METRIC REPRESENTATIONS OF NETWORKS: A UNIQUENESS RESULT

Santiago Segarra, T. Mitchell Roddenberry, Facundo Mémoli, and Alejandro Ribeiro

ABSTRACT

In this paper, we consider the problem of projecting networks onto metric spaces. Networks are structures that encode relationships between pairs of elements or nodes. However, these relationships can be independent of each other, and need not be defined for every pair of nodes. This is in contrast to a metric space, which requires that a distance between *every* pair of elements in the space be defined. To understand how to project networks onto metric spaces, we take an axiomatic approach: we first state two axioms for projective maps from the set of all networks to the set of finite metric spaces, then show that only one projection satisfies these requirements. The developed technique is shown to be an effective method for finding approximate solutions to combinatorial optimization problems. Finally, we illustrate the use of metric trees for efficient search in projected networks.

***Index Terms*—** Networks, metric space, combinatorial optimization, axiomatic framework, nearest-neighbor search.

1. 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, including interactions between proteins or organisms in biology [1, 2], individuals or institutions in sociology [3, 4], and neurons or brain regions in neuroscience [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. 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. 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 is approximable in polynomial time to within a factor of 3/2 in metric spaces [10]. 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 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 onto the subset of networks that represent metric spaces. This is the problem addressed in this paper.

Related work and contributions. The traditional way of mapping a generic dissimilarity function between pairs of points to a metric space is through multidimensional scaling (MDS) [11]. Different problem formulations give rise to the definition of different types of MDS with a basic distinction between metric MDS, where the input consists of quantitative dissimilarities [12, 13], and non-metric MDS where dissimilarities can be ordinal [14, 15]. However, all these techniques have in common that the end goal is to facilitate visualization of the data [16]. Thus, MDS embeds the input dissimilarities onto familiar and low-dimensional metric spaces such as $\mathbb{R}^2$ or $\mathbb{R}^3$. Similarly, there is a whole gamut of *node embedding* methods [17–21] that embed the node set into $\mathbb{R}^d$ for some relatively small d based on the structure encoded by the network and, possibly, features on the nodes.

In this paper, we propose a new perspective to the problem of mapping generic dissimilarity networks to metric spaces. In particular, we follow an *axiomatic approach*, i.e., we encode as axioms desirable properties of the projections. This approach was originally inspired by [22, 23], where the authors propose three axioms to study hierarchical clustering of metric spaces. Given the relation between hierarchical clustering and ultrametrics [24–26], such a problem can be recast as projecting networks onto finite ultrametric spaces. In our current work, however, we study the projections of networks onto (regular) metric spaces.

The contributions of this paper are threefold. First, we introduce an axiomatic approach for the study of metric projections and show that the proposed axioms imply a unique admissible way of performing these projections, which is based on shortest paths in the network. Second, we show that this (canonical) projection is also optimal in a well-defined sense, thus helping to approximate certain classes of combinatorial problems in graphs. Lastly, we illustrate how the canonical metric representation of a network can be used to perform fast (approximate) search in the network.

2. NETWORKS AND METRIC SPACES

We consider weighted, undirected graphs or *networks*. A *graph* $G = (V, E, W)$ is a triplet formed by a finite set of n nodes or vertices V, a set of edges $E \subseteq V \times V$ where $(x, y) \in E$ represents an edge from $x \in V$ to $y \in V$, and a map $W\colon E \to \mathbb{R}_{++}$ from the set of edges to the strictly positive real numbers. Since G is assumed to be undirected, $(x, y) \in E$ implies $(y, x) \in E$, and similarly $W(x, y) = W(y, x)$. Moreover, the graphs considered do not contain self-loops, i.e., $(x, x) \notin E$ for all $x \in V$. Denote by $\mathcal{N}$

S. Segarra and T.M. Roddenberry are with the Department of Electrical and Computer Engineering, Rice University. F. Mémoli is with the Department of Mathematics and the Department of Computer Science and Engineering, Ohio State University. A. Ribeiro is with the Department of Electrical and Systems Engineering, University of Pennsylvania. Emails: segarra@rice.edu, mitch@rice.edu, memoli@math.osu.edu, aribeiro@seas.upenn.edu.

the set of all networks, where networks in $\mathcal{N}$ can have different node sets V, different edge sets E, or different weights W.

Given a finite set X, a *metric* $d\colon X \times X \to \mathbb{R}_+$ is a function 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: $d(x, y) \leq d(x, z) + d(z, y)$.

The ordered pair (X, d) is said to be a *finite metric space* [27], and the set of all finite metric spaces is denoted as $\mathcal{M}$. In this paper we treat finite metric spaces (X, d) as networks of the form $(X, X \times X, d)$. One can observe that $\mathcal{M} \subset \mathcal{N}$.

To define metric projections for networks, it is important to define paths and path lengths. For a network $G = (V, E, W)$ and $x, x' \in V$, a *path* $P_{xx'}$ is an ordered sequence of nodes in V,

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

which starts at x and finishes at x' and $e_i = (x_i, x_{i+1}) \in E$ for $i = 0, \ldots, l-1$. In this paper, we only consider connected graphs, i.e., there exists at least one path $P_{xx'}$ for all $x, x' \in V$.

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

$$s(x, x') = \min_{P_{xx'}} \; h(P_{xx'}). \tag{2}$$

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

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

3. AXIOMS OF PROJECTION AND TRANSFORMATION

The given definition of a metric projection $\mathcal{P}$ allows for maps with undesirable behavior. Take, for instance, the map that projects networks onto a discrete metric space, i.e., for any graph $G = (V, E, W)$, consider the projection onto the finite metric space (V, d) where $d(x, y) = 1$ and $d(x, x) = 0$ for all $x \neq y \in V$. Such a projection clearly does not incorporate the original network structure, although the output is a valid metric space. To enforce the incorporation of network structure in a valid projection, we state two axioms to guide the design of such projections.

Recall from Section 2 that the set of all finite metric spaces is a subset of the set of networks, i.e., $\mathcal{M} \subset \mathcal{N}$. Motivated by this, we state the following axiom for metric projections of networks.

Axiom 1 (Axiom of Projection) *Any metric space is a fixed point of a projection map* $\mathcal{P}$, *i.e.,* $\mathcal{P}(M) = M$ for all $M \in \mathcal{M}$.

Given that our goal is to represent networks with more structured metric spaces, if we already have a metric space there is no justification to change it. This concept is illustrated in Fig. 1.

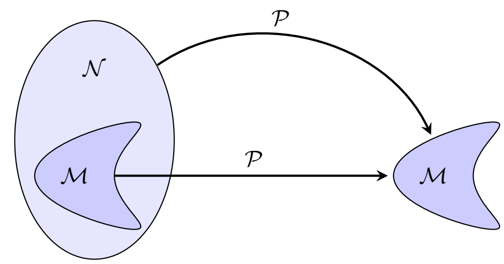


Fig. 1: Axiom of Projection. The metric space $\mathcal{M}$ is an invariant set of the projection map $\mathcal{P}$.

We also wish to enforce intuitive behavior for transformations of networks. Given two networks $G = (V, E, W), G' = (V', E', W')$, an injective map $\phi\colon V \to V'$ is *dissimilarity-reducing* if, for all $(x, x') \in E$, the following properties hold:

Dissimilarity-reducing: $W(x, x') \geq W'(\phi(x), \phi(x'))$,
Edge-preserving: $(\phi(x), \phi(x')) \in E'$.

Intuitively, dissimilarity-reducing maps reduce the path lengths between nodes in the domain of the map, as shown in Fig. 2. Thus, we expect nodes to be closer together in their metric projection after undergoing a dissimilarity-reducing transformation, which we enforce with the following axiom.

Axiom 2 (Axiom of Transformation) *Consider any two networks* $G = (V, E, W), G' = (V', E', W')$, *and any dissimilarity-reducing map* $\phi\colon V \to V'$. *Then, for all* $x, x' \in V$, *the metric projections* $(V, d) = \mathcal{P}(G)$ *and* $(V', d') = \mathcal{P}(G')$ *satisfy*

$$d(x, x') \geq d'(\phi(x), \phi(x')). \tag{3}$$

Axiom 2 encodes the natural requirement that if we reduce dissimilarities in network G to obtain G', then the metric space onto which G' is mapped cannot have distances larger than those on the metric space associated with G. We say that a projection $\mathcal{P}$ is *admissible* if it satisfies Axioms 1 and 2. The landscape of admissible projections is characterized next.

4. THE CANONICAL METRIC PROJECTION

In the framework established by Axioms 1 and 2, we first seek to establish the *existence* of an admissible projection operator. To this end, we define the *canonical projection* $\mathcal{P}^*\colon \mathcal{N} \to \mathcal{M}$, where for a given graph $G = (V, E, W)$, the metric space $(V, d^*) = \mathcal{P}^*(G)$ is defined by

$$d^*(x, x') = s(x, x') \text{ for } x, x' \in V. \tag{4}$$

That is, the canonical projection map yields a metric corresponding to the geodesic distance or shortest path length between nodes. It can be shown that the canonical projection satisfies both axioms, as stated in the following proposition.

Proposition 1 (Admissibility) *The canonical metric projection map* $\mathcal{P}^*$ *is admissible. I.e.,* d^* *defined by* (4) *is a metric for all graphs* G *and* $\mathcal{P}^*$ *satisfies the Axioms of Projection and Transformation.*

The proof of this can be found in Appendix A. Given that the canonical projection is admissible, two questions arise: (i) Are there any other admissible projections?; and (ii) Is the canonical projection special in any sense? Both questions are answered by the following uniqueness theorem.

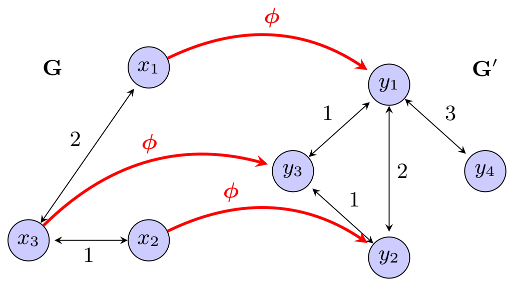


Fig. 2: Dissimilarity-reducing map. The injective map ϕ takes every edge in network G to an edge in network G' of less or equal weight.

Theorem 1 (Uniqueness) *Let $\mathcal{P} : \mathcal{N} \to \mathcal{M}$ be a metric projection, and $\mathcal{P}^*$ be the canonical projection with output metric as defined in* (4). *If $\mathcal{P}$ is admissible, then $\mathcal{P} \equiv \mathcal{P}^*$.*

The proof of this can be found in Appendix B. By Theorem 1, there is only one admissible projection from the set of networks to the set of finite metric spaces, namely the canonical projection.

Notice that the proposed canonical metric projection is nothing more than the well-known *shortest path metric* in the graph. However, the proposed axioms provide a clean framework for the validity of this widely used metric. If a practitioner agrees with the axioms, then shortest path *must* be the right metric between nodes in the graph. Conversely, if a different metric is used, then a practitioner can be now aware that at least one of the axioms must be violated. Apart from providing a clean theoretical framework, the proposed axioms allow us to establish desirable properties of the (canonical) shortest path metric, as we study next.

5. OPTIMALITY

A myriad of combinatorial optimization problems exist, where the goal is to find subsets of edges of a network that are optimal in some sense. Examples include the traveling salesman problem – finding a path that visits each node exactly once with smallest length [28] – and the minimum bisection problem [29] – separating the network into two equally-sized blocks so that the sum of the weights in the edges that connect the blocks is minimal. In this section, we focus on problems 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. We can leverage the fact that combinatorial problems are often simpler to solve in metric spaces to efficiently obtain lower bounds for $f(G)$ [30, 31]. More specifically, we restrict our attention to cost functions f that do not decrease with increasing edge weights, i.e., for networks with a shared set of nodes $f(G') \geq f(G)$ if the identity map is dissimilarity-reducing from G' to G. Then, if we project an arbitrary 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 $\mathcal{P}^*$ as we show in the following proposition.

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

$$\mathcal{P}^* = \underset{\mathcal{P}}{\operatorname{argmin}}\, f(G) - f(\mathcal{P}(G)) \qquad (5)$$
$$\text{s.t.} \qquad d(x, x') \leq W(x, x') \quad \text{for all } (x, x') \in E.$$

The proof of this can be found in Appendix C. Proposition 2 states that of all projections $\mathcal{P} : \mathcal{N} \to \mathcal{M}$ that do not increase dissimilarities [cf. constraint in (5)], the canonical projection $\mathcal{P}^*$ decreases the edge weights the least while imposing a metric structure. Thus, since f is non-decreasing in the edge weights, $\mathcal{P}^*$ perturbs $f(G)$ the least, for every network $G \in \mathcal{N}$. Given that $\mathcal{P}^*(G)$ and G are close, algorithms for efficiently solving combinatorial optimization problems in metric spaces can be applied to find a lower bound to $f(G)$ that is optimal over all metric projections of G.

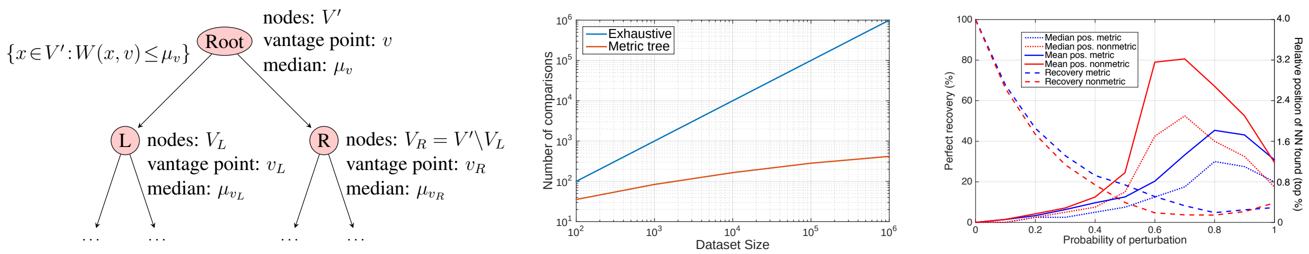
6. EXPERIMENTS: FAST APPROXIMATE SEARCH

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. An additional point $z \in V \backslash V'$ is then revealed, and we seek to find the node $x \in V'$ closest to z, i.e., the node x for which $W(x, z)$ is minimized. The described setting occurs frequently in practice, e.g., in the implementation of k nearest neighbor (k-NN) methods [32, 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 the 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 leveraging metric trees [8–10]. In this section, we propose an efficient search strategy in networks by first projecting a general network onto a metric space 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 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 then 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 with a single point in V'; see Fig. 3a. 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. We then 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



(a) (b) (c)

Fig. 3: (a) Vantage point tree. The whole node 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 node to the vantage point. The process is repeated iteratively to construct the whole tree. (b) Number of comparisons needed to find the nearest neighbor of a node in a metric space as a function of the size for exhaustive search and search aided using metric trees. (c) 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 $\mathcal{P}^*$ (blue).

node in the vp-tree, say node L in Fig. 3a 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 with 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

$$\begin{cases} i)\, W(z,v) \leq \mu_v - \tau \;\Rightarrow\; \text{visit only the left child,} \\ ii)\, \mu_v - \tau < W(z,v) \leq \mu_v + \tau \;\Rightarrow\; \text{visit left \& right child,} \\ iii)\, \mu_v + \tau < W(z,v) \;\Rightarrow\; \text{visit only right child.} \end{cases} \tag{6}$$

Even though statements $i)$ and $iii)$ in (6) 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 in $\mathbb{R}^2$ and consider their Euclidean distance as the dissimilarities W. In Fig. 3b we plot the average number of comparisons 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. 3b, a possible way to search a non-metric network G is to first project it onto a metric space M via the canonical projection $M = \mathcal{P}^*(G)$ and then construct a vp-tree on M. 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. Notice that we do not consider the complexity of implementing the projection $\mathcal{P}^*$ into the computational cost of the search since this can be done once and offline, and then reused for fast online NN queries.

In order to illustrate the search performance, 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. 3c we illustrate the search performance over 1000 queries as a function of r (blue curves). 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 (red curves), i.e., when we apply the aforementioned construction scheme and navigation rules [cf. (6)] 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 curves) consistently outperforms the one constructed on G (red curves). 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 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 on both G and M.

7. CONCLUSION

We analyzed how to project networks onto finite metric spaces. We defined the Axioms of Projection and Transformation as desirable properties of projections and showed that there is a unique canonical way of projecting any network onto the set of finite metric spaces. In particular, this axiomatic framework provided theoretical support for the computation of shortest path distances between nodes of networked datasets. Moreover, we showed that 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. Future directions include the study of projections onto generalizations of metric spaces including their directed counterparts [33].

8. 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] 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.

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

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

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

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

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

[17] H. Cai, V. W. Zheng, and K. C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, Sep. 2018.

[18] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD Intl. Conf. on Knowledge Disc. and Data Mining*, 2014, pp. 701–710.

[19] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD Intl. Conf. on Knowledge Disc. and Data Mining*, 2016, pp. 855–864.

[20] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Neural Info. Proc. Syst. (NIPS)*, 2017, pp. 1024–1034.

[21] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Intl. Conf. on Learning Repr. (ICLR)*, 2017.

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

[23] ——, "Classifying clustering schemes," *Found. Comp. Math.*, vol. 13, no. 2, pp. 221–252, 2013.

[24] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," in *IEEE Intl. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, May 2013, pp. 5219–5223.

[25] ——, "Admissible hierarchical clustering methods and algorithms for asymmetric networks," *IEEE Trans. Signal and Info. Process. over Networks*, vol. 3, no. 4, pp. 711–727, Dec 2017.

[26] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Hierarchical clustering of asymmetric networks," *Adv. Data Analysis and Classification*, vol. 12, no. 1, pp. 65–105, Mar 2018.

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

[28] 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.

[29] 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.

[30] 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.

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

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

[33] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Hierarchical quasi-clustering methods for asymmetric networks," *International Conference on Machine Learning (ICML)*, vol. 32, no. 1, pp. 352–360, 2014.

A. PROOF OF PROPOSITION 1

Proof: 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

$$P_{xx'} \oplus P_{x'x''} = [x = x_0, \dots, x_l = x' = x'_0, \dots, x'_{l'} = x'']. \quad (7)$$

We first prove that d^* is indeed a metric on the node space V. That $d^*(x, x') = d^*(x', x)$ follows from combining the facts that the original graph G is symmetric and that the norm $\| \cdot \|_1$ is symmetric in the elements of the vectors. Moreover, that $d^*(x, x') = 0$ if and only if $x = x'$ is a consequence of the positiveness property of the norm. To verify that the triangle inequality holds, let $P_{xx'}$ and $P_{x'x''}$ be paths that achieve the minimum in (4) for $d^*(x, x')$ and $d^*(x', x'')$, respectively. Then, from definition (4) it follows that

$$\begin{aligned} d^*(x, x'') &= \min_{P_{xx''}} h(P_{xx''}) \leq h(P_{xx'} \oplus P_{x'x''}) \\ &= h(P_{xx'}) + h(P_{x'x''}) \\ &= d^*(x, x') + d^*(x', x''), \end{aligned} \quad (8)$$

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 $d^*(x, x'')$ minimizes the norm across all such paths.

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

$$d^*(x, x') = \min_{P_{xx'}} h(P_{xx'}) \leq h([x, x']) = d(x, x'), \quad (9)$$

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 $P^*_{xx'} = [x = x_0, x_1, \dots, x_l = x']$ the path achieving the minimum in (9), then we may leverage the fact that d satisfies the triangle inequality to write

$$d(x, x') \leq \sum_{i=0}^{l-1} d(x_i, x_{i+1}) = h(P^*_{xx'}) = d^*(x, x'). \quad (10)$$

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

To show fulfillment of the Axiom of Transformation, consider two networks $G = (V, E, W)$ and $G' = (V', E', W')$ and a dissimilarity-reducing map $\phi : V \to V'$. Let $(V, d) = \mathcal{P}^*(G)$ and $(V', d') = \mathcal{P}^*(G')$ 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 $P^*_{xx'} = [x = x_0, \dots, x_l = x']$ a path that achieves the minimum in (4) so as to write

$$d(x, x') = h(P^*_{xx'}). \quad (11)$$

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

$$h(P^*_{\phi(x)\phi(x')}) \leq d(x, x'). \quad (12)$$

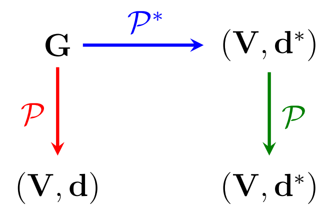


Fig. 4: Diagram of maps between spaces for the proof of Theorem 1.

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

$$d'(\phi(x), \phi(x')) \leq h(P^*_{\phi(x)\phi(x')}). \quad (13)$$

Upon replacing (12) into (13), it follows that $d'(\phi(x), \phi(x')) \leq d(x, x')$, as desired. ■

B. PROOF OF THEOREM 1

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

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

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

We begin by showing that $d(x, x') \leq d^*(x, x')$ for all $x, x' \in V$. Consider an arbitrary pair of points x and x' and let $P_{xx'} = [x = x_0, \dots, x_l = x']$ be a path achieving the minimum in (4), so that we can write

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

Focus now on a series of two-node networks $G_i = (V_i, E_i, W_i)$ for $i = 0, \dots, l-1$, 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 metric – in fact, any two-node network is a valid metric – and the method $\mathcal{P}$ satisfies the Axiom of Projection, if we define $(\{z, z'\}, d_i) = \mathcal{P}(G_i)$ 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 $\mathcal{P}$.

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 that

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

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

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

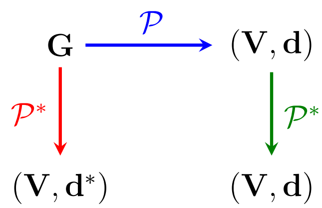


Fig. 5: Diagram of maps between spaces for the proof of Prop. 2.

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

We now show that $d(x, x') \geq d^*(x, x')$ 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

$$d^*(x, x') = \min_{P_{xx'}} h(P_{xx'}) \leq W(x, x'), \tag{18}$$

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. 4. 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 of Projection. Since the aforementioned identity map ϕ is dissimilarity reducing, we can use the fact that $\mathcal{P}$ satisfies the Axiom of Transformation to say that

$$d(x, x') \geq d^*(\phi(x), \phi(x')) = d^*(x, x'), \tag{19}$$

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

C. PROOF OF PROPOSITION 2

Proof: That $\mathcal{P}^*$ is feasible, meaning that its output metric $(V, d^*) = \mathcal{P}^*(G)$ satisfies the constraint in problem (5), can be shown using the same argument used to write expression (18). To see that $\mathcal{P}^*$ is optimal, denote by $\mathcal{P}$ a feasible metric projection with output $(V, d) = \mathcal{P}(G)$. The diagram in Fig. 5 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 metric and that $\mathcal{P}^*$ satisfies the Axiom of Projection (*cf.* Prop. 1). Moreover, notice that d satisfying the constraint in (5) guarantees that the identity map $\phi : V \to V$ from G to (V, d) is dissimilarity-reducing. Consequently, we combine the fact that $\mathcal{P}^*$ fulfills the Axiom of Transformation (*cf.* Prop. 1) with the relations between spaces in Fig. 5 to write

$$d^*(x, x') \geq d(\phi(x), \phi(x')) = d(x, x'), \tag{20}$$

for all $x, x' \in V$. Combining (20) with the constraint in problem (5), we can write that

$$d(x, x') \leq d^*(x, x') \leq W(x, x'), \tag{21}$$

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