CUBEREP: LEARNING RELATIONS BETWEEN DIFFERENT VIEWS OF DATA

Rishi Sonthalia
Department of Mathematics
University of California, Los Angeles
rsonthal@math.ucla.edu

Anna C. Gilbert
Department of Mathematics and Statistics
Yale University
anna.gilbert@yale.edu

Matthew Durham
Department of Mathematics
University of California, Riverside
mdurham@ucr.edu

ABSTRACT

Multi-view learning tasks typically seek an aggregate synthesis of multiple views or perspectives of a single data set. The current approach assumes that there is an ambient space X in which the views are images of X under certain functions and attempts to learn these functions via a neural network. Unfortunately, such an approach neglects to consider the geometry of the ambient space. Hierarchically hyperbolic spaces (HHSes) do, however, provide a natural multi-view arrangement of data; they provide geometric tools for the assembly of different views of a single data set into a coherent global space, a *CAT(0) cube complex*. In this work, we provide the first step toward theoretically justifiable methods for learning embeddings of multi-view data sets into CAT(0) cube complexes. We present an algorithm which, given a finite set of finite metric spaces (views) on a finite set of points (the objects), produces the key components of an HHS structure. From this structure, we can produce a *CAT(0) cube complex* that encodes the hyperbolic geometry in the data while simultaneously allowing for Euclidean features given by the detected relations among the views.

1 INTRODUCTION

There are many data analysis tasks for which we have multiple views of the same aggregate data set and for which we must synthesize the different views. These views might be images of objects from different angles or in different lighting (Geusebroek et al., 2005), or the translations of a single text into multiple languages (Christodoulopoulos & Steedman, 2015). In many cases, these views encode similar information but they also have complementary information that is important to capture. Aggregating all of the information into a coherent global framework is crucial for obtaining the best performance on downstream tasks.

The predominant current approach to multi-view learning is to assume that there is an ambient space X such that the views are images of X under certain functions. Such an approach, while successful, has drawbacks. Notably, the choice of the geometry of the space in which we represent our aggregate data set is currently made without theoretical considerations. These methods (Xu et al. (2018); Sun et al. (2019); Li et al. (2019); Zhao et al. (2017); Wu et al. (2019); Han et al. (2021)) also assume the maps from the ambient space to the views are obtained via a neural network, which is used simply as a blackbox and thus prevents mathematical insight and explicability.

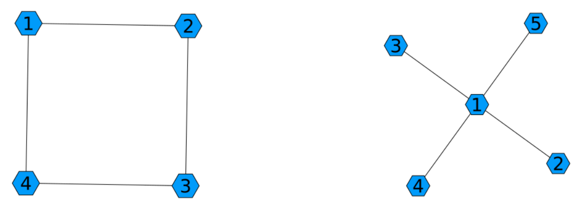
Multi-view arrangements of data naturally arise in *geometric group theory* in the context of *hierarchically hyperbolic spaces (HHSes)* (Behrstock et al. (2017)). This hierarchical approach builds on work in several areas of low dimensional topology, including mapping class groups (Masur & Minsky (2000), Teichmüller spaces (Brock (2003); Rafi (2014); Durham (2016)), and hyperbolic 3-manifolds (Minsky (2010); Brock et al. (2012)). HHSes provide geometric tools for the assembly of different views of a single data set into a coherent global space, a *CAT(0) cube complex*.

CAT(0) cube complexes are combinatorial objects that are capable of blending a mixture of both hyperbolic (e.g., tree-like) and Euclidean geometries (Sageev (2014)). We note that previous work such as Billera et al. (2001); John (2017) exploited the cubical structure of the space of phylogenetic trees to improve the reconstruction of phylogeny from genetic data.

In this work, we take the first step toward theoretically justifiable methods for learning embeddings of multi-view data sets into CAT(0) cube complexes. Specifically, given a finite collection of objects $x_1, \ldots, x_n$, each occurring as points in a finite collection of views $V_1, \ldots, V_K$, one can produce a finite collection $T_1, \ldots, T_K$ of finite trees with leaves the x_i. One method for learning these trees is TREEREP (Sonthalia & Gilbert, 2020). We develop algorithms which learn certain *relations* among the views and *relative projection* data, from which one can produce an HHS structure on $\{x_1, \ldots, x_n\}$. From this HHS structure, we can produce a *CAT(0) cube complex* that encodes the hyperbolic geometry in the T_i while simultaneously allowing for Euclidean features given by the detected relations among the V_i (Behrstock et al. (2021)).

In this paper, our main purpose is to present an algorithm which, given a finite set of finite metric spaces (views) on a finite set of points (the objects), produces the key components of an HHS structure.

2 BACKGROUND



(a) 4 cycle (b) Cross

Figure 1: Here are two simple graphs in which: (a) all of the views are independent and (b) all of the views are related to each other.

Fix n objects $x_1, \ldots, x_n$ and K metric spaces $V_1, \ldots, V_K$, with each x_i representing a point in each V_j. We think of the x_i as being global objects and their representatives in each view V_j as a sort of projection. For example, take a set of points $\{x_i\}_{i=1}^n \subset \mathbb{R}^K$ and $V_1, \ldots, V_k$ the K coordinate axes. Then, the representatives of the x_i in the j^{th} copy of $\mathbb{R}$ are simply the j^{th} coordinate of each x_i.

An HHS is defined axiomatically by relations among its constituent metric spaces (Behrstock et al. (2017)). We begin with an ambient metric space X with a finite collection of Gromov hyperbolic geodesic metric spaces $\mathcal{V}$. We note that HHSes were developed to study infinite groups where $\mathcal{V}$ is infinite, but we can and will assume $\mathcal{V}$ is finite for simplicity of this discussion.

To each space $V \in \mathcal{V}$, there is an associated projection map $\pi_V : X \to V$. The guiding philosophy is that these projections to hyperbolic spaces encode most of the coarse geometry of X. This manifests in several ways, which we will now describe.

First, we can construct an product space $\prod_{\mathcal{V}} V$ by the map $\Phi : X \to \prod_{\mathcal{V}} V$ given by $x \mapsto (\pi_V(x))_{V \in \mathcal{V}}$. The image of Φ is controlled by certain *relations* among the elements of $\mathcal{V}$. Two spaces V, W can be *orthogonal* $V \perp W$, *transverse* $V \pitchfork W$, or *nested* $V \sqsubseteq W$, though we will not consider nesting in this iteration of our algorithm.

Orthogonality and transversality are symmetric relations that constrain the coordinates of the image $X \to \prod_{\mathcal{V}} V$. The orthogonality relation $V \perp W$ imposes no constraints; i.e. the composition $X \to \prod_{\mathcal{V}} V \to V \times W$ given by $x \mapsto (\pi_V(x), \pi_W(x))$ is surjective (that is, the composition sends X onto $V \times W$).

On the other hand, transversality $V \pitchfork W$ provides certain *consistency inequalities* (CI). These are defined in terms of *relative projections* $\rho_W^V : V \to W, \rho_V^W : W \to V$ for $V \pitchfork W$, whose images have bounded diameter in W and V, respectively, as well as a family of thresholds $\theta_V > 0$ for each $V \in \mathcal{V}$, which function to differentiate the various projections V.

(CI) For any $x, y \in X$, if $d_V(\pi_V(x), \pi_V(y)) > \theta_V$ and $D_W(\pi_W(x), \pi_W(y) > \theta_W$, then up to switching the roles of x, y, either $d_V(x, \rho_V^W) < \theta_V$ or $d_W(y, \rho_W^V) < \theta_W$.

Note that this is an exclusive or, that is, exactly one of the $d_V(x, \rho_V^W) < \theta_V$ or $d_W(y, \rho_W^V) < \theta_W$ is true.

For each pair $x, y \in X$, these consistency inequalities can be converted into a *partial order* $<_{x,y}$ on $\mathcal{L}(x, y)$. The partial orders across the pairs are further constrained by the following *bounded geodesic image* (BGI) property:

(BGI) If $x, y \in X$ and $V, W \in \mathcal{L}(x, y)$ with $V \pitchfork W$, then any geodesic $[x, y]_V$ in V between x, y satisfies $d_V([x, y]_V, \rho_V^W) < \theta_V$.

The idea is that if we have $V_i <_{x,y} V_j$, then when moving efficiently from x to y in the ambient space X, we first have to move in V_i and then in V_j.

We now describe how to convert an HHS structure

$$\mathcal{L}(x, y) := \{V \in \mathcal{V} : d_V(x, y) > \theta_V\}.$$

One essential tool is the *distance formula*, which says that

$$d_X(x, y) \asymp \sum_{V \in \mathcal{L}(x,y)} d_V(\pi_V(x), \pi_V(y)),$$

where $\asymp$ denotes equality up to scaling by a linear function Masur & Minsky (2000); Behrstock et al. (2017).

As a consequence, the map $\Phi : X \to \prod_{\mathcal{V}} V$ behaves like a *rough isometric embedding*, up to ignoring the (possibly infinitely-many) components of $\prod_{\mathcal{V}} V$ where pairs of points in the image are close. In fact, the set of views $\mathcal{V}$ is finite in our setting, making Φ a rough isometric embedding.

Let us look at the examples in Figure 1 to understand these definitions. First, consider the 4-cycle as X in (a). Let us assume that we have two views $V_{(1,2)}, V_{(1,4)}$ given by the projections onto the edges $(1, 2)$ and $(1, 4)$. Now we can see that the map $X \to V_{(1,2)} \times V_{(1,4)}$ is surjective (i.e., onto) and, hence, the views are independent. On the other hand, consider the cross in part (b), with views given by the edges $V_{(1,2)}$ and $V_{(1,3)}$. Then we see that $X \to V_{(1,2)} \times V_{(1,3)}$ is not surjective. Further, when going from node 2 to node 3, we first have to traverse the edge $(2, 1)$ and then the edge $(1, 3)$. Thus, we see that $V_{(1,2)} <_{2,3} V_{(1,3)}$ and that the views are transverse.

We briefly remark on the CAT(0) cube complexes produced from the HHS structure on these synthetic examples. In the case when the graph is a tree, the resulting CAT(0) space is again the tree. And when the graph contains cycles, we get a simplicial complex whose 1-skeleton is the underlying graph, with higher dimensional cells corresponding to pairwise-orthogonal collections of views.

3 ALGORITHM

For each pair of points x_s, x_t, we want to learn the relations between the views of the data as we travel from x_s to x_t in the ambient space X. To do this, we need to learn the two quantities:

1. The thresholds θ_{V_k} for $V_k \in \mathcal{V}$.
2. A directed acyclic graph G_{st} that represents the partial ordering given by the (BGI) property.

Given threshold values θ_{V_k}, and a pair of points x_s and x_t, we add the directed edge (V_i, V_j) in G_{st} if there exists two points x_{r_1} and x_{r_2} such that

$$d_{V_i}(x_{r_1}, x_t) \leq \theta_{V_i},\ d_{V_j}(x_{r_2}, x_t) > \theta_{V_j},\ d_{V_i}(x_{r_1}, x_s) > \theta_{V_i},\ d_{V_j}(x_{r_2}, x_s) \leq \theta_{V_j} \tag{1}$$

and for every new pair of points $x_{s'}, x_{t'}$, we have that

$$\begin{aligned} d_{V_j}(x_{s'}, x_{t'}) > \theta_{V_j} &\Rightarrow d_{V_i}(x_{r_2}, [x_{s'}, x_{t'}]_{V_i}) < \theta_{V_i} \\ d_{V_i}(x_{s'}, x_{t'}) > \theta_{V_i} &\Rightarrow d_{V_j}\left(x_{r_1}, [x_{s'}, x_{t'}]_{V_j}\right) < \theta_{V_j} \end{aligned} \tag{2}$$

Here we assume that $\rho_{V_j}^{V_i}(V_i) = x_{r_2}$ and $\rho_{V_i}^{V_j}(V_j) = x_{r_1}$. Hence satisfying Equation 1 implies that Property (CI) is satisfied and satisfying Equations 2 that Property (BGI) is satisfied.

For a given set of thresholds, this directed graph could have cycles. Hence we want to measure how far the directed graph is from being acyclic. To do so, we compute $L(G) = (\mathrm{Tr}(\exp(G) - K))^2$. From Wei et al. (2020), we know that $L(G)$ is zero if and only if G is a directed acyclic graph. Thus, our overall loss function is

$$L(\theta_{V_1}, \dots, \theta_{V_K}) := \sum_{s \neq t} L(G_{st}).$$

One idea would be optimize for the threshold values θ_{V_i} via gradient descent, but due to the combinatorial nature of the objects the gradient of the loss function with respect to the thresholds will always be 0. Instead we will do a random walk on a mesh in $\mathbb{R}^K$ as follows. Given a view V_i, let $d_1^{(i)} < \dots < d_{M_i}^{(i)}$ be the unique distances in d_{V_i}. Now for any θ_{V_i} in the interval $(d_a^{(i)}, d_b^{(i)})$, the constructed graphs are the same. Hence we can pick $M_i - 1$ values $(\tau_l^{(i)})_{l=1,\dots M_i-1}$ as the midpoints of these intervals, as well as one point $\tau_{M_i}^{(i)}$ bigger than $d_{M_i}^{(i)}$ and one point $\tau_0^{(i)}$ smaller than $d_1^{(i)}$. Thus, we will think of our thresholds as living in the folllowing mesh.

$$\Theta := (\theta_{V_1}, \dots, \theta_{V_K}) \in \{\tau_0^{(1)}, \dots, \tau_{M_1}^{(1)}\} \times \dots \times \{\tau_0^{(K)}, \dots, \tau_{M_K}^{(K)}\}.$$

For the random walk, given Θ_{old}, we randomly pick a view V_k and then update θ_{V_k} to a uniformly random value in $\{\tau_0^{(k)}, \dots, \tau_{M_k}^{(k)}\}$ to get Θ_{new}. Then for both sets of thresholds we compute the loss values $L(\Theta_{old}), L(\Theta_{new})$. If $L(\Theta_{new}) < L(\Theta_{old})$, we move to the new set of thresholds, otherwise we move to the new state with probability $\left(\frac{L(\Theta_{old})}{L(\Theta_{new})}\right)^2$. Thus, our algorithm is a Markov chain Monte Carlo method.

4 Experiments

We test our algorithms on synthetic data to demonstrate that we find the correct relations. We first look at simple undirected graphs as the ambient space. The nodes of the graph are our data points and our views are the projections onto the edges of the graph. The idea is to learn an HHS structure on this information which will allow us to rebuild the graph.

For the first test, we start with two very simple examples seen in Figure 1. We run our method a 1000 times to see which relations persistently appear. For the 4-cycle, we never get any relations. Thus, we always get that the views are independent. On the other hand, for the cross, we get that every view is related to every other view. Thus, we see that our method works in this setting.

In the first generalization, we can look at graphs that are even cycles. For these graphs we should never have any relations between the views. We can also look at graphs that are trees. For these graphs, we should have that every pair of views is related. Running our method on these graphs provides relations that are consistent with our expectations.

Finally, we look at graphs that have both cycles and tree like parts. Specifically, we look at bipartite graphs for which no cycle has a chord. Examples of such graphs can be seen in Figure 2. Let us look at the graph on the left of Figure as an example. For this graph, we expect that the edges (views) in the cycle, $(1, 3), (3, 7), (7, 5), (5, 1)$, are independent of each other and that all other pairs of views are transverse. This is exactly the set of relations that we get when we run our method. Thus, we see that on these simple instances our method learns the correct relations between the views.

We note that the relations and relative projections produced by our algorithms are sufficient for rebuilding the graph. While this is possible to do directly from the initial edge and projection data, we believe that these synthetic experiments are a good proof of concept.

5 Future Work and Conclusion

We considered using multi-view metric data to reconstruct an ambient space that is consistent with the views of the data. We developed an algorithm that uses ideas from geometric group theory to

Figure 2: Examples of bipartite graphs with no chords.

learn relations between the views. Our algorithm is an MCMC algorithm that use DAG learning inspired loss functions. Using our algorithm, we show that for certain types of undirected graphs, we exactly recover the correct relations between the views.

We have shown so far that our method accurately recovers the relations between the views. There is, however, more work to be done: (1) develop an algorithm that converts a learned HHS structure into the corresponding CAT(0) space, (2) test the algorithm on other types of synthetic data, as well as real world data, and (3) speed up the learning algorithm.

References

Jason Behrstock, Mark Hagen, and Alessandro Sisto. Hierarchically hyperbolic spaces, i: Curve complexes for cubical groups. *Geometry & Topology*, 21(3):1731–1804, 2017.

Jason Behrstock, Mark F Hagen, and Alessandro Sisto. Quasiflats in hierarchically hyperbolic spaces. *Duke Mathematical Journal*, 1(1):1–88, 2021.

Louis J. Billera, Susan P. Holmes, and Karen Vogtmann. Geometry of the space of phylogenetic trees. *Advances in Applied Mathematics*, 27(4):733–767, 2001. ISSN 0196-8858. doi: https://doi.org/10.1006/aama.2001.0759. URL `https://www.sciencedirect.com/science/article/pii/S0196885801907596`.

Jeffrey Brock. The weil-petersson metric and volumes of 3-dimensional hyperbolic convex cores. *Journal of the American Mathematical Society*, 16(3):495–535, 2003.

Jeffrey F Brock, Richard D Canary, and Yair N Minsky. The classification of kleinian surface groups, ii: The ending lamination conjecture. *Annals of Mathematics*, pp. 1–149, 2012.

Christos Christodoulopoulos and Mark Steedman. A massively parallel corpus: the bible in 100 languages. *Language Resources and Evaluation*, 49:375 – 395, 2015.

Matthew Gentry Durham. The augmented marking complex of a surface. *Journal of the London Mathematical Society*, 94(3):933–969, 2016.

Jan-Mark Geusebroek, Gertjan J. Burghouts, and Arnold W. M. Smeulders. The amsterdam library of object images. 61(1), 2005. ISSN 0920-5691. doi: 10.1023/B:VISI.0000042993.50813.60. URL `https://doi.org/10.1023/B:VISI.0000042993.50813.60`.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=OOsR8BzCnl5`.

Katherine St John. Review paper: The shape of phylogenetic treespace. *Systematic Biology*, 66:e83 – e94, 2017.

Y. Li, M. Yang, and Z. Zhang. A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 31(10):1863–1883, oct 2019. ISSN 1558-2191. doi: 10.1109/TKDE.2018.2872063.

Howard A Masur and Yair N Minsky. Geometry of the complex of curves II: Hierarchical structure. *Geometric and Functional Analysis*, 10(4):902–974, 2000.

Yair Minsky. The classification of kleinian surface groups, i: Models and bounds. *Annals of Mathematics*, pp. 1–107, 2010.

Kasra Rafi. Hyperbolicity in teichmüller space. *Geometry & Topology*, 18(5):3025–3053, 2014.

Michah Sageev. $\mathrm{CAT}(0)$ cube complexes and groups. In *Geometric group theory*, volume 21 of *IAS/Park City Math. Ser.*, pp. 7–54. Amer. Math. Soc., Providence, RI, 2014.

Rishi Sonthalia and Anna Gilbert. Tree! i am no tree! i am a low dimensional hyperbolic embedding. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 845–856. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/093f65e080a295f8076b1c5722a46aa2-Paper.pdf`.

Shiliang Sun, Liang Mao, Ziang Dong, and Lidan Wu. Multiview machine learning. In *Springer Singapore*, 2019.

Dennis Wei, Tian Gao, and Yue Yu. Dags with no fears: A closer look at continuous optimization for learning bayesian networks. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 3895–3906. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper/2020/file/28a7602724ba16600d5ccc644c19bf18-Paper.pdf`.

Chuhan Wu, Fangzhao Wu, Mingxiao An, Jianqiang Huang, Yongfeng Huang, and Xing Xie. Neural news recommendation with attentive multi-view learning. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 3863–3869. International Joint Conferences on Artificial Intelligence Organization, 7 2019. doi: 10.24963/ijcai.2019/536. URL `https://doi.org/10.24963/ijcai.2019/536`.

Cai Xu, Ziyu Guan, Wei Zhao, Yunfei Niu, Quan Wang, and Zhiheng Wang. Deep multi-view concept learning. In *IJCAI*, pp. 2898–2904, 2018. URL `https://doi.org/10.24963/ijcai.2018/402`.

Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2017.02.007. URL `https://www.sciencedirect.com/science/article/pii/S1566253516302032`.