Characterizing Cycle-Complete Dissimilarities in Terms of Associated Indexed 2-Hierarchies

Characterizing Cycle-Complete Dissimilarities in Terms of Associated Indexed 2-Hierarchies*

Kazutoshi Ando[1][0000−0002−1415−062X] and Kazuya Shoji[2]

[1] Faculty of Engineering, Shizuoka University, Hamamatsu, Shizuoka 432-8561, Japan
ando.kazutoshi@shizuoka.ac.jp

[2] Graduate School of Integrated Science and Technology, Shizuoka University, Hamamatsu, Shizuoka 432-8561, Japan

Abstract. 2-ultrametrics are a generalization of the ultrametrics and it is known that there is a one-to-one correspondence between the set of 2-ultrametrics and the set of indexed 2-hierarchies (which are a generalization of indexed hierarchies). Cycle-complete dissimilarities, recently introduced by Trudeau, are a generalization of ultrametrics and form a subset of the 2-ultrametrics; therefore the set of cycle-complete dissimilarities corresponds to a subset of the indexed 2-hierarchies. In this study, we characterize this subset as the set of indexed *acyclic* 2-hierarchies, which in turn allows us to characterize the cycle-complete dissimilarities. In addition, we present an $\mathrm{O}(n^2 \log n)$ time algorithm that, given an arbitrary cycle-complete dissimilarities of order n, finds the corresponding indexed acyclic 2-hierarchy.

Keywords: Hierarchical classification · Quasi-hierarchy · Quasi-ultrametric · Cluster analysis.

1 Introduction

Ultrametrics appear in a wide variety of research fields, including phylogenetics [10], cluster analysis [9], and cooperative game theory [2]. They have, among others, two important properties: there is a one-to-one correspondence between the set of ultrametrics and the set of indexed hierarchies [6, 8, 3], and every dissimilarity has a corresponding subdominant ultrametric [7].

2-ultrametrics [7] are a generalization of the ultrametrics and maintain their important properties: there is a one-to-one correspondence between the set of the 2-ultrametrics and the set of indexed 2-hierarchies [7] (which are a generalization of indexed hierarchies), and every dissimilarity has a corresponding subdominant 2-ultrametric [7].

Motivated by the work of Trudeau [11], Ando et al. [1] introduced the concept of cycle-complete dissimilarities. These form a subset of the 2-ultrametrics, so there is a corresponding subset of the indexed 2-hierarchies. In this study, we characterize this subset as the set of indexed *acyclic* 2-hierarchies, which in

* This work was supported by JSPS KAKENHI Grant Number 18K11180.

turn allows us to characterize the cycle-complete dissimilarities. In addition, we present an $\mathrm{O}(n^2 \log n)$ time algorithm that, given an arbitrary cycle-complete dissimilarity of order n, finds the corresponding indexed acyclic 2-hierarchy.

The rest of this paper is organized as follows. In Section 2, we review 2-ultrametrics and 2-hierarchies and the one-to-one correspondence between them. In Section 3, we characterize the cycle-complete dissimilarities in terms of indexed 2-hierarchies. In Section 4, we present an $\mathrm{O}(n^2 \log n)$ time algorithm for finding the indexed 2-hierarchy corresponding to a given cycle-complete dissimilarities. Finally, in Section 5, we conclude this paper.

2 2-ultrametrics and indexed 2-hierarchies

Let X be a finite set. A mapping $d\colon X \times X \to \mathbb{R}_+$ is called a *dissimilarity* on X if for all $x, y \in X$ we have

$$d(x,y) = d(y,x) \text{ and } d(x,x) = 0. \tag{1}$$

A dissimilarity d on X is *proper* if $d(x,y) = 0$ implies $x = y$ for all $x, y \in X$. In addition, it is called a *quasi-ultrametric* [5] if for all $x, y, z, t \in X$ we have

$$\max\{d(x,z), d(y,z)\} \le d(x,y) \Longrightarrow d(z,t) \le \max\{d(x,t), d(y,t), d(x,y)\}. \tag{2}$$

A family $\mathcal{K}$ of subsets of X is called a *quasi-hierarchy* on X if $\mathcal{K}$ satisfies the following conditions.

(i) $X \in \mathcal{K}$, $\emptyset \notin \mathcal{K}$,
(ii) $\{x\} \in \mathcal{K}$ for all $x \in X$,
(iii) $\forall A, B \in \mathcal{K} : A \cap B \in \mathcal{K} \cup \{\emptyset\}$,
(iv) $\forall A, B, C \in \mathcal{K} : A \cap B \cap C \in \{A \cap B, B \cap C, C \cap A\}$.

For any quasi-hierarchy $\mathcal{K}$ on X, a mapping $f\colon \mathcal{K} \to \mathbb{R}_+$ satisfying the following two conditions is called an *index* of $\mathcal{K}$ and the pair $(\mathcal{K}, f)$ is called an *indexed quasi-hierarchy* on X.

(1) $\forall x \in X\colon f(\{x\}) = 0$,
(2) $\forall A, B \in \mathcal{K}\colon A \subset B \Longrightarrow f(A) < f(B)$.

A quasi-hierarchy $(X, \mathcal{K})$ is said to be a *2-hierarchy* if it also satisfies

(v) $\forall A, B \in \mathcal{K} : A \cap B \notin \{A, B\} \implies |A \cap B| \le 1$.

Likewise, a dissimilarity d on X is called a *2-ultrametric* [7] if for all $x, y, z, t \in X$, we have

$$d(x,y) \le \max\{d(x,z), d(y,z), d(x,t), d(y,t), d(z,t)\}. \tag{3}$$

Let d be a dissimilarity on X and σ be a positive real number. Then, the undirected graph $G_d^\sigma = (X, E_d^\sigma)$ defined by

$$E_d^\sigma = \{\{x,y\} \mid x, y \in X, x \ne y, d(x,y) \le \sigma\} \tag{4}$$

is called the *threshold graph* of d at the threshold σ. We denote the set of all the maximal cliques of threshold graphs of d's by $\mathcal{K}_d$, i.e.,

$$\mathcal{K}_d = \bigcup_{\sigma \geq 0} \{K \mid K \text{ is a maximal clique of } G_d^\sigma\}. \tag{5}$$

In addition, for each $K \in \mathcal{K}_d$ we define $\mathrm{diam}_d(K)$ as

$$\mathrm{diam}_d(K) = \max\{d(x,y) \mid x, y \in K\} \tag{6}$$

and call it the *diameter* of K with respect to d.

With these definitions in place, we can now present the following useful lemma, followed by two propositions that clarify the relationships between quasi-ultrametrics and indexed quasi-hierarchies and between 2-ultrametrics and indexed 2-hierarchies.

Lemma 1. *Let d be a dissimilarity on X. If $K \in \mathcal{K}_d$, then K is a maximal clique of G_d^σ for $\sigma = \mathrm{diam}_d(K)$.*

Proof. Let $K \in \mathcal{K}_d$ be arbitrary and $\sigma = \mathrm{diam}_d(K)$. Since $d(x,y) \leq \mathrm{diam}_d(K) = \sigma$ for all $x, y \in K$, K is a clique of G_d^σ. Also, K is not a clique of $G_d^{\sigma'}$ for any σ' such that $\sigma' < \sigma$ since $d(x,y) = \sigma$ for some $x, y \in K$. Therefore, K is a maximal clique of $G_d^{\sigma''}$ for some σ'' such that $\sigma \leq \sigma''$. However, since for such a σ'', every clique of G_d^σ is a clique of $G_d^{\sigma''}$, it follows that K must be a maximal clique of G_d^σ. □

Proposition 1 (Diatta and Fichet [5]). *A proper dissimilarity d on X is a quasi-ultrametric if and only if $(\mathcal{K}_d, \mathrm{diam}_d)$ is an indexed quasi-hierarchy on X.*

Proposition 2 (Jardin and Sibson [7]). *A proper dissimilarity d on X is a 2-ultrametric if and only if $(\mathcal{K}_d, \mathrm{diam}_d)$ is an indexed 2-hierarchy on X.*

3 Characterizing cycle-complete dissimilarities in terms of their associated indexed 2-hierarchies

Let d be a dissimilarity on X. First, we introduce the complete weighted graph K_X, whose vertex set is X and whose edges $\{x, y\}$ have weight $d(x,y) = d(y,x)$.

We call a sequence

$$F : x_0, x_1, \cdots, x_{l-1}, x_l \tag{7}$$

of elements in X a *cycle* in K_X if all the x_i $(i = 0, \cdots, l-1)$ are distinct and $x_0 = x_l$. A dissimilarity d on X is called *cycle-complete* [1] if for each cycle (7) in K_X and each chord $\{x_p, x_q\}$ of F, we have

$$d(x_p, x_q) \leq \max_{i=1}^{l} d(x_{i-1}, x_i). \tag{8}$$

Proposition 3. *Let d be a dissimilarity on X. If d is cycle-complete, then it is also a 2-ultrametric.*

Proof. Let x, y, z, t be arbitrary distinct elements of X. If d is cycle-complete, then we have

$$d(x,y) \leq \max\{d(x,z), d(z,y), d(y,t), d(t,x)\} \tag{9}$$
$$\leq \max\{d(x,z), d(z,y), d(x,t), d(y,t), d(z,t)\}. \tag{10}$$

□

If a dissimilarity d on X is not cycle-complete, then there must exist a cycle $F : x_0, x_1, \cdots, x_{l-1}, x_l (= x_0)$ of K_X and a chord $\{x_p, x_q\}$ of F such that (8) does not hold. We call such a cycle an *invalid cycle* in K_X.

Lemma 2. *Let d be a dissimilarity on X that is not cycle-complete and*

$$F\colon x_0, x_1, \cdots, x_l (= x_0) \tag{11}$$

be an invalid cycle in K_X of minimum length l. If $l \geq 5$, then for all $0 \leq p \leq l-3$ and $2 \leq q \leq l-1$ such that $2 \leq q-p \leq l-2$, we have

$$\max_{i=1}^{l} d(x_{i-1}, x_i) < d(x_p, x_q) = \text{const}. \tag{12}$$

Proof. Let F be an invalid cycle (11) of minimum length l, where $l \geq 5$. Let

$$\delta = \max\{d(x_p, x_q) \mid \{x_p, x_q\} \text{ is a chord of } F\} \tag{13}$$

and $\delta = d(x_p, x_q)$ for some chord $\{x_p, x_q\}$ of F. We can assume without loss of generality that $0 \leq p$ and $p+3 \leq q \leq l-1$. Let

$$Y = \{p, p+1, \cdots, q\},$$
$$W = \{q, q+1, \cdots, l-1, 0, \cdots, p\}.$$

Let $\{x_i, x_j\}$ be a chord of F such that $\{i, j\} \subseteq Y$. If $d(x_i, x_j) < \delta$, then

$$F'\colon x_0, x_1, \cdots, x_{i-1}, x_i, x_j, x_{j+1}, \cdots, x_{l-1}, x_l (= x_0) \tag{14}$$

is an invalid cycle with a length less than l, contradicting the initial choice of F. Hence, we must have $d(x_i, x_j) = \delta$. Similarly, for a chord $\{x_i, x_j\}$ of F such that $\{i, j\} \subseteq W$, we have $d(x_i, x_j) = \delta$.

Next, let $\{x_i, x_j\}$ be a chord of F such that $i \in Y - W$ and $j \in W - Y$. If $i = p+1$, then, since $d(x_{p+1}, x_q) = \delta$, we have $d(x_{p+1}, x_j) = \delta$ by the same argument as above. If $i > p+1$, then, since $\{x_p, x_{p+2}\}$ is a chord of F such that $\{p, p+2\} \subseteq Y$, we have $d(p, p+2) = \delta$. Then, we again have that $d(x_i, x_j) = \delta$ by the same argument as above. □

For a family $\mathcal{K}$ of subsets of X, a sequence

$$C_0, C_1, \cdots, C_{l-1}, C_l \tag{15}$$

of subsets in $\mathcal{K}$ is called a *cycle* in $\mathcal{K}$ if we have

(i) $C_{i-1} \cap C_i \notin \{C_{i-1}, C_i, \emptyset\}$ for $i = 1, \cdots, l$,
(ii) $C_i \cap C_j = \emptyset$ for $0 \le i \le l-3$ and $2 \le j \le l-1$ with $2 \le j-i \le l-2$, and
(iii) $C_0 = C_l$,

where $l \ge 3$. If $\mathcal{K}$ has no cycle, we call it *acyclic* .

Theorem 1. *A proper dissimilarity d on X is cycle-complete if and only if $(\mathcal{K}_d, \mathrm{diam}_d)$ is an indexed acyclic 2-hierarchy on X.*

Proof. Here, we treat the "if" and "only if" parts separately.

(The "only if" part:) If we assume d is cycle-complete, that means it is a 2-ultrametric (Proposition 3), and hence, $(\mathcal{K}_d, \mathrm{dim}_d)$ is an indexed 2-hierarchy (Proposition 2). Thus, it only remains to show that $\mathcal{K}_d$ is acyclic.

Suppose, to the contrary, that there is a cycle

$$K_0, K_1, \cdots, K_{l-1}, K_l(= K_0) \tag{16}$$

in $\mathcal{K}_d$. Then, let

$$\delta = \max\{\mathrm{diam}_d(K_i) \mid i = 0, \cdots, l-1\} \tag{17}$$

and $i^* = 0, \cdots, l-1$ such that $\mathrm{diam}_d(K_{i^*}) = \delta$. If

$$d(x, y) \le \delta \text{ for all } x, y \in \bigcup_{i=0}^{l-1} K_i, \tag{18}$$

then $\cup_{i=0}^{l-1} K_i$ would be a clique of G_d^δ. However, this is impossible since K_{i^*} is a maximal clique of G_d^δ (Lemma 1). Hence, there would have to exist $x, y \in \cup_{i=0}^{l-1} K_i$ such that $d(x, y) > \delta$. Without loss of generality, suppose that $x \in K_a$ and $y \in K_b$ for $0 \le a < b \le l-1$ and choose $x_i \in K_i \cap K_{i+1}$ for $i = 0, \cdots, l-1$. For the sake of simplicity, we assume that $x, y \notin K_i \cap K_{i+1}$ for $i = 0, \cdots, l-1$. Then, we could construct an invalid cycle F in K_X via

$$F\colon x_0, \cdots, x_{a-1}, x, x_a, \cdots, x_{b-1}, y, x_b, \cdots, x_{l-1}, x_l(= x_0), \tag{19}$$

contradicting the cycle-completeness of d.

(The "if" part:) Here, we assume $(\mathcal{K}_d, \mathrm{diam}_d)$ is an indexed acyclic 2-hierarchy on X and show that the mapping d is cycle-complete. By Proposition 2, d is a 2-ultrametric. If d is not cycle-complete, then there would have to exist an invalid cycle in K_X. Let $F : x_0, x_1, \cdots, x_{l-1}, x_l(= x_0)$ be such a cycle of minimum length l.

First, we consider the case where $l \ge 5$. By Lemma 2, we have

$$d(x_p, x_q) > \max_{i=1}^{l} d(x_{i-1}, x_i) \text{ for all chord } \{x_p, x_q\} \text{ of } F. \tag{20}$$

For each $i = 0, \cdots, l-1$, let us choose a maximal clique K_i of G_d^σ such that $\{x_i, x_{i+1}\} \subseteq K_i$, where $\sigma = \max_{i=1}^{l} d(x_{i-1}, x_i)$. By (20), we would have

$$K_i \cap \{x_0, x_1, \cdots, x_{l-1}\} = \{x_i, x_{i+1}\} \quad (i = 0, \cdots, l-1). \tag{21}$$

In particular, all K_i $(i = 0, \cdots, l-1)$ would be pairwise distinct. Also, since each K_i is a maximal clique of G_d^σ, we would have

$$K_i \cap K_{i+1} \notin \{K_i, K_{i+1}, \emptyset\} \quad (i = 0, \cdots, l-1). \tag{22}$$

Let i and j be such that $0 \leq i,\ j \leq l-1$ and $2 \leq j - i \leq l-2$. We now show that $K_i \cap K_j = \emptyset$. To the contrary, suppose that $x \in K_i \cap K_j$. Then, we would have

$$d(x_i, x) \leq \sigma \text{ and } d(x, x_{j+1}) \leq \sigma. \tag{23}$$

From this, it would follow that

$$F' \colon x_0, \cdots, x_i, x, x_{j+1}, \cdots, x_l$$

is an invalid cycle of length less than l, contradicting the choice of F. Thus, $K_i \cap K_j = \emptyset$, so we would have shown that $K_0, K_1, \cdots, K_{l-1}, K_l (= K_0)$ is a cycle in $\mathcal{K}_d$, a contradiction.

Next, we consider the case where $l = 4$. Let

$$F \colon x_0, x_1, x_2, x_3, x_4 (= x_0) \tag{24}$$

be an invalid cycle in K_X and $\sigma = \max\{d(x_{i-1}, x_i) \mid i = 1, 2, 3, 4\}$. We assume, without loss of generality, that $d(x_0, x_2) > \sigma$ and show that $d(x_1, x_3) > \sigma$. Suppose, to the contrary, that $d(x_1, x_3) \leq \sigma$. Then, there would exist maximal cliques K and K' of G_d^σ such that $\{x_0, x_1, x_3\} \subseteq K$ and $\{x_1, x_2, x_3\} \subseteq K'$, and hence, $\{x_1, x_3\} \subseteq K \cap K'$. This contradicts the assumption that $\mathcal{K}_d$ is a 2-hierarchy since $K \neq K'$ by $d(x_0, x_2) > \sigma$. Then, by defining K_i as a maximal clique of G_d^σ such that $\{x_i, x_{i+1}\} \subseteq K_i$ for $i = 0, 1, 2, 3$, we would have (21) and (22), similar to the $l \geq 5$ case.

Now, suppose that for some $x \in X - \{x_0, x_1, x_2, x_3\}$ we have $x \in K_0 \cap K_2$. Then, there would have to exist a maximal clique K of G_d^σ such that $\{x_0, x, x_3\} \subseteq K$. It would then follow that $K \cap K_0 \supseteq \{x_0, x\}$ and $K \neq K_0$, contradicting the assumption that $\mathcal{K}_d$ is a 2-hierarchy. Therefore, we have that $K_0 \cap K_2 = \emptyset$ and similarly that $K_1 \cap K_3 = \emptyset$. Then, $K_0, K_1, K_2, K_3, K_4 (= K_0)$ would be a cycle in $\mathcal{K}_d$, contradicting the assumption that $\mathcal{K}_d$ is acyclic. $\square$

Corollary 1. *The mapping* $d \mapsto (\mathcal{K}_d, \mathrm{diam}_d)$ *is a one-to-one correspondence between the set of proper cycle-complete dissimilarities on* X *and the set of indexed acyclic* 2*-hierarchies on* X*.*

4 Algorithm

A vertex v of a connected graph G is called a *cut vertex* if $G - v$ is not connected. A graph is called *2-connected* if it is connected and has no cut vertex. Note that a graph with only one vertex is 2-connected. A maximal 2-connected subgraph of a graph G is called a *2-connected component* of G.

Lemma 3. *Let* d *be a cycle-complete dissimilarity on* X*. Then, for all* $\sigma \geq 0$*, the vertex set of a 2-connected component of* G_d^σ *is a clique of* G_d^σ*.*

Input : Proper cycle-complete dissimilarity d on X.
Output: Indexed acyclic 2-hierarchy $(\mathcal{K}_d, \mathrm{diam}_d)$.

Let

$$0 < \sigma_1 < \cdots < \sigma_l$$

be the distinct values of $d(x, y)$ $(x, y \in X, x \neq y)$;
$\mathcal{K} \leftarrow \mathcal{K}^{(0)} \leftarrow \{\{x\} \mid x \in X\}$;
$f(\{x\}) \leftarrow 0$ $(x \in X)$;
for $p = 1$ **to** l **do**
    Let $\mathcal{K}^{(p)}$ be the vertex sets of the 2-connected components of $G_d^{\sigma_p}$;
    $\mathcal{L} \leftarrow \mathcal{K}^{(p)} - \mathcal{K}^{(p-1)}$;
    $\mathrm{diam}_d(K) \leftarrow \sigma_p$ $(K \in \mathcal{L})$;
    $\mathcal{K} \leftarrow \mathcal{K} \cup \mathcal{L}$;
end
return $(\mathcal{K}, f)$;

Algorithm 1: Outline of the algorithm for computing $(\mathcal{K}_d, \mathrm{diam}_d)$.

Proof. Let $Q \subseteq X$ be the vertex set of a 2-connected component of G_d^σ. If $|Q| \leq 2$, then Q is a clique of G_d^σ by the definition of a 2-connected component, so we assume $|Q| \geq 3$. Suppose, to the contrary, that there exist distinct vertices $x, y \in Q$ such that $\{x, y\} \notin E_d^\sigma$. By the definition of Q, there are two openly disjoint paths P_1 and P_2 in G_d^σ connecting x and y. By concatenating P_1 and P_2, we can create a cycle in K_X, where all the edges have weights of at most σ. Since $\{x, y\}$ is a chord of this cycle, it follows from the cycle-completeness of d that $d(x, y) \leq \sigma$, and hence $\{x, y\} \in E_d^\sigma$, a contradiction. □

The set of maximal cliques of the threshold graph of a cycle-complete dissimilarity is characterized as follows.

Lemma 4. *Let d be a cycle-complete dissimilarity on X and $\sigma \geq 0$. Then, $K \subseteq X$ is a maximal clique of $G_d^\sigma = (X, E_d^\sigma)$ if and only if K is the vertex set of some 2-connected component of G_d^σ.*

Proof. Assume that $K \subseteq X$ is a maximal clique of $G_d^\sigma = (X, E_d^\sigma)$. Since K corresponds to a 2-connected subgraph of G_d^σ, it is a subset of the vertex set Q of some 2-connected component of G_d^σ. However, since Q is a clique (Lemma 3), we must have $K = Q$ by the maximality of K. Conversely, if $Q \subseteq X$ is the vertex set of a 2-connected component of G_d^σ, then Q is a clique of G_d^σ (Lemma 3). If this clique is not maximal, then there must exist a vertex $x \in X - Q$ such that $\{x, y\} \in E_d^\sigma$ for all $y \in Q$, contradicting the assumption that Q is the vertex set of a 2-connected component of G_d^σ. □

Based on Lemma 4, we have designed an algorithm for constructing the indexed acyclic 2-hierarchy $(\mathcal{K}_d, \mathrm{diam}_d)$ for a given proper cycle-complete dissimilarity d, as outlined in Algorithm 1. The validity of the algorithm follows straightforwardly from the propositions presented above.

Input : Proper cycle-complete dissimilarity d on X.
Output: Indexed acyclic 2-hierarchy $(\mathcal{K}_d, \mathrm{diam}_d)$.

Let $e_1, \ldots, e_m$ be the edges of K_X ordered in nondecreasing order of d, where $m = \frac{n(n-1)}{2}$;
$\mathcal{K} \leftarrow \{\{x\} \mid x \in X\}$;
$f(\{x\}) \leftarrow 0 \ (x \in X)$;
$\mathcal{L} \leftarrow \emptyset$;
for $i = 1$ **to** m **do**
  $\{x, y\} \leftarrow e_i$;
  if *x and y are in different 2-connected components of G_{i-1}* **then**
    if *x and y are in the same component* **then**
      Let P be a path connecting x and y in G_{i-1};
      Let $Q_1, \ldots, Q_l$ be the vertex sets of the 2-connected components of G_{i-1} which contain at least two vertices of P;
      $Q \leftarrow \bigcup_{k=1}^{l} Q_k$;
      $\mathcal{L} \leftarrow \mathcal{L} \cup \{Q\} - \{Q_1, \ldots, Q_l\}$;
    else
      $Q \leftarrow \{x, y\}$;
      $\mathcal{L} \leftarrow \mathcal{L} \cup \{Q\}$;
    end
  end
  if $d(e_i) < d(e_{i+1})$ *or* $i = m$ **then**
    $\mathcal{K} \leftarrow \mathcal{K} \cup \mathcal{L}$;
    $f(K) \leftarrow d(e_i) \ (K \in \mathcal{L})$;
    $\mathcal{L} \leftarrow \emptyset$;
  end
end
return $(\mathcal{K}, f)$;

Algorithm 2: More detailed description of the algorithm for computing $(\mathcal{K}_d, \mathrm{diam}_d)$.

It is not immediately clear how to implement Algorithm 1 efficiently, however. To achieve this, we need to able to identify the 2-connected components of a threshold graph efficiently. Let $e_1, \ldots, e_m$ be the edges of K_X arranged in nondecreasing order of d, where $m = \frac{n(n-1)}{2}$. Then, we construct the vertex sets of the 2-connected components of the undirected graph $G_i = (X, E_i)$ incrementally for $i = 0, 1 \cdots, m$, where E_i is defined by $E_i = \{e_1, \cdots, e_i\}$. A more detailed description of the algorithm is given in Algorithm 2.

Let $G = (X, E)$ be an undirected graph whose vertex set is X. Let A and $\mathcal{Q}$ be the set consisting of all the cut vertices and the set of the 2-connected components of G, respectively. The block forest (cf. [4]) of G is the bipartite graph $B = (A, \mathcal{Q}; F)$ defined by $F = \{(a, Q) \mid a \in A, Q \in \mathcal{Q}, a \in Q\}$, as shown in Figure 1.

Theorem 2. *Given a proper cycle-complete dissimilarity d on X, Algorithm 2 correctly produces the indexed acyclic 2-hierarchy $(\mathcal{K}_d, \mathrm{diam}_d)$ and terminates in $\mathrm{O}(n^2 \log n)$ time, where $n = |X|$.*

Proof. First, we show that the algorithm is valid. In Lines 6–17, it finds the vertex set Q of the 2-connected component of G_i formed by adding the edge $e_i = \{x, y\}$ to G_{i-1}, if it exists. This set is either $Q_1 \cup \cdots \cup Q_l$ or $e_i = \{x, y\}$, depending on whether or not x and y are in the same component. Then, the algorithm adds Q to the list $\mathcal{L}$, removing $Q_1, \cdots, Q_l$ in the first case. Then, the collection $\mathcal{L}$ of vertex sets in Line 19 is exactly the same as $\mathcal{K}^{(p)} - \mathcal{K}^{(p-1)}$ in Line 6 of Algorithm 1, where $d(e_i) = \sigma_p$.

Next, we consider the algorithm's time complexity. It takes $\mathrm{O}(n^2 \log n)$ time to sort the edges of K_X using any standard sorting algorithm, so the complexity must be at least that. Here, we show that the other operations in Algorithm 2 only require $\mathrm{O}(n^2)$ time. To achieve this bound, we represent the 2-connected components of G_i as block forest B_i, and assume that each of the trees in the forest B_i is rooted at some vertex for $i = 0, 1 \cdots, m$. In addition, we use a mapping $q\colon X - A \to \mathcal{Q}$ that associates each $x \in X - A$ with the unique 2-connected component $q(x)$ of G_{i-1} to which x belongs. With this, given arbitrary $x, y \in X$, we can determine whether or not x and y are in the same 2-connected component of G_{i-1} in $\mathrm{O}(1)$ time. We can also find the 2-connected components $Q_1, \cdots, Q_l$ (Line 10) in $\mathrm{O}(n)$ time by searching for the path P' in the forest B_i connecting the nodes corresponding to x and y, as shown in Figure 1(b). The block forest can be updated in $\mathrm{O}(n)$ time by reducing the 2-connected components $Q_1, \cdots, Q_l$ on the path P' to a single 2-connected component Q. See Figure 2(b). The mapping q can also be updated in $\mathrm{O}(n)$ time. Since the number of i's for which x and y are in different 2-connected components is $\mathrm{O}(n)$ [1, Lemma 3.5], it follows that the total time taken to compute Lines 8–16 is $\mathrm{O}(n^2)$. □

5 Conclusions

It is known [5] that the mapping $d \mapsto (\mathcal{K}_d, \mathrm{diam}_d)$ gives a one-to-one correspondence between the set of quasi-ultrametrics and the set of indexed quasi-hierarchies on X, where $\mathcal{K}_d$ is the set of all the maximal cliques of threshold graphs of d and the function $\mathrm{diam}_d\colon \mathcal{K}_d \to \mathbb{R}_+$ gives the diameter of each clique in $\mathcal{K}_d$. This leads to a similar one-to-one correspondence between the set of 2-ultrametrics and the set of indexed 2-hierarchies on X [7]. The cycle-complete dissimilarities [1] form a subset of the 2-ultrametrics, so the mapping $d \mapsto (\mathcal{K}_d, \mathrm{diam}_d)$ gives a correspondence between these and a subset of the indexed 2-hierarchies on X. In this paper, we have characterized this subset as the set of indexed acyclic 2-hierarchies on X, which has then allowed us to characterize the cycle-complete dissimilarities. In addition, we have presented an algorithm for finding the indexed acyclic 2-hierarchy $(\mathcal{K}_d, \mathrm{diam}_d)$ on X corresponding to a cycle-complete dissimilarity d on X and shown that runs in $\mathrm{O}(n^2 \log n)$ time, where $n = |X|$.

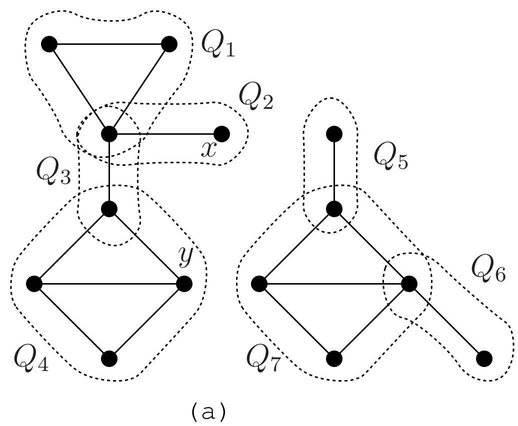

Fig. 1. (a) All 2-connected components of a graph G. (b) Block forest of G, where the cut vertices are indicated by rectangles, and the path P' between Q_2 and Q_4 is indicated by a wavy line.

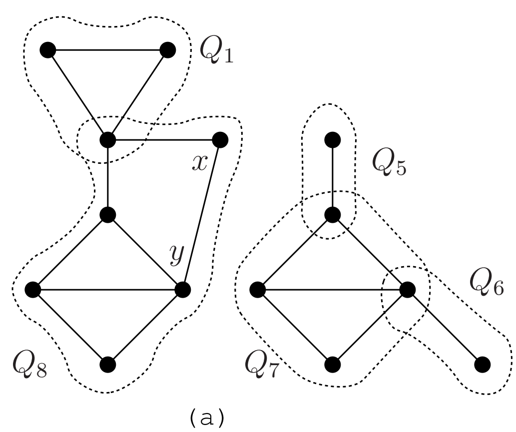

Fig. 2. (a) All 2-connected components of the graph $G + \{x, y\}$, where G is the graph in Figure 1(a). (b) Block forest of $G + \{x, y\}$, where the cut vertices are indicated by rectangles. Here, Q_2, Q_3 and Q_4 in Figure 1(b) have been reduced to form Q_8.

Acknowledgments

The authors are grateful to the anonymous referees for useful comments which improved the presentation of the original version of this paper.

References

1. K. Ando, R. Inagaki and K. Shoji: Efficient algorithms for subdominant cycle-complete cost functions and cycle-complete solutions. *Discrete Applied Mathematics* **225** (2017) 1–10.
2. K. Ando and S. Kato: Reduction of ultrametric minimum cost spanning tree games to cost allocation games on rooted trees. *Journal of the Operations Research Society of Japan* **53** (2010) 62–68.
3. J.-P. Benzécri: *L'analyse des Données* (Dunod, 1973).
4. J.A. Bondy and U. S. R. Murty: *Graph Theory* (Springer, 2008).
5. J. Diatta and B. Fichet: Quasi-ultrametrics and their 2-ball hypergraphs. *Discrete Mathematics* **192** (1998) 87–102.
6. C.J. Jardin, N. Jardin and R. Sibson: The structure and construction of taxonomic hierarchies. *Mathematics Biosciences* **1** (1967) 173–179.
7. N. Jardin and R. Sibson: *Mathematical Taxonomy* (Wiley, New York, 1971).
8. S.C. Johnson: Hierarchical clustering schemes. *Psychometrika* **32** (1967) 241–254.
9. G.W. Milligan: Ultrametric hierarchical clustering algorithms. *Psychometrika* **44** (1979) 343–346.
10. C. Semple and M. Steel: *Phylogenetics* (Oxford University Press, 2003).
11. C. Trudeau: A new stable and more responsive cost sharing solution for minimum cost spanning tree problems. *Games and Economic Behavior* **75** (2012) 402–412.