two splits $S_1, S_2 \in \mathrm{supp}(\mu)$, some $x \in X$ such that $S_1, S_2 \notin v(x)$ holds.

Next, one defines

$$\Delta_\mu(u,v) := \bigcup_{x \in X} u(x) \triangle v(x)$$

and

$$D_\mu(u,v) := \mu_+(\Delta_\mu(u,v))$$

for all maps $u, v : X \to \mathcal{P}(\mathrm{supp}(\mu))$ and, noting that

$$u(x) \triangle v(x) = (\Delta_\mu(x,y) \triangle u(y)) \triangle (\Delta_\mu(x,y) \triangle v(y)) = u(y) \triangle v(y)$$

holds for all $x, y \in X$ and $u, v \in N_\mu$, one sees that

$$\Delta_\mu(u,v) := u(x) \triangle v(x)$$

and

$$D_\mu(u,v) := \mu_+(u(x) \triangle v(x)) = \sum_{S \in u(x) \triangle v(x)} \mu(S)$$

holds for every $x \in X$ and any two maps $v, u \in N_\mu$. And one defines $\psi_\mu : X \to N_\mu$ by associating, to any $x \in X$, the map

$$\psi_\mu(x)(S) : X \to \mathrm{supp}(\mu) : y \mapsto \Delta_\mu(x,y).$$

Using these constructions, it can be shown that the $\mathbb{R}_{\geq 0}$-weighted split systems over X "classify" the *injective objects* in `X—NET`, i.e., the X-nets $\mathcal{N}$ that "coincide" with their injective hull. More precisely, the following holds for any finite set X:

Theorem 4.2. (i) *Given any two $\mathbb{R}_{\geq 0}$-weighted split systems μ and μ' over X, there exists a morphism φ from $\mathcal{N}_\mu$ into $\mathcal{N}_{\mu'}$ in* `X—NET` *if and only if $\mu' \leq \mu$ holds (i.e., if and only if $\mu'(S) \leq \mu(S)$ holds for every X-split S in $\mathcal{S}^*(X)$); in particular, two X-nets of the form $\mathcal{N}_\mu$ and $\mathcal{N}_{\mu'}$ are isomorphic if and only if $\mu = \mu'$ holds.*

(ii) *Given any X-net $\mathcal{N}$, its injective hull $\mathcal{N}^*$ is canonically isomorphic to the X-net $\mathcal{N}_\mu$ for the $\mathbb{R}_{\geq 0}$-weighted split systems $\mu := \mu_{\mathcal{N}}$; in particular, $\mathcal{N}$ is an injective object in* `X—NET` *if and only if it is isomorphic to an X-net of the form $\mathcal{N}_\mu$ for some $\mathbb{R}_{\geq 0}$-weighted split system μ over X in which case there is only one such $\mathbb{R}_{\geq 0}$-weighted split system μ,* viz., *the $\mathbb{R}_{\geq 0}$-weighted split system $\mu_{\mathcal{N}}$; in particular,*

1. *given any X-net $\mathcal{N}$, its injective hull $\mathcal{N}^*$ is canonically isomorphic to the X-net $\mathcal{N}_{\mu_{\mathcal{N}}}$,*

(2) *two X-nets $\mathcal{N}$ and $\mathcal{N}'$ have isomorphic injective hulls if and only if $\mu_{\mathcal{N}} = \mu_{\mathcal{N}'}$ holds,*
(3) *there exists a morphism from an X-net $\mathcal{N}$ into the injective hull of an X-net $\mathcal{N}'$ if and only if $\mu_{\mathcal{N}} \leq \mu_{\mathcal{N}'}$ holds.*

(ii) *Given any X-net $\mathcal{N}$, its injective hull $\mathcal{N}^*$ is canonically isomorphic to the X-net $\mathcal{N}_\mu$ for the $\mathbb{R}_{\geq 0}$-weighted split systems $\mu := \mu_{\mathcal{N}}$; in particular, $\mathcal{N}$ is an injective object in* `X—NET` *if and only if it is isomorphic to an X-net of the form $\mathcal{N}_\mu$ for some $\mathbb{R}_{\geq 0}$-weighted split system μ over X in which case there is only one such $\mathbb{R}_{\geq 0}$-weighted split system μ,* viz., *the $\mathbb{R}_{\geq 0}$-weighted split system $\mu := \mu_{\mathcal{N}}$; in particular,*

(1) *given any X-net $\mathcal{N}$, its injective hull $\mathcal{N}^*$ is canonically isomorphic to the X-net $\mathcal{N}_{\mu_{\mathcal{N}}}$,*
(2) *two X-nets $\mathcal{N}$ and $\mathcal{N}'$ have isomorphic injective hulls if and only if $\mu_{\mathcal{N}} = \mu_{\mathcal{N}'}$ holds,*
(3) *there exists a morphism from an X-net $\mathcal{N}$ into the injective hull of an X-net $\mathcal{N}'$ if and only if $\mu_{\mathcal{N}} \leq \mu_{\mathcal{N}'}$ holds.*

5. **Examples**

Here are some "real-life" examples of X-nets. The first two belong to the group of altogether more than 10 phylogenetic trees ever published. They were carefully drawn by Ernst Haeckel who published his work on XXX in 1866, just 7 years after Charles Darwin published "The Origin of Species".

Text for Figure 1: Two phylogenetic trees from Ernst Haeckel's book on XXX.

The next two figures present *proper* networks, constructed using the program *SplitsTrees* based on data provided by XXX Helms in his thesis on the perception of colour similarity, Hamburg, 1980.

Text for Figure 2: An X-net constructed by applying the program *SplitsTrees* to data regarding the perception of colour similarity (Helms XXX, Hamburg, 1980). The set X consists of 10 distinct colours.

Text for Figure 3: Using data from the same source, now regarding a colour-blind subject's perception of colour similarity (also from data published by Helms XXX, Hamburg, 1980): The X-net constructed by *SplitsTrees* now degenerates into an X-tree.

Also the next three figures present networks that do not refer to a biological context (even though manuscript copying has a number of interesting analogies with sequence evolution).

Text for Figure 4: An X-net constructed by applying *SplitsTrees* to data from a German Road Atlas. The set X consists of 10 distinct German cities.

Text for Figure 5: An X-net constructed by Peter Robinson et al by applying *SplitsTrees* to data that he derived by comparing distinct hand-written copies of the Prologue of XXX Chaucer's "The Wife of Bath". The set X consists of more than 40 such manuscripts. And the resulting X-net is, not unexpectedly, rather "tree-ish".

Text for Figure 6: An X-net constructed by Mihai Albu by applying *Neighbour Net* to data from *The World Atlas of Language Structure* depicting the overall structural (dis)similarity of 20 world languages. XXX

The last four figures deal with proper biological data. Figure 7 deals with 16S rRNA sequences from all three *Kingdoms of Life*, the Eucariots, the Procariots, and the Archeae. Figures 8 and 9 deal with data regarding various variants of the AIDS virus, including HIV1, HIV2, and HIV sequences discovered in other primates. The same data have been analysed in two distinct ways, the first figure being based on (dis)similarity taking into account the first position, only, from each coding triple, the second one taking into account all positions, yet registering only the difference between purins and purimidins, neglecting transitions between nucleotides. The resulting, rather tree-ish structures are surprisingly similar, corroborating each other, and demonstrating that it is very unlikely that they both are just artefacts of the respective methods of quantifying (dis)similarity applied for deriving these two X-nets. Figure 10 deals with Human mitochondrial DNA data, and the resulting X-net clearly supports Allan Wilson's "Out of Africa" hypothesis (as African sequences can be found all over in that net while all other groups are clearly localized).

Text for Figure 7: An X-net constructed by applying the program *SplitsTrees* to data regarding the (dis)similarity of 16 distinct 16S rRNA sequences from Eucariots, Procariots, and Archeae. A proper net, though almost a tree as the length of edges not fitting in the "obvious" underlying tree structure is negligibly small. The underlying tree structure clearly supports Karl Woese's thesis claiming the existence of one common ancestor for all currently existing forms of Archeae.

Text for Figure 8: An X-net constructed by applying the program *SplitsTrees* to data regarding the (dis)similarity of HIV sequences, taking into account the first position, only, from each coding triple.

Text for Figure 9: An X-net constructed by applying the program *SplitsTrees* to data regarding the (dis)similarity of HIV sequences, taking into account the difference between purins and purimidins, only, neglecting all *transitions* between nucleotides.

Text for Figure 10: An X-net constructed by applying the program *Neighbour-Net* to data regarding the (dis)similarity of Human mitochondrial DNA sequences. The resulting X-net clearly supports Allan Wilson's "Out of Africa" hypothesis as African sequences can be found all over in that net while all other groups are clearly localized.

References

[1] H. -J. Bandelt, Recognition of tree metrics, *SIAM J. Disc. Math.* **3** (1990) 1-6.

[2] H. -J. Bandelt, A. Dress, Weak hierarchies associated with similarity measures – an additive clustering technique, *Bul. Math. Biol.* **51** (1989) 133-166.

[3] H. -J. Bandelt, A. Dress, A canonical split decomposition theory for metrics on a finite set, *Adv. Math.* **92** (1992) 47-105.

[4] H .-J. Bandelt, A. Dress, Split decomposition: a new and useful approach to phylogenetic analysis of distance data, Molecular Phylogenetics and Evolution **1**(3) (1992b) 242-252.

[5] J.-P. Barthélemy and A. Guénoche, Trees and proximity representations (Wiley, 1991).

[6] P. Buneman, The recovery of trees from measures of dissimilarity. In F. Hodson et al., Mathematics in the Archeological and Historical Sciences, Edinburgh University Press, 1971, pp. 387-395.

[7] A. Dress, Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces, Adv. in Math. **53** (1984) 321-402.

[8] A. Dress, K.T. Huber and V. Moulton, Some variations on a theme by Buneman, Ann. Combin. **1** (1997) 339-352.

[9] A. Dress, K.T. Huber and V. Moulton, An explicit computation of the injective hull of certain finite metric spaces in terms of their associated Buneman complex, Adv. in Math. **168** (2002) 1-28.

[10] A. Dress, V. Moulton and W. Terhalle, *T*-theory: An overview, Europ. J. Combinatorics 17 (1996) 161-175.

[11] D. Huson, SplitsTree: a program for analyzing and visualizing evolutionary data, *Bioinformatics* **14** (1998) 68-73; see also *http://www-ab.informatik.uni-tuebingen.de/software/splits/welcome_en.h tml.*

[12] J. Isbell, Six theorems about metric spaces, *Comment. Math. Helv.* **39** (1964) 65-74.

[13] P. J. Lockhart, A. E. Meyer, D. Penny, Testing the phylogeny of Swordtail fishes using Split Decomposition and Spectral Analysis, J. Mol. Evol. **41** (1995) 666-674.

[14] C. Semple, M. Steel, Phylogenetics, Oxford University Press, 2003.