Metric Representations of Network Data

Santiago Segarra, Gunnar Carlsson, Facundo Mémoli and Alejandro Ribeiro

Abstract—Networks are structures that encode relationships between pairs of elements or nodes. However, there is no imposed connection between these relationships, i.e., the relationship between two nodes can be independent of every other one in the network, and need not be defined for *every* possible pair of nodes. This is not true for metric spaces, where the triangle inequality imposes conditions that must be satisfied by triads of distances in the space, and these distances *must* be defined for every pair of nodes. In this paper, we study how to project networks into q-metric spaces, a generalization of metric spaces that encompasses a larger class of structured representations. In order to do this, we encode as axioms two intuitively desirable properties of the mentioned projections. We show that there is only one way of projecting networks onto q-metric spaces satisfying these axioms. Moreover, for the special case of (regular) metric spaces, this method boils down to computing the shortest path between every node and, for the case of ultrametric spaces, it coincides with single linkage hierarchical clustering. Furthermore, we show that the projection method satisfies two properties of practical relevance: optimality, which enables its utilization for the efficient estimation of combinatorial optimization problems, and nestedness, which entails consistency of the structure induced when projecting onto different q-metric spaces. Finally, we illustrate how metric projections can be used to efficiently search networks aided by metric trees.

I. INTRODUCTION

Networks are data structures that encode relationships between elements and can be thought of as signals that, instead of having values associated with elements, have values associated with pairs of elements. As such, they play an important role in our current scientific understanding of problems in which relationships between elements are important. These problems include interactions between proteins or organisms in biology [1], [2], individuals or institutions in sociology [3], [4], and neurons or regions in the brain [5]–[7].

Despite their pervasive presence, tools to analyze networks and algorithms that exploit network data are not as well developed as tools and algorithms for processing of conventional signals. Although some of this lag can be attributed to different developmental stages, there is also the matter of the complexity of network data. To understand this latter point it is instructive to observe that particular cases of signals that encode relationships between elements are well understood and pose little challenge for analysis and algorithm design. E.g., a correlation matrix is a representation of the proximity between components of a random signal and a finite metric space defines distances between elements of a space. Both

can be considered as particular cases of networks and the understanding of both is on par with the understanding of signals. Since correlation matrices and metric spaces have been studied for longer, the relative lag of network analysis can be again attributed to different developmental stages. However, the available evidence suggests otherwise.

Indeed, consider a problem of proximity search in which we are given a network and an element whose dissimilarity to different nodes of the network can be determined. We are asked to find the element that is least dissimilar. In an arbitrary network finding the least dissimilar node requires comparison against all nodes and incurs a complexity that is linear in the size of the network. In a metric space, however, the triangle inequality encodes a transitive notion of proximity. If two points are close to each other in a metric space and one of them is close to a third point, then the other one is also close to this third point. This characteristic can be exploited to design efficient search methods using metric trees whose complexity is logarithmic in the number of nodes [8]–[10]. Likewise, many hard combinatorial problems on graphs are known to be approximable in metric spaces but not approximable in generic networks. The traveling salesman problem, for instance, is not approximable in generic graphs but it is approximable in polynomial time to within a factor of 3/2 in metric spaces [11]. In either case, the advantage of the metric space is that the triangle inequality endows it with a structure that an arbitrary network lacks. It is this structure that makes network analysis and algorithm design tractable.

If metric spaces are easier to handle than arbitrary networks, a possible route for network analysis is to design projection operators to map arbitrary networks into the subset of networks that represent metric spaces. This is the problem addressed in this paper.

A. Related work and contributions

The concept of an abstract metric space, introduced in the early 20th century [12], encompasses a wide variety of scientific and engineering constructions where the notion of distance is present. The main difference with weighted networks is that distances in metric spaces are governed by the triangle inequality whereas weights in networks need not satisfy any similar requirement. During the first half of the past century, metric spaces were regarded as mere presentations of underlying topological spaces and a lot of effort was put on the study of embedding general metric spaces into more familiar ones [13]. However, in the late sixties, there was a partial shift in the focus of analysis, and the first formal studies of metric spaces *as such* – not seen as representations of some underlying topological space – appeared [14], specifically in the field of category theory [15]. We leverage the fact that the

fundamental understanding of metric spaces is more developed than that of networks in order to gain insight on the latter by projecting them onto the former and using analytical tools designed for the study of metric spaces.

The traditional way of mapping a generic dissimilarity function between pairs of points to a metric space is through multidimensional scaling (MDS) [16]. Different problem formulations give rise to the definition of different types of MDS where there is a basic distinction between metric MDS, where the input consists of quantitative similarities [17], [18], and non-metric MDS where dissimilarities can be ordinal [19], [20]. However, all these techniques have in common that the end goal is to facilitate visualization of the data [21]. Thus, MDS embeds the input dissimilarities into familiar and low-dimensional metric spaces such as $\mathbb{R}^2$ or $\mathbb{R}^3$.

The advantages of metric embeddings exceed visualization. Since our interest exceeds data visualization, we propose an alternative procedure to map generic dissimilarity networks to metric spaces. In order to do this, we plan to follow an axiomatic approach, i.e., we encode as axioms desirable properties of the projections.

In Section II, we formally introduce networks and q-metric spaces, the latter being an extension of metric spaces that enables the generalization of our framework to a variety of structured spaces. For instance, both the (regular) metric spaces and the ultrametric spaces [22] can be seen as particular cases of q-metric spaces. In Section III, the two axioms of the proposed framework are introduced: a projection axiom – networks that are already q-metric remain unchanged after projection – and a transformation axiom – smaller networks have smaller metric projections. We deem as admissible a projection method that satisfies these two axioms. The apparent weakness of these axioms contrast with the stringent theoretical consequences. More specifically, in Section IV we show that there is a unique admissible way of projecting networks onto q-metric spaces, that we call the *canonical q-metric projection*. For the particular case of (regular) metric spaces, this implies that the shortest path between two nodes is the unique admissible distance between them. Moreover, when focusing on projections onto ultrametric spaces, single-linkage clustering [23, Ch. 4] is identified as the only admissible hierarchical clustering method. Furthermore, in Section V we show that the axioms of Projection and Transformation confer two practical properties to canonical projections: optimality and nestedness. The former implies that the canonical projection of a network can be used to approximate the solution of combinatorial optimization problems on the network. Nestedness implies that the q-metric space obtained when projecting an arbitrary network is invariant to intermediate projections onto other q'-metric spaces with laxer structure. Finally, in Section VI, we propose an efficient strategy for approximate search in networks by first projecting a general network onto a q-metric space and then leveraging this structure for search via the construction of a metric tree [8]–[10]. We show that the parameter q can be used to control the tradeoff between computational gain and search performance.

II. NETWORKS AND q-METRIC SPACES

In the present paper we consider weighted and undirected graphs or *networks*. Formally, we define a graph $G = (V, E, W)$ as a triplet formed by a finite set of n nodes or vertices V, a set of edges $E \subset V \times V$ where $(x, y) \in E$ represents an edge from $x \in V$ to $y \in V$, and a map $W : E \to \mathbb{R}_{++}$ from the set of edges to the strictly positive reals, representing weights $W(x, y) > 0$ associated with each edge (x, y). The weights are related to dissimilarities, i.e. the larger the weight the less similar the nodes are. The lack of direction in G implies that if $(x, y) \in E$ then $(y, x) \in E$ and $W(x, y) = W(y, x)$. The graphs considered here are further assumed not to contain self-loops, i.e., $(x, x) \notin E$ for all $x \in V$. Denote by $\mathcal{N}$ the set of all networks. Networks in $\mathcal{N}$ can have different node sets V, different edge sets E, or different weights W.

Given a set X, a metric $d : X \times X \to \mathbb{R}_+$ is a function from the space of pairs of elements $X \times X$ to the non-negative reals satisfying the following three properties for every $x, y, z \in X$:

Symmetry: $d(x, y) = d(y, x)$.
Identity: $d(x, y) = 0$ if and only if $x = y$.
Triangle inequality: $d(x, y) \le d(x, z) + d(z, y)$.

The ordered pair $M = (X, d)$ is said to be a *finite metric space* [24] and the space of all finite metric spaces is defined as $\mathcal{M}$.

In this paper we also consider tighter metric spaces that are termed *q-metric spaces* and are parametrized by a constant $q \in [1, \infty]$. For strictly finite $q < \infty$, a q-metric space is a pair $M = (X, d)$ where the function d satisfies the symmetry and identity properties but a q-triangle inequality in lieu of the regular triangle inequality. This q-triangle inequality is such that for all $x, y, z \in X$, it holds,

q-triangle inequality: $d(x, y)^q \le d(x, z)^q + d(z, y)^q$.

When $q = \infty$ we define an ∞-metric space as a pair $M = (X, d)$ that satisfies the symmetry and identity properties as well as the ∞-triangle inequality

∞-triangle inequality: $d(x, y) \le \max[d(x, z), d(z, y)]$,

for all $x, y, z \in X$. Notice that the regular definition of metric space is recovered when $q = 1$. When $q = \infty$, the ∞-triangle inequality is equivalent to the strong triangle inequality that characterizes ultrametric spaces – which are equivalent representations of dendrograms, the outputs of hierarchical clustering methods [25], [26]. Another instance of interest is that of 2-metric spaces, in which case all of the triangles in the space are acute angled triangles; see Section IV-C.

In this paper, we interpret q-metrics as particular cases of networks which we can do if we associate the network $M = (X, X \times X, d)$ with the q-metric space $M = (X, d)$. The space of all q-metric spaces is denoted here as $\mathcal{M}_q$ and the space of metrics in general is denoted as $\mathcal{M} \equiv \mathcal{M}_1$. Notice that for all $q > q' > 1$ we must have

$$\mathcal{M}_q \subset \mathcal{M}_{q'} \subset \mathcal{M}_1 \equiv \mathcal{M} \subset \mathcal{N}. \tag{1}$$

A closely related definition is that of a norm [24]. In this case, for a given vector space Y, a norm $\| \cdot \|$ is a function $\| \cdot \| : Y \to \mathbb{R}_+$ from Y to the non-negative reals such that, for all vectors $v, w \in Y$ and scalar constant β, it satisfies:

Positiveness: $\|v\| \geq 0$ with equality if and only if $v = \vec{0}$.
Positive homogeneity: $\|\beta\, w\| = |\beta|\, \|w\|$.
Subadditivity: $\|v + w\| \leq \|v\| + \|w\|$.

A commonly used family of norms for vectors in $\mathbb{R}^l$ are the p-norms $\| \cdot \|_p$ for real $p \in [1, \infty]$, where the norm of a vector $v = [v_1, \ldots, v_l]$ is given by

$$\|v\|_p = \begin{cases} \left(\sum_{i=1}^{l} |v_i|^p \right)^{1/p} & \text{for } p < \infty, \\ \max_i |v_i| & \text{for } p = \infty. \end{cases} \quad (2)$$

In the definition of metric projections, the concepts of path and path length are important. Given a network $G = (V, E, W)$ and $x, x' \in V$, a path $P_{xx'}$ is an *ordered* sequence of nodes in V,

$$P_{xx'} = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x'], \quad (3)$$

which starts at x and finishes at x' and $e_i = (x_i, x_{i+1}) \in E$ for $i = 0, \ldots, l-1$. We say that $P_{xx'}$ links or connects x to x'. The links e_i of a path are the edges connecting consecutive nodes of the path in the direction given by the path. In the present paper we refer exclusively to connected graphs, i.e., graphs where there exist paths $P_{xx'}$ for all $x, x' \in V$. Given two paths $P_{xx'} = [x = x_0, x_1, ..., x_l = x']$ and $P_{x'x''} = [x' = x'_0, x'_1, ..., x'_{l'} = x'']$ such that the end point x' of the first one coincides with the starting point of the second one, we define the *concatenated path* $P_{xx'} \oplus P_{x'x''}$ as

$$P_{xx'} \oplus P_{x'x''} = [x = x_0, \ldots, x_l = x' = x'_0, \ldots, x'_{l'} = x'']. \quad (4)$$

For a given norm $\| \cdot \|$, we define the *length* of a given path $P_{xx'} = [x = x_0, \ldots, x_l = x']$ as $h_{\|\cdot\|}(P_{xx'}) = \|[W(x_0, x_1), \ldots, W(x_{l-1}, x_l)]\|$, i.e., the norm of the vector that consists of the weights associated to the links in the path. For notational consistency, we utilize the convention that $h_{\|\cdot\|}([x, x]) = 0$ for all $x \in V$. We define the minimum length function $s_{\|\cdot\|} : V \times V \to \mathbb{R}_+$ where the minimum length $s_{\|\cdot\|}$ between nodes $x, x' \in V$ is given by

$$s_{\|\cdot\|}(x, x') = \min_{P_{xx'}} h_{\|\cdot\|}(P_{xx'}). \quad (5)$$

The connectedness of G ensures that $s_{\|\cdot\|}(x, x')$ is well-defined for every pair of nodes $x, x' \in V$.

In the present paper we study the design of q-metric projections $\mathcal{P}_q$ with the objective of representing networks as q-metric spaces. Formally, for all node sets V we define a q-metric projection $\mathcal{P}_q : \mathcal{N} \to \mathcal{M}_q$ as a map that projects every network onto a q-metric space while preserving V. We say that two q-metric projections $\mathcal{P}_q$ and $\mathcal{P}'_q$ are *equivalent*, and we write $\mathcal{P}_q \equiv \mathcal{P}'_q$, if and only if $\mathcal{P}_q(G) = \mathcal{P}'_q(G)$, for all $G \in \mathcal{N}$.

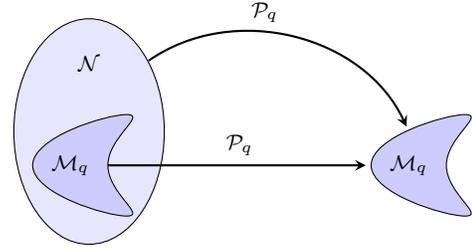

Fig. 1: Axiom of Projection. The q-metric space $\mathcal{M}_q$ is an invariant set of the projection map $\mathcal{P}_q$.

III. Axioms of Projection and Transformation

The broad definition of $\mathcal{P}_q$ as a node-preserving and q-metric inducing map presented in Section II admits maps of undesirable behavior. E.g., we may define the map $\mathcal{P}_q^1$ such that for any $G = (V, E, W)$ it outputs the space $(V, d) = \mathcal{P}_q^1(G)$ where d is defined as $d(x, x') = 1$ and $d(x, x) = 0$ for all $x \neq x' \in V$. It is immediate to see that the output space is q-metric for all q, since it is ∞-metric or ultrametric. However, the defined q-metric d completely ignores the edge structure in E and the edge weights in W, which is undesirable. In order to discard unreasonable projections like $\mathcal{P}_q^1$, we encode in the form of the axioms of Projection and Transformation, desirable properties that a map $\mathcal{P}_q$ should satisfy.

The space of q-metric spaces $\mathcal{M}_q$ is a subspace of the space of all networks $\mathcal{N}$; see Section II. Formally, we associate every q-metric space $(V, d) \in \mathcal{M}_q$ to the complete network (V, E, W_d) where the edge set contains every possible edge except self-loops, i.e. $(x, y) \in E$ for all $x \neq y \in V$. Furthermore, the edge weights W_d are given by the q-metric function d, i.e., $W_d(x, y) = d(x, y)$. Throughout the paper, we represent q-metric spaces as (V, d) or its network equivalent (V, E, W_d) interchangeably. Based on this observation, we define the following axiom:

(A1) Axiom of Projection. The q-metric space $M \in \mathcal{M}_q$ is a fixed point of the projection map $\mathcal{P}_q$, i.e. $\mathcal{P}_q(M) = M$.

Given that our goal is the design of maps that transform general networks into more structured q-metric spaces, if we already have a q-metric space there is no justification to change it; see Fig. 1. This concept is captured in axiom (A1) where we define $\mathcal{M}_q$ as the fixed set of $\mathcal{P}_q$. Equivalently, we say that the map $\mathcal{P}_q$ restricted to $\mathcal{M}_q$ is the identity map. It is immediate that axiom (A1) implies idempotency of $\mathcal{P}_q$, which is a requirement of projection maps, hence, its denomination as Axiom of Projection.

The second restriction on the space of allowable maps $\mathcal{P}_q$ formalizes our expectations for the behavior of $\mathcal{P}_q$ when confronted with a transformation of the underlying node space V, edge space E, and edge weights W. Consider networks $G = (V, E, W)$ and $G' = (V', E', W')$, and denote by $M = \mathcal{P}_q(G) = (V, d)$ and $M' = \mathcal{P}_q(G') = (V', d')$ the corresponding q-metric space outputs when applying $\mathcal{P}_q$ to G and G', respectively. If we can map all the nodes of the network G into the nodes of the network G' in such
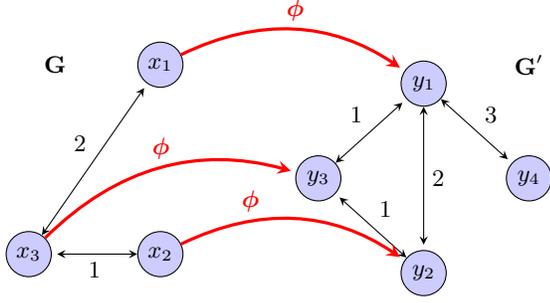
Fig. 2: Dissimilarity-reducing map. The injective map ϕ takes every edge in network G to an edge in network G' of less or equal weight.

a way that no edge weight is increased, then we expect the points in M' to be closer to each other than those in M. In order to formalize this notion, we introduce the following concept: given two networks $G = (V, E, W)$ and $G' = (V', E', W')$, the *injective* map $\phi : V \rightarrow V'$ is called *dissimilarity-reducing* map if it holds that $(\phi(x), \phi(x')) \in E'$ and $W(x, x') \geq W(\phi(x), \phi(x'))$ for all $(x, x') \in E$; see Fig. 2. A dissimilarity-reducing map matches every edge in G with an edge in G' of less or equal weight.

Intuitively, if we look at any path $P_{xx'}$ between nodes x and x' in G, the existence of the dissimilarity-reducing map ϕ ensures that there is an associated path $P_{\phi(x)\phi(x')}$ in G' between the nodes $\phi(x)$ and $\phi(x')$ such that the weight of every link in this second path is not greater than the corresponding links in the first one. On top of this, there might exist additional paths between $\phi(x)$ and $\phi(x')$ in G' that are not the image under ϕ of any path between x and x' in G. Thus, it is expected for nodes $\phi(x)$ and $\phi(x')$ to be closer to each other in the output q-metric spaces.

The Axiom of Transformation that we introduce next is a formal statement of the intuition described above:

(A2) Axiom of Transformation. Consider any two networks $G = (V, E, W)$ and $G' = (V', E', W')$ and any dissimilarity-reducing map $\phi : V \rightarrow V'$. Then, for all $x, x' \in V$, the output q-metric spaces $(V, d) = \mathcal{P}_q(G)$ and $(V', d') = \mathcal{P}_q(G')$ satisfy

$$d(x, x') \geq d'(\phi(x), \phi(x')). \qquad (6)$$

We say that a projection $\mathcal{P}_q$ is *admissible* if it satisfies axioms (A1) and (A2). In the next section, we study the space of admissible projections $\mathcal{P}_q$ for every q.

Remark 1 (Related axiomatic constructions) In [25], [27] the authors propose three axioms to study hierarchical clustering of metric spaces. Given the relation between hierarchical clustering and ultrametrics [25], [26], such a problem can be reformulated in terms of our notation as the definition of maps from $\mathcal{M}_1$ to $\mathcal{M}_\infty$. In our current work, the domain set is $\mathcal{N}$ (which is a superset of $\mathcal{M}_1$) and we use a unified framework to study projections onto all $\mathcal{M}_q$, of which a particular case is $q = \infty$. The axioms of Projection and Transformation are related to two of the axioms in [25], [27]. In Section IV-B we

show that, using our more general framework, we recover the unicity result in [25], [27] with less stringent axioms. In [26], the authors use a similar axiomatic framework to study the projection of asymmetric networks onto $\mathcal{M}_\infty$. The network model in [26] is different to the one in the current paper. It is more restrictive in the sense that complete graphs are assumed but, at the same time, more general since directed graphs are allowed. Extensions of the axiomatic framework here presented for the projection of *directed* networks onto general q-metric spaces are left as future work.

IV. UNIQUENESS OF METRIC PROJECTIONS

After posing the axioms of Projection and Transformation, we first seek to answer if *any* map $\mathcal{P}_q$ satisfies them. In this direction, given a graph $G = (V, E, W)$, for each q we define the *canonical q-metric projection* $\mathcal{P}_q^*$ with output $(V, d_q^*) = \mathcal{P}_q^*(G)$ where the q-metric d_q^* between points x and x' is given by

$$d_q^*(x, x') = s_{\|\cdot\|_q}(x, x') = \min_{P_{xx'}} h_{\|\cdot\|_q}(P_{xx'}). \qquad (7)$$

In (7), to find the distance between two points, we look for the path that links these nodes while minimizing its q-norm, as defined in (2).

For the method $\mathcal{P}_q^*$ to be a properly defined q-metric projection for all q, we need to establish that (V, d_q^*) is a valid q-metric space. Furthermore, it can also be shown that $\mathcal{P}_q^*$ satisfies axioms (A1)-(A2). We prove both assertions in the following proposition.

Proposition 1 *The canonical q-metric projection map $\mathcal{P}_q^*$ is valid and admissible. I.e., d_q^* defined by (7) is a q-metric for all graphs G and $\mathcal{P}_q^*$ satisfies the axioms of Projection (A1) and Transformation (A2).*

Proof: We first prove that d_q^* is indeed a q-metric on the node space V. That $d_q^*(x, x') = d_q^*(x', x)$ follows from combining the facts that the original graph G is symmetric and that the norms $\|\cdot\|_q$ are symmetric in the elements of the vectors for all q. Moreover, that $d_q^*(x, x') = 0$ if and only if $x = x'$ is a consequence of the positiveness property of the q-norms. To verify that the q-triangle inequality holds, let $P_{xx'}$ and $P_{x'x''}$ be paths that achieve the minimum in (7) for $d_q^*(x, x')$ and $d_q^*(x', x'')$, respectively. Then, from definition (7) it follows that

$$\begin{aligned} d_q^*(x, x'')^q &= \min_{P_{xx''}} h_{\|\cdot\|_q}(P_{xx''})^q \leq h_{\|\cdot\|_q}(P_{xx'} \oplus P_{x'x''})^q \\ &= h_{\|\cdot\|_q}(P_{xx'})^q + h_{\|\cdot\|_q}(P_{x'x''})^q \\ &= d_q^*(x, x')^q + d_q^*(x', x'')^q, \qquad (8) \end{aligned}$$

where the inequality follows from the fact that the concatenated path $P_{xx'} \oplus P_{x'x''}$ is *a particular* path between x and x'' while the definition of $d_q^*(x, x'')$ minimizes the norm across all such paths.

To see that the Axiom of Projection (A1) is satisfied, pick an arbitrary q-metric space $M = (V, d) \in \mathcal{M}_q$ and denote by $(V, d_q^*) = \mathcal{P}_q^*(M)$ the output of applying the canonical q-metric projection to M. For an arbitrary pair of nodes $x, x' \in$

V, we have that

$$d_q^*(x,x') = \min_{P_{xx'}} h_{\|\cdot\|_q}(P_{xx'}) \le h_{\|\cdot\|_q}([x,x']) = d(x,x'), \quad (9)$$

for all q, where the inequality comes from specializing the path $P_{xx'}$ to the path $[x,x']$ with just one link from x to x'. Moreover, if we denote by $P_{xx'}^* = [x = x_0, x_1, \ldots, x_l = x']$ the path achieving the minimum in (9), then we may leverage the fact that d satisfies the q-triangle inequality to write

$$d(x,x') \le \left(\sum_{i=0}^{l-1} d(x_i,x_{i+1})^q\right)^{1/q} = h_{\|\cdot\|_q}(P_{xx'}^*) = d_q^*(x,x'). \quad (10)$$

Upon substituting (10) into (9), we obtain that all the inequalities are, in fact, equalities, implying that $d_q^*(x,x') = d(x,x')$. Since nodes x,x' were chosen arbitrarily, it must be that $d \equiv d_q^*$ which implies that $\mathcal{P}_q^*(M) = M$, as wanted.

To show fulfillment of axiom (A2), consider two networks $G = (V,E,W)$ and $G' = (V',E',W')$ and a dissimilarity-reducing map $\phi : V \to V'$. Let $(V,d_q) = \mathcal{P}_q^*(G)$ and $(V',d_q') = \mathcal{P}_q^*(G')$ be the outputs of applying the canonical projection to networks G and G', respectively. For an arbitrary pair of nodes $x,x' \in V$, denote by $P_{xx'}^* = [x = x_0, \ldots, x_l = x']$ a path that achieves the minimum in (7) so as to write

$$d_q(x,x') = h_{\|\cdot\|_q}(P_{xx'}^*). \quad (11)$$

Consider the transformed path $P_{\phi(x)\phi(x')}^* = [\phi(x) = \phi(x_0), \ldots, \phi(x_l) = \phi(x')]$ in the space V'. Since the transformation ϕ does not increase dissimilarities, we have that for all links in this path $W'(\phi(x_i),\phi(x_{i+1})) \le W(x_i,x_{i+1})$. Combining this observation with (11) we obtain,

$$h_{\|\cdot\|_q}(P_{\phi(x)\phi(x')}^*) \le d_q(x,x'). \quad (12)$$

Further note that $P_{\phi(x)\phi(x')}$ is a particular path joining $\phi(x)$ and $\phi(x')$ whereas the metric d_q' is given by the minimum across all such paths. Therefore,

$$d_q'(\phi(x),\phi(x')) \le h_{\|\cdot\|_q}(P_{\phi(x)\phi(x')}^*). \quad (13)$$

Upon replacing (12) into (13), it follows that $d_q'(\phi(x),\phi(x')) \le d_q(x,x')$, as required by the Axiom of Transformation. ∎

Given that we have shown that the canonical q-metric projection $\mathcal{P}_q^*$ satisfies axioms (A1)-(A2), two questions arise: i) are there other projections satisfying (A1)-(A2)?; and ii) is the projection $\mathcal{P}_q^*$ special in any sense? Both questions are answered by the following uniqueness theorem.

Theorem 1 *Let $\mathcal{P}_q : \mathcal{N} \to \mathcal{M}_q$ be a q-metric projection, and $\mathcal{P}_q^*$ be the canonical projection with output q-metric as defined in (7). If $\mathcal{P}_q$ satisfies the axioms of Projection (A1) and Transformation (A2) then $\mathcal{P}_q \equiv \mathcal{P}_q^*$, for all q.*

Proof: Given an arbitrary network $G = (V,E,W)$ and a fixed q, denote by $(V,d) = \mathcal{P}_q(G)$ and $(V,d^*) = \mathcal{P}_q^*(G)$ the output q-metric spaces when applying a generic admissible q-metric projection and the canonical q-metric projection, respectively. We will show that

$$d^*(x,x') \le d(x,x') \le d^*(x,x'), \quad (14)$$

for all $x,x' \in V$. Given that G was chosen arbitrarily, this implies that $\mathcal{P}_q \equiv \mathcal{P}_q^*$, as wanted.

We begin by showing that $d(x,x') \le d^*(x,x')$ for all $x,x' \in V$. Consider an arbitrary pair of points x and x' and let $P_{xx'} = [x = x_0, \ldots, x_l = x']$ be a path achieving the minimum in (7) so that we can write

$$d^*(x,x') = \left(\sum_{i=0}^{l-1} W(x_i,x_{i+1})^q\right)^{1/q}. \quad (15)$$

Focus now on a series of two-node networks $G_i = (V_i,E_i,W_i)$ for $i = 0, \ldots, l-1$, such that $V_i = \{z,z'\}$ and $E_i = \{(z,z'),(z',z)\}$ for all i but with different weights given by $W_i(z,z') = W_i(z',z) = W(x_i,x_{i+1})$. Since every network G_i is already a q-metric – in fact, any two-node network is a valid q-metric for all q – and the method $\mathcal{P}_q$ satisfies the Axiom of Projection (A1), if we define $(\{z,z'\},d_i) = \mathcal{P}_q(G_i)$ we must have that $d_i(z,z') = W(x_i,x_{i+1})$, i.e., every graph G_i is a fixed point of the map $\mathcal{P}_q$.

Consider transformations $\phi_i : \{z,z'\} \to V$ given by $\phi_i(z) = x_i$, $\phi_i(z') = x_{i+1}$ so as to map z and z' in G_i to subsequent points in the path $P_{xx'}$ used in (15). This implies that maps ϕ_i are dissimilarity-reducing since they are injective and the only edge in G_i is mapped to an edge of the exact same weight in G for all i. Thus, it follows from the Axiom of Transformation (A2) that

$$d(\phi_i(z),\phi_i(z')) = d(x_i,x_{i+1}) \le d_i(z,z') = W(x_i,x_{i+1}). \quad (16)$$

To complete the proof we use the fact that since d is a q-metric and $P_{xx'} = [x = x_0, \ldots, x_l = x']$ is a path joining x and x', the q-triangle inequality dictates that

$$d(x,x') \le \left(\sum_{i=0}^{l-1} d(x_i,x_{i+1})^q\right)^{1/q} \le \left(\sum_{i=0}^{l-1} W(x_i,x_{i+1})^q\right)^{1/q}, \quad (17)$$

where we used (16) for the second inequality. The proof that $d(x,x') \le d^*(x,x')$ follows from substituting (15) into (17).

We now show that $d(x,x') \ge d^*(x,x')$ for all $x,x' \in V$. To do this, first notice that for an arbitrary pair of points x and x', if the edge $(x,x') \in E$ then we have that

$$d^*(x,x') = \min_{P_{xx'}} h_{\|\cdot\|_q}(P_{xx'}) \le W(x,x'), \quad (18)$$

where the inequality comes from considering the particular path $P_{xx'}$ with only two points $[x,x']$. Hence, the identity map $\phi : V \to V$ such that $\phi(x) = x$ for all $x \in V$ is a dissimilarity reducing map from G to (V,d^*), since it is injective and every existing edge in G is mapped to an edge with smaller or equal weight. Consequently, we can build the diagram of relations between spaces depicted in Fig. 3. The top (blue) and left (red) maps in the figure are given by the definitions at the beginning of this proof while the relation on the right (green) is a consequence of the axiom (A1). Since the aforementioned identity map ϕ is dissimilarity reducing, we can use the fact that $\mathcal{P}_q$ satisfies axiom (A2) to say that

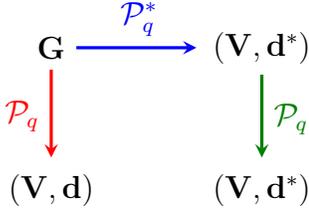$$d(x,x') \ge d^*(\phi(x),\phi(x')) = d^*(x,x'), \quad (19)$$

Fig. 3: Diagram of maps between spaces for the proof of Theorem 1.

for all $x, x' \in V$, concluding the proof. ∎

According to Theorem 1, for each q there is one and only one projection map from the space of networks onto $\mathcal{M}_q$ that satisfies the axioms of Projection and Transformation, and this map is $\mathcal{P}_q^*$. Any other conceivable map into $\mathcal{M}_q$ must violate at least one of the axioms. E.g., if we have $q > q'$ then $\mathcal{P}_q^*$ can be viewed as a map into $\mathcal{M}_{q'}$ since $\mathcal{M}_q \subset \mathcal{M}_{q'}$. However, map $\mathcal{P}_q^*$ violates axiom (A1) when viewed as a map into $\mathcal{M}_{q'}$, since any q'-metric space $M \in \mathcal{M}_{q'}$ that is *not* in $\mathcal{M}_q$ would not be a fixed point of $\mathcal{P}_q^*$.

A. Metric spaces

For an arbitrary network $G = (V, E, W)$, we may particularize our analysis to the case of (regular) metric spaces, i.e. $q = 1$. In this case, the canonical projection $\mathcal{P}_1^*$ outputs the metric space $(V, d_1^*) = \mathcal{P}_1^*(G)$ where d_1^* is given by [cf. (7)]

$$d_1^*(x, x') = \min_{P_{xx'}} \sum_{i=0}^{l-1} W(x_i, x_{i+1}), \qquad (20)$$

for all $x, x' \in V$. Equivalently, $\mathcal{P}_1^*$ sets the distance between two nodes to the length of the shortest path between them.

By specializing Theorem 1 to the case $q = 1$, we obtain the following corollary.

Corollary 1 *The shortest path between every pair of nodes is the only admissible metric in networks, where admissibility is given by axioms (A1)-(A2).*

From the comparison of networks [28] to the determination of node importance [29], [30], shortest paths constitute a basic feature of network analysis and their application is ubiquitous. Moreover, efficient algorithms for the computation of every shortest path in a network exist [31], [32]. In Section V-A, we discuss the utility of projecting graphs onto metrics for the approximation of otherwise NP-hard graph theoretical problems that have guaranteed error for metric data [33], [34].

B. Ultrametric spaces

The equivalence between dendrograms and ultrametric spaces is a well-established fact [25], [26]. Hence, hierarchical clustering of networks [23], [35], [36], i.e., the mapping of networks onto dendrograms, can be posed as a projection problem of networks onto ultrametric spaces. When setting

$q = \infty$, the canonical projection induces the ultrametric d_∞^* given by [cf. (7)]

$$d_\infty^*(x, x') = \min_{P_{xx'}} \max_i W(x_i, x_{i+1}), \qquad (21)$$

for all $x, x' \in V$. Expression (21) implies that the canonical projection $\mathcal{P}_\infty^*$ is equivalent to single linkage hierarchical clustering [23, Ch. 4]. Moreover, for the case $q = \infty$, the ensuing corollary follows from Theorem 1.

Corollary 2 *Single linkage hierarchical clustering is the only admissible hierarchical clustering method for networks, where admissibility is given by axioms (A1)-(A2).*

Single linkage was previously shown to posses desirable theoretical features. In [25], single linkage is shown to be the only admissible method from metric spaces $\mathcal{M}_1$ to ultrametric spaces $\mathcal{M}_\infty$ satisfying three axioms, two of which can be derived from the axioms of Projection and Transformation here presented. This implies that, when specializing our framework for $q = \infty$, there are two main advantages between the axiomatic framework derived here and that in [25]: i) single linkage is the only admissible method from $\mathcal{N}$ to $\mathcal{M}_\infty$, and not just from $\mathcal{M}_1$ to $\mathcal{M}_\infty$; and ii) the third axiom considered in [25] is redundant, since the uniqueness result can be derived based solely on the first two.

In practice, single linkage has shown to have some undesirable features like the so-called chaining effect [37]. Nevertheless, our construction is of utility for the practitioner who prefers other hierarchical clustering methods. More specifically, by clearly stating our desired properties as axioms, it is made clear that at least one of the axioms (A1)-(A2) must be violated when picking a method different from single linkage.

C. 2-metric spaces

Metric and ultrametric spaces are the two most common examples of q-metric spaces. However, for intermediate values of q, i.e., between 1 and ∞, spaces with other special characteristics arise. In particular, when $q = 2$, we obtain a space in which every triangle is acute. More specifically, when the distances between points satisfy the 2-triangle inequality, it can be shown that every angle in every triangle is not greater than 90 degrees. Mimicking the reasoning in Sections IV-A and IV-B, it follows that the canonical projection $\mathcal{P}_2^*$ with an associated induced distance between points x and x' given by

$$d_2^*(x, x') = \min_{P_{xx'}} \sqrt{\sum_{i=0}^{l-1} W(x_i, x_{i+1})^2}, \qquad (22)$$

is the only admissible way of inducing a structure space in a network where every resulting triangle is acute.

V. PROPERTIES OF THE CANONICAL PROJECTION

The axioms of Projection and Transformation (A1)-(A2) uniquely determine the family of canonical projections $\mathcal{P}_q^*$ for different q. Moreover, additional practical properties can be extracted from the aforementioned axioms. In this section we discuss the properties of optimality (Section V-A) and nestedness (Section V-B).
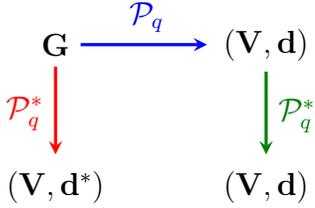
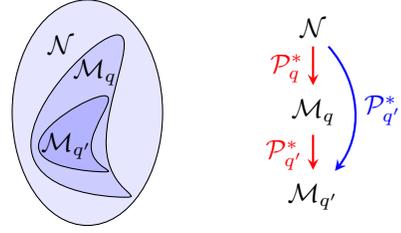Fig. 4: Diagram of maps between spaces for the proof of Proposition 2.

Fig. 5: The canonical projection of a network in $\mathcal{N}$ onto $\mathcal{M}_{q'}$ (blue) is invariant to intermediate canonical projections onto spaces $\mathcal{M}_q$ (red) for $q \leq q'$.

A. Optimality

A myriad of combinatorial optimization problems exist, where the goal is to find subsets of nodes or edges of a network that are optimal in some sense. Examples of the former are graph coloring – finding a partition of non-adjacent nodes with smallest cardinality [38] – and the maximum independent set problem – finding a set of non-adjacent nodes with maximal cardinality [39]. Examples of the latter include the traveling salesman problem – finding a path that visits each node exactly once with smallest length [40] – and the minimum bisection problem [41] – separating the network in two pieces with the same number of nodes so that the sum of the weights in the edges that connect the pieces is minimal. In this section we focus on this second category, where the problems are characterized by an objective function that depends on the weights of the edges of the network.

Define the function $f : \mathcal{N} \to \mathbb{R}$ that maps every network G to the minimum cost $f(G)$ of an optimization problem that depends on the structure of G. For the traveling salesman problem, $f(G)$ is the length of the optimal salesman's trajectory in G. For minimum bisection, $f(G)$ is the sum of the weights in the optimal bisection of G. Traveling salesman and minimum bisection are known to be NP-hard and also hard to approximate in general. This means that not only is the problem of finding optimal solutions computationally intractable, but the problem of finding approximate solutions in polynomial time is impossible as well – impossibility unless P=NP is known for the traveling salesman problem [42] and undetermined for the minimum bisection problem [43]. However, when the network under consideration is metric, both problems are approximable in polynomial time [11], [33]. These two examples are not isolated, there are many other combinatorial problems that are approximable when we restrict our attention to metric networks [34].

We can leverage the fact that combinatorial problems are simpler to solve in metric spaces to efficiently obtain lower bounds for $f(G)$. More specifically, we restrict our attention to cost functions f that do not decrease with increasing edge weights, i.e. for networks with the same number of nodes $f(G') \geq f(G)$ if the identity map is dissimilarity reducing from G' to G. Then, if we project a network G onto a metric space M where no dissimilarity is increased, we may compute the lower bound $f(M)$ efficiently. The optimal choice for this projection is the canonical map $\mathcal{P}_1^*$ as we show next in a more general proposition for q-metric spaces.

Proposition 2 *Given an arbitrary network $G = (V, E, W)$, let $\mathcal{P}_q : \mathcal{N} \to \mathcal{M}_q$ be a generic q-metric projection with output $(V, d) = \mathcal{P}_q(G)$. Then, for any cost function f non-decreasing in the edge weights of G, the canonical projection $\mathcal{P}_q^*$ satisfies*

$$\mathcal{P}_q^* = \underset{\mathcal{P}_q}{\arg\min}\, f(G) - f(\mathcal{P}_q(G)) \qquad (23)$$

$$\text{s.to} \qquad d(x, x') \leq W(x, x') \quad \text{for all } (x, x') \in E.$$

Proof: That $\mathcal{P}_q^*$ is feasible, meaning that its output q-metric $(V, d^*) = \mathcal{P}_q^*(G)$ satisfies the constraint in problem (23), can be shown using the same argument used to write expression (18). To see that $\mathcal{P}_q^*$ is optimal, denote by $\mathcal{P}_q$ a feasible q-metric projection with output $(V, d) = \mathcal{P}_q(G)$. The diagram in Fig. 4 summarizes the relations between G, (V, d), and (V, d^*). The top (blue) and left (red) maps represent the definitions of the metric projections. The right (green) map is justified by the fact that (V, d) is, by definition, a q-metric and that $\mathcal{P}_q^*$ satisfies the Axiom of Projection (cf. Prop. 1). Moreover, notice that d satisfying the constraint in (23) guarantees that the identity map $\phi : V \to V$ from G to (V, d) is dissimilarity-reducing. Consequently, we combine the fact that $\mathcal{P}_q^*$ fulfills the Axiom of Transformation (cf. Prop. 1) with the relations between spaces in Fig. 4 to write

$$d^*(x, x') \geq d(\phi(x), \phi(x')) = d(x, x'), \qquad (24)$$

for all $x, x' \in V$. Combining (24) with the constraint in problem (23), we can write that

$$d(x, x') \leq d^*(x, x') \leq W(x, x'), \qquad (25)$$

for all $(x, x') \in E$, and optimality of $\mathcal{P}_q^*$ follows from the non-decreasing nature of the cost function f. ∎

The optimality result in Proposition 2 provides an efficient way to bound the minimum cost of a combinatorial optimization problem. First, an *upper* bound can be achieved by finding a series of feasible solutions to the problem, e.g., particular circuits for the traveling salesman or cuts for the bisection problem, possibly aided by heuristics designed for the particular problem of interest. Second, we apply the canonical projection $\mathcal{P}_1^*$ to the network under analysis and solve efficiently the combinatorial problem in the metric space [34], with the guarantee that obtained *lower* bound is the tightest among those achievable via a metric projection that does not increase the dissimilarities.
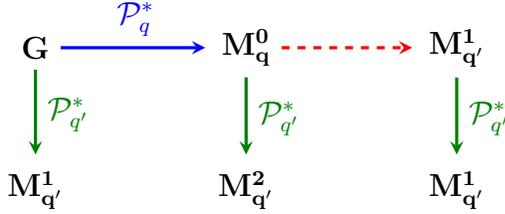
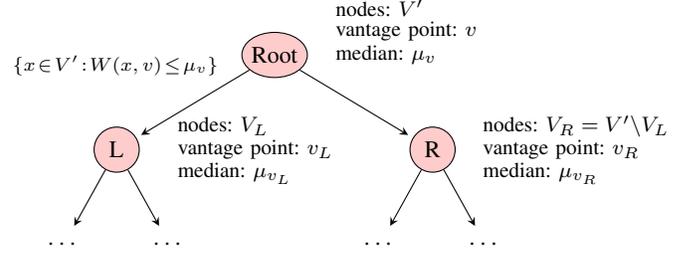Fig. 6: Diagram of maps between spaces for the proof of Proposition 3.

Fig. 7: Vantage point tree. The whole point set V' is associated with the root of the tree. A vantage point v is chosen at random and V' is partitioned into the left and right children of the root depending on the distance of each point to the vantage point. The process is repeated iteratively to construct the whole tree.

B. Nestedness

From the increasing structure imposed by the q-triangle inequality as q increases, it follows that $\mathcal{M}_{q'} \subseteq \mathcal{M}_q$ for $q' \geq q$. This implies that the canonical projection of any network onto a q'-metric space can be alternatively achieved by a direct application of $\mathcal{P}_{q'}^*$ or by first applying $\mathcal{P}_q^*$ to the network and then applying $\mathcal{P}_{q'}^*$ to the resulting q-metric space; see Fig. 6. Both approaches are equivalent, as we formally state next.

Proposition 3 *Given an arbitrary network $G = (V, E, W) \in \mathcal{N}$, we have that*

$$\mathcal{P}_{q'}^*(G) = \mathcal{P}_{q'}^*(\mathcal{P}_q^*(G)), \qquad (26)$$

for $q' \geq q$.

Proof: Define the q'-metric spaces $M_{q'}^1 = \mathcal{P}_{q'}^*(G) = (V, d_1)$ and $M_{q'}^2 = \mathcal{P}_{q'}^*(\mathcal{P}_q^*(G)) = (V, d_2)$, as well as the q-metric space $M_q^0 = \mathcal{P}_q^*(G) = (V, d_0)$. To prove the proposition, we must show that $d_1(x, x') = d_2(x, x')$ for all $x, x' \in V$. Consider the diagram in Fig. 6, where the blue map and the two left-most green maps represent the aforementioned definitions and the remaining green map is justified by the Axiom of Projection. Notice that the proposition statement is equivalent to declaring that the red dashed map is also equal to $\mathcal{P}_{q'}^*$. Assume that the identity map $\phi: V \to V$ is dissimilarity reducing from G to M_q^0 and that the same identity map is also dissimilarity reducing from M_q^0 to $M_{q'}^1$ (top row of Fig. 6). Then, from the fact that $\mathcal{P}_{q'}^*$ satisfies the Axiom of Transformation, it would follow that $d_1(x, x') \geq d_2(x, x') \geq d_1(x, x')$ for all $x, x' \in V$ (bottom row of Fig. 6), showing the desired equality. Thus, to complete the proof we need to show that the identity maps ϕ are effectively dissimilarity reducing. That ϕ is dissimilarity reducing from G to M_q^0 was shown in the proof of Theorem 1 [cf. (18)]. Finally, to see that ϕ is dissimilarity reducing from M_q^0 to $M_{q'}^1$, pick an arbitrary pair of nodes $x, x' \in V$ and let $P_{xx'} = [x = x_0, \ldots, x_l = x']$ be a path achieving the minimum in (7) for d_0 so that we can write

$$d_0(x, x') = \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^q \right)^{1/q} \qquad (27)$$

$$\geq \left(\sum_{i=0}^{l-1} W(x_i, x_{i+1})^{q'} \right)^{1/q'} \geq d_1(x, x'),$$

where the first inequality is obtained by combining the facts that $q' \geq q$ and the weights W are non-negative, and the second inequality comes from the fact that $d_1(x, x')$ is obtained by minimizing over all paths from x to x' and here we are considering a particular path $P_{xx'}$. From (27) it follows that the identity map is dissimilarity reducing from M_q^0 to $M_{q'}^1$, concluding the proof. ∎

Proposition 3 shows that the q'-metric space associated with a given network G is independent of any intermediate canonical projections to q-metric spaces for $q \leq q'$. Intuitively, this result implies that the intermediate map $\mathcal{P}_q^*$ induces part of the structure imposed by $\mathcal{P}_{q'}^*$. Thus, it is equivalent to induce the whole structure in one step by applying $\mathcal{P}_{q'}^*$ or doing it gradually by applying maps $\mathcal{P}_q^*$ for $q \leq q'$.

A direct consequence of (26) is that if one is interested in, e.g., computing the single linkage hierarchical clustering output of a given network G (cf. Section IV-B) then there is no gain (or loss) in first projecting the network G onto a metric space and then computing the clustering output of the resulting space.

VI. SEARCH IN METRIC STRUCTURES

Given a network $G = (V, E, W)$, assume that we have access only to a subset of the network $G' = (V', E', W')$ where $V' \subset V$ and E' and W' are the restrictions of E and W to V', respectively. We are then revealed an additional point $z \in V \setminus V'$ and are interested in finding the node $x \in V'$ closest to z, i.e., the node x for which $W(z, x)$ is minimized. The described setting occurs frequently in practice, e.g., in the implementation of k nearest neighbor (k-NN) methods [44, Chapter 2] where G is the dataset of interest and G' is the training set. The complexity of the mentioned task depends on how structured network G is. When no structure is present, an exhaustive search is the only option and z must be compared with every node in V'. By contrast, when G is a metric space, then the NN of z can be found efficiently by using metric trees [8]–[10]. In this section, we propose an efficient search strategy in networks by first projecting a general network onto a metric network and then leveraging this structure for search via the construction of a metric tree.

Intuitively, if z is far away from a node x in a metric space, i.e. $W(z, x)$ is large, then the triangle inequality implies that z
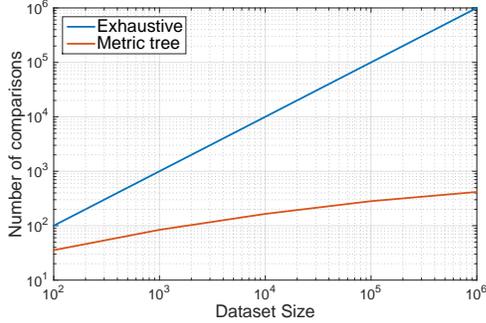
Fig. 8: Number of comparisons needed to find the nearest neighbor of a point in a metric space as a function of the space size for exhaustive search and search aided using metric trees.

will also be far away from any node x' close to x, thus, there is no need to consider nodes x' as potential candidates for the NN of z. Metric trees formally leverage this intuition by constructing hierarchical structures of V' in order to accelerate search. In this paper we focus on the vantage point tree (vp-tree) [8], one of the most popular types of metric tree. The implementation of a metric tree is a two-step process: we first construct the tree and then utilize it for (possibly multiple) queries.

To construct a vp-tree given G', we begin by associating the whole node set V' to the root of the tree and we pick a node (the vantage point) at random, say $v \in V'$. We then compute the median μ_v of the distances $W(v, x)$ from the vantage point to every other node $x \in V'$ and partition V' into two blocks: one containing the nodes whose distance to v is smaller than or equal to μ_v and the other one containing the rest of V'. The nodes in the first block are assigned to the left child of the root of the vp-tree while the right child consists of the nodes in the second block. We iteratively repeat this procedure within each of the children until every leaf in the vp-tree is associated to a single point in V'; see Fig. 7. For more details, see [8].

To efficiently search a vp-tree for the NN of a query point z, we traverse the nodes of the tree and compare z *only with the vantage point* of the current node of the vp-tree. Moreover, we leverage the triangle inequality to discard branches of the vp-tree without even traversing them, reducing the number of measurements needed to find the NN of z. More specifically, assume that we are searching at an intermediate node in the vp-tree, say node L in Fig. 7 and the current best estimate of the NN is at distance τ from z, which can be initialized as $\tau = \infty$ for the root of the vp-tree. We then compute the distance $W(z, v_L)$ between z and the vantage point v_L associated to the current node in the vp-tree. If $W(z, v_L) < \tau$, we then update our estimate of τ. In order to continue traversing the vp-tree, we follow the ensuing rules where v is the vantage point of the current node in the vp-tree

$$\begin{cases} i)\, W(z,v) \leq \mu_v - \tau \Rightarrow \text{visit only the left child,} \\ ii)\, \mu_v - \tau < W(z,v) \leq \mu_v + \tau \Rightarrow \text{visit left \& right child,} \\ iii)\, \mu_v + \tau < W(z,v) \Rightarrow \text{visit only right child.} \end{cases}$$

(28)

Even though statements $i)$ and $iii)$ in (28) entail that we discard part of the nodes in V' during our search, the way the metric tree is constructed guarantees that the NN of z is not contained among the discarded nodes.

The construction of the vp-tree, a one-time computational effort, can be shown to have complexity $\mathcal{O}(n \log n)$ where n is the cardinality of V'. However, once it is built it can be used to reduce the complexity of a brute force linear search from $\mathcal{O}(n)$ to an expected cost of $\mathcal{O}(\log n)$ [8]. To corroborate this, we construct metric spaces of varying sizes by embedding points in a square area of $\mathbb{R}^2$ and consider their Euclidean distance as the dissimilarity values W. In Fig. 8 we plot the average number of comparisons – values of W – needed to find the nearest neighbor of a query point in this metric space as a function of n for exhaustive and metric-tree search. This average is computed across 1000 queries. As expected, exhaustive search complexity grows linearly with n whereas vp-tree's complexity grows logarithmically. Notice that there is a marked difference in the number of measurements required, e.g., for $n = 10^6$ the metric tree search can be performed with an expected cost of 500 measurements.

Motivated by the computational gain depicted in Fig. 8, a possible way to search a non-metric network G is to first project it onto a metric space M via the canonical projection $M = \mathcal{P}_1^*(G)$ and then construct a vp-tree on M. Notice that this construction guarantees an efficient search of the NN in M of a given query. However, we are interested in finding the NN in G, thus, potentially committing an error. Intuitively, the furthest away the structure of G is from being metric, the larger the error in the NN found. In order to illustrate this effect, we generate metric spaces obtained by randomly embedding 1000 points in $\mathbb{R}^{100}$ and considering their Euclidean distances as dissimilarities between them. We then obtain (non-metric) perturbed versions of each metric space by multiplying a subset of the dissimilarities by $1 + \delta$ where δ is a random variable uniformly distributed in $[0, 10]$. The subset of dissimilarities to modify is chosen randomly with probability of perturbation r. In Fig. 9a we illustrate the search performance over 1000 queries as a function of r (blue lines). The dashed line illustrates the percentage of perfect recovery (left y-axis), i.e., the proportion of the 1000 queries in which the node found coincides with the actual NN of the query point. The solid and the pointed lines represent, respectively, the mean and median relative positions of the actual node found (right y-axis). E.g., a value in 0%–1% indicates that the node found is actually contained among the 10 nearest nodes (1% of 1000) to the query. Finally, to illustrate the value of the projection method proposed, we also illustrate the search performance when the vp-tree is constructed directly on G, i.e. when we apply the aforementioned construction scheme and navigation rules [cf. (28)] to G even though it is non-metric. First of all, notice that when $r = 0$, both schemes work perfectly since $G = M$ corresponds to a metric space. For other values of r, the vp-tree constructed on M (blue lines) consistently outperforms the one constructed on G (red lines). E.g., for $r = 0.6$ the median and mean relative positions of the nodes found on M are in the top 0.5% and 0.8%, respectively, which contrast with the ones found on G which are in the
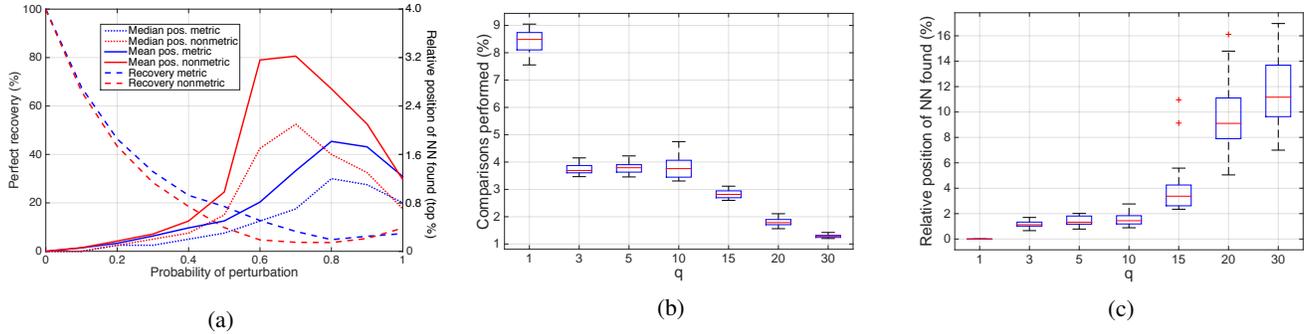
Fig. 9: (a) Percentage of perfect recovery (dashed lines, left y-axis), mean and median relative positions of search result (solid and pointed lines, right y-axis) as a function of the probability of perturbation in a metric network when the tree search is performed in the resulting non-metric space (red) and when the space is previously projected using $\mathcal{P}_1^*$ (blue). (b) Number of comparisons needed to search a metric space when projected first onto a q-metric space via $\mathcal{P}_q^*$ as a function of q. Larger q correspond to more efficient searches. (c) Search performance as indicated by the relative position of the NN found when performing the searches in (b). Smaller q correspond to more accurate searches.

top 1.7% and 3.2%, respectively. Finally, notice that for large values of r when most of the edges in G are perturbed, the structure becomes more similar to a metric space and, thus, there is an improvement in the search performance both on G and M.

When $q > 1$, q-metric spaces contain a more stringent structure than regular metric spaces and this additional structure can be used to speed up search, as we show next.

Proposition 4 *When using a vp-tree for nearest neighbor search in a q-metric space, the following rules ensure optimality of the node found [cf. (28)]*

$$\begin{cases} i)\, W(z,v)^q \leq \mu_v^q - \tau^q \Rightarrow \textit{visit only the left child,} \\ ii)\, \mu_v^q - \tau^q < W(z,v)^q \leq \mu_v^q + \tau^q \Rightarrow \textit{visit left \& right child,} \\ iii)\, \mu_v^q + \tau^q < W(z,v)^q \Rightarrow \textit{visit only right child,} \end{cases}$$
(29)

where z is the query point and v is the vantage point of the current node of the vp-tree.

Proof: We need to show that by following the rules in (29) we are not discarding any node that could be the NN of our query point z. Assume that $i)$ is true, then the q-triangle inequality implies that for all $t \in V$

$$W(v,t)^q \leq W(v,z)^q + W(z,t)^q \leq \mu_v^q - \tau^q + W(z,t)^q, \quad (30)$$

where the second inequality follows from $i)$. Furthermore, for every t belonging to the right child of the current node of the vp-tree we have (by construction) that $W(v,t) > \mu_v$. Combining this fact with (30) it follows that $W(z,t) > \tau$ for every t in the right child. This implies that every node t in the right child is at a distance from z larger than the current best estimate τ and, thus, can be discarded. Similarly, if we assume that $iii)$ is true, we may leverage the q-triangle inequality to write that for all $t \in V$

$$\mu_v^q + \tau^q < W(z,v)^q \leq W(z,t)^q + W(t,v)^q. \quad (31)$$

If we combine (31) with the fact that for every t in the left child $W(t,v) \leq \mu_v$, it follows that $W(z,t) > \tau$ for every t in the left child, concluding the proof. ∎

Notice that when $q = 1$, the conditions in (29) boil down to those in (28), as expected. Further, the range encompassed by (29)-$ii)$ is smaller than that in (28)-$ii)$, and decreases in size with larger q. To see this, compute the q-th of (28)-$ii)$ and, from the fact that μ_v and τ are positive, it follows that $\mu_v^q + \tau^q \leq (\mu_v + \tau)^q$ showing that the upper bound in (29)-$ii)$ is tighter. Similarly, whenever $0 \leq \tau \leq \mu_v$ – otherwise, both lower bounds are effectively zero –, we have that $(\mu_v - \tau)^q \leq \mu_v^q - \tau^q$, showing that the lower bound in (29)-$ii)$ is tighter as well compared to the one in (28)-$ii)$. This implies that in more structured spaces – larger q – it is more likely to discard parts of the vp-tree – satisfying conditions $i)$ or $iii)$ – speeding up the search. Based on this observation, one can project a metric space onto q-metric spaces with $q > 1$ in order to increase search speed with the cost of decreasing search performance due to the deformation introduced when projecting the original metric space. Figs. 9b and 9c illustrate this tradeoff. We generate 20 metric spaces with 1000 points each and perform 100 queries in each metric space. In Fig. 9b we illustrate the distribution of the 20 averages of the number of comparisons needed (as a percentage of 1000) to perform the search when first projecting the metric space applying $\mathcal{P}_q^*$ for varying q. In Fig. 9c we plot the relative position of the NN found as a function of q. Notice that when $q = 1$, the node found is always the correct one (0%) since the original space was chosen metric. Furthermore, when, e.g., $q = 3$ the number of measurements needed to search the vp-tree is reduced to around half of those needed for metric spaces (cf. Fig. 9b) while the neighbors found are within the top 2% candidates among the 1000 points (cf. Fig. 9c). For large values of q, the reduction in computation is noticeable – around 8 times for $q = 30$ – but the detriment in performance is also large. Depending on the application, the value of q can be tuned to find the correct equilibrium between computational efficiency and search performance.

VII. Conclusion

We analyzed how to project networks onto q-metric spaces, a generalization of metric spaces that encompasses a larger class of structured representations. We defined the axioms of Projection and Transformation as desirable properties of the projection methods and showed that for each q-metric space there is a unique canonical way of projecting any network onto it. In particular, this axiomatic framework provided theoretical support for the computation of shortest path distances between nodes as well as for single linkage hierarchical clustering of networked datasets. Moreover, we showed that q-metric projections can be used in practice to approximate combinatorial optimization problems in graphs and to efficiently search a network for the nearest neighbor of a given query point.

References

[1] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, H. Lu, J. Zhang, S. Sun, L. Ling, and N. Zhang, "Topological structure analysis of the protein–protein interaction network in budding yeast," *Nucleic Acids Res.*, vol. 31, no. 9, pp. 2443–2450, 2003.

[2] E. Lieberman, C. Hauert, and M. Nowak, "Evolutionary dynamics on graphs," *Nature*, vol. 433, no. 7023, pp. 312–316, 2005.

[3] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Phys. Rev. E*, vol. 74, no. 3, p. 036104, 2006.

[4] J. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[5] ——, "Complex networks and decentralized search algorithms," in *Proc. Intl. Cong. Math. (ICM)*, vol. 3, 2006, pp. 1019–1044.

[6] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," in *Proc. ACM Symp. Theo. Comp. (STOC)*, 2004, pp. 561–568.

[7] N. Lynch, *Distributed Algorithms*. Morgan Kaufmann Publishers Inc., 1996.

[8] P. N. Yianilos, "Data structures and algorithms for nearest neighbor search in general metric spaces," in *Proc. ACM-SIAM Symp. Disc. Algo. (SODA)*, 1993, pp. 311–321.

[9] J. K. Uhlmann, "Satisfying general proximity/similarity queries with metric trees," *Info. Process. Lett.*, vol. 40, no. 4, pp. 175 – 179, 1991.

[10] C. Traina, A. Traina, B. Seeger, and C. Faloutsos, "Slim-trees: High performance metric trees minimizing overlap between nodes," in *Adv. Database Tech. (EDBT)*, ser. Lecture Notes in Computer Science, C. Zaniolo, P. Lockemann, M. Scholl, and T. Grust, Eds. Springer Berlin Heidelberg, 2000, vol. 1777, pp. 51–65.

[11] N. Christofides, "Worst-case analysis of a new heuristic for the travelling salesman problem," DTIC Document, Tech. Rep., 1976.

[12] M. Frechet, "Sur quelques points du calcul fonctionnel," *Rendiconti del Circolo Matematico di Palermo*, vol. 22, no. 1, pp. 1–72, 1906.

[13] L. M. Blumenthal, *Distance Geometry: A Study of the Development of Abstract Metrics*. University of Missouri Studies, 1938, vol. 13.

[14] J. R. Isbell, "Six theorems about injective metric spaces," *Commentarii Mathematici Helvetici*, vol. 39, no. 1, pp. 65–76, 1964.

[15] M. Barr and C. Wells, *Category Theory for Computing Science*. Prentice Hall, 1990.

[16] M. A. A. Cox and T. F. Cox, "Multidimensional scaling," in *Handbook of Data Visualization*, ser. Springer Handbooks Comp.Statistics. Springer Berlin Heidelberg, 2008, pp. 315–347.

[17] W. S. Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[18] S. J. Messick and R. P. Abelson, "The additive constant problem in multidimensional scaling," *Psychometrika*, vol. 21, no. 1, pp. 1–15, 1956.

[19] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964.

[20] R. N. Shepard, "The analysis of proximities: Multidimensional scaling with an unknown distance function. I." *Psychometrika*, vol. 27, no. 2, pp. 125–140, 1962.

[21] J. B. Kruskal and M. Wish, *Multidimensional Scaling*. Sage, 1978.

[22] N. Jardine and R. Sibson, *Mathematical Taxonomy*. John Wiley & Sons Ltd., 1971.

[23] A. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice Hall Inc., 1988.

[24] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*. American Mathematical Soc., 2001.

[25] G. Carlsson and F. Mémoli, "Characterization, stability and convergence of hierarchical clustering methods," *J. Mach. Learn. Res. (JMLR)*, vol. 11, pp. 1425–1470, 2010.

[26] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," *arXiv preprint arXiv:1301.7724*, 2014.

[27] G. Carlsson and F. Mémoli, "Classifying clustering schemes," *Found. Comp. Math.*, vol. 13, no. 2, pp. 221–252, 2013.

[28] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059 – 1069, 2010.

[29] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[30] S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," *arXiv preprint arXiv:1410.5119*, 2014.

[31] R. W. Floyd, "Algorithm 97: shortest path," *Commun. ACM*, vol. 5, no. 6, p. 345, 1962.

[32] S. Warshall, "A theorem on boolean matrices," *J. ACM*, vol. 9, no. 1, pp. 11–12, 1962.

[33] W. Fernandez de la Vega, M. Karpinski, and C. Kenyon, "Approximation schemes for metric bisection and partitioning," in *Proc. ACM-SIAM Symp. Disc. Algo. (SODA)*, 2004, pp. 506–515.

[34] P. Indyk, "Sublinear time algorithms for metric space problems," in *Proc. ACM Symp. Theo. Comp. (STOC)*, 1999, pp. 428–434.

[35] G. N. Lance and W. T. Williams, "A general theory of classificatory sorting strategies 1. Hierarchical systems," *Computer J.*, vol. 9, no. 4, pp. 373–380, 1967.

[36] Y. Zhao and G. Karypis, "Hierarchical clustering algorithms for document datasets," *Data Min. Knowl. Discov.*, vol. 10, pp. 141–168, 2005.

[37] R. Sibson, "SLINK: an optimally efficient algorithm for the single-link cluster method," *Computer J.*, vol. 16, no. 1, pp. 30–34, 1973.

[38] T. R. Jensen and B. Toft, *Graph Coloring Problems*. John Wiley & Sons, 2011.

[39] R. E. Tarjan and A. E. Trojanowski, "Finding a maximum independent set," *SIAM J. Comput.*, vol. 6, no. 3, pp. 537–546, 1977.

[40] E. Lawler, J. Lenstra, and A. Rinnooy Kan, "The traveling salesman problem: a guided tour of combinatorial optimization," *Wiley Intersc. Ser. in Disc. Math.*, 1985.

[41] T. Bui, S. Chaudhuri, F. Leighton, and M. Sipser, "Graph bisection algorithms with good average case behavior," *Combinatorica*, vol. 7, no. 2, pp. 171–191, 1987.

[42] S. Sahni and T. Gonzalez, "P-complete approximation problems," *J. ACM*, vol. 23, no. 3, pp. 555–565, 1976.

[43] U. Feige and R. Krauthgamer, "A polylogarithmic approximation of the minimum bisection," *SIAM J. Comput.*, vol. 31, no. 4, pp. 1090–1118, 2002.

[44] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006.