Discrete Applied Mathematics 226 (2017) 51–57

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

On minimum spanning tree-like metric spaces[☆]

Momoko Hayamizu[a,b,*], Kenji Fukumizu[a]

[a] The Institute of Statistical Mathematics, Tokyo 190-8562, Japan

[b] JST PRESTO, Tokyo 190-8562, Japan

ARTICLE INFO

Article history:
Received 5 December 2015
Received in revised form 29 March 2017
Accepted 2 April 2017
Available online 4 May 2017

Keywords:
Minimum spanning tree
Tree-like metric space
Median graph

ABSTRACT

We attempt to shed new light on the notion of 'tree-like' metric spaces by focusing on an approach that does not use the four-point condition. Our key question is: Given metric space M on n points, when does a fully labelled positive-weighted tree T exist on the same n vertices that precisely realises M using its shortest path metric? We prove that if a spanning tree representation, T, of M exists, then it is isomorphic to the unique minimum spanning tree in the weighted complete graph associated with M, and we introduce a *fourth-point condition* that is necessary and sufficient to ensure the existence of T whenever each distance in M is unique. In other words, a finite median graph, in which each geodesic distance is distinct, is simply a tree. Provided that the tie-breaking assumption holds, the fourth-point condition serves as a criterion for measuring the goodness-of-fit of the minimum spanning tree to M, i.e., the spanning tree-likeness of M. It is also possible to evaluate the spanning path-likeness of M. These quantities can be measured in $O(n^4)$ and $O(n^3)$ time, respectively.

1. Introduction

Historically, graphs as finite metric spaces have been studied extensively [5]. Although we approach them differently, we would like to emphasise, amongst others [12,16–18], the classical result provided by Buneman [6]. In short, a metric on a finite set can be realised by the shortest path metric in a positive-weighted tree if and only if it satisfies the four-point condition. The four-point condition is not only frequently utilised in the context of evolutionary trees [15], but is also known for its direct connection to the theory of Gromov hyperbolic metric spaces [9]. It is also widely known that there is a unique tree representation for every metric satisfying the four-point condition [7,11].

Given this background, a metric space that satisfies the four-point condition is commonly considered tree-like. However, an important caveat should be addressed: the four-point condition is necessary and sufficient to ensure the existence of a *partially labelled* tree that realises a given metric [5,11,15]. For example, a complete graph with a uniform edge length clearly satisfies the four-point condition, but it only becomes tree-like after an extra vertex is added. In this case, the four-point condition does not ensure that a metric is realised by a *fully labelled* tree on *the same* set. The same applies to *block graphs* (*i.e.*, unweighted graphs in which all bi-connected components are complete subgraphs) [2]. Thus, it does not characterise the distance within trees but rather the shortest path metrics induced by a more general class of graphs.

This may not create an issue in conventional phylogenetics, but considering the recent surge of renewed biological interest in minimum spanning tree (MST)-based tree estimation [14], determining when a metric space is realised by a

[☆] This work is part of the first author's doctoral thesis approved by The Graduate University for Advanced Studies.

[*] Corresponding author at: The Institute of Statistical Mathematics, Tokyo 190-8562, Japan.
E-mail addresses: hayamizu@ism.ac.jp (M. Hayamizu), fukumizu@ism.ac.jp (K. Fukumizu).

http://dx.doi.org/10.1016/j.dam.2017.04.001

positive weighted tree on the same set is not only a natural undertaking but also a meaningful one. Thus far, this problem has not been properly recognised, much less addressed. The only two exceptions to this are the recent work described in [1] and in [10]. This question seems to be non-trivial, not only because it cannot be answered using Buneman's theorem, but also because it is equivalent to determining a method for recognising a special case of the metric travelling salesman problem (TSP). If an input – a metric on a set of cities – is the shortest path metric in a tree on the city set, the length of the optimal tour must equal twice the length of the MST.

In this paper, we examine the sub-type of tree metrics that does not rely on the four-point condition. Our work is based on three components: the tie-breaking assumption, which has been popular in algorithmic applications since the work provided by Kruskal in [13]; what we call the fourth-point condition, which can typically be found in the definition of median metric spaces [8]; and a simple trick for metric-preserving edge removal, which applies to any finite metric space. These concepts, which are part of our original results, are defined and discussed in Section 2.

As expected, if it exists, a fully labelled positive-weighted tree that realises a finite metric space is the unique MST in its associated weighted complete graph (Proposition 2.15). Our goal is to prove the following: A finite metric space under the tie-breaking rule is realised by the MST if and only if it satisfies the fourth point condition (Theorem 3.1). This implies that every finite median graph, in which the shortest path lengths between all pairs of vertices are distinct, is necessarily a tree (Corollary 3.3). This result also yields a stronger condition for understanding when a finite metric space is realised, especially by a spanning path graph (Corollary 3.5). We define and discuss the notion of the spanning tree-likeness of a finite metric space in Section 4.

2. Preliminaries

We apply the metric-related terminology provided in [8] throughout this paper. Let (X, d_M) be a *finite metric space*, that is, a finite set X equipped with metric d_M. For two distinct points x and x' in X, the *closed metric interval* between them is defined as the set

$$I(x, x') := \{i \in X : d_M(x, x') = d_M(x, i) + d_M(i, x')\}.$$

All graphs considered in this paper will be simple, undirected, *fully labelled* (*i.e.*, each vertex is labelled), and *positive weighted* (*i.e.*, each edge has a positive length). A graph is denoted by $(V, E; w)$ for a set V of labelled vertices and a set E of edges that are associated with a positive edge-weighting function, $w : E \to \mathbb{R}^+$. Given a graph G, the sets of vertices and edges are denoted by $V(G)$ and by $E(G)$, respectively. Moreover, a graph G is said to be a graph *on* $V(G)$. Vertices may be renamed as needed, assuming no confusion arises, and a vertex labelled 'x' is referred to as vertex x. The distance in a graph G is defined as the shortest path metric and is represented using d_G.

For any finite metric space $M := (X, d_M)$, there is a graph that realises M by its shortest path metric. Indeed, the weighted complete graph $K_M := (X, \binom{X}{2}; d_M)$ associated with M is a trivial realisation of M. An edge of K_M that joins two distinct vertices x and x' is denoted by $e(x, x')$. This paper uses the terms '*points*' and '*vertices*' interchangeably, because there is a one-to-one correspondence between X and $V(K_M)$ for any finite metric space M.

2.1. The tie-breaking rule

Given connected graph G, a subtree that connects all vertices of G is said to be a *spanning tree* in G. In particular, a spanning tree whose length (*i.e.*, the sum of all edge weights) is shortest amongst all spanning trees is called a *minimum spanning tree* (MST). The problem of finding an MST in a connected graph is known as the MST problem, which is efficiently solved by greedy algorithms, such as Kruskal's method. In fact, one can easily find an MST in K_M by selecting edges so as not to create a cycle in ascending order of the value of d_M. Although K_M can have one or more MSTs in general, its MST is uniquely determined if the following assumption holds.

Definition 2.1. A finite metric space (X, d_M) is said to satisfy the *tie-breaking rule* if the values of d_M are distinct for all pairs in X.

The tie-breaking rule has been widely known since it was introduced by Borůvka [4] (cited in [13]) and by Kruskal [13]. This assumption is strong enough to ensure the uniqueness of the MST, and it is reasonable in many practical situations and can be quickly checked in $O(|X|^2)$ time. The present paper explores another of its benefits through a discussion of the relation between an MST and a finite metric space.

2.2. The fourth point condition

Definition 2.2 (Fig. 1)**.** A finite metric space (X, d_M) is said to satisfy the *fourth-point condition* if, for every (not necessarily distinct) three points $x, y, z \in X$, there exists a point $p^* \in X$ such that

$$d_M(x, p^*) + d_M(y, p^*) + d_M(z, p^*) = \frac{1}{2}\{d_M(x, y) + d_M(y, z) + d_M(z, x)\}.$$