Curvature-Dimension Tradeoff for Generalization in Hyperbolic Space

Nicolás Alvarado
nfalvarado@mat.uc.cl

Hans Löbel
halobel@ing.puc.cl

Mircea Petrache
mpetrache@mat.uc.cl

Abstract

The inclusion of task-relevant geometric embeddings in deep learning models is an important emerging direction of research, particularly when using hierarchical data. For instance, negatively curved geometries such as hyperbolic spaces are known to allow low-distortion embedding of tree-like hierarchical structures, which Euclidean spaces do not afford. Learning techniques for hyperbolic spaces, such as Hyperbolic Neural Networks (HNNs), have shown empirical accuracy improvement over classical Deep Neural Networks in tasks involving semantic or multi-scale information, such as recommender systems or molecular generation. However, no research has investigated generalization properties specific to such geometries. In this work, we introduce generalization bounds for learning tasks in hyperbolic spaces, marking the first time such bounds have been proposed. We highlight a previously unnoticed and important difference with Euclidean embedding models, namely, under embeddings into spaces of negative curvature $-\kappa < 0$ and dimension d, only the product $\sqrt{\kappa}\, d$ influences generalization bounds. Hence, the curvature parameter of the space can be varied at fixed d with the same effect on generalization as when varying d.

1 Introduction

Data representations are a crucial element in current Machine Learning techniques, with a major impact on the efficiency of the algorithms. It has been observed [11],[26], [30], [27] that hyperbolic embeddings of much lower dimensions than Euclidean ones allow equivalent performance for tree-like types of data, e.g. hyperbolic spaces of dimension five can outperform usual 200-dimensional spaces for embedding taxonomic data. An increasing variety of classes of datasets is being shown to have a tree-like, or hyperbolic structure, starting from the first popularization of the subject by Krioukov [23] in the context of complex networks [32], [36].

Recall that hyperbolic spaces are homogeneous spaces of constant curvature equal to -1. The more relevant and crucial theoretical property of hyperbolic spaces and of spaces of negative curvature [10] in general is that they are able to embed graphs such as trees with arbitrarily low distortion of the natural metrics. This has been first observed by Gromov [21], who introduced a much larger class of spaces, called δ-hyperbolic spaces, which are shown to be almost isometric to trees [20], including cases of graphs with control on the diameter of cycles [31], whereas euclidean and positively curved spaces do not allow to embed trees with bounded distortion of the metric [8, 22, 12, 1, 29, 7]. Hyperbolic embeddings have also shown promise for routing [17], clustering [13, 24], biological networks [2], phylogenetic trees [6, 25], neuroscience [3], text embedding [18, 4], knowledge graphs [33].

Practically, many formulas used in classical Deep Neural Networks (DNN) have direct counterparts applicable in hyperbolic spaces, originally developed for relativity applications [35]. These formulas enable the creation of Hyperbolic Neural Networks (HNN), designed for operations within hyperbolic embeddings [19], [26], [30]. HNNs utilize hyperbolic distance between neurons instead of

Mathematics of Modern Machine Learning Workshop at NeurIPS 2023.

Euclidean distance, facilitating geometry-aware information processing. The motivation stems from the connection to tree-like structures in real-world datasets and the natural hyperbolic space structure inherent in entailment relations, as exemplified by the negative curvature of the space of Gaussian variables with the Fisher Information metric [16]. For an in-depth exploration of HNN models and applications, refer to [28], [38].

While, as said above, it is well understood empirically that low-dimension hyperbolic spaces work well for learning in hierarchical datasets and that the link to embedding theory has a long history, to the best of our knowledge, no theoretical bounds for generalization in hyperbolic spaces has been yet developed. This article contributes to filling this gap in the literature with the following main contributions:

- We prove PAC-learning bounds for arbitrary learning models in hyperbolic and negatively curved geometries. The techniques of proof are new, as hyperbolic spaces are not doubling, and a large part of previous literature on PAC-learning was set up on doubling metric spaces. In particular, we prove new covering number bounds in hyperbolic spaces, which are interesting in themselves.
- We show that curvature $-\kappa < 0$ and dimension d mix up in the quantity $\sqrt{\kappa}d$, which is the one relevant for generalization. This means that the effects of increasing dimension at fixed $-\kappa \leq 0$ or decreasing $-\kappa < 0$ while keeping d fixed have interchangeable roles. As $\kappa = 0$ corresponds to Euclidean spaces, this gives the first rigorous statement for the observation that hyperbolic learning techniques require lower dimensions than Euclidean ones for equivalent performance.

2 Preliminaries

Learning model. We consider a model space $\widetilde{\mathcal{F}}$ composed of Lipschitz functions $\widetilde{f} : \mathcal{X} \to \mathcal{Y}$, in which $\mathcal{X}, \mathcal{Y}$ will be metric spaces. As previously mentioned, we focus on the case of $\mathcal{X}, \mathcal{Y}$ of curvature bounded from above by $-c \leq 0$. We also consider a *loss function* $l : \mathcal{X} \times \mathcal{Y} \to [-M, M]$, which will also be assumed to be Lipschitz. We assume that our learning task is modeled by a *data distribution* $\mathcal{D}$, which is a probability measure over $\mathcal{X} \times \mathcal{Y}$, and the goal of our learning algorithms is to approximate, i.e. minimize the average loss, of $\mathcal{D}$ via graphs of functions from $\widetilde{\mathcal{F}}$: $\min_{\widetilde{f}\in\widetilde{\mathcal{F}}} \mathbb{E}_{Z=(X,Y)\sim\mathcal{D}}[l(\widetilde{f}(X), Y)]$. Here the distribution $\mathcal{D}$ of a random variable $Z = (X, Y)$ represents the desired optimal rule that we want our optimum models $\widetilde{f}^*$ to learn to use in order to assign label $y \sim \widetilde{f}^*(x)$ to each $x \in \mathcal{X}$.

We will simplify notation below and consider only the space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, and focus on the function spaces $\mathcal{F}$ defined in terms of $\widetilde{\mathcal{F}}$ and $l(y, y')$ as follows:

$$\mathcal{F} := \{f : \mathcal{X} \times \mathcal{Y} \to [-M, M] : \exists \widetilde{f} \in \widetilde{\mathcal{F}}, f(x, y) = l(\widetilde{f}(x), y)\}.$$

We also assume that all elements of $\mathcal{F}$ are L-Lipschitz for some constant $L > 0$. This assumption is guaranteed if the loss function is Lipschitz and $\widetilde{\mathcal{F}}$ is composed of Lipschitz functions. This is commonly satisfied for applications of learning models, such as DNN.

A function $f\colon X \to Y$ defined between two metric spaces (X, d_X), (Y, d_Y) is L-Lipschitz when $d_Y(f(x), f(x')) < L d_X(x, x')$ for all pairs of points $x, x' \in X$.

Metric PAC-Learning bounds. Based on covering growth bounds, classical concentration inequalities can be used to control the sampling complexity of the loss function over a given class. We follow the standard approach, such as that outlined in [9]. Classical generalization bounds guarantee that the empirically measured error is not much lower than the actual error. Thus the relevant concentration bounds are as follows:

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}} \mathsf{GenErr}(f, Z^{(n)}, \mathcal{D}) \geq \mathcal{R}(\mathcal{F}_Z) + \epsilon\right] \leq 2\exp\left(-\frac{\epsilon^2 n}{2M}\right). \tag{2.1}$$

In the above, for an i.i.d. random sample $Z^{(n)} = \{Z_i = (X_i, Y_i) \sim \mathcal{D}, 1 \leq i \leq n\}$, we define $\mathsf{GenErr}(f, Z^{(n)}, \mathcal{D}) := \mathbb{E}_{Z=(X,Y)\sim\mathcal{D}}[f(Z)] - \frac{1}{n}\sum_{i=1}^n f(Z_i)$, and $\mathcal{R}(\mathcal{F}_Z)$ denotes the Rademacher complexity: $\mathcal{R}(\mathcal{F}_Z) := \mathbb{E}_\sigma \sup_{\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(Z_i)$, where $\sigma = (\sigma_1, \dots, \sigma_n)$ denotes the so-called Rademacher variables, which is a uniformly distributed random variable over $\{-1, 1\}^n$, independent of the Z_i's. Intuitively, Rademacher complexity measures how well a set of functions or a hypothesis class can fit random noise, which is a way of assessing the flexibility of the class of functions.

Covering and packing numbers. If (X, d) is a metric space, we define the covering number $N(X, \epsilon, d)$ as the smallest number N such that X can be covered by N balls of radius ϵ. The packing number $M(X, \epsilon, d)$ in the largest M such that X contains M disjoint balls of radius ϵ. We will see below that generalization error bounds for $\mathcal{F}$ will be performed in terms of integrals (with respect to ϵ) of $N(\mathcal{Z}, \epsilon, d)$, thus it is important to study their growth.

PAC-bounds via covering numbers. We can bound $\mathcal{R}(\mathcal{F}_Z)$ via the Dudley entropy integral or the chaining bound (see e.g. [37, Thm. 17] or [5, Thm. 1.1]). These bounds are usually proved for $\mathcal{Z}$ a normed space, which is not the case for M^d_κ. We can prove that for $\mathcal{Z}$ equal to the unit ball of M^d_κ we have the following:

$$\mathcal{R}(\mathcal{F}_Z) \leq 4 \inf_{\alpha > 0} \left(\alpha + \frac{3}{\sqrt{n}} \int_\alpha^2 \sqrt{\ln N(\mathcal{F}, t, \| \cdot \|_\infty)} dt \right), \tag{2.2}$$

where N is the covering number of $\mathcal{F}$ in supremum norm $\|f - g\|_\infty := \sup_{z \in \mathcal{Z}} |f(z) - g(z)|$).

It is known that by rescaling the fundamental estimate due to Kolmogorov-Tikhomirov [34, eq. 238, with $s = 1$, and eq. 1], and under the mild assumption that $\mathcal{F}$ is composed of L-Lipschitz functions on $\mathcal{Z}$ with values in an interval $[-C, C]$, for a centralizable[1] metric space $\mathcal{Z}$, the following holds

$$N(\mathcal{Z}, 2\epsilon) \leq \log_2 N(\mathcal{F}, \epsilon, \| \cdot \|_\infty) \leq \log_2 \left(\frac{2C}{\epsilon} + 1 \right) + N(\mathcal{Z}, \epsilon/2). \tag{2.3}$$

Hyperbolic spaces. A differentiable manifold X with an inner product $g_p(\cdot, \cdot)$ on each tangent space $T_pX \simeq \mathbb{R}^d$ is a Riemannian manifold. The only Riemannian manifold of constant negative curvature -1 and dimension d is the hyperbolic space $\mathbb{H}^d$, which can be identified (in the so-called Poincaré model) with the unit ball of $\mathbb{R}^d$ with the non-euclidean distance: $\rho(x, y) = \mathsf{arccosh}\left(1 + 2(\|x - y\|^2)/(1 - \|x\|^2)(1 - \|y\|^2)\right)$.

Model spaces. For $\kappa > 0$ and an integer $d \geq 2$ we define the homogeneous model space of dimension d and constant curvature $-\kappa$, M^d_κ, as the space obtained from the hyperbolic space $\mathbb{H}^d$ by multiplying the metric by the constant $1/\sqrt{-\kappa}$

Growth of balls, model spaces. A metric space (X, d) is called *doubling* if there exists a constant $\lambda > 0$ such that for all $R > 0$, every ball of radius R in X can be covered by no more than λ balls of radius $R/2$. This is a common assumption, and it is true for $\mathbb{R}^d$ and for compact Riemannian manifolds, as well as for discrete spaces such as finite groups endowed with the word metric.

However, for hyperbolic spaces $\mathbb{H}^d$, and more generally for model spaces M^d_κ, the doubling condition does not hold, as shown in Theorem 3.1. To our knowledge the covering bound in the following result is new, and is of independent interest.

3 Growth of balls and generalization bounds in negative curvature

Now, we present a formula to bound the covering number of a ball of radius R in the model space with negative curvature.

Theorem 3.1. *Let $\mathcal{Z}$ be a ball of radius $R > 0$ in the model space M^d_κ. Then, in the regime $\epsilon \lesssim 1$ and $R \gtrsim 1$ we have*

$$N(\mathcal{Z}, \epsilon, \rho) \simeq \frac{e^{d\sqrt{\kappa}\, R}}{(\sqrt{\kappa}\, \epsilon)^d}. \tag{3.1}$$

[1] This mild condition signifies that for any open set U of diameter at most $2r$ there exists a point x^0 so that U is contained in $B(x^0, r)$, which is true for the spaces M^d_κ.

Proof. We use the following bound (see e.g. [14, eq. III.4.1]):

$$\mathsf{Vol}(\mathcal{Z}) = \mathrm{Vol}^d_\kappa(R) := \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^R \left(\frac{\sinh(\sqrt{\kappa} t)}{\sqrt{\kappa}}\right)^{d-1} dt. \tag{3.2}$$

Next, note that $\sinh(t) \simeq t$ for $t \lesssim 1$ and $\sinh(t) \simeq \frac{1}{2}e^t$ for $t \gtrsim 1$. Therefore we have, in the regime $\epsilon \lesssim 1, R \gtrsim 1$,

$$\frac{\mathrm{Vol}^d_\kappa(R)}{\mathrm{Vol}^d_\kappa(\epsilon)} = \frac{\int_0^R (\sinh(\sqrt{\kappa}t)^{d-1}\, dt}{\int_0^\epsilon (\sinh(\sqrt{\kappa}t)^{d-1}\, dt} = \frac{\int_0^{\sqrt{\kappa}R} (\sinh(t))^{d-1}\, dt}{\int_0^{\sqrt{\kappa}\epsilon} (\sinh(t))^{d-1}\, dt} \simeq \frac{e^{\sqrt{\kappa}\ d\ R}}{\kappa^{d/2}\epsilon^d},$$

where the implicit constants can be chosen to be independent of d, κ, ϵ, R. If a cover of $\mathcal{Z}$ by balls of radius ϵ has cardinality N, then we have $\mathrm{Vol}(\mathcal{Z}) \le N\mathrm{Vol}^d_\kappa(\epsilon)$, which shows the lower bound for $N(\mathcal{Z}, \epsilon, \rho)$ in (3.1).

For the upper bound, we observe that a maximal $\epsilon/2$-packing of $\mathcal{Z}$, i.e., a maximal finite set in $\mathcal{Z}$ such that open balls of radius $\epsilon/2$ centered at points in the set are disjoint, is an ϵ-covering. If it were not, then there would be an extra point at distance $\geq \epsilon$ from all the rest, and thus an $\epsilon/2$-ball at the extra point would be disjoint from the rest, a contradiction to assumed maximality. The maximum packing number $M(\mathcal{Z}, \epsilon/2, \rho) \geq N(\mathcal{Z}, \epsilon, \rho)$ has been studied in more detail than minimum covering number, and it is known that it scales like (3.1), see e.g. [15, Sec. 4] for the case $\kappa = 1$, and the technique generalizes to general κ. This proves the upper bound in (3.1) and concludes the proof. □

Corollary 3.2. *Under the same hypotheses $\epsilon \lesssim 1, R \gtrsim 1$, as in Theorem 3.1, we have:*

$$\ln N(\mathcal{F}, \epsilon, \|\cdot\|_\infty) \simeq \frac{e^{d\sqrt{\kappa}\ R}}{(\sqrt{\kappa}\ \epsilon)^d}. \tag{3.3}$$

Proof. We use (2.3) in combination with Theorem 3.1, as follows. Note that for $\epsilon \lesssim 1$ the dependence in ϵ is by a power law, implying that covering numbers with radius 2ϵ or $\epsilon/2$ are comparable. Furthermore, due to the result of Theorem 3.1, for $R \gtrsim 1$ the logarithmic factor from (2.3) is not leading and can be absorbed. □

Using the above bounds, we obtain from (2.2) the following bound, provided m is large enough:

Corollary 3.3. *Assuming that $m \geq \frac{\exp(\sqrt{\kappa}\ R)}{2^d\sqrt{\kappa}}$, we have*

$$\mathcal{R}(\mathcal{F}_Z) \lesssim \frac{\exp(\sqrt{\kappa}\ R)}{\sqrt{\kappa}} m^{-1/d}. \tag{3.4}$$

Proof. The optimal choice of α in (2.2) is the one for which $m \simeq \ln N(\mathcal{F}, \alpha, \|\cdot\|_\infty)$, and thus, in view of the control given in Corollary (3.2), we choose

$$\alpha := \frac{C}{m^{1/d}}, \quad \text{for} \quad C := \frac{\exp(\sqrt{\kappa}\ R)}{\sqrt{\kappa}}.$$

This choice is allowed due to our hypothesis on m. Substituting this in (2.2), we thus find

$$\mathcal{R}(\mathcal{F}_Z) \lesssim \frac{C}{m^{1/d}} + \frac{C^{d/2}}{m^{1/2}} \int_{C/m^{1/d}}^2 \frac{dt}{t^{d/2}} = \begin{cases} \frac{d}{d-2}\left(\frac{C}{m^{1/d}} - 2^{-d/2+1}\frac{C^{d/2}}{m^{1/2}}\right), & \text{if } d > 2, \\ \frac{\exp(\sqrt{\kappa}R)}{\sqrt{\kappa}\sqrt{m}}\left(1 + \sqrt{\kappa}\ R - \ln(2\sqrt{\kappa}\sqrt{m})\right) & \text{if } d = 2. \end{cases}$$

Due to the hypothesis on m we can absorb all but the first term in each parenthesis, and for $d \geq 3$ we can bound $d/(d-2) \lesssim 1$, obtaining the same bound in both cases $d > 2, d = 2$, up to universal factor. □

Note that as a consequence of the above we have the following:

Corollary 3.4. *Under the above assumptions, and if $Z^{(n)} = \{Z_i\}_{i=1}^n$ is an i.i.d. sample from data distribution $\mathcal{D}$, in order for the generalization error to satisfy with probability larger that $1-\delta$ the bound*

$$\sup_{f\in\mathcal{F}} \mathsf{GenErr}(f, Z^{(n)}, \mathcal{D}) \le \epsilon,$$

the following requirement on the number of samples n is sufficient:

$$n \gtrsim \max\left\{\frac{\exp(\sqrt{\kappa}\ d\ R)}{\epsilon^d \kappa^{d/2}},\ \frac{2M\log(2/\delta)}{\epsilon^2}\right\}.$$

In particular we see that the leading dependence on dimension is given by the factor $\exp(\sqrt{\kappa}dR)$, in which only the quantity $\sqrt{\kappa}d$ mixes the roles of κ and d.

4 Conclusions and future work

PAC-learning bounds quantify the minimum training data required for a learning algorithm to achieve a desired level of generalization accuracy. These bounds depend on the learning model's complexity, training data size, and desired accuracy level. In our study, we derived PAC-learning bounds for models operating in negatively curved geometries. While our initial focus was on hyperbolic neural networks, our findings yielded broader insights regarding the tradeoff between the model space's curvature and its dimension, particularly in the parameter $\sqrt{\kappa}d$. This suggests potential optimization opportunities for learning algorithms by carefully balancing model space curvature and dimensionality.

The proposed framework is versatile, extending beyond graphs to encompass all Gromov-hyperbolic spaces, including structures with bounded-length cycles as highlighted by Sarkar [31]. This broader applicability also encompasses parameterized probability measures' geometry, like Gaussian measures ([16]), which exhibit negative curvature. This framework is suitable for learning entailment relations without relying on fixed graphs.

As a future work we will to consider implications for more general Gromov-hyperbolic geometries, including metric trees. Furthermore, another area of future work is to explore the use of empirical methods to study the bounds for the Rademacher complexity. While our theoretical results provide useful upper bounds, it would be interesting to see how these bounds play out in practice and how they can be used to guide the design of machine learning algorithms.

Overall, our work contributes to a deeper understanding of the relationship between geometry, complexity, and learning in spaces with negative curvature, and opens up new avenues for research in this important area.

Acknowledgments and Disclosure of Funding

The authors acknowledge funding support from the Millennium Institute for Foundational Research on Data (IMFD ANID–Millennium Science Initiative Program–Code ICN17_002) and the National Center of Artificial Intelligence (CENIA FB210017, Basal ANID).

The third author acknowledge funding Fondecyt Regular grant number 1210462 entitled "Rigidity, stability and uniformity for large point configurations"

References

[1] Charu C Aggarwal, Alexander Hinneburg, and Daniel A Keim. On the surprising behavior of distance metrics in high dimensional space. In *Database Theory—ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings 8*, pages 420–434. Springer, 2001.

[2] Réka Albert, Bhaskar DasGupta, and Nasim Mobasheri. Topological implications of negative curvature for biological and social networks. *Physical Review E*, 89(3):032811, 2014.

[3] Antoine Allard and M Ángeles Serrano. Navigable maps of structural brain networks across species. *PLoS computational biology*, 16(2):e1007584, 2020.

[4] Ivana Balazevic, Carl Allen, and Timothy Hospedales. Multi-relational poincaré graph embeddings. In *Proceedings of NeurIPS*, pages 4463–4473, 2019.

[5] Peter Bartlett. Covering numbers, chaining & dudley's entropy integral (lecture 24). class notes on statistical learning theory, 2006.

[6] 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.

[7] Michele Borassi, Alessandro Chessa, and Guido Caldarelli. Hyperbolicity measures democracy in real-world networks. *Physical Review E*, 92(3):032812, 2015.

[8] Jean Bourgain. On lipschitz embedding of finite metric spaces in hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985.

[9] Olivier Bousquet, Stéphane Boucheron, and Gábor Lugosi. Introduction to statistical learning theory. *Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures*, pages 169–207, 2004.

[10] Martin R Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.

[11] BP Chamberlain, J Clough, and MP Deisenroth. Neural embeddings of graphs in hyperbolic space. arxiv 2017. *arXiv preprint arXiv:1705.10359*.

[12] Ines Chami. *Representation Learning and Algorithms in Hyperbolic Spaces*. Stanford University, 2021.

[13] Ines Chami, Albert Gu, Vaggos Chatziafratis, and Christopher Ré. From trees to continuous embeddings and back: Hyperbolic hierarchical clustering. *Advances in Neural Information Processing Systems*, 33:15065–15076, 2020.

[14] Isaac Chavel. *Riemannan Geometry: A modern introduction*. Cambridge University Press, 2006.

[15] Henry Cohn and Yufei Zhao. Sphere packing bounds via spherical codes. 2014.

[16] Sueli IR Costa, Sandra A Santos, and Joao E Strapasson. Fisher information distance: A geometrical reading. *Discrete Applied Mathematics*, 197:59–69, 2015.

[17] Andrej Cvetkovski and Mark Crovella. Hyperbolic embedding and routing for dynamic graphs. In *IEEE INFOCOM 2009*, pages 1647–1655. IEEE, 2009.

[18] Bhuwan Dhingra, Christopher J Shallue, Mohammad Norouzi, Andrew M Dai, and George E Dahl. Embedding text in hyperbolic spaces. *arXiv preprint arXiv:1806.04313*, 2018.

[19] Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in neural information processing systems*, 31, 2018.

[20] M. Gromov. Hyperbolic Groups. In S. M. Gersten, editor, *Essays in Group Theory*, Mathematical Sciences Research Institute Publications, pages 75–263. Springer, New York, NY, 1987.

[21] Mikhael Gromov. *Hyperbolic groups*. Springer, 1987.

[22] Piotr Indyk, Ji ˇ rı Matou ˇ sek, and Anastasios Sidiropoulos. 8 LOW-DISTORTION EMBEDDINGS OF FINITE METRIC SPACES. 2017.

[23] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, September 2010. Publisher: American Physical Society.

[24] John Lamping, Ramana Rao, and Peter Pirolli. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*, pages 401–408, Denver, Colorado, United States, 1995. ACM Press.

[25] Hirotaka Matsumoto, Takahiro Mimori, and Tsukasa Fukunaga. Novel metric for hyperbolic phylogenetic tree embeddings. *Biology Methods and Protocols*, 6(1):bpab006, 2021.

[26] Maximillian Nickel and Douwe Kiela. Poincaré Embeddings for Learning Hierarchical Representations. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[27] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pages 3779–3788. PMLR, 2018.

[28] Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. Hyperbolic deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10023–10044, 2021.

[29] Marco A Rodríguez-Flores and Fragkiskos Papadopoulos. Hyperbolic mapping of human proximity networks. *Scientific Reports*, 10(1):20244, 2020.

[30] Frederic Sala, Chris De Sa, Albert Gu, and Christopher Ré. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, pages 4460–4469. PMLR, 2018.

[31] Rik Sarkar. Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane. In Marc Van Kreveld and Bettina Speckmann, editors, *Graph Drawing*, volume 7034, pages 355–366. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. Series Title: Lecture Notes in Computer Science.

[32] Yuval Shavitt and Tomer Tankel. Hyperbolic embedding of internet graph for distance estimation and overlay construction. *IEEE/ACM Transactions on Networking*, 16(1):25–36, 2008.

[33] Zequn Sun, Muhao Chen, Wei Hu, Chengming Wang, Jian Dai, and Wei Zhang. Knowledge association with hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2010.02162*, 2020.

[34] Vladimir M. Tikhomirov. ε-entropy and ε-capacity of sets in functional spaces. *Selected Works of AN Kolmogorov: Volume III: Information Theory and the Theory of Algorithms*, pages 86–170, 1993.

[35] Abraham Albert Ungar. *Analytic hyperbolic geometry and Albert Einstein's special theory of relativity*. World Scientific, 2008.

[36] Kevin Verbeek and Subhash Suri. Metric embedding, hyperbolic space, and social networks. In *Proceedings of the thirtieth annual symposium on Computational geometry*, pages 501–510, 2014.

[37] Ulrike von Luxburg and Olivier Bousquet. Distance-based classification with lipschitz functions. *J. Mach. Learn. Res.*, 5(Jun):669–695, 2004.

[38] Menglin Yang, Min Zhou, Zhihao Li, Jiahong Liu, Lujia Pan, Hui Xiong, and Irwin King. Hyperbolic graph neural networks: A review of methods and applications. *arXiv preprint arXiv:2202.13852*, 2022.