Hyperbolic matrix factorization reaffirms the negative curvature of the native biological space

Aleksandar Poleksic

Department of Computer Science, University of Northern Iowa, Cedar Falls, IA 50613

Abstract

Past research in systems biology has taken for granted the Euclidean geometry of biological space. This has not only drawn parallels to other fields but has also been convenient due to the ample statistical and numerical optimization tools available to address the core task and downstream machine learning problems. However, emerging theoretical studies now demonstrate that biological databases exhibit hierarchical topology, characterized by heterogeneous degree distribution and a high degree of clustering, thus contradicting the flat geometry assumption. Namely, since the number of nodes in hierarchical structures grows exponentially with node depth, the biological networks naturally reside in a hyperbolic space where the circle circumference and disk area are the exponential functions of the radius. To test these claims and assess potential benefits of the applications grounded in the above hypothesis, we have developed a mathematical framework and an accompanying computational procedure for matrix factorization and implied biological relationship inference in hyperbolic space. Not only does our study demonstrate a significant increase in the accuracy of hyperbolic embedding compared to Euclidean embedding, but it also shows that the latent dimension of an optimal hyperbolic embedding is by more than an order of magnitude smaller than the latent dimension of an optimal Euclidean embedding. We see this as additional evidence that hyperbolic geometry, rather than Euclidean, underlines the biological system.

Introduction

Although biological databases have seemingly high dimensionality, their intrinsic dimension is relatively low. For instance, many genes in a genomic database are co-expressed and a significant portion of variations in these databases can be explained by a few variables, including the cell type, the number of detected transcripts, or a gene program [1]. In a different example, the low intrinsic dimension of databases of adverse drug reactions is due to associations of side-effects to particular chemical substructures and their combinations. In other words, it is well established that drugs sharing chemical substructures are likely to give rise to the same side-effects [2,3]. Thus, representing these and other biomedical entities as points in a low dimensional space, in a way that preserves distances between those entities, is essential for many systems biology tasks, including classification, clustering, link prediction, and data visualization.

Matrix factorization is the method of choice for dimensionality reduction and relationship inference [4-21]. It is the key technique used to predict drug-disease associations, drug-target interactions, drug-side effect associations, and disease-associated genes, to name a few. Given a matrix $R_{m\times n}$ of relationships between the elements from two biological domains A and B, the goal is to find the best approximation $R \approx UV^T$ as the product of two lower dimensional matrices $U_{m\times d}$ and $V_{n\times d}$. The rows of U and V can be seen as points in $\mathbb{R}^d$ and are called the *feature* or *latent* representations of elements from A and B, respectively, while the entries of UV^T can be thought of as the predicted association scores between those elements. In a different interpretation, if we view the elements of A and B as the network nodes and elements of R as the network edges, then the rows of U and V represent the solutions of the heterogeneous network embedding problem, namely, the problem of embedding a simple two-domain network.

The research on matrix factorization and network embedding has historically focused on computational and statistical aspects of the particular problem of interest, taking the Euclidean geometry of the latent space for granted. However, studies have recently emerged suggesting that the native space of complex networks is not the Euclidean space $\mathbb{R}^d$, but the hyperbolic space of negative Gaussian curvature [22-28]. A compelling argument in support of this hypothesis is that complex networks resemble hierarchical, tree-like organization [29] and that discrete hyperbolic geometry is the

geometry of trees. Several filed specific applications have also been developed, demonstrating the benefits of data representation in a negatively curved space [1,30-37].

Just like the Internet and social networks [22,38], biological networks are organized into many small topological modules, grouped together in a hierarchical manner to form a hierarchical network [25,39]. Embedding such a network in a flat Euclidean space leads to distortion in spatial arrangement of the feature vectors representing individual biological entities. In contrast, the hyperbolic space can accommodate exponential growth of the number of relevant network features since the circumference of a hyperbolic circle and the disk area are exponential functions of the radius. Nevertheless, in spite of this simple, convincing argument, the above assumption on the nature of the biological space has never been tested in practice nor it has been built upon to create practical and useful biological applications.

In this study we lay out the theoretical foundation for biological native space representation and develop the corresponding computational procedure for hyperbolic matrix factorization. More specifically, we take advantage of recent advances in the development of probabilistic models and numerical optimization, including the work on pseudo-hyperbolic Gaussian distribution [40] and a new technique for gradient descent in hyperbolic space [41,42] to learn the latent node representation and, in turn, compute probabilities of associations between biological entities. Our benchmarking data clearly and consistently demonstrates a significant advantage in accuracy and a drastic reduction in latent space dimensionality of hyperbolic matrix factorization.

Results

We carried out a head-to-head comparison of the Euclidean and hyperbolic matrix factorization using the probabilistic logistic matrix factorization framework. Originally proposed for item recommendations in the media industry [43,44], logistic matrix factorization has become the state-of-the-art method for biological relationship inference [45-53]. Given an incomplete binary matrix $R = (r_{i,j})$ of observed relationships between elements of two domains A and B, the probability of $a^i \in A$ and $b^j \in B$ being related $(r_{i,j} = 1)$ is modeled as

$$p\left(r_{i,j} = 1 \middle| u^i, v^j\right) = exp\left(\langle u^i, v^j \rangle + \theta\right) / \left(1 + exp\left(\langle u^i, v^j \rangle + \theta\right)\right) \quad (1)$$

where $u^i \in \mathbb{R}^d$ and $v^j \in \mathbb{R}^d$ are the latent vector representations of a^i and b^j, respectively, $\langle \cdot, \cdot \rangle$ represents the Euclidean inner product, and θ is a trainable parameter [44]. The latent vectors are learned through the maximum a posteriori estimation using the alternating gradient descent technique.

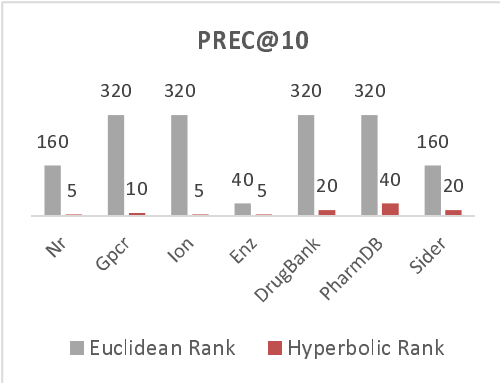
We developed a hyperbolic variant of the above method by considering u^i and v^j as the points in d-dimensional hyperbolic space $\mathbb{H}^d$ and by replacing the Euclidean inner product $\langle u^i, v^j \rangle$ in (1) by the Lorentzian product $\langle u^i, v^j \rangle_{\mathcal{L}}$ (details in Methods section).

Table 1. Comparison of "bare-bones" (no side information) hyperbolic and Euclidean matrix factorization using 10 rounds of 5-fold cross validation.

	AUPR		***AUC***	
	Hyperbolic	***Euclidean***	***Hyperbolic***	***Euclidean***
Nr	**0.48(0.05)**	0.35(0.00)	**0.84(0.01)**	0.76(0.01)
Gpcr	**0.63(0.00)**	0.59(0.01)	**0.91(0.00)**	0.87(0.01)
Ion	**0.87(0.00)**	0.84(0.00)	**0.97(0.00)**	0.97(0.00)
Enzyme	**0.82(0.00)**	0.74(0.00)	**0.97(0.00)**	0.95(0.00)
DrugBank	**0.52(0.00)**	0.44(0.00)	**0.95(0.00)**	0.92(0.00)
PharmDB	**0.35(0.00)**	0.27(0.01)	**0.88(0.00)**	0.78(0.00)
Sider	**0.50(0.05)**	0.49(0.00)	**0.95(0.00)**	0.94(0.00)

We trained the parameters of each method (Euclidean and hyperbolic) separately on each benchmarking data set, using the gradient descent in hyperbolic space and an extensive grid search. Table 1 shows the comparison in accuracy of Euclidean and hyperbolic factorization, as measured by the AUC and the AUPR scores on seven biological databases: drug-target interaction databases Nr, Gpcr, Ion, Enz, [54], and DrugBank [55], drug-disease association database PharmacoDB [56], and the SIDER database of adverse drug reactions [57]. Fig. 1 shows the latent dimension of an optimal Euclidean factorization as well as the smallest hyperbolic dimension that yields better or equal AUC and AUPR scores.

As seen in table 1 and Fig. 1., the hyperbolic matrix factorization routinely outperforms the Euclidean factorization in every benchmark and across fundamentally different classification measures. Moreover, the hyperbolic matrix factorization achieves superior accuracy at latent dimensions that are by an order of magnitude smaller compared to dimensions needed for optimal Euclidean embedding. Specifically, optimal Euclidean factorization is most often achieved at ranks exceeding 300. In contrast, hyperbolic factorization needs only 5 or 10 latent features to achieve the same or better classification scores.



	AUPR		AUC	
	Hyperbolic	***Euclidean***	***Hyperbolic***	***Euclidean***
Nr	**5**	**160**	**5**	**320**
Gpcr	**10**	**320**	**5**	**320**
Ion	**10**	**320**	**20**	**320**
Enzyme	**5**	**320**	**20**	**320**
DrugBank	**10**	**320**	**10**	**320**
PharmDB	**10**	**320**	**20**	**320**
Sider	**40**	**320**	**10**	**320**

Fig. 1. Optimal Euclidean rank (latent space dimension) and the hyperbolic rank yielding the same or better classification scores with respect to PREC@10 metric (left) and AUC and AUPR (right).

We obtained similar results on MovieLens [58] and Epinions [59] databases (data not shown). While our focus is on biomedical applications, the benchmarking data on movie rating and social networking datasets further confirms the hyperbolic nature of the geometry of complex networks and superiority of hyperbolic factorization.

Methods

Hyperbolic geometry can be modeled on the n-dimensional hyperboloid in the Lorentzian space $\mathbb{R}^{n,1}$, where $\mathbb{R}^{n,1}$ is a copy of $\mathbb{R}^{n+1}$ equipped with a bilinear form $\langle\cdot,\cdot\rangle_{\mathcal{L}}$ defined as

$$\langle x, y\rangle_{\mathcal{L}} = x_1 y_1 + \cdots + x_n y_n - x_{n+1} y_{n+1} \tag{2}$$

Hyperbolic space is represented by one sheet of the two-sheeted hyperboloid

$$\{x \in \mathbb{R}^{n,1} | \langle x, x\rangle_{\mathcal{L}} = -1\} \tag{3}$$

(which can be thought of as a sphere of radius $i = \sqrt{-1}$) namely,

$$\mathbb{H}^n = \{x \in \mathbb{R}^{n,1} | \langle x, x\rangle_{\mathcal{L}} = -1, x_{n+1} > 0\} \tag{4}$$

It can be shown that the bilinear form $\langle\cdot,\cdot\rangle_{\mathcal{L}}$ restricted on the tangent space at a point $p \in \mathbb{H}^n$, defined by

$$T_p\mathbb{H}^n = \{x \in \mathbb{R}^{n,1} | \langle p, x\rangle_{\mathcal{L}} = 0\} \tag{5}$$

is positive definite, thereby representing a genuine Riemannian metric on $\mathbb{H}^n$. The distance between two points $x, y \in \mathbb{H}^n$ is given by

$$d_{\mathbb{H}^n}(x, y) = \operatorname{arccosh}(-\langle x, y\rangle_{\mathcal{L}}) \tag{6}$$

An interesting and in the biological context insightful property of the hyperbolic space is that the shortest path between two random points in $\mathbb{H}^{\boldsymbol{n}}$ that are far away from the vertex $\boldsymbol{\mu_0}$ as almost the same length as the path through the vertex (Fig. 2). This resembles the property of the distance function on trees, where the shortest path between two randomly selected nodes deep in the tree is almost of the same length as the path through the root. A good biological analogue is a phylogenetic tree, where the

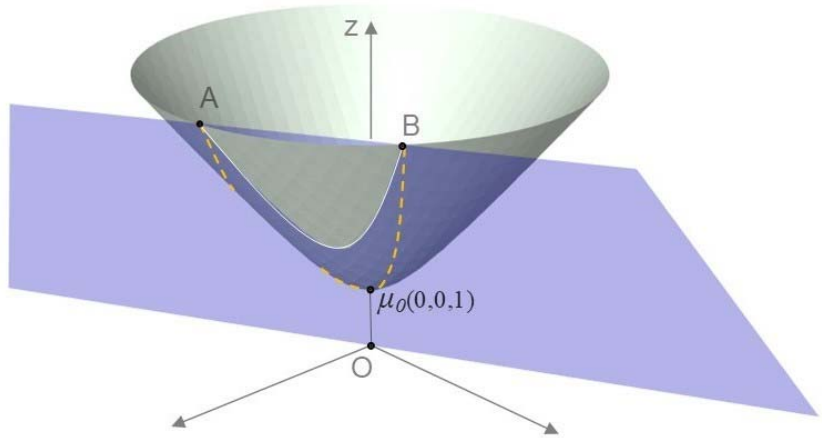


Fig. 2. Hyperboloid model of $\mathbb{H}^2$. The shortest path between points A and B is the line of intersection of the hyperboloid with the plane (blue) through A, B and O. As A and B are moving away from the vertex μ_0, the length of this geodesic line (white) is almost the same as the length of the path through μ_0 (orange).

shortest path between two arbitrary selected organisms is almost the path through their most ancient ancestor (root).

The theoretical foundation

While the general theory, presented below, can be tailored to different matrix factorization techniques (or, at least to those in which the objective function is a function of the hyperbolic inner product), we illustrate it in the framework of logistic matrix factorization. Our choice is due to the explicit statistical formulation of logistic matrix factorization, simplicity of presentation, and its high accuracy in biological applications [47-53].

Let A and B be two biological domains consisting of entities (nodes) $\{a^i\}_{i=1}^m$ and $\{b^j\}_{j=1}^n$, respectively. Denote by $R = (r_{i,j})_{m\times n}$ the matrix of relationships (edges) between the elements of A and B. Specifically, $r_{i,j} = 1$ if a^i is related to b^j and $r_{i,j} = 0$ otherwise (no relationship or unknown). Let u^i, $v^j \in \mathbb{H}^d$ be the latent vector representations of a^i and b^j, respectively, where $d \ll \max(m,n)$. Denote by $e_{i,j}$ the event that a^i and b^j are related. Following the Euclidean case, we model the probability of $e_{i,j}$ as the logistic function in the Lorentz space $\mathbb{R}^{d,1}$

$$p_{i,j} = p(r_{i,j} = 1|u^i, v^j) = \exp\left(-d_{\mathcal{L}}^2(u^i, v^j)\right) / \left(1 + \exp\left(-d_{\mathcal{L}}^2(u^i, v^j)\right)\right) \tag{7}$$

where $d_{\mathcal{L}}^2(x,y)$ denotes the squared *Lorentzian distance* [60] between the points $x, y \in \mathbb{H}^d$, namely

$$d_{\mathcal{L}}^2(x,y) = \|x-y\|_{\mathcal{L}}^2 = \langle x-y, x-y\rangle_{\mathcal{L}} = -2 - 2\langle x, y\rangle_{\mathcal{L}} \tag{8}$$

Denote by $W = (w_{i,j})_{m\times n}$ our confidence in the entries $r_{i,j}$ of the interaction matrix R. In many practical applications, $w_{i,j} = 1$ if $r_{i,j} = 0$ and $w_{i,j} = c$ if $r_{i,j} = 1$, where $c > 1$ is a constant. In general, the idea is to assign higher weights to trustworthy pairs i.e., the pairs of biological concepts (a^i, b^j) for which we have higher confidence of association. Given the weights $w_{i,j}$, the likelihood of $r_{i,j}$ given u^i and v^j is

$$p(r_{i,j}|u^i, v^j) = p_{i,j}^{w_{i,j} r_{i,j}} (1 - p_{i,j})^{w_{i,j}(1-r_{i,j})} \tag{9}$$

Assuming the independence of events $e_{i,j}$, it follows that

$$p(R|U,V) = \prod_{i,j} \left(p_{i,j}^{r_{i,j}} (1-p_{i,j})^{1-r_{i,j}}\right)^{w_{i,j}} \tag{10}$$

where U and V represent the matrices of latent preferences of elements from A and B, respectively. In other words, the ith row of U is the vector u^i and i^{th} row of V is v^i.

Computing the prior distribution

Our goal is to derive the probability $p(U,V|R)$ from (10) through the Bayesian inference, which requires placing a prior on latent vectors (rows of U and V).

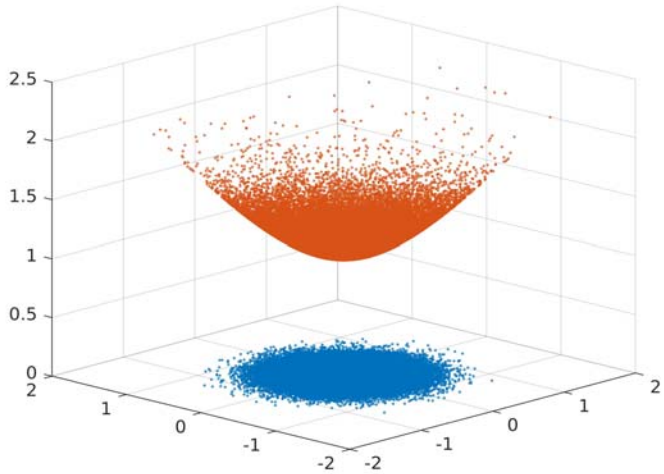


Fig. 3. 100,000 samples from $\mathcal{N}(0,\Sigma)$ (blue) and the corresponding samples from $\mathcal{G}(\mu_0,\Sigma)$ (red), $\Sigma = 0.1 \cdot I_{2\times 2}$.

Borrowing from the recent work on wrapped normal distribution in hyperbolic space [40], we define the prior distributions

$$p(U|\sigma_U^2) = \prod_{i=1}^{m} \mathcal{G}(u_i|\mu_0, \sigma_U^2 I),$$
$$p(V|\sigma_V^2) = \prod_{j=1}^{n} \mathcal{G}(v_j|\mu_0, \sigma_V^2 I) \qquad (11)$$

where $\mathcal{G}(\mu,\Sigma)$ is the pseudo-hyperbolic Gaussian distribution and $\mu_0 = (0,\dots,0,1)$ is the vertex of the hyperboloid i.e. the origin of the hyperbolic space.

The pseudo-hyperbolic Gaussian distribution extends the notion of Gaussian distribution on the hyperbolic space (Fig. 3). In short, for $\mu \in \mathbb{H}^d$ and positive definite Σ, sampling from $\mathcal{G}(\mu,\Sigma)$ can be thought of as a three step process: (a) sample a vector $x \in T_{\mu_0}\mathbb{H}^d$ from $\mathcal{N}(0,\Sigma)$, (b) transport x along the geodesic joining the points $\mu_0 \in \mathbb{H}^d$ and $\mu \in \mathbb{H}^d$ to $y \in T_\mu \mathbb{H}^d$, (c) project y to $z \in \mathbb{H}^d$.

The step (b) is carried out using the parallel transport $g_{\mu_0\to\mu}: T_{\mu_0}\mathbb{H}^d \to T_\mu \mathbb{H}^d$ (Fig. 4):

$$g_{\mu_0\to\mu}(x) = x + \frac{\langle \mu + \langle \mu_0,\mu\rangle_{\mathcal{L}} \cdot \mu_0, x\rangle_{\mathcal{L}}}{1-\langle \mu_0,\mu\rangle_{\mathcal{L}}}(\mu_0+\mu) \qquad (12)$$

while the step (c) uses the exponential map $Exp_\mu: T_\mu\mathbb{H}^d \to \mathbb{H}^d$ (Fig. 4), defined by

$$Exp_\mu(y) = \cosh(\|y\|_{\mathcal{L}})\,\mu + sinh\,(\|y\|_{\mathcal{L}})\frac{y}{\|y\|_{\mathcal{L}}} \qquad (13)$$

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

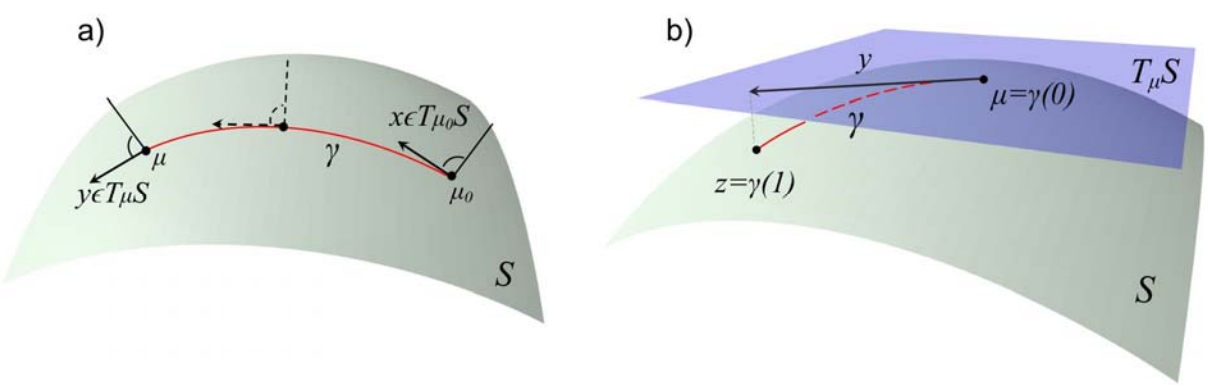


Fig. 4. (a) Parallel transport of $x \in T_{\mu_0}S$ to $y \in T_\mu S$ along the geodesic γ, where $\mu_0 = \gamma(0)$ and $\mu = \gamma(1)$. (b) The exponential map.

It is not difficult to show that the length of the geodesic joining μ to $Exp_\mu(y)$ on $\mathbb{H}^d$ is equal to $\|y\|_{\mathcal{L}}$, i.e., $d_{\mathbb{H}^d}\left(\mu, Exp_\mu(y)\right) = \|y\|_{\mathcal{L}}$. The relationship between the probability densities $X \sim \mathcal{N}(0,\Sigma)$ and $Z \sim \mathcal{G}(\mu,\Sigma)$ is

$$p(x) = p(z)\det(J_f) \qquad (14)$$

where $f = Exp_\mu \circ g_{\mu_0\to\mu}$ and $\det(J_f)$ denotes the determinant of the Jacobian $J_f = \left|\frac{\partial f}{\partial x}\right|$ [40]. It can also be shown that

$$\ln p(z) = \ln p(x) - (d-1)\ln\frac{\sinh(r)}{r} \tag{15}$$

where $r = \operatorname{arccosh}(-\langle \mu, z\rangle_{\mathcal{L}})$ [40].

The loss function

With the prior placed on U and V, we return to calculating the posterior probability $p(U,V|R)$ through the Bayesian inference:

$$p(U,V|R) \propto p(R|U,V)p(U|\sigma^2)p(V|\sigma^2) \tag{16}$$

Taking the negative logarithm of the posterior distribution (16), we arrive at the closed form expression for the loss function

$$\begin{aligned} L = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}\left[\ln\left(1+e^{-d_{\mathcal{L}}^2(u^i,v^j)}\right) + r_{i,j}d_{\mathcal{L}}^2(u^i,v^j)\right] \\ -\sum_{i=1}^{m}\left[\ln p\left(\overline{u^i}\right) - (d-1)\ln\frac{\sinh\left(\left\|\overline{u^i}\right\|_{\mathcal{L}}\right)}{\left\|\overline{u^i}\right\|_{\mathcal{L}}}\right] - \sum_{j=1}^{n}\left[\ln p\left(\overline{v^j}\right) - (d-1)\ln\frac{\sinh\left(\left\|\overline{v^j}\right\|_{\mathcal{L}}\right)}{\left\|\overline{v^j}\right\|_{\mathcal{L}}}\right] \end{aligned} \tag{17}$$

In the expression above, p is the probability density function of the normal distribution $\mathcal{N}(0,\sigma^2 I)$ in the tangent space $T_{\mu_0}\mathbb{H}^d$ at the vertex $\mu_0 = (0,\dots,0,1)$. For $x = (x_1,\dots,x_d,x_{d+1}) \in \mathbb{H}^d$

$$\overline{x} = Exp_{\mu_0}^{-1}x = \frac{\operatorname{arccosh}(-\langle\mu_0,x\rangle_{\mathcal{L}})}{\sqrt{\langle\mu_0,x\rangle_{\mathcal{L}}^2-1}}(x+\langle\mu_0,x\rangle_{\mathcal{L}}\cdot\mu_0) = \frac{\operatorname{arccosh}(x_{d+1})}{\sqrt{x_{d+1}^2-1}}(x_1,\dots,x_d,0) \tag{18}$$

It follows that

$$\ln p(\overline{x}) = -\frac{1}{2\sigma^2}\operatorname{arccosh}^2(x_{d+1}) + C_1 \tag{19}$$

where C_1 is a constant. Since

$$\|\overline{x}\|_{\mathcal{L}} = \operatorname{arccosh}(-\langle\mu_0,x\rangle_{\mathcal{L}}) = \operatorname{arccosh}(x_{d+1}) \tag{20}$$

we have

$$\frac{\sinh(\|\overline{x}\|_{\mathcal{L}})}{\|\overline{x}\|_{\mathcal{L}}} = \frac{\sqrt{x_{d+1}^2-1}}{\operatorname{arccosh}(x_{d+1})} \tag{21}$$

Hence, our loss function will have the form:

$$\begin{aligned} L = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}\left[\ln\left(1+e^{-d_{\mathcal{L}}^2(u^i,v^j)}\right) + r_{i,j}d_{\mathcal{L}}^2(u^i,v^j)\right] \\ +\sum_{i=1}^{m}\left[\alpha_U\operatorname{arccosh}^2\left(u_{d+1}^i\right) + (d-1)\ln\frac{\sqrt{\left(u_{d+1}^i\right)^2-1}}{\operatorname{arccosh}\left(u_{d+1}^i\right)}\right] \\ +\sum_{j=1}^{n}\left[\alpha_V\operatorname{arccosh}^2\left(v_{d+1}^j\right) + (d-1)\ln\frac{\sqrt{\left(v_{d+1}^j\right)^2-1}}{\operatorname{arccosh}\left(v_{d+1}^j\right)}\right] + C \end{aligned} \tag{22}$$

where $\alpha_U = \frac{1}{2\sigma_U^2}$, $\alpha_V = \frac{1}{2\sigma_V^2}$ are trainable parameters and C is a constant.

Alternating gradient descent in hyperbolic space

Finding a point of (local) minimum of a real function defined in a d-dimensional Euclidean space $\mathbb{R}^d$ is routinely accomplished using the gradient descent. We adopt a similar technique for finding the point $u \in \mathbb{H}^d$ of a local minimum of any real valued function $f: \mathbb{H}^d \to \mathbb{R}$ [41]. For this strategy to work, the function f has to be defined is in the ambient space $\mathbb{R}^{d,1}$ of $\mathbb{H}^d$, as well as on $\mathbb{H}^d$. Specifically, given the

initial value $u = u^{(0)}$ and a step size η, the gradient descent in hyperbolic space can be carried out by repeating the following steps:

1. Compute the gradient $\nabla_u^{\mathbb{R}^{d,1}} f$
2. Project $\nabla_u^{\mathbb{R}^{d,1}} f$ orthogonally to vector $\nabla_u^{\mathbb{H}^d} f \in T_u \mathbb{H}^d$
3. Set $u^{new} = Exp_u(-\eta \nabla_u^{\mathbb{H}^d} f)$

The gradient $\nabla_u^{\mathbb{R}^{d,1}} f$ of f in the ambient space $\mathbb{R}^{d,1}$ is a vector of partial derivatives

$$\nabla_u^{\mathbb{R}^{d,1}} f = \left(\frac{\partial L}{\partial x_1}|_u, \dots, \frac{\partial L}{\partial x_n}|_u, -\frac{\partial L}{\partial x_{d+1}}|_u \right) \tag{23}$$

The representation above (note the negative sign of the last vector's component) follows directly from the definition of the gradient:

$$\forall v \in \mathbb{R}^{d,1},\ \langle \nabla_u^{\mathbb{R}^{d,1}} f, v \rangle_{\mathcal{L}} = D_v f(u) \tag{24}$$

The orthogonal projection from the ambient onto the tangent space in (step 2 above) is given by

$$\nabla_u^{\mathbb{H}^d} f = \nabla_u^{\mathbb{R}^{d,1}} f + \langle u, \nabla_u^{\mathbb{R}^{d,1}} f \rangle_{\mathcal{L}} u \tag{25}$$

We will use the "alternating gradient descent" method to minimize the error function $L_{A,B}$ given in (22). The partial derivatives of $L_{A,B}$ are relatively easy to compute:

$$\frac{\partial L}{\partial u_k^i} = 2 \sum_{j=1}^n w_{i,j} \left(p_{i,j} - r_{i,j} \right) v_k^j \tag{26}$$

$$\frac{\partial L}{\partial u_{d+1}^i} = 2 \sum_{j=1}^n w_{i,j} \left(r_{i,j} - p_{i,j} \right) v_{d+1}^j + 2\alpha_U \frac{\operatorname{arccosh}\left(u_{d+1}^i\right)}{\sqrt{\left(u_{d+1}^i\right)^2 - 1}} + (d-1) \frac{u_{d+1}^i \operatorname{arccosh}\left(u_{d+1}^i\right) - \sqrt{\left(u_{d+1}^i\right)^2 - 1}}{\left[\left(u_{d+1}^i\right)^2 - 1\right] \operatorname{arccosh}(u_{d+1}^i)} \tag{27}$$

ALGORITHM *HyperbolicAlternatingGradientDescent*
***//* Input:** Relationship matrix R, weight matrix W, number of epochs E,
// learning rate η, parameters σ_U, σ_V, latent space dimension d
***//* Output:** The matrices of latent preferences $U = (u_k^i)$ and $V = (v_k^j)$

```
for i = 1 to m
        sample u^i from G(μ_0, σ_U^2 I)
end for
for j = 1 to n
        sample v^j from G(μ_0, σ_V^2 I)
end for
for e = 1 to E
        for i = 1 to m
                compute ∇_{u^i}^{R^{d,1}} L
                project ∇_{u^i}^{R^{d,1}} L orthogonally to vector ∇_{u^i}^{H^d} L ∈ T_{u^i} H^d
                u^i = Exp_{u^i}(−η ∇_{u^i}^{H^d} L)
        end for
        for i = j to n
                compute ∇_{v^j}^{R^{d,1}} L
                project ∇_{v^j}^{R^{d,1}} L orthogonally to vector ∇_{v^j}^{H^d} L ∈ T_{v^j} H^d
                v^j = Exp_{v^j}(−η ∇_{v^j}^{H^d} L)
        end for
end for
```

Generalization to spaces of arbitrary curvature and field specific applications

There are several directions for future research that can enhance the utility of hyperbolic matrix factorization. One approach is to generalize the theory to spaces of arbitrary negative curvature. The hyperbolic space of curvature $-1/\beta$, where $\beta > 0$, is defined as

$$\mathbb{H}^{d,\beta} = \{x \in \mathbb{R}^{d,1} | \langle x, x \rangle_{\mathcal{L}} = -\beta, x_{d+1} > 0\} \tag{28}$$

Curvature specific matrix factorization can be carried out by generalizing (7) to

$$p_{i,j} = p\left(r_{i,j} = 1 | u^i, v^j\right) = \exp\left(-d^2_{\mathcal{L},\beta}(u^i, v^j)\right) / \left(1 + \exp\left(-d^2_{\mathcal{L},\beta}(u^i, v^j)\right)\right) \tag{29}$$

where $d^2_{\mathcal{L},\beta}(x, y)$ denotes the generalized Lorentzian distance [60] between the points $x, y \in \mathbb{H}^{d,\beta}$:

$$d^2_{\mathcal{L},\beta}(x, y) = \langle x - y, x - y \rangle_{\mathcal{L}} = -2\beta - 2\langle x, y \rangle_{\mathcal{L}} \tag{30}$$

However, extending the factorization procedure to $\mathbb{H}^{d,\beta}$ requires the notion of pseudo-hyperbolic Gaussian distribution from $\mathbb{H}^{d,1}$ to $\mathbb{H}^{d,\beta}$.

In certain applications, such as recommender systems, the matrix factorization is supplemented with other techniques to improve the accuracy of relationship inference. In one approach, the *neighborhood regularization* is used to ensure that similar entities from the domain A are in relationship with similar entities from the domain B. The existing techniques for incorporating side information [45,46,47,61] can be carried out in $\mathbb{H}^{d,\beta}$ by adding

$$\beta_U \sum_{i=1}^{m} \sum_{j=1}^{m} s_{i,j}\, d^2_{\mathcal{L},\beta}(u^i, u^j) + \beta_V \sum_{i=1}^{n} \sum_{j=1}^{n} t_{i,j}\, d^2_{\mathcal{L},\beta}(v^i, v^j) \tag{31}$$

to the loss function L, where β_U and β_V are trainable parameters.

Similarly, the arrival of a new entity in one of the domains, coined as the *cold-start* problem, can be addressed with the hyperbolic variant of the weighted-profile method [45,46,62]. In Euclidean case, the latent representation $u^i \in \mathbb{R}^d$ of the element $a^i \in A$ that is not in an observed relationship with any $b^j \in B$ is often computed as the weighted combination of the rows $u^j \in U$ most similar to u^i:

$$u^i = \frac{1}{SM} \sum_{j=1}^{J} s_{i,j}\, u^j \tag{32}$$

where $SM = \sum_{j=1}^{J} s_{i,j}$. We note that the point u^i in (28) represents the center of mass of the points $\{u^j\}_{j=1}^{J}$ in $\mathbb{R}^d$. Specifically, u^i is the solution to the minimization problem

$$\min_u \sum_{j=1}^{J} s_{i,j} d^2_{\mathbb{R}^d}(u, u^j)$$

(33)

where $d^2_{\mathbb{R}^d}(x, y)$ is the squared Euclidean distance between the points x and y.

The corresponding hyperbolic weighted-profile method can utilize a closed form solution for the center of mass in the hyperbolic space, i.e. the solution of the minimization problem

$$\min_u \sum_{j=1}^{J} s_{i,j} d^2_{\mathcal{L},\beta}(u, u^j) \tag{34}$$

given recently by Law *et al*. [63], namely

$$u^i = \sqrt{\beta} \frac{\sum_{j=1}^{J} s_{i,j} u^j}{\left| \left\| \sum_{j=1}^{J} s_{i,j} u^j \right\|_{\mathcal{L},\beta} \right|} \tag{35}$$

where $\left| \|x\|_{\mathcal{L},\beta} \right| = \sqrt{\left| \|x\|^2_{\mathcal{L},\beta} \right|}$.

Conclusion

Matrix factorization is one of the main techniques used in systems biology for predicting the relationships between the elements from a pair of biological domains. The key idea is to complete an input matrix of observed relationships using its close low-rank approximation i.e., the product of two lower dimensional matrices. The technique results in the representation of biological objects as points in

a low dimensional Euclidean space. This latent space representation is essential in nearly all downstream machine learning tasks, such as classification, clustering, visualization, and the overall understanding of the structure and dynamics of a biological system.

Past research in systems biology has taken the Euclidean geometry of the biological space for granted. This has been convenient due to the availability of advanced analytic, numerical, statistical and machine learning procedures in the Euclidean space. However, recent theoretical studies suggest that the hyperbolic geometry, rather than Euclidean, underpins all complex networks in general and biological networks in particular. Therefore, a radical shift in data representation is necessary to obtain an undistorted view of the biological space and, in turn, ensure further progress in systems biology and related fields.

We have developed and benchmarked a technique for a probabilistic hyperbolic matrix factorization and resulting hyperbolic space embedding. We demonstrate that the Lorentzian model of hyperbolic space allows for a closed form expression of the key transformations and techniques required for latent space dimensionality reduction. Our method builds upon some recent advances in the development of probabilistic models and numerical optimization in hyperbolic space to learn an optimal hyperbolic embedding and to compute the probabilities of relationships between biological entities. The benchmarking tests performed on different types of biomedical data clearly demonstrate a significant increase in accuracy and a drastic reduction in latent space dimensionality of hyperbolic embedding compared to Euclidean embedding. These findings confirm the negative curvature of the native biological space.

The techniques presented in this study can be generalized to hyperbolic spaces of any given curvature. Finding the proper curvature of the space that underlines a particular biological database will result in a more accurate latent representation and, in turn, more accurate biological relationship prediction.

References

[1] Ding, J., & Regev, A. (2019). Deep generative model embedding of single-cellRNA-Seq profiles on hyperspheres and hyperbolic spaces. *BioRxiv*, 853457.

[2] Scheiber, J., Jenkins, J. L., Sukuru, S. C. K., Bender, A., Mikhailov, D., Milik, M., Azzaoui, K., Whitebread, S., Hamon, J., Urban, L., & Glick, M. (2009). Mapping adverse drug reactions in chemical space. *Journal of medicinal chemistry*, *52*(9), 3103-3107.

[3] Mizutani, S., Pauwels, E., Stoven, V., Goto, S., & Yamanishi, Y. (2012). Relating drug–protein interaction network with drug side effects. Bioinformatics, 28(18), i522-i528.

[4] Brunet, J. P., Tamayo, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101(12), 4164-4169.

[5] Devarajan, K. (2008). Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. PLoS Comput Biol, 4(7), e1000029.

[6] Tamayo, P., Scanfeld, D., Ebert, B. L., Gillette, M. A., Roberts, C. W., & Mesirov, J. P. (2007). Metagene projection for cross-platform, cross-species characterization of global transcriptional states. Proceedings of the National Academy of Sciences, 104(14), 5959-5964.

[7] Gao, Y., & Church, G. (2005). Improving molecular cancer class discovery through sparse non-negative matrix factorization. Bioinformatics, 21(21), 3970-3975.

[8] Kim, P. M., & Tidor, B. (2003). Subsystem identification through dimensionality reduction of large-scale gene expression data. Genome research, 13(7), 1706-1718.

[9] Chagoyen, M., Carmona-Saez, P., Shatkay, H., Carazo, J. M., & Pascual-Montano, A. (2006). Discovering semantic features in the literature: a foundation for building functional associations. BMC bioinformatics, 7(1), 41.

[10] Hiisila, H., & Bingham, E. (2004, November). Dependencies between transcription factor binding sites: Comparison between ICA, NMF, PLSA and frequent sets. In Fourth IEEE International Conference on Data Mining (ICDM'04) (pp. 114-121). IEEE.

[11] Jung, I., Lee, J., Lee, S. Y., & Kim, D. (2008). Application of nonnegative matrix factorization to improve profile-profile alignment features for fold recognition and remote homolog detection. Bmc Bioinformatics, 9(1), 298.

[12] Okun, O., & Priisalu, H. (2006). Fast nonnegative matrix factorization and its application for protein fold recognition. EURASIP Journal on Advances in Signal Processing, 2006(1), 071817.

[13] Sharan, R., Ulitsky, I., & Shamir, R. (2007). Network-based prediction of protein function. Molecular systems biology, 3(1), 88.

[14] Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10), P10008.

[15] Singh, R., Xu, J., & Berger, B. (2008). Global alignment of multiple protein interaction networks with application to functional orthology detection. Proceedings of the National Academy of Sciences, 105(35), 12763-12768.

[16] Zeng, Z., Vo, A. H., Mao, C., Clare, S. E., Khan, S. A., & Luo, Y. (2019). Cancer classification and pathway discovery using non-negative matrix factorization. *Journal of biomedical informatics*, *96*, 103247.

[17] Zhang, P., Zheng, C. H., Li, B., & Wen, C. G. (2008, September). Tumor classification using non-negative matrix factorization. In *International Conference on Intelligent Computing* (pp. 236-243). Springer, Berlin, Heidelberg.

[18] Stein-O'Brien, G. L., Arora, R., Culhane, A. C., Favorov, A. V., Garmire, L. X., Greene, C. S., Goff, L.A., Li, Y., Ngom, A., Ochs, M.F., & Xu, Y. (2018). Enter the matrix: factorization uncovers knowledge from omics. Trends in Genetics, 34(10), 790-805.

[19] Zheng, C. H., Huang, D. S., Zhang, L., & Kong, X. Z. (2009). Tumor clustering using nonnegative matrix factorization with gene selection. *IEEE Transactions on Information Technology in Biomedicine*, *13*(4), 599-607.

[20] Zhang, W., Yue, X., Lin, W., Wu, W., Liu, R., Huang, F., & Liu, F. (2018). Predicting drug-disease associations by using similarity constrained matrix factorization. *BMC bioinformatics*, *19*(1), 1-12.

[21] Yang, J., Li, Z., Fan, X., & Cheng, Y. (2014). Drug–disease association and drug-repositioning predictions in complex diseases using causal inference–probabilistic matrix factorization. *Journal of chemical information and modeling*, *54*(9), 2562-2569.

[22] Krioukov, D., Papadopoulos, F., Vahdat, A., & Boguná, M. (2009). Curvature and temperature of complex networks. *Physical Review E*, *80*(3), 035101.

[23] Papadopoulos, F., Psomas, C., & Krioukov, D. (2014). Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking*, *23*(1), 198-211.

[24] Nickel, M., & Kiela, D. (2017). Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems* (pp. 6338-6347).

[25] Albert, R., DasGupta, B., & Mobasheri, N. (2014). Topological implications of negative curvature for biological and social networks. *Physical Review E*, *89*(3), 032811.

[26] Alanis-Lobato, G., Mier, P., & Andrade-Navarro, M. A. (2016). Efficient embedding of complex networks to hyperbolic space via their Laplacian. *Scientific reports*, *6*, 30108.

[27] De Sa, C., Gu, A., Ré, C., & Sala, F. (2018). Representation tradeoffs for hyperbolic

embeddings. *Proceedings of machine learning research*, *80*, 4460.

[28] Bose, A. J., Smofsky, A., Liao, R., Panangaden, P., & Hamilton, W. L. (2020). Latent Variable Modelling with Hyperbolic Normalizing Flows. *arXiv preprint arXiv:2002.06336*.

[29] Gromov, M. (2007). Metric structures for Riemannian and non-Riemannian spaces. Springer Science & Business Media.

[30] Dhingra, B., Shallue, C. J., Norouzi, M., Dai, A. M., & Dahl, G. E. (2018). Embedding text in hyperbolic spaces. *arXiv preprint arXiv:1806.04313*.

[31] Chamberlain, B. P., Clough, J., & Deisenroth, M. P. (2017). Neural embeddings of graphs in hyperbolic space. *arXiv preprint arXiv:1705.10359*.

[32] Leimeister, M., & Wilson, B. J. (2018). Skip-gram word embeddings in hyperbolic space. *arXiv preprint arXiv:1809.01498*.

[33] Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G., & Cannistraci, C. V. (2017). Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nature communications*, *8*(1), 1-19.

[34] Monath, N., Zaheer, M., Silva, D., McCallum, A., & Ahmed, A. (2019, July). Gradient-based hierarchical clustering using continuous representations of trees in hyperbolic space. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 714-722).

[35] Tran, L. V., Tay, Y., Zhang, S., Cong, G., & Li, X. (2020, January). HyperML: A Boosting Metric Learning Approach in Hyperbolic Space for Recommender Systems. In *WSDM* (pp. 609-617).

[36] Schmeier, T., Chisari, J., Garrett, S., & Vintch, B. (2019, September). Music recommendations in hyperbolic space: an application of empirical bayes and hierarchical poincaré embeddings. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 437-441).

[37] Yu, K., Visweswaran, S., & Batmanghelich, K. (2020). Semi-Supervised Hierarchical Drug Embedding in Hyperbolic Space. *arXiv preprint arXiv:2006.00986*.

[38] Verbeek, K., & Suri, S. (2014, June). Metric embedding, hyperbolic space, and social networks. In *Proceedings of the thirtieth annual symposium on Computational geometry* (pp. 501-510).

[39] Ravasz, E. (2009). Detecting hierarchical modularity in biological networks. In *Computational Systems Biology* (pp. 145-160). Humana Press.

[40] Nagano, Y., Yamaguchi, S., Fujita, Y., & Koyama, M. (2019). A wrapped normal distribution on hyperbolic space for gradient-based learning. *arXiv preprint arXiv:1902.02992*.

[41] Nickel, M., & Kiela, D. (2018). Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417*.

[42] Wilson, B., & Leimeister, M. (2018). Gradient descent in hyperbolic space. *arXiv preprint arXiv:1805.08207*.

[43] Steck, H. Training and testing of recommender systems on data missing not at random. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge discovery and data mining,* 713–722 (2010).

[44] Johnson, C. C. Logistic matrix factorization for implicit feedback data. In *Advances in Neural Information Processing Systems* 27: Distributed Machine Learning and Matrix Computations Workshop 2014.

[45] Lim, H., Gray, P., Xie, L., & Poleksic, A. (2016). Improved genome-scale multi-target virtual screening via a novel collaborative filtering approach to cold-start problem. *Scientific reports*, *6*(1),

1-11.

[46] Poleksic, A., & Xie, L. (2018). Predicting serious rare adverse reactions of novel chemicals. *Bioinformatics*, *34*(16), 2835-2842.

[47] Liu, Y., Wu, M., Miao, C., Zhao, P. & Li, X. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction. PLoS Comput. Biol. 12, e1004760 (2016).

[48] Hao, M., Bryant, S. H., & Wang, Y. (2017). Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Scientific reports*, *7*(1), 1-11.

[49] Li, Y., Li, J., & Bian, N. (2019). DNILMF-LDA: prediction of lncRNA-disease associations by dual-network integrated logistic matrix factorization and Bayesian optimization. *Genes*, *10*(8), 608.

[50] Hao, M., Bryant, S. H., & Wang, Y. (2019). Open-source chemogenomic data-driven algorithms for predicting drug–target interactions. *Briefings in Bioinformatics*, *20*(4), 1465-1474.

[51] Zhao, Q., Zhang, Y., Hu, H., Ren, G., Zhang, W., & Liu, H. (2018). IRWNRLPI: integrating random walk and neighborhood regularized logistic matrix factorization for lncRNA-protein interaction prediction. *Frontiers in genetics*, *9*, 239.

[52] Ban, T., Ohue, M., & Akiyama, Y. (2019). NRLMFβ: Beta-distribution-rescored neighborhood regularized logistic matrix factorization for improving the performance of drug–target interaction prediction. *Biochemistry and biophysics reports*, *18*, 100615.

[53] Yan, C., Wang, J., Ni, P., Lan, W., Wu, F. X., & Pan, Y. (2017). DNRLMF-MDA: predicting microRNA-disease associations based on similarities of microRNAs and diseases. *IEEE/ACM transactions on computational biology and bioinformatics*, *16*(1), 233-243.

[54] Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug–target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics*, *24*(13), i232-i240.

[55] Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., Gautam, B., & Hassanali, M. (2008). DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, *36*(suppl_1), D901-D906.

[56] Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, *6*, e26726.

[57] Kuhn, M., Letunic, I., Jensen, L. J., & Bork, P. (2016). The SIDER database of drugs and side effects. *Nucleic acids research*, *44*(D1), D1075-D1079.

[58] Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, *5*(4), 1-19.

[59] Richardson, M., Agrawal, R., & Domingos, P. (2003, October). Trust management for the semantic web. In *International semantic Web conference* (pp. 351-368). Springer, Berlin, Heidelberg.

[60] Ratcliffe, J. G., Axler, S., & Ribet, K. A. (2006). *Foundations of hyperbolic manifolds* (Vol. 149). New York: Springer.

[61] Yao, Y., Tong, H., Yan, G., Xu, F., Zhang, X., Szymanski, B. K., & Lu, J. (2014, November). Dual-regularized one-class collaborative filtering. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (pp. 759-768).

[62] Poleksic A. Overcoming sparseness of biological networks to identify drug repositioning candidates. bioRxiv doi: 10.1101/2020.06.07.138966 (2020).

[63] Law, M., Liao, R., Snell, J., & Zemel, R. (2019, May). Lorentzian distance learning for hyperbolic representations. In International Conference on Machine Learning (pp. 3672-3681).