Hyperbolic Disk Embeddings for Directed Acyclic Graphs

Ryota Suzuki [1] **Ryusuke Takahama** [1] **Shun Onoda** [1]

Abstract

Obtaining continuous representations of structural data such as directed acyclic graphs (DAGs) has gained attention in machine learning and artificial intelligence. However, embedding complex DAGs in which both ancestors and descendants of nodes are exponentially increasing is difficult. Tackling in this problem, we develop *Disk Embeddings*, which is a framework for embedding DAGs into quasi-metric spaces. Existing state-of-the-art methods, Order Embeddings and Hyperbolic Entailment Cones, are instances of Disk Embedding in Euclidean space and spheres respectively. Furthermore, we propose a novel method *Hyperbolic Disk Embeddings* to handle exponential growth of relations. The results of our experiments show that our Disk Embedding models outperform existing methods especially in complex DAGs other than trees.

1. Introduction

Methods for obtaining appropriate feature representations of a symbolic objects are currently a major concern in machine learning and artificial intelligence. Studies exploiting embedding for highly diverse domains or tasks have recently been reported for graphs (Grover & Leskovec, 2016; Goyal & Ferrara, 2018; Cai et al., 2018), knowledge bases (Nickel et al., 2011; Bordes et al., 2013; Wang et al., 2017), and social networks (Hoff et al., 2002; Cui et al., 2018).

In particular, studies aiming to embed linguistic instances into geometric spaces have led to substantial innovations in natural language processing (NLP) (Mikolov et al., 2013; Pennington et al., 2014; Kiros et al., 2015; Vilnis & McCallum, 2015). Currently, word embedding methods are indispensable for NLP tasks, because it has been shown that the use of pre-learned word

[1]LAPRAS Inc., Tokyo, Japan. Correspondence to: Ryota Suzuki <suzuki@lapras.com>.

embeddings for model initialization leads to improvements in various of tasks (Kim, 2014). In these methods, symbolic objects are embedded in low-dimensional Euclidean vector spaces such that the symmetric distances between semantically similar concepts are small.

In this paper, we focus on modeling asymmetric transitive relations and directed acyclic graphs (DAGs). Hierarchies are well-studied asymmetric transitive relations that can be expressed using a tree structure. The tree structure has a single root node, and the number of children increases exponentially according to the distance from the root. Asymmetric relations that do not exhibit such properties cannot be expressed as a tree structure, but they can be represented as DAGs, which are used extensively to model dependencies between objects and information flow. For example, in genealogy, a family tree can be interpreted as a DAG, where an edge represents a parent child relationship and a node denotes a family member (Kirkpatrick, 2011). Similarly, the commit objects of a distributed revision system (e.g., Git) also form a DAG[1]. In citation networks, citation graphs can be regarded as DAGs, with a document as a node and the citation relationship between documents as an edge (Price, 1965). In causality, DAGs have been used to control confounding in epidemiological studies (Robins, 1987; Merchant & Pitiphat, 2002) and as a means to study the formal understanding of causation (Spirtes et al., 2000; Pearl, 2003; Dawid, 2010).

Recently, a few methods for embedding asymmetric transitive relations have been proposed. Nickel & Kiela reported pioneering research on embedding symbolic objects in hyperbolic spaces rather than Euclidean spaces and proposed *Poincaré Embedding* (Nickel & Kiela, 2017; 2018). This approach is based on the fact that hyperbolic spaces can embed any weighted tree while primarily preserving their metric (Gromov, 1987; Bowditch, 2005; Sarkar, 2011). Vendrov et al. developed *Order Embedding*, which attempts to embed the partially ordered structure of a hierarchy in Euclidean spaces (Vendrov et al., 2016). This method embeds relations among symbolic objects in reserved product orders in Euclidean spaces, which is a type of partial ordering. Objects are represented as inclusive relations of orthants in Euclidean spaces.

Furthermore, inheriting Poincaré Embeddings and Order Embeddings, Ganea & Hofmann proposed *Hyperbolic Entailment Cones*, which embed symbolic objects as convex cones in hyperbolic spaces (Ganea et al., 2018). The authors used polar coordinates of the Poincaré ball and designed their cones such that the number of descendants increases exponentially as one moves going away from the origin. Although the researchers reported that their approach can handle any DAG, their method is not suitable for embedding complex graphs in which the number of both ancestors and descendants grows exponentially, and their experiments were conducted using only hierarchical graphs. Recently, Dong et al. suggested a concept of embedding hierarchy to n-disks (balls) in the Euclidean space (Dong et al., 2019). However, their method cannot be applied to generic DAGs, as it is a bottom up algorithm that can only be applied to trees, which is previously known to be expressed by disks in a Euclidean plane (Stapleton et al., 2011).

In this paper, we propose *Disk Embedding*, a general framework for embedding DAGs in metric spaces, or more generally, quasi-metric spaces. Disk Embedding can be considered as a generalization of the aforementioned state-of-the-art methods: Order Embedding (Vendrov et al., 2016) and Hyperbolic Entailment Cones (Ganea et al., 2018) are equivalent to Disk Embedding in Euclidean spaces and spheres, respectively. Moreover, extending this framework to a hyperbolic geometry, we propose *Hyperbolic Disk Embedding*. Because this method maintains the translational symmetry of hyperbolic spaces and uses exponential growth as a function of the radius, it can process data with exponential increase in both ancestors and descendants. Furthermore, we construct a learning theory for general Disk Embedding using the Riemannian stochastic gradient descent (RSGD) and derive closed-form expressions of the RSGD for frequently used geometries, including Euclidean, spherical, and hyperbolic geometries.

Experimentally, we demonstrate that our Disk Embedding models outperform all of the baseline methods, especially for DAGs other than trees. We used three methods to investigate the efficiency of our approach: Poincaré Embeddings (Nickel & Kiela, 2017), Order Embeddings (Vendrov et al., 2016) and Hyperbolic Entailment Cones (Ganea & Hofmann, 2017).

Our contributions are as follows:

- To embed DAGs in metric spaces, we propose a general framework, Disk Embedding, and systemize its learning method.

- We prove that the existing state-of-the-art methods can be interpreted as special cases of Disk Embedding.

- We propose Hyperbolic Disk Embedding by extending our Disk Embedding to graphs with bi-directional exponential growth.

- Through experiments, we confirm that Disk Embedding models outperforms the existing methods, particularly for general DAGs.

2. Mathematical preliminaries

2.1. Partially ordered sets

As discussed by (Nickel & Kiela, 2018), a concept hierarchy forms a partially ordered set (poset), which is a set X equipped with reflexive, transitive, and antisymmetric binary relations $\preceq$. We extend this idea for application to a general DAG. Considering the reachability from one node to another in a DAG, we obtain a partial ordering.

Partial ordering is essentially equivalent to the inclusive relation between certain subsets called the *lower cone* (or *lower closure*), $C_{\preceq}(x) = \{y \in X | y \preceq x\}$.

Proposition 1 *Let* $(X, \preceq)$ *be a poset. Then,* $x \preceq y$ *holds if and only if* $C_{\preceq}(x) \subseteq C_{\preceq}(y)$.

Embedding DAGs in continuous space can be interpreted as mapping DAG nodes to lower cones with a certain volume that contains its descendants.

Order isomorphism. A pair of poset $(X, \preceq), (X', \preceq')$ is *order isomorphic* if there exists a bijection $f : X \to X'$ that preserves the ordering, i.e., $x \preceq y \Leftrightarrow f(x) \preceq' f(y)$. We further consider that two embedded expressions of a DAGs are equivalent if they are order isomorphic.

2.2. Metric and quasi-metric spaces

A metric space is a set in which a metric function $d : X \times X \to \mathbb{R}$ satisfying the following four properties is defined: **non-negativity**: $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$, **identity of indiscernibles**: $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{y}$, **subadditivity**: $d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{z})$, **symmetry**: $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$.

For generalization, we consider a *quasi-metric* as a natural extension of a metric that satisfies the above properties except for the symmetry (Goubault-Larrecq, 2013a;b). In a Euclidean space $\mathbb{R}^n$, we can obtain various quasi-metrics as follows:

Proposition 2 (Polyhedral quasi-metric) *Let* $W = \{\boldsymbol{w}_i\}$ *be a finite set of vectors in* $\mathbb{R}^n$ *such that* $\mathrm{coni}(W) := \{\sum_i a_i \boldsymbol{w}_i | a_i \geq 0\} = \mathbb{R}^n$. *Let*

$$d_W(\boldsymbol{x}, \boldsymbol{y}) := \max_i \left\{ \boldsymbol{w}_i^\top (\boldsymbol{x} - \boldsymbol{y}) \right\}. \tag{1}$$

Then d_W *is a quasi-metric in* $\mathbb{R}^n$.
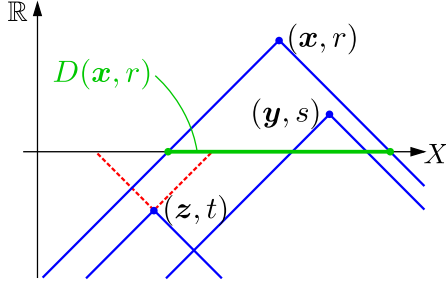
Figure 1. Lower cones (solid blue lines) of generalized formal disks in $\mathbf{B}(X)$. $(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{y}, s)$, $(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{z}, t)$ holds and $(\boldsymbol{z}, t)$ has a negative radius $t < 0$.

The assumption $\mathrm{coni}(W) = \mathbb{R}^n$ in Prop. 2 is equivalent to the condition that convex polytope spanned by W contains the origin in its interior. The shape of disk $d_W(\boldsymbol{p}, \boldsymbol{x}) \leq r$ for fixed $\boldsymbol{p}$ and r is a polytope whose facets are perpendicular to the vectors of W. For instance, let $\{\boldsymbol{e}_k\}$ be a standard basis and $W = \{\pm\boldsymbol{e}_k\}_{k=1}^n$; then, d_W becomes a uniform distance $d_\infty(\boldsymbol{x}, \boldsymbol{y}) = \max_k |x_k - y_k|$, whose disks form hypercubes. An example of d_W whose disks are $(n{-}1)$-simplices appears in Theorem 1 in Sec. 4.1. Except for it, the word quasi-metric in this paper can be replaced by metric since metric is quasi-metric.

2.3. Formal disks

Let (X, d) be a quasi-metric space and consider a *closed disk* $D(\boldsymbol{x}, r) = \{\boldsymbol{p} \in X | d(\boldsymbol{p}, \boldsymbol{x}) \leq r\}$, which is a closed set in the sense of topology induced by d. In considering a Euclidean space with ordinary Euclidean distances, the inclusive relation of two closed disks is characterized as follows:

$$D(\boldsymbol{x}, r) \supseteq D(\boldsymbol{y}, s) \iff d(\boldsymbol{x}, \boldsymbol{y}) \leq r - s. \quad (2)$$

Because this relation is a set inclusion relation, it forms a poset, as discussed above.

In a general metric space, we introduce *formal disks (balls)*[2], which were first introduced by Weihrauch & Schreiber and studied as a computational model of metric space (Blanck, 1997; Edalat & Heckmann, 1998; Heckmann, 1999). Formal disks are also naturally extended to quasi-metric spaces (Goubault-Larrecq, 2013b). Let (X, d) be a quasi-metric space, where a formal disk is defined as a pair $(\boldsymbol{x}, r) \in \boldsymbol{B}^+(X) = X \times \mathbb{R}_+$ of *center* $\boldsymbol{x}$ and *radius* r. The binary relation $\sqsupseteq$ of formal disks holds such that

$$(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{y}, s) \iff d(\boldsymbol{x}, \boldsymbol{y}) \leq r - s \quad (3)$$

is a partial ordering (Edalat & Heckmann, 1998).

---

[2] This is generally called a formal ball, but we call it a formal disk in this paper, even in high-dimensional spaces for clarity.

When defining a partial order with (3), it is straightforward to determine that the radii of formal disks need not be positive. We define $\boldsymbol{B}(X) = X \times \mathbb{R}$ as a collection of *generalized* formal disks (Tsuiki & Hattori; Goubault-Larrecq, 2013a) that are allowed to have *negative* radius. Lower cones $C_\sqsubseteq(\boldsymbol{x}, r)$ of formal disks are shown in Figure 1. The cross section $\{(\boldsymbol{x}', 0) \in C_\sqsubseteq(\boldsymbol{x}, r)\}$ of the lower cone $C_\sqsubseteq(\boldsymbol{x}, r)$ at $r = 0$, which is described as a solid green segment in Fig. 1, is a closed disk in X. The negative radius of generalized formal disk can be regarded as the radius of the cross section of the upper cone (dashed red line in Fig. 1).

The following properties hold for generalized formal disks.

Translational symmetry. For all $a \in \mathbb{R}$,
$$(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{y}, s) \Leftrightarrow (\boldsymbol{x}, r + a) \sqsupseteq (\boldsymbol{y}, s + a). \quad (4)$$

Reversibility. If d is symmetric (i.e., d is metric),
$$(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{y}, s) \Leftrightarrow (\boldsymbol{x}, -r) \sqsubseteq (\boldsymbol{y}, -s). \quad (5)$$

These properties reflect the reversibility of the graph and symmetry between generations, which are important when embedding DAGs in posets of formal disks.

2.4. Riemannian manifold

A Riemannian manifold is a manifold $\mathcal{M}$ with a collection of inner products $g_{\boldsymbol{x}} : T_{\boldsymbol{x}}\mathcal{M} \times T_{\boldsymbol{x}}\mathcal{M} \to \mathbb{R}$, called a Riemann metric. Let $\gamma : [0, 1] \to \mathcal{M}$ be a smooth curve, where a length of γ is calculated by:

$$\ell(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} \, dt.$$

The infimum of the curve length $d_\mathcal{M}(\boldsymbol{x}, \boldsymbol{y}) = \inf_\gamma \ell(\gamma)$ from $\boldsymbol{x}$ to $\boldsymbol{y}$, which becomes metric (and, consequently, quasi-metric) at $\mathcal{M}$.

Geodesic: A geodesic is a Riemannian analog of straight lines in Euclidean spaces and is defined as a locally length-minimizing curve. If Riemannian manifold is complete, for every two points $\boldsymbol{x}$ and $\boldsymbol{y}$, there exists a geodesic that connects $\boldsymbol{x}$ and $\boldsymbol{y}$ with a minimal length.

Exponential map: An exponential map $\exp_{\boldsymbol{x}} : T_{\boldsymbol{x}}\mathcal{M} \to \mathcal{M}$ is defined as $\exp_{\boldsymbol{x}}(\mathbf{v}) = \gamma(1)$, where γ is a unique geodesic satisfying $\gamma(0) = x$ with an initial tangent vector $\gamma'(0) = \mathbf{v}$. The map can also be interpreted as the destination reached after a unit time when traveling along the geodesic from x at an initial velocity of $\mathbf{v}$. Therefore, $d(\boldsymbol{x}, \exp_{\boldsymbol{x}}(\mathbf{v})) = \sqrt{g_{\boldsymbol{x}}(\mathbf{v}, \mathbf{v})}$ holds in a sufficiently small neighborhood of $\mathbf{v} = 0$.

2.5. Hyperbolic geometry

A hyperbolic geometry is a uniquely characterized complete and simply connected geometry with constant

negative curvature. Hyperbolic geometries with curvature of -1 are described by several models that are identical up to isometry. The frequently used models in machine learning include the Poincaré ball model (Nickel & Kiela, 2017) and the Lorentz model (Nickel & Kiela, 2018).

Poincaré ball model $\mathbb{D}^n$ is a Riemannian manifold defined on the interior of an n-dimensional unit ball, where distances are calculated as

$$d_{\mathbb{D}}(\boldsymbol{x}, \boldsymbol{y}) = \operatorname{arcosh}\left(1 + 2\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{(1 - \|\boldsymbol{x}\|^2)(1 - \|\boldsymbol{y}\|^2)}\right), \tag{6}$$

and its geodesics are arcs that perpendicularly intersect the boundary sphere.

Lorentz model $\mathcal{L}^n$ is defined on n-dimensional hyperboloid $\{(x_0, \cdots, x_n) | -x_0^2 + \sum_{k=1}^n x_k^2 = -1\}$ embedded in $(n+1)$-dimensional Euclidean space, where distance is calculated as

$$d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y}) = \operatorname{arcosh}(-\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\mathcal{L}}), \tag{7}$$

where $-\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\mathcal{L}} = -x_0 y_0 + \sum_{k=1}^n x_k y_k$.

Translational symmetry Hyperbolic geometry is symmetric with respect to translations, i.e., geometric properties including distances do not change even if all points are translated in the same time. In the Poincaré ball model, a translation that maps the origin to $\boldsymbol{y}$ is obtained as

$$\boldsymbol{x} \mapsto \frac{(1 - \|\boldsymbol{y}\|^2)\,\boldsymbol{x} + (1 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{x}\|^2)\,\boldsymbol{y}}{1 + 2\langle \boldsymbol{x}, \boldsymbol{y}\rangle + \|\boldsymbol{y}\|^2\|\boldsymbol{x}\|^2}, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is an ordinary Euclidean inner product.

3. Disk Embedding models

In this section, we introduce *Disk Embeddings* as a general platform for embedding DAGs in quasi-metric spaces. We first define Disk Embeddings and evaluate its representability. Second, we introduce Hyperbolic Disk Embeddings and discuss its qualitative nature. Third, we derive loss functions and an optimization method for Disk Embeddings and finally obtain closed-form expressions for some commonly used geometries.

3.1. Disk Embeddings and its representability

Let $C = \{c_i\}_{i=1}^N$ be a set of entities with partial ordering relations $\preceq$, and let (X, d) be a quasi-metric space. Disk Embeddings are defined as a map $c_i \mapsto (\boldsymbol{x}_i, r_i) \in \mathbf{B}(X)$ such that $c_i \preceq c_j$ if and only if $(\boldsymbol{x}_i, r_i) \sqsubseteq (\boldsymbol{x}_j, r_j)$.

We can use various type of quasi-metric spaces (X, d), where X can represent spherical or hyperbolic spaces,
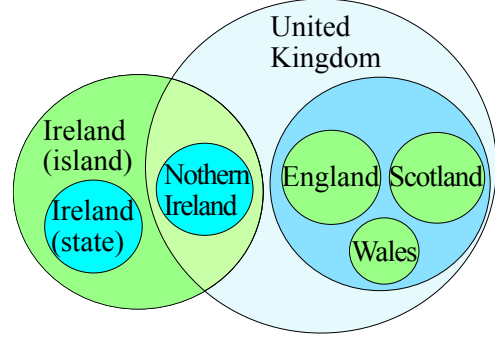

Figure 2. Disk Embeddings in 2D Euclidean space with non-negative radii are equivalent to 2D Euler diagrams of circles.

and d can be a unique quasi-metric. Assuming that X is an ordinary 2D Euclidean plane $\mathbb{R}^2$ and that formal disks have positive radii, Disk Embeddings become an inclusive relation of closed disks, which is equivalent to an Euler diagram with circles, except that only inclusions are meaningful and intersections do not make sense (Figure 2). Stapleton et al. studied the drawability of 2D Euler diagrams and demonstrated that a certain class of diagrams, termed *pierced diagrams*, can be drawn with circles (Stapleton et al., 2011). This result suggests that our Disk Embeddings will be effective for certain classes of problems, even for only three dimensions (two for centers and one for radii). As Dong et al. mentioned, the tree structure can be obviously embedded in n-disks because it is the simplest *pierced* diagram, in which none of circles intersect each other (Dong et al., 2019). In higher Euclidean spaces, Disk Embeddings have a greater representability than 2D because all graphs that are representable in 2D can also be embedded in higher dimensions via a natural injection $\iota : \mathbb{R}^2 \to \mathbb{R}^n$, s.t. $\iota(x_1, x_2) = (x_1, x_2, 0, \cdots, 0)$.

As a related work, embedding DAGs in Minkowski spacetime is introduced by (Clough & Evans, 2017), in which the related nodes are timelike separated. The lower (upper) cones in Disk Embeddings can be interpreted to the *light cones* of Minkowski spacetime.

Since we focus on transitive relation of DAGs, non-transitive details of DAGs are not reserved in any models based on order isomorphism including Disk Embeddings, Hyperbolic Cones (Ganea et al., 2018) and Order Embeddings (Vendrov et al., 2016). For instance, a DAG with three node x, y, z and edges $x \to y$ and $y \to z$ will have a same representation as a DAG having an additional edge $x \to z$.

3.2. Hyperbolic Disk Embeddings

We now introduce Disk Embeddings for a hyperbolic geometry. In a hyperbolic geometry of two or more

dimensions, it is known that the area of the disk increases as an the exponential function of the radius. Thus, the lower cone section, shown in Figure 1, increases exponentially as the generation moves from parent to child in the DAG. In addition, it should be noted that both the inner and outer sides of the lower cones becomes wider. For instance, when $(\boldsymbol{x}, r) \sqsupseteq (\boldsymbol{y}, s)$, the region $C_{\sqsubseteq}(\boldsymbol{x}, r) - C_{\sqsubseteq}(\boldsymbol{y}, s)$ also increases exponentially. This property of the Hyperbolic Disk Embeddings is suitable for embedding graphs in which the number of descendants increases rapidly.

Considering the reversibility of Disk Embeddings (5), the above property is established for not only the descendant but also the ancestor direction. In other words, not only lower cones but also upper cones show an exponential extension. In addition, considering translational symmetry in hyperbolic spaces (8) and along radial axis (4), the same result holds for lower and upper cones starting from any node in the graph. Thus, in Hyperbolic Disk Embeddings, complex threads that repeatedly intersect and separate from each other in a complex DAG can be accurately expressed.

3.3. Loss functions

We define a *protrusion parameter* between two formal disks $(\boldsymbol{x}_i, r_i), (\boldsymbol{x}_j, r_j) \in \mathbf{B}(X)$ as

$$l_{ij} = l(\boldsymbol{x}_i, r_i; \boldsymbol{x}_j, r_j) = d(\boldsymbol{x}_i, \boldsymbol{x}_j) - r_i + r_j. \quad (9)$$

Because $(\boldsymbol{x}_i, r_i) \sqsupseteq (\boldsymbol{x}_j, r_j)$ implies $l_{ij} \leq 0$, an appropriate representation can be obtained by learning such that l_{ij} is small for positive examples and large for negative examples.

Although various loss functions can be used, we adopt the following margin loss in this study, similar to that adopted by (Vendrov et al., 2016; Ganea et al., 2018):

$$L = \sum_{(i,j) \in \mathcal{T}} h_+(E_{ij}) + \sum_{(i,j) \in \mathcal{N}} h_+(\mu - E_{ij}), \quad (10)$$

where $\mathcal{T}$ and $\mathcal{N}$ are sets of positive and negative samples, μ is a constant margin, $h_+(x) = \max(0, x)$ and $E_{ij} = E(\boldsymbol{x}_i, r_i; \boldsymbol{x}_j, r_j)$ is an arbitrary energy function such that $E_{ij} > 0$ if and only if $l_{ij} > 0$. The shape of (10) is shown in Figure 3. Naturally, we can simply set

$$E_{ij} = l_{ij}, \quad (11)$$

in which case, the loss function corresponds to ones used by (Vendrov et al., 2016) and (Ganea et al., 2018), except they used $h_+(l_{ij})$ instead, causing that the gradient vanishes at $l_{ij} < 0$ in negative samples (dashed line in Fig. 3).
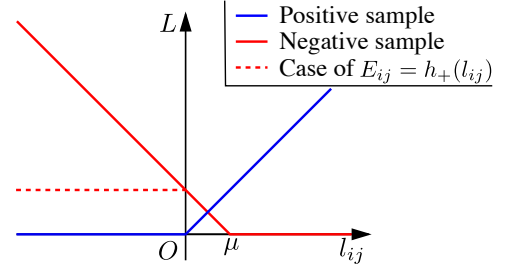

Figure 3. Loss function. solid lines are the case of (11) and dashed line is the case of $E_{ij} = h_+(l_{ij})$.

3.4. Riemannian optimization

Given that X is a general Riemannian manifold, we must account for the Riemannian metric when optimizing the loss function. In this case, we utilize the RSGD method (Bonnabel, 2013), which is similar to the ordinary SGD except that the Riemannian gradient is used instead of the gradient and an exponential map is used for updates.

To execute the RSGD on $\mathbf{B}(X) = X \times \mathbb{R}$, we must determine the Riemannian metric in the product space and calculate the corresponding Riemannian gradient. To maintain translational symmetry Eq.(4) along the radial axis, the Riemannian gradient on $\mathbf{B}(X)$ should have the following form

$$\nabla_{\mathbf{B}(X)} f(\boldsymbol{x}, r) = \left(\lambda \nabla_{X;\boldsymbol{x}} f(\boldsymbol{x}, r), \nu \frac{\partial}{\partial r} f(\boldsymbol{x}, r) \right), \quad (12)$$

where $\nabla_{X;\boldsymbol{x}}$ is the Riemannian gradient on X for variable $\boldsymbol{x}$ and λ, ν are positive constants. Then the update formulae of RSGD for the parameters $(\boldsymbol{x}_i, r_i)$ are given as follows:

$$\boldsymbol{x}_i^{(t+1)} = \exp_{\boldsymbol{x}_i^{(t)}} \left[-\eta \lambda \nabla_{X;\boldsymbol{x}_i} L^{(t)} \right], \quad (13)$$

$$r_i^{(t+1)} = r_i^{(t)} - \eta \nu \frac{\partial L^{(t)}}{\partial r_i}, \quad (14)$$

where η is a learning ratio.

A specific form of (13) is directly derived from the Riemannian gradient of quasi-metric $d(\boldsymbol{x}, \boldsymbol{y})$ and the exponential map for each geometry via equations (9) to (11). Especially, if the distance function d_X induced from the Riemannian metric in X is used as the quasi-metric d for formal disks in (9), the Riemannian gradient of d has a special form:

$$\nabla_{X;\boldsymbol{x}} d(\boldsymbol{x}, \boldsymbol{y}) = \nabla_{X;\boldsymbol{x}} d(\boldsymbol{y}, \boldsymbol{x}) = -\gamma'(0) \quad (15)$$

where $\gamma'(0)$ is the initial tangent vector of the unit speed geodesic connecting from $\boldsymbol{x}$ to $\boldsymbol{y}$. In addition, for frequently used geometries, we here present closed form expressions as follows.
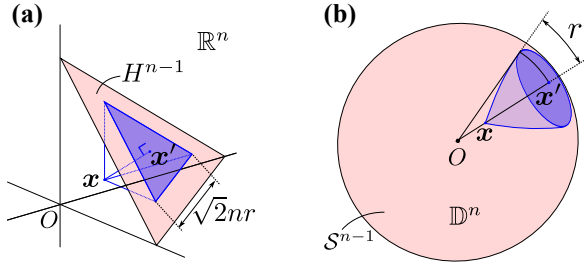
Figure 4. Equivalence of Disk Embeddings and existing methods. **(a)** Order Embeddings (Vendrov et al., 2016) and **(b)** Hyperbolic Entailment Cones (Ganea et al., 2018).

Euclidean geometry. The RSGD in Euclidean spaces is equivalent to the standard SGD, where the gradient of the distance is $\nabla_{\boldsymbol{x}} d(\boldsymbol{x}, \boldsymbol{y}) = \frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|}$ and the exponential map is $\exp_{\boldsymbol{x}}(\mathbf{v}) = \boldsymbol{x} + \mathbf{v}$.

Spherical geometry. The RSGD on an n-sphere $\mathcal{S}^n$ with spherical distance $d_{\mathcal{S}}$ is conducted using the following formulae,

$$\nabla_{\mathcal{S};\boldsymbol{x}} d_{\mathcal{S}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{-\boldsymbol{h}}{\|\boldsymbol{h}\|}, \quad \boldsymbol{h} = \boldsymbol{y} - \langle \boldsymbol{x}, \boldsymbol{y} \rangle \boldsymbol{x},$$

$$\exp_{\mathcal{S};\boldsymbol{x}}(\mathbf{v}) = \boldsymbol{x} \cos(\|\mathbf{v}\|) + \frac{\mathbf{v}}{\|\mathbf{v}\|} \sin(\|\mathbf{v}\|).$$

Hyperbolic geometry. We use the Lorentz model (Nickel & Kiela, 2018) for Hyperbolic Disk Embeddings because the RSGD in the Lorentz model involves considerably simpler formulae and has a greater numerical stability compared to the Poincaré ball model. In the Lorentz model, the gradient of the distance and the exponential map are computed as follows,

$$\nabla_{\mathcal{L};\boldsymbol{x}} d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y}) = \frac{-\boldsymbol{h}}{\|\boldsymbol{h}\|_{\mathcal{L}}}, \quad \boldsymbol{h} = \boldsymbol{y} + \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{L}} \boldsymbol{x},$$

$$\exp_{\mathcal{L};\boldsymbol{x}}(\mathbf{v}) = \boldsymbol{x} \cosh(\|\mathbf{v}\|_{\mathcal{L}}) + \frac{\mathbf{v}}{\|\mathbf{v}\|_{\mathcal{L}}} \sinh(\|\mathbf{v}\|_{\mathcal{L}}),$$

where $\|\mathbf{v}\|_{\mathcal{L}} = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle_{\mathcal{L}}}$.

4. Equivalence of Disk Embedding models

In this section, we illustrate the relationship between our Disk Embeddings and the current state-of-the-art methods. We demonstrate that the embedding methods for DAGs in metric spaces are equivalent to Disk Embeddings by projecting lower cones into appropriate $n-1$ subspaces.

4.1. Order Embeddings (Vendrov et al., 2016)

Vendrov et al. proposed Order Embeddings, in which a partial ordering relation is embedded in reversed product

order on $\mathbb{R}_+^n$,

$$\boldsymbol{x} \succeq_{\mathrm{ord}} \boldsymbol{y} \Leftrightarrow x_k \leq y_k \text{ for } k = 1, \cdots, n. \quad (16)$$

In Order Embeddings, the shape of the lower cone $C_{\preceq}(\boldsymbol{x}) = [x_1, \infty) \times \cdots \times [x_n, \infty)$ is orthant.

As shown in Figure 4(**a**), we consider a projection onto a hyperplane $H^{n-1} = \left\{ (x_1, \cdots, x_n) \big| \frac{1}{n} \sum_{k=1}^{n} x_k = a \right\}$, that is isometric to $\mathbb{R}^{n-1}$. The shape of the cross-section of the lower cone with H^{n-1} is $(n-1)$-simplex[3] Thus, we can consider the relation (16) as an inclusive relation between corresponding simplices. By using a polyhedral quasi-metric d_W (1), we show that this is equivalent to Disk Embeddings in H_n with an additional constraint $r < a - d_W(\mathbf{0}, \boldsymbol{x})$, which forces all entities to be descendants of the origin.

Theorem 1 *Order Embeddings* (16) *is order isomorphic to Euclidean Disk Embeddings with quasi-metric* d_W *with* $W = \{P\boldsymbol{e}_k\}_{k=1}^{n}$ *via a smooth map* $\phi_{\mathrm{ord}} : \mathbb{R}_+^n \to \mathbf{B}(H^{n-1})$,

$$\phi_{\mathrm{ord}}(\boldsymbol{x}) = (\boldsymbol{x}', r) = \left(P\boldsymbol{x}, \ a - \frac{1}{n} \sum_{k=1}^{n} x_k \right), \quad (17)$$

where $P = I - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ *is a projection matrix onto* H^{n-1}, $\mathbf{1}^\top = (1, \cdots, 1)$.

The energy function used for their loss function is

$$E^{\mathrm{ord}}(\boldsymbol{x}, \boldsymbol{y}) = \|h_+(\boldsymbol{x} - \boldsymbol{y})\|^2, \quad (18)$$

which cannot be directly expressed using l_{ij} defined in (9), but can be well approximated by a lower bound.

Theorem 2 *Energy function* (18) *has a lower bound:*

$$E_{ij}^{\mathrm{ord}} \geq h_+(l_{ij})^2, \quad (19)$$

and the equality holds iff $\left| \left\{ k \big| (\boldsymbol{x}_j - \boldsymbol{x}_i)_k > 0 \right\} \right| \leq 1$.

In contrast to (11), (19) has a quadratic form, which may cause optimization being exponential decay of l_{ij} instead of crossing zero and a vanishing gradient in $l_{ij} < 0$ even for negative samples, making the optimization inefficient.

4.2. Hyperbolic Entailment Cones (Ganea et al., 2018)

Ganea et al. developed a method for embedding cones extending in the radial direction of the Poincaré ball model. The embedding relation is expressed as follows

$$\boldsymbol{x} \succeq_{\mathrm{hyp}} \boldsymbol{y} \Leftrightarrow \psi(\boldsymbol{x}) \geq \Xi(\boldsymbol{x}, \boldsymbol{y}), \quad (20)$$

where $\Xi(\boldsymbol{x}, \boldsymbol{y})$ is the angle $\pi - \angle Oxy$ and $\psi(\boldsymbol{x}) = \arcsin\left(K \frac{1 - \|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|} \right)$ is the generatrix angle of the cone.

---

[3] 1-simplex is a line segment, 2-simplex is a regular triangle, 3-simplex is a regular tetrahedron, and so on.

The authors focused on the polar coordinates of the Poincaré ball, in which rotational symmetry around the origin is assumed and the position in the hierarchical structure is mapped to the radial component. Thus, they implicitly assumed a non-trivial separation of the hyperbolic space into $\mathcal{S}^{n-1} \times \mathbb{R}_+$. To illustrate this point, we consider projections of entailment cones onto the border of a Poincaré ball $\partial \mathbb{D} \cong \mathcal{S}^{n-1}$. As shown in Figure 4(**b**), the projections of the entailment cones form disks in an $(n-1)$-sphere, and relation (20) is represented as the inclusion of corresponding disks.

Theorem 3 *Hyperbolic Cones* (20) *are order isomorphic to Disk Embeddings on $(n-1)$-sphere $\mathcal{S}^{n-1}$ via a smooth map:* $\phi_{\mathrm{hyp}} : \mathbb{D}^n \to \mathbf{B}(\mathcal{S}^{n-1})$,

$$
\begin{aligned}
\phi_{\mathrm{hyp}}(\boldsymbol{x}) &= (\boldsymbol{x}', r) \\
&= \left(\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \ \arcsin\left(\frac{1 + \|\boldsymbol{x}\|^2}{2\|\boldsymbol{x}\|} \sin\theta_0 \right) - \theta_0 \right),
\end{aligned}
\tag{21}
$$

where $\theta_0 = \arctan 2K$.

Their energy function is given as

$$
E^{\mathrm{hyp}}(\boldsymbol{x}, \boldsymbol{y}) = h_+ \left(\Xi\left(\boldsymbol{x}, \boldsymbol{y} \right) - \psi\left(\|\boldsymbol{x}\| \right) \right), \tag{22}
$$

and is also represented in the format of Disk Embeddings, by using only distances between centers and the radii of formal disks.

Theorem 4 *Energy function* (22) *has the following form:*

$$
E^{\mathrm{hyp}}_{ij} = h_+ \left(\arcsin\left(q(d_{ij}, r_i, r_j) \cdot 2\sin\left(\frac{l_{ij}}{2} \right) \right) \right), \tag{23}
$$

where

$$
\begin{aligned}
q(d_{ij}, r_i, r_j) =& \frac{\cos\left(\frac{r_i + r_j - d_{ij}}{2} + \theta_0 \right)}{\cos\theta_0 \ \sin\theta_0} \\
& \times \sqrt{\frac{\sin r_i \ \sin(r_i + 2\theta_0)}{s_i^2 + s_j^2 - 2s_i s_j \cos d_{ij} - \sin^2 d_{ij}}},
\end{aligned}
\tag{24}
$$

and $s_i = \frac{\sin(r_i + \theta_0)}{\sin\theta_0}, s_j = \frac{\sin(r_j + \theta_0)}{\sin\theta_0}$.

As can be easily seen from (23), the energy E^{hyp}_{ij} is linearly approximated by:

$$
E^{\mathrm{hyp}}_{ij} = q(r_i - r_j, r_i, r_j) \cdot h_+(l_{ij}) + \mathcal{O}\left(d_{ij}^2 \right), \tag{25}
$$

around $d_{ij} \approx r_i - r_j$ for fixed r_i, r_j. Equation (25) is similar to (11) except that the coefficient $q(r_i - r_j, r_i, r_j)$ and gradient vanishing at $l_{ij} < 0$ as observed for (19).

Table 1. Dataset statistics. The average number of ancestors (descendants) of leaf (root) nodes are shown.

Dataset	Nodes	Edges	Ancestors	Descendants
WordNet	82,115	743,086	9.2	82114.0

4.3. Vector Embedding models

Ordinary embedding models based on similarity, e.g, word2vec (Mikolov et al., 2013) and Poincaré Embeddings (Nickel & Kiela, 2017; 2018), can be seen as a application of Disk Embeddings in which radius information is neglected.

Nickel & Kiela argued that when embedding in hyperbolic spaces, general concepts can be obtained closer to the origin by learning with loss function distances between points. However, because a hyperbolic space has translational symmetry, there are no special point and any point can be the origin. Thus, simultaneously translating all of the points does not change the loss function, and which node is closer to the origin is determined only by the initial vectors. Furthermore, an approach in which distances from the origin are interpreted as levels in a hierarchy is not suitable for complex DAGs in which both ancestors and descendants grow exponentially, with no single root.

5. Experiments

In this section we evaluate Disk Embedding models for various metric spaces including Euclidean, spherical and hyperbolic geometries.

5.1. Datasets

For evaluation we use the *WordNet* ® (Miller, 1995)[4], a large lexical database that provides hypernymy relations. In our experiments, we used a noun closure for evaluating hierarchical data. The statistics for the dataset are shown in Table 1. The WordNet noun dataset is an example of a tree-like hierarchical network, characterized by a highly limited number of ancestors, whereas the number of descendants is considerably large. We also used data obtained by inverting the relations in WordNet noun dataset. Because the reverse of a DAG is a DAG, this data is considered to be an example of non-tree structural DAG.

5.2. Training and evaluation details

We conducted learning by using pairs of nodes connected by edges of these graph data as positive examples $(i, j) \in \mathcal{T}$ such that $c_i \preceq c_j$. Because these data only have positive pairs, we randomly sampled negative pairs for each iteration of RSGD. For evaluating the learned expression,

---

[4] https://wordnet.princeton.edu/

Table 2. Test F1 results for various models. Hyperbolic Entailment Cones is proposed by (Ganea et al., 2018), Order Embeddings is proposed by (Vendrov et al., 2016) and Poincaré Embeddings is proposed by (Nickel & Kiela, 2017).

	Embedding Dimension $= 5$				Embedding Dimension $= 10$			
	Percentage of Transitive Closure (Non-basic) Edges in Training							
	0%	10%	25%	50%	0%	10%	25%	50%
WordNet nouns								
Our Euclidean Disk Embeddings	35.6%	38.9%	42.5%	45.1%	**45.6**%	54.0%	65.8%	72.0%
Our Hyperbolic Disk Embeddings	32.9%	69.1%	81.3%	83.1%	36.5%	79.7%	90.5%	**94.2**%
Our Spherical Disk Embeddings	**37.5**%	**84.8**%	**90.5**%	**93.4**%	42.0%	**86.4**%	**91.5**%	93.9%
Hyperbolic Entailment Cones	29.2%	80.0%	87.1%	92.8%	32.4%	84.9%	90.8%	93.8%
Order Embeddings	34.4%	70.6%	75.9%	82.1%	43.0%	69.7%	79.4%	84.1%
Poincaré Embeddings	28.1%	69.4%	78.3%	83.9%	29.0%	71.5%	82.1%	85.4%
WordNet nouns reversed								
Our Euclidean Disk Embeddings	**35.4**%	38.7%	42.3%	44.6%	**46.6**%	55.9%	67.3%	70.6%
Our Hyperbolic Disk Embeddings	30.8%	49.0%	66.8%	78.5%	32.1%	53.7%	79.1%	88.2%
Our Spherical Disk Embeddings	34.8%	**59.0**%	**76.8**%	**84.9**%	38.0%	60.6%	**83.1**%	**90.1**%
Hyperbolic Entailment Cones	17.3%	57.5%	71.8%	75.7%	20.5%	**61.9**%	73.1%	75.8%
Order Embeddings	32.9%	33.8%	34.8%	35.8%	34.7%	36.7%	38.8%	41.4%
Poincaré Embeddings	26.0%	48.4%	48.8%	51.4%	27.4%	49.7%	50.9%	51.9%

we use an F1 score for a binary classification of whether a randomly selected pair (i, j) satisfies the transitive relation $c_i \succeq c_j$, in other words, to assess whether there exists a directed path from i to j in the DAG.

5.3. Baselines

We also evaluated Poincaré Embeddings (Nickel & Kiela, 2017), Order Embeddings (Vendrov et al., 2016) and Hyperbolic Entailment Cones (Ganea et al., 2018) as baseline methods. For these baseline methods, we used the implementation reported by Ganea et al.[5]. In addition, experimental conditions such as hyperparameters are designed to be nearly similar to those of the experiments conducted by Ganea et al.. Considering that Hyperbolic Cones (Ganea et al., 2018) use Poincaré Embeddings (Nickel & Kiela, 2017) for pretraining, we also apply this approach to Spherical Disk Embeddings (which is equivalent to Hyperbolic Cones as shown in Theorem 3) to make a fair comparison. Although Poincaré Embeddings (Nickel & Kiela, 2017) is a method used for learning symmetric relations based on similarity, it can also be used to estimate asymmetric relations by using a heuristic score $S(\boldsymbol{x}, \boldsymbol{y}) = (1 + \lambda(\|\boldsymbol{x}\| - \|\boldsymbol{y}\|)) \, d_{\mathbb{D}}(\boldsymbol{x}, \boldsymbol{y})$, where the parameter λ is determined by maximizing the F1 score for validation data which are sampled separately from the test data.

5.4. Results and discussion

Table 2 shows the F1 score on each dataset. As shown in Section 4, we proved equivalence between our Spherical

Disk Embeddings and Hyperbolic Entailment Cones. It is observed that our Spherical Disk Embeddings reaches almost the same result of Hyperbolic Entailment Cones with **WordNet nouns**. The slight improvement of our model can be explained the change of the loss function. **WordNet nouns reversed** is generated from WordNet by reversing directions of all edges, nevertheless, it is an example of DAGs. In this data, our Disk Embeddings models obviously outperformed other existing methods because our methods maintain reversibility (Eq. (5)) while existing methods implicitly assume hierarchical structure.

6. Conclusion

We introduced Disk Embeddings, which is a new framework for embedding DAGs in quasi-metric spaces to generalize the state-of-the-art methods. Furthermore, extending this framework to a hyperbolic geometry, we propose Hyperbolic Disk Embedding. Experimentally we demonstrate that our methods outperform all of the baseline methods, especially for DAGs other than tree.

For future work, large-scale experiments for complex DAGs such such as citation networks is desired, in which both ancestors and descendants increase rapidly, and exponential nature of Hyperbolic Disk Embedding will demonstrate its core.

For reproducibility, our source code for the experiments are publicly available online[6]. The datasets we used are also available online. See Section 5.1.

---

[5] https://github.com/dalab/hyperbolic_cones

[6] https://github.com/lapras-inc/disk-embedding

Acknowledgements

References

Blanck, J. Domain Representability of Metric Spaces. *Annals of Pure and Applied Logic*, 83(3):225–247, 1997. ISSN 0168-0072. doi: 10.1016/S0168-0072(96) 00017-6.

Bonnabel, S. Stochastic Gradient Descent on Riemannian Manifolds. *IEEE Transactions on Automatic Control*, 58 (9):2217–2229, 2013.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. Translating Embeddings for Modeling Multi-Relational Data. In *Advances in Neural Information Processing Systems 26 (NeurIPS)*, pp. 2787–2795, 2013.

Bowditch, B. H. A Course on Geometric Group Theory, 2005.

Cai, H., Zheng, V. W., and Chang, K. A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

Clough, J. R. and Evans, T. S. Embedding graphs in Lorentzian spacetime. *PLOS ONE*, 12(11), 2017.

Cui, P., Wang, X., Pei, J., and Zhu, W. A Survey on Network Embedding. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

Dawid, A. P. Beware of the DAG! In *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, volume 6, pp. 59–86, 2010.

Dong, T., Cremers, O., Jin, H., Li, J., Bauckhage, C., Cremers, A. B., Speicher, D., and Zimmermann, J. Imposing Category Trees Onto Word-Embeddings Using A Geometric Construction. In *Proceedings of the Seventh International Conference on Learning Representations (ICLR)*, 2019.

Edalat, A. and Heckmann, R. A Computational Model for Metric Spaces. *Theoretical Computer Science*, 193 (1):53–73, 1998. ISSN 0304-3975. doi: 10.1016/ S0304-3975(96)00243-5.

Ganea, O.-E. and Hofmann, T. Deep Joint Entity Disambiguation with Local Neural Attention. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.

2619–2629. Association for Computational Linguistics, 2017.

Ganea, O.-E., Bécigneul, G., and Hofmann, T. Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Goubault-Larrecq, J. *Metrics, Quasi-metrics, Hemimetrics*, pp. 203–259. New Mathematical Monographs. Cambridge University Press, 2013a. doi: 10.1017/ CBO9781139524438.006.

Goubault-Larrecq, J. A Few Pearls in the Theory of Quasi-Metric Spaces. In *Proceedings of the 28th Summer Conference on Topology and Its Applications*, Nipissing University, North Bay, Ontario, Canada, 2013b.

Goyal, P. and Ferrara, E. Graph Embedding Techniques, Applications, and Performance: A Survey. *Knowledge-Based Systems*, 151:78–94, 2018.

Gromov, M. Hyperbolic Groups. In *Essays in Group Theory*, pp. 75–263. Springer, 1987.

Grover, A. and Leskovec, J. node2vec: Scalable Feature Learning for Networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 855–864. ACM, 2016.

Heckmann, R. Approximation of Metric Spaces by Partial Metric Spaces. *Applied Categorical Structures*, 7(1): 71–83, 1999. ISSN 1572-9095. doi: 10.1023/A: 1008684018933.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. Latent Apace Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

Kim, Y. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, 2014.

Kirkpatrick, B. B. Haplotypes Versus Genotypes on Pedigrees. *Algorithms for Molecular Biology*, 6(1):10, 2011.

Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., and Fidler, S. Skip-Thought Vectors. In *Advances in Neural Information Processing Systems 27 (NeurIPS)*, pp. 3294–3302, 2015.

Merchant, A. T. and Pitiphat, W. Directed Acyclic Graphs (DAGs): An Aid to Assess Confounding in Dental

Research. *Community dentistry and oral epidemiology*, 30(6):399–404, 2002.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26 (NeurIPS)*, pp. 3111–3119. 2013.

Miller, G. A. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11):39–41, 1995.

Nickel, M. and Kiela, D. Poincaré Embeddings for Learning Hierarchical Representations. In *Advances in Neural Information Processing Systems 30 (NeurIPS)*, pp. 6338–6347, 2017.

Nickel, M. and Kiela, D. Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

Nickel, M., Tresp, V., and Kriegel, H.-P. A Three-Way Model for Collective Learning on Multi-Relational Data. In *Proceedings of the 28th International Conference on Machine Learning (ICML)*, volume 11, pp. 809–816, 2011.

Pearl, J. Causality: Models, Reasoning, and Inference. *Econometric Theory*, 19(675-685):46, 2003.

Pennington, J., Socher, R., and Manning, C. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014.

Price, D. J. D. S. Networks of Scientific Papers. *Science*, pp. 510–515, 1965.

Robins, J. A Graphical Approach to the Identification and Estimation of Causal Parameters in Mortality Studies with Sustained Exposure Periods. *Journal of chronic diseases*, 40:139S–161S, 1987.

Sarkar, R. Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane. In *International Symposium on Graph Drawing*, pp. 355–366. Springer, 2011.

Spirtes, P., Glymour, C. N., Scheines, R., Heckerman, D., Meek, C., Cooper, G., and Richardson, T. *Causation, Prediction, and Search*. MIT press, 2000.

Stapleton, G., Zhang, L., Howse, J., and Rodgers, P. Drawing Euler Diagrams with Circles: The Theory of Piercings. 17(7):1020–1032, 2011. ISSN 1077-2626. doi: 10.1109/TVCG.2010.119.

Tsuiki, H. and Hattori, Y. Lawson Topology of the Space of Formal Balls and the Hyperbolic Topology. *Theoretical Computer Science*, 405(1):198–205. ISSN 0304-3975. doi: 10.1016/j.tcs.2008.06.034.

Vendrov, I., Kiros, R., Fidler, S., and Urtasun, R. Order-Embeddings of Images and Language. In *Proceedings of the 4th International Conference on Learning Representations (ICLR)*, 2016.

Vilnis, L. and McCallum, A. Word Representations via Gaussian Embedding. In *Proceedings of the Seventh International Conference on Learning Representations (ICLR)*, 2015.

Wang, Q., Mao, Z., Wang, B., and Guo, L. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, 2017.

Weihrauch, K. and Schreiber, U. Embedding Metric Spaces into CPO's. *Theoretical Computer Science*, 16(1):5–24, 1981. ISSN 0304-3975. doi: 10.1016/0304-3975(81) 90027-X.