Lecture 8 : Tree-metric theorem

MATH285K - Spring 2010 *Lecturer: Sebastien Roch*

References: [SS03, Chapter 7]

Previous class

DEF 8.1 (Four-point condition) *A dissimilarity map δ satisifies that* four-point condition (4PC) *if for all $x, y, w, z \in X$ (not necessarily distinct)*

$$\delta(w, x) + \delta(y, z) \leq \max\{\delta(w, y) + \delta(x, z), \delta(w, z) + \delta(x, y)\}. \tag{1}$$

DEF 8.2 (Ultrametric) *A dissimilarity map δ on X is an* ultrametric *if for every three distinct $x, y, z \in X$,*

$$\delta(x, y) \leq \max\{\delta(x, z), \delta(y, z)\}. \tag{2}$$

DEF 8.3 (Gromov Product) *Let δ be a dissimilarity map on X and fix $r \in X$. The* Gromov product *of x and y in $X \backslash \{r\}$ is*

$$\delta_r(x, y) = \begin{cases} \frac{1}{2}(\delta(x, y) - \delta(r, x) - \delta(r, y)) & x \neq y \\ 0 & \textit{o.w.} \end{cases} \tag{3}$$

LEM 8.4 *A nonnegative dissmilarity map δ satisfies the 4PC if and only if δ_r is an ultrametric for all $r \in X$.*

We prove the nontrivial direction of the Tree-Metric Theorem:

THM 8.5 (Tree-Metric Theorem) *Let δ be a nonnegative dissimilarity map. Then, δ is a tree metric if and only if δ saitisfies the 4PC.*

1 Equidistant Representation

We begin with a result about ultrametrics.

DEF 8.6 (Equidistant Representation) *Let δ be a dissimilarity map on X. An* equidistant representation *of δ is a rooted phylogenetic tree $\mathcal{T} = (T, \phi)$ with $T = (V, E)$ and root ρ, and an edge weight function $w : E \to \mathbb{R}$ such that:*

1. *For all $x, y \in X$*
$$d_{T,w}(\rho, \phi(x)) = d_{T,w}(\rho, \phi(y)).$$
(By definition of a path metric, the equality then holds with ρ replaced with $u \in V$ as long as $u \leq_T \phi(x), \phi(y)$, that is, u is a common ancestor of $\phi(x)$ and $\phi(y)$.)

2. *If $u \leq_T v \leq_T \phi(x)$ for $u, v \in \mathring{V}$ and $x \in X$ then*
$$d_{T,w}(\phi(x), v) \leq d_{T,w}(\phi(x), u).$$
(In particular, all interior edge weights are non-negative.)

It is straighforward to check that a dissimilarity map admitting an equidistant representation is an ultrametric. There is also a converse:

THM 8.7 *If δ is an ultrametric on X, then it has an equidistant representation.*

Proof: The proof is based on a simple reconstruction algorithm.

DEF 8.8 (Cherry) *A* cherry *is a pair of leaves (u, v) with a common neighbour.*

Let δ be an ultrametric on X. Consider the following recursive procedure:

function EQUIDISTANT

Input: Dissimilarity map δ on X

Output: Equidistant representation $(\mathcal{T}, w)$ of δ

- If $|X| = 2$, return a cherry with edge weights $\frac{1}{2}\delta(a, b)$.
- Otherwise:
  - Find $a, b \in X$ minimizing $\delta(a, b)$.
  - Set $\delta^{(ab)}$ to be δ restricted to $X \backslash \{b\}$.
  - Compute $(\mathcal{T}^{(ab)}, w^{(ab)}) = \text{EQUIDISTANT}(\delta^{(ab)})$.
  - Let $l^{(ab)} = \phi^{(ab)}(a)$. Let $\mathcal{T}$ be $\mathcal{T}^{(ab)}$ where $l^{(ab)}$ is replaced with a new cherry (l_a, l_b) with $l_a = \phi(a)$ and $l_b = \phi(b)$ and edge weights $\frac{1}{2}\delta(a, b)$. Let $e^{(ab)}$ be the edge adjacent to $\phi^{(ab)}(a)$ in $\mathcal{T}^{(ab)}$. Let $\mathring{e}$ be interior edge of $\mathcal{T}$ adjacent to the common neighbour of l_a and l_b. Set $w_{\mathring{e}} = w_{e^{(ab)}} - \frac{1}{2}\delta(a, b)$.

- $*$ Return $(\mathcal{T}, w)$.

The correctness of this procedure follows by induction on $|X| \geq 2$. The case $|X| = 2$ is trivial. Assume the reconstruction is correct for $|X| - 1$. The choice of a, b above guarantees that

$$\delta(a,b) \leq \delta(a,x) = \delta(b,x)$$

for all $x \in X \backslash \{a, b\}$ and $w_{\hat{e}} \geq 0$. ■

2 Proof of Tree-Metric Theorem

Proof:(of Theorem 8.5) Choose $r \in X$. Since δ satisfies the 4PC, δ_r is an ultra-metric on $X' = X \backslash \{r\}$ and there exists an equidistant representation $(\mathcal{T}', w')$ of δ_r with $\mathcal{T}' = (T', \phi')$ and root ρ'. Define

$$p = -d_{T',w'}(\rho', \phi'(x)),$$

which is independent of $x \in X'$.

To obtain a tree metric representation $(\mathcal{T}, w)$ of δ, we add a leaf edge e_r to ρ' with a new leaf r. Guided by the formula

$$\delta(x,y) = \delta(r,x) + \delta(r,y) + 2\delta_r(x,y),$$

for $r \notin \{x, y\}$, we set

$$w_e = \begin{cases} 2w'_e & \text{if } e \in \mathring{E}(T') \\ 2w'_e + \delta(r,x) & \text{if } e = \{\phi'(x), u\} \text{ for some } u \in V(T') \text{ and } x \in X' \\ 2p & \text{if } e = e_r. \end{cases} \quad (4)$$

The choice of w_{e_r} in (4) is justified by

$$d_{T,w}(x,r) = 2p + 2d_{T',w'}(\phi'(x), \rho') + \delta(r,x) = \delta(r,x).$$

To see that the weights in (4) are non-negative, note that, since δ satisfies the 4PC, it satisfies the triangle inequality so

$$\delta_r(x,y) = \frac{1}{2}(\delta(x,y) - \delta(r,x) - \delta(r,y)) \leq 0.$$

Hence, taking $x, y \in X'$ such that the path between $\phi'(x)$ and $\phi'(y)$ goes through ρ' in T', we have

$$2p = -\delta_r(x,y) \geq 0.$$

Similarly, since δ satisfies the triangle inequality, leaf edges in any tree metric representation of δ must have non-negative weight (see the proof of the Uniqueness of Tree Representation Theorem). Finally, by definition of an equidistant representation, $w'(e) \geq 0$ for $e \in \mathring{E}(T')$.

Contracting zero-weight edges, we obtain a tree metric representation with positive edge weights. ■

Further reading

The definitions and results discussed here were taken from Chapter 7 of [SS03]. Much more on the subject can be found in that excellent monograph. See also [SS03] for the relevant bibliographic references.

References

[SS03] Charles Semple and Mike Steel. *Phylogenetics*, volume 24 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2003.