The Combinatorics of Finite Metric Spaces and the (Re-)Construction of Phylogenetic Trees.
Extended abstract and some relevant references

Andreas Dress,
CAS-MPG Partner Institute for Computational Biology,
Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences, Shanghai 200031, P.R. China
and Max-Planck-Institut fuer Mathematik in den Naturwissenschaften,
Inselstrasse 22-26, D-04103 Leipzig, Germany

2. Oktober 2007

Over the last thirty years, various combinatorial aspects of finite metric spaces have been studied in the context of phylogenetic-tree reconstruction. A starting point was the following observation:

Given a metric $D : X \times X \to \mathbb{R}$ representing the approximative genetic distances $D(x, y), \ldots$ between the members $x, y, \ldots$ of a finite collection $X$ of taxa, it was shown in [15] that the following assertions relating to the object of desire, a “phylogenetic $X$-tree”, all are equivalent:

(i) The “tight span”

$$T_D := \{ f \in \mathbb{R}^X : \forall x \in X \ f(x) = \sup_{y \in X} (D(x, y) - f(y)) \}$$

of $D$ is an $\mathbb{R}$-tree.

(ii) There exists an edge-weighted $X$-tree $(V, E; \ell)$ — i.e., a finite tree $(V, E)$ with vertex set $V$ and edge set $E$ such that $V$ contains $X$ and every vertex in $V - X$ has degree at least 3, and a (positive) edge weighting
\( \ell : E \to \mathbb{R}_{>0} \) that assigns a positive length \( \ell(e) \) to every edge \( e \) in \( E \) — such that \( D \) is the restriction to \( X \) of the \textit{shortest-path metric} induced by \( \ell \) on \( V \).

(iii) There exists a map \( w : \mathcal{S}(X) \to \mathbb{R}_{\geq 0} \) from the set \( \mathcal{S}(X) \) of all bipartitions or \textit{splits} \( S = \{A, B\} \) of \( X \) into the set \( \mathbb{R}_{\geq 0} \) of non-negative real numbers such that

- given any two splits \( S = \{A, B\} \) and \( S' = \{A', B'\} \) in \( \mathcal{S}(X) \) with \( w(S), w(S') \neq 0 \), at least one of the four intersections \( A \cap A' \), \( B \cap A' \), \( A \cap B' \), and \( B \cap B' \) is empty and

- \( D(x, y) = \sum_{S \in \mathcal{S}(X : x \leftrightarrow y)} w(S) \) holds where

\[
\mathcal{S}(X : x \leftrightarrow y) := \{ \{A, B\} \in \mathcal{S}(X) : x \in A, y \in B \}
\]

denotes the set of those splits \( S = \{A, B\} \in \mathcal{S}(X) \) that \textit{separate} \( x \) and \( y \).

(iv) \( D(x, y) + D(u, v) \leq \max \{ D(x, u) + D(y, v), D(x, v) + D(y, u) \} \) holds for all \( x, y, u, v \in X \)

Moreover, the metric space \( T_D \) actually coincides in this case with the \( \mathbb{R} \)-tree that is canonically associated with an edge-weighted tree \((V, E; \ell)\).

This observation suggested to further investigate (1) the tight-span construction and (2) representations of metrics by weighted split systems with various specific properties, even if the metric in question would not satisfy the very special properties described above. These investigations have, in turn, given rise to a full-fledged research program dealing many diverse aspects of these two topics (see the list of references below).

In my lecture, I will focus on the rather new developments relating to \textit{block decomposition} and \textit{virtual cut points} of metric spaces reported in [28] to [31] that allow to canonically decompose any given finite metric space into a sum of pairwise compatible \textit{block metrics} thus providing a far-reaching generalization of the result recalled above.
Literatur


