Mixed-Curvature Representation Learning for Biological Pathway Graphs

Daniel McNeela [1 2] **Frederic Sala*** [3] **Anthony Gitter*** [1 2 3]

Abstract

Models that embed graphs in non-Euclidean spaces have shown substantial benefits in a variety of contexts, but their application has not been studied extensively in the biological domain, particularly with respect to biological pathway graphs. Such graphs exhibit a variety of complex network structures, presenting challenges to existing embedding approaches. Learning high-quality embeddings for biological pathway graphs is important for researchers looking to understand the underpinnings of disease and train high-quality predictive models on these networks. In this work, we investigate the effects of embedding pathway graphs in non-Euclidean mixed-curvature spaces and compare against traditional Euclidean models. We then train a supervised model using the learned embeddings to predict missing protein-protein interactions in pathway graphs. We find large reductions in distortion and boosts in in-distribution edge prediction performance from using mixed-curvature embeddings and their corresponding graph neural network models. Furthermore, mixed-curvature modeling provides some utility for out-of-distribution edge prediction.

1. Introduction

Machine learning methods for embedding graphs enable learning on data ranging from social media networks, to proteins and molecules, to phylogenies and knowledge graphs. These embeddings then enable useful node classification and edge prediction models, which can perform tasks as diverse as predicting whether a molecule is active against a given drug target or whether two users are likely to share a preference for a given product.

*Equal contribution [1]Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA [2]Morgridge Institute for Research, Madison, WI, USA [3]Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI, USA. Correspondence to: Anthony Gitter <gitter@biostat.wisc.edu>.

The 2023 ICML Workshop on Computational Biology. Honolulu, Hawaii, USA, 2023.

While traditional graph learning methods employ Euclidean representations (Grover & Leskovec, 2016), it is known that for certain graphs Euclidean representations are unable to perfectly preserve graph structure, regardless of what algorithm is used (Bourgain, 1985). As a result, recent works have studied whether lower graph distortion can be achieved by embedding in non-Euclidean spaces, such as hyperbolic space (Sala et al., 2018). Generally speaking, lower distortion correlates with better downstream task performance.

The works of Gu et al. (2018) and Giovanni et al. (2022) have examined embeddings into mixed-curvature products of spaces and heterogeneous manifolds, respectively. While these methods have been evaluated on standard graph benchmarking datasets, their hypotheses have not been validated for specialized graphs found in biological pathways and networks, which may have properties and topologies that differ from general graphs in other domains. The product space approach is appealing because deciding a priori which space to embed into can be challenging.

Biological pathways are graphs that encode cellular processes. Typically, the nodes in such graphs are entities such as genes, proteins, or metabolites, and the edges designate relationships between them. For example, the presence of an edge connecting nodes A and B might indicate that the presence of protein A controls the transcription of gene B. Pathways are an important object of study in network biology as they can be used to infer subcellular relationships and understand the mechanisms underlying disease.

Embedding biological pathways is difficult because no canonical methods exist. Pathways exhibit a high degree of complexity. Some, due to a lack of study, are sparse while others exhibit high inter-connectivity. Their complex network structures suggest that non-Euclidean representations might provide significant benefits. However, no systematic study of embedding methods applied to biological pathway graphs has been undertaken. Only Euclidean embedding methods have been applied to pathway graphs (M A Basher & Hallam, 2020; Pershad et al., 2020), and because pathway graphs differ from standard graphs used to benchmark non-Euclidean embedding models, it is unknown to what extent these models would work for pathway graphs.

In this work, we study non-Euclidean embeddings of biological pathway graphs and their performance relative to stan-

dard Euclidean embeddings. We perform a large-scale test of a variety of embedding methods to pathway graphs taken from PathBank and embed into a number of different combinations of spaces. For each pathway graph, we determine a best embedding space, as measured by the lowest graph distortion, and learn the node embeddings for that graph. Although biological pathway databases are of high quality, they are incomplete and only capture a fraction of the knowledge about the relevant biological processes (Hanspers et al., 2020). Therefore, we investigate the downstream performance of our graph embeddings by predicting potentially missing pathway edges. We find that mixed-curvature representations outperform, as measured by distortion, Euclidean representations in all cases. Furthermore, the positive impact of improved embeddings extends to downstream tasks such as edge prediction of held-out edges, where we find improvements in area under the receiver operating characteristic curve (AUC) for tree-like graphs.

2. Background and Related Work

2.1. Non-Euclidean Embeddings and Machine Learning

Much of the research into non-Euclidean embeddings in machine learning originated in studies of graphs and networks, where they were originally used to embed concept ontologies. For example, Nickel & Kiela (2017) developed a method for embedding into the Poincaré model of hyperbolic space and used it to generate node embeddings for the WORDNET ontology. Sala et al. (2018) expanded on this work by determining the precision-dimensionality tradeoffs inherent in hyperbolic embeddings. Ganea et al. (2018) and Chami et al. (2019) then produced generalized Hyperbolic Neural Networks and Hyperbolic Graph Convolutional Neural Networks to perform prediction directly in hyperbolic space on data of various types. Finally, Gu et al. (2018) extended graph representation learning to a Cartesian product of hyperbolic, spherical, and Euclidean spaces, while Giovanni et al. (2022) further generalized this to a product of manifolds of heterogeneous curvature. Our approach follows that of Gu et al. (2018).

2.2. Pathway Graphs and Embeddings

Pathway graphs have been well-studied from a biological perspective, but embedding them to facilitate downstream prediction tasks is relatively new. For example, M A Basher & Hallam (2020) developed a method called pathway2vec, which combines five different Euclidean embedding methods to learn embeddings of biological pathways. Similarly, Pershad et al. (2020) used node2vec to generate embeddings of protein-protein interaction (PPI) networks and used the resulting embeddings as one component of a method to predict response to psychiatric drugs. Pathway embeddings are a crucial input to models that operate on pathway network structure to make predictions. Euclidean graph neural networks have been broadly applied to biological pathways to predict cancer subtypes (Lee et al., 2020; Hayakawa et al., 2022), synthetic lethality (Lai et al., 2021), PPIs (Pham & Dang, 2021), cancer survival (Liang et al., 2022), textual pathway descriptions (Yang et al., 2022), and subcellular localization (Magnano & Gitter, 2023). However, there has been no systematic study investigating the use of non-Euclidean and mixed-curvature embedding models for pathway graphs.

2.3. PPI Prediction

PPIs can be predicted based on combinations of proteins' sequence, expression, functional, evolutionary, or 3D structural features (Durham et al., 2023). One PPI prediction formulation is as an edge or link prediction task in a network of known PPIs (Li et al., 2022). This can be accomplished using features from the original graph or by learning node embeddings as features for the edge prediction task. For example, Feng et al. (2020) predict signaling cascades from PPI graphs by integrating transcriptomics and copy-number data into a graph neural network. Jiang et al. (2020) use graph embeddings to predict links that indicate an enzymatic reaction between pairs of molecules from the KEGG database. Finally, Zhang & Kabuka (2019) and Liu et al. (2020) leverage a combination of sequence and network information to make predictions of PPIs.

3. Methods

3.1. Data Processing

We downloaded PathBank pathways (Wishart et al., 2020) from Pathway Commons (Rodchenkov et al., 2020) v12, using the .txt files containing interaction participants, edge types, and associated metadata. For each pathway, we created a NetworkX graph object and ignored Pathway Commons edge types, treating each edge as undirected with no additional annotations. We then generated edge lists for the undirected graphs and learned embeddings using the mixed-curvature embedding Python package (Gu et al., 2018).

For the edge prediction test set, we downloaded a list of candidate edges (PPIs) from STRING (Szklarczyk et al., 2023). We identified PPI network nodes present in the pathway graphs by mapping both pathway nodes and STRING nodes to Ensembl IDs. We only included STRING edges with experimental evidence and discarded all edges with a score of less than 500 out of 1000.

3.2. Learning Embeddings

Let $\mathbb{S}^d_K, \mathbb{H}^d_K$ be the spherical and hyperbolic spaces of dimension d and curvature $K, -K$, respectively, and $\mathbb{E}^d$ the

Graph ID	Nodes	Edges	H dim	H copies	E dim	E copies	S dim	S copies	Best Euclidean Distortion	Best Overall Distortion	% Reduction in Distortion
623	10	11	14	2	14	3	14	2	0.0243	0.0015	93.83
701	6	11	25	0	25	1	25	3	0.0016	0.0001	93.75
622	8	10	12	2	12	3	12	3	0.0102	0.0008	92.16
620	8	10	12	2	12	3	12	3	0.0102	0.0008	92.16
621	8	10	12	2	12	3	12	3	0.0102	0.0008	92.16
858	9	8	12	3	12	3	12	2	0.015	0.0015	90.00
571	35	94	12	2	12	3	12	3	0.0314	0.0041	86.94
584	14	17	16	2	16	1	16	3	0.0217	0.0032	85.25
482	47	64	14	2	14	3	14	2	0.0327	0.0113	65.44
484	47	64	14	2	14	3	14	2	0.0327	0.0113	65.44

Table 1: Top ten PathBank pathway graphs as measured by % reduction in distortion over a purely Euclidean embedding.

Euclidean space of dimension d. We describe our main embedding space: for sequences of dimensions $s_1, \ldots, s_m$, $h_1, \ldots, h_n$, and $e_1, \ldots, e_p$, we write

$$\mathcal{P} = \mathbb{S}^{s_1} \times \cdots \times \mathbb{S}^{s_m} \times \mathbb{H}^{h_1} \times \cdots \times \mathbb{H}^{h_n} \times \mathbb{E}^{e_1} \times \cdots \times \mathbb{E}^{e_p},$$

a product manifold with $m + n + p$ component spaces and total dimension $\sum_i s_i + \sum_j h_j + \sum_k e_k$. We refer to each $\mathbb{S}^{s_i}, \mathbb{H}^{h_i}, \mathbb{E}^{e_i}$ as *components* or *factors*. We refer to the decomposition, e.g., $(\mathbb{H}^2)^2 = \mathbb{H}^2 \times \mathbb{H}^2$, as the *signature*. For convenience, let $M_1, \ldots, M_{m+n+p}$ refer to the factors in the product.

We learn an embedding function $f : \mathcal{G} \to \mathcal{P} = M_1 \times \cdots \times M_{m+n+p}$ where $\mathcal{G}$ is the space of pathway graphs and $\mathcal{P}$ is the Cartesian product of the factors. Distances in the product space decompose as sums of distances in the factor spaces (Lee, 2018), i.e.

$$d_{\mathcal{P}}(u, v) = \sum_i d_{M_i}(\pi_i(u), \pi_i(v))$$

for $u, v \in \mathcal{P}$, where π_i denotes projection onto the ith factor. The learning takes place according to the method given in Gu et al. (2018), namely embeddings are learned in a product manifold $\mathcal{P}$ via optimization of the following loss function

$$\mathcal{L}(x) = \sum_{1 \le i \le j \le n} \left| \left(\frac{d_{\mathcal{P}}(x_i, x_j)}{d_G(X_i, X_j)} \right)^2 - 1 \right|$$

where $d_G(X_i, X_j)$ denotes the graph distance between nodes X_i and X_j and $d_{\mathcal{P}}(x_i, x_j)$ denotes the product manifold geodesic distance between their learned embeddings.

The main metric used to evaluate our embeddings is the average graph distance *distortion*. We define the distortion, $\mathcal{D}$, of f relative to a graph $G = (V, E)$ to be

$$\mathcal{D}(f) = \frac{1}{|V|^2} \sum_{u,v \in V} \frac{|d_{\mathcal{P}}(f(u), f(v)) - d_G(u, v)|}{d_G(u, v)}$$

where graph distance $d_G(u, v)$ is the length of the shortest path connecting nodes u and v in G.

3.3. Hyperparameters

We create 252 embeddings for each pathway graph corresponding to different combinations of the number of hyperbolic, spherical, and Euclidean spaces, the learning rate, and the dimensionalities of each space (Appendix). We test having different numbers of copies of each space, where the number of copies of each type ranges from 0 to 3. We also constrain the total dimensionality of all spaces to sum to 100. This is to keep the comparison across different combinations fair by ensuring they each have the same representational capacity as governed by the number of dimensions.

4. Experiments

4.1. PathBank Pathway Embeddings

For each pathway graph, we determine the best combination of hyperbolic, Euclidean, and spherical components as determined by lowest distortion. Figure 1 demonstrates the reductions in distortion gained from learning the non-Euclidean embeddings over all 881 PathBank graphs. We observe that mixed-curvature product spaces provide marked reductions in distortion relative to the standard Euclidean embedding, with many graphs achieving a greater than 50% reduction in distortion. Table 1 summarizes the ten graphs that show the greatest benefit from a mixed-curvature embedding as measured by percent reduction in distortion over the Euclidean value. We also characterize each pathway graph via a number of common graph properties and learn a linear regression model that predicts the percent reduction in distortion over the Euclidean embeddings from these graph features, achieving an $R^2 = 0.37$ (Appendix). The goal in learning a regressor is to be able to predict, from a set of easily-generated graph features, which graphs are likely to benefit from a non-Euclidean representation.

4.2. PathBank Edge Prediction

We train on individual pathway graphs to predict a validation set of held-out edges from the PathBank graphs, then

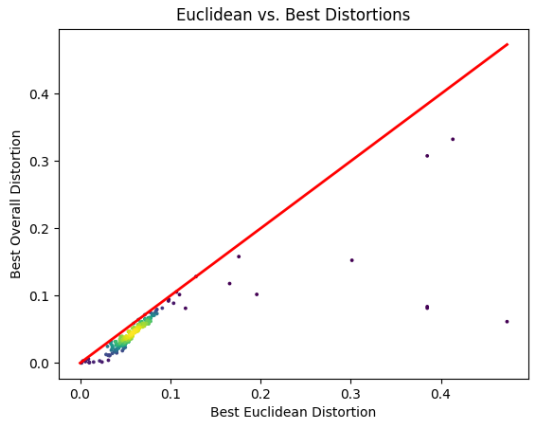

Figure 1. Scatterplot of distortion in the Euclidean embedding vs. distortion in the mixed-curvature embedding. Points are colored by local density, with yellow indicating the highest density.

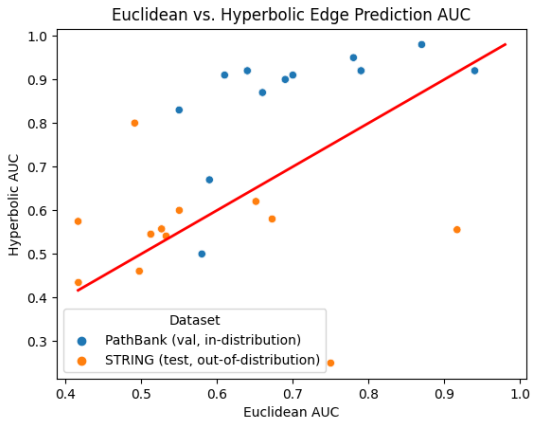

Figure 2. AUCs for the Hyperbolic vs Euclidean models on 24 tree-like PathBank graphs. Blue points represent the validation set of held-out edges from PathBank graphs, whereas orange points represent the test set of out-of-distribution edges from STRING.

test our prediction models on the test set of experimentally validated edges from STRING. We compare two edge prediction methods on 24 pathway graphs that exhibit tree-like structure indicating that they would be good candidates for a hyperbolic embedding. For the first, we initialize the embedding layer of a Euclidean Graph Convolutional Network (Kipf & Welling, 2017) with the pretrained 100-dimensional Euclidean embeddings. For the second, we initialize the embedding layer of a Hyperbolic Neural Network (Chami et al., 2019) with 100-dimensional hyperbolic embeddings. Although the hyperbolic model greatly improves edge predictions in almost all cases on the validation set, the relative performances on the test set is mixed (Figure 2)[4].

5. Discussion and Conclusion

We find that performing representation learning in non-Euclidean and mixed-curvature spaces yields notable improvements in distortion and downstream in-distribution edge prediction performance. In all cases, a mixed-curvature representation yields an embedding with lower distortion than a simple Euclidean embedding. However, the exact decomposition of the product space into its mixture of component factors depends on the graph topology. Thus, it is beneficial to perform a hyperparameter sweep over the number and types of factors, as well as their dimensionalities, when learning a representation for a biological pathway graph. Such a sweep need not be highly time intensive as biological pathway graphs are generally of a modest size.

Interestingly, for the edge prediction task, the full structure of the graph is not known when the node embeddings are learned because the test set edges from STRING are not included in the training set. This leads to a number of questions about how to further generalize this approach and improve out-of-distribution performance. For example, we may ask how close the full graph structure must be to the observed structure in order to produce node embeddings that are usable for edge prediction via our method.

Another future direction would be to perform similar analyses on other types of biological networks and pathways. Pathway Commons includes pathways from Reactome, KEGG, NetPath, HumanCyc, PID, PANTHER, and INOH that provide fertile ground for future exploration for the application of non-Euclidean geometry. Furthermore, edge prediction is not the only useful task for biological pathways that can be improved by non-Euclidean representation learning. We plan to investigate how our pathway embeddings benefit other downstream tasks such as node classification, for example, predicting the type (gene, protein, small molecule, metabolite, etc.) of a pathway node.

Non-Euclidean embedding models have not been applied to pathway graphs, perhaps due to a lack of awareness among network biology researchers of their utility in reducing distortion of graph distances and improving downstream predictive performance. We demonstrate that pathway graphs benefit from the incorporation of non-Euclidean geometries into embeddings and prediction models. We encourage researchers to consider making use of these non-standard geometries when learning embeddings and making downstream predictions. To this end, we also provide code at `https://github.com/mcneela/Mixed-Curvature-Pathways`

[4]Figure 2 originally mistakenly showed values on the validation set as if they were obtained on the test set. It has now been corrected with the addition of the correct values obtained on the test set.

Acknowledgements

We thank David Merrell, Christopher Magnano, and Sam Gelman for their help in processing Pathway Commons graphs and generating visualizations. This research was supported by NIH award R01GM135631 and the Wisconsin Alumni Research Foundation.

References

Bourgain, J. On lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1): 46–52, 1985. doi: 10.1007/BF02776078. URL https://doi.org/10.1007/BF02776078.

Chami, I., Ying, Z., Ré, C., and Leskovec, J. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 4869–4880, 2019.

Durham, J., Zhang, J., Humphreys, I. R., Pei, J., and Cong, Q. Recent advances in predicting and modeling protein–protein interactions. *Trends in Biochemical Sciences*, 48(6):527–538, June 2023. ISSN 0968-0004. doi: 10.1016/j.tibs.2023.03.003. URL https://www.cell.com/trends/biochemical-sciences/abstract/S0968-0004(23)00079-8.

Feng, J., Zeng, A., Chen, Y., Payne, P., and Li, F. Signaling interaction link prediction using deep graph neural networks integrating protein-protein interactions and omics data. *bioRxiv*, pp. 2020.12.23.424230, December 2020. doi: 10.1101/2020.12.23.424230. URL https://www.biorxiv.org/content/10.1101/2020.12.23.424230v1.

Ganea, O.-E., Bécigneul, G., and Hofmann, T. Hyperbolic Neural Networks. *arXiv:1805.09112*, June 2018. doi: 10.48550/arXiv.1805.09112. URL http://arxiv.org/abs/1805.09112.

Giovanni, F. D., Luise, G., and Bronstein, M. M. Heterogeneous manifolds for curvature-aware graph embedding. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022. URL https://openreview.net/forum?id=rtUxsN-kaxc.

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*, 2016.

Gu, A., Sala, F., Gunel, B., and Ré, C. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2018.

Hagberg, A. A., Schult, D. A., and Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. In *Proceedings of the 7th Python in Science Conference*, pp. 11–15, Pasadena, CA USA, 2008.

Hanspers, K., Riutta, A., Summer-Kutmon, M., and Pico, A. R. Pathway information extracted from 25 years of pathway figures. *Genome Biology*, 21(1):273, November 2020. ISSN 1474-760X. doi: 10.1186/s13059-020-02181-2. URL https://doi.org/10.1186/s13059-020-02181-2.

Hayakawa, J., Seki, T., Kawazoe, Y., and Ohe, K. Pathway importance by graph convolutional network and Shapley additive explanations in gene expression phenotype of diffuse large B-cell lymphoma. *PLOS ONE*, 17(6):e0269570, June 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0269570. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0269570.

Jiang, J., Liu, L.-P., and Hassoun, S. Learning graph representations of biochemical networks and its application to enzymatic link prediction. *Bioinformatics*, 37(6):793–799, 10 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa881. URL https://doi.org/10.1093/bioinformatics/btaa881.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Lai, M., Chen, G., Yang, H., Yang, J., Jiang, Z., Wu, M., and Zheng, J. Predicting Synthetic Lethality in Human Cancers via Multi-Graph Ensemble Neural Network. In *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1731–1734, November 2021. doi: 10.1109/EMBC46164.2021.9630716.

Lee, J. M. *Introduction to Riemannian Manifolds*, volume 176 of *Graduate Texts in Mathematics*. Springer International Publishing, Cham, 2018. ISBN 978-3-319-91754-2 978-3-319-91755-9. doi: 10.1007/978-3-319-91755-9. URL http://link.springer.com/10.1007/978-3-319-91755-9.

Lee, S., Lim, S., Lee, T., Sung, I., and Kim, S. Cancer subtype classification and modeling by pathway attention and propagation. *Bioinformatics*, 36(12):3818–3824, June 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa203. URL https://doi.org/10.1093/bioinformatics/btaa203.

Li, M. M., Huang, K., and Zitnik, M. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6(12):1353–1369, December 2022. ISSN 2157-846X. doi: 10.1038/

s41551-022-00942-x. URL https://www.nature.com/articles/s41551-022-00942-x.

Liang, B., Gong, H., Lu, L., and Xu, J. Risk stratification and pathway analysis based on graph neural network and interpretable algorithm. *BMC Bioinformatics*, 23(1): 394, September 2022. ISSN 1471-2105. doi: 10.1186/s12859-022-04950-1. URL https://doi.org/10.1186/s12859-022-04950-1.

Liu, L., Zhu, X., Ma, Y., Piao, H., Yang, Y., Hao, X., Fu, Y., Wang, L., and Peng, J. Combining sequence and network information to enhance protein–protein interaction prediction. *BMC Bioinformatics*, 21(16):537, 2020. doi: 10.1186/s12859-020-03896-6. URL https://doi.org/10.1186/s12859-020-03896-6.

M A Basher, A. R. and Hallam, S. J. Leveraging heterogeneous network embedding for metabolic pathway prediction. *Bioinformatics*, 37(6):822–829, 10 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa906. URL https://doi.org/10.1093/bioinformatics/btaa906.

Magnano, C. S. and Gitter, A. Graph algorithms for predicting subcellular localization at the pathway level. In *Pacific Symposium on Biocomputing*, pp. 145–156. World Scientific, 2023. ISBN 9789811270604. doi: 10.1142/9789811270611_0014. URL https://www.worldscientific.com/doi/10.1142/9789811270611_0014.

Nickel, M. and Kiela, D. Poincaré embeddings for learning hierarchical representations. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 30*, pp. 6338–6347. Curran Associates, Inc., 2017.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, volume 32, pp. 8024–8035. Curran Associates, Inc., 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. ISSN 1533-7928. URL http://jmlr.org/papers/v12/pedregosa11a.html.

Pershad, Y., Guo, M., and Altman, R. B. Pathway and network embedding methods for prioritizing psychiatric drugs. *Pacific Symposium on Biocomputing*, 25:671–682, 2020.

Pham, C. and Dang, T. Link Prediction for Biomedical Network. In *The 12th International Conference on Advances in Information Technology*, IAIT2021, pp. 1–5, New York, NY, USA, July 2021. Association for Computing Machinery. ISBN 978-1-4503-9012-5. doi: 10.1145/3468784.3471608. URL https://dl.acm.org/doi/10.1145/3468784.3471608.

Rodchenkov, I., Babur, O., Luna, A., Aksoy, B. A., Wong, J. V., Fong, D., Franz, M., Siper, M. C., Cheung, M., Wrana, M., Mistry, H., Mosier, L., Dlin, J., Wen, Q., O'Callaghan, C., Li, W., Elder, G., Smith, P. T., Dallago, C., Cerami, E., Gross, B., Dogrusoz, U., Demir, E., Bader, G. D., and Sander, C. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Research*, 48(D1):D489–D497, January 2020. ISSN 0305-1048. doi: 10.1093/nar/gkz946. URL https://doi.org/10.1093/nar/gkz946.

Sala, F., De Sa, C., Gu, A., and Re, C. Representation tradeoffs for hyperbolic embeddings. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4460–4469. PMLR, July 2018. URL https://proceedings.mlr.press/v80/sala18a.html.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., and von Mering, C. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1):D638–D646, January 2023. ISSN 0305-1048. doi: 10.1093/nar/gkac1000. URL https://doi.org/10.1093/nar/gkac1000.

Wishart, D. S., Li, C., Marcu, A., Badran, H., Pon, A., Budinski, Z., Patron, J., Lipton, D., Cao, X., Oler, E., Li, K., Paccoud, M., Hong, C., Guo, A. C., Chan, C., Wei, W., and Ramirez-Gaona, M. PathBank: a comprehensive pathway database for model organisms. *Nucleic Acids Research*, 48(D1):D470–D478, January 2020. ISSN 0305-1048. doi: 10.1093/nar/gkz861. URL https://doi.org/10.1093/nar/gkz861.

Yang, J., Liu, Z., Zhang, M., and Wang, S. Pathway2Text: Dataset and Method for Biomedical Pathway Description Generation. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pp. 1441–1454, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-naacl.

108. URL https://aclanthology.org/2022.findings-naacl.108.

Zhang, D. and Kabuka, M. Multimodal deep representation learning for protein interaction identification and protein family classification. *BMC Bioinformatics*, 20(16):531, 2019. doi: 10.1186/s12859-019-3084-y. URL https://doi.org/10.1186/s12859-019-3084-y.

Appendix

A. Training Details

A.1. Hyperparameters

Learning Rates	Hyperbolic Copies	Euclidean Copies	Spherical Copies
0.001	0	0	0
0.01	1	1	1
0.1	2	2	2
1	3	3	3

Table 2: Range of values for the hyperparameter sweep. Space dimensions were calculated automatically (to sum to 100) based on number of copies of each space. For example, if there were 2 hyperbolic copies, 1 Euclidean copy, and 3 spherical copies, then there would be floor(100/6) = 16 dimensions assigned to each copy.

A.2. Graph IDs and Pathway Names

Graph ID	PathBank Pathway Name
623	BTG Family Proteins and Cell Cycle Regulation
701	Eumelanin Biosynthesis
622	Multiple Carboxylase Deficiency, Neonatal or Early Onset Form
620	Biotin Metabolism
621	Biotinidase Deficiency
858	Methadone Metabolism Pathway
571	Degradation of Superoxides
584	D4-GDI Signaling Pathway
482	Homocysteine Degradation
484	gamma-Cystathionase Deficiency (CTH)

Table 3: Mapping of Graph IDs to Pathway Names from Table 1

Graph ID	PathBank Pathway Name
365	Glycogenosis, Type IC
362	Glycogen Storage Disease Type 1A (GSD1A) or Von Gierke Disease
93	Tyrosinemia Type I
94	Tyrosinemia, Transient, of the Newborn
339	Ethanol Degradation
449	Betaine Metabolism
645	Estrone Metabolism
627	Alternative Complement Pathway
408	Vitamin A Deficiency
407	Retinol Metabolism
11	3-Methylcrotonyl-CoA Carboxylase Deficiency Type I
575	Vitamin B6 Metabolism
87	Monoamine Oxidase-A Deficiency (MAO-A)
8	3-Hydroxy-3-methylglutaryl-CoA Lyase Deficiency
360	Gluconeogenesis
90	Tyrosine Metabolism
641	Phosphatidylethanolamine Biosynthesis
458	Spermidine and Spermine Biosynthesis
115	Thioguanine Action Pathway
83	Alkaptonuria
12	3-Methylglutaconic Aciduria Type III
14	3-Methylglutaconic Aciduria Type IV
22	Isovaleric Aciduria
25	Maple Syrup Urine Disease

Table 4: Tree-like pathways shown in Figure 2

B. Data

B.1. Pathway Commons

We use the Pathway Commons v12 PathBank pathways provided in this file: https://www.pathwaycommons.org/archives/PC2/v12/PathwayCommons12.pathbank.hgnc.txt.gz

Pathway Commons provides a number of different data formats, including text, SIF, JSON, and BioPAX. We use the text format as it provides the simplest graph representation of the most important interactions in PathBank pathways.

C. Modeling

C.1. Initializing Graph Neural Networks with Embeddings

We save the embeddings learned by the embedding model in a PyTorch .pt file (Paszke et al., 2019). We then load these embeddings and use them to initialize the weights of the embedding layer for the models in the edge prediction task, namely the Euclidean Graph Convolutional Network and Hyperbolic Neural Network. These embeddings are then further trained via backpropagation during the edge prediction task.

C.2. Linear Regression Model

We learn an off-the-shelf linear regression model, with the default hyperparameters, from scikit-learn (Pedregosa et al., 2011) v1.1.1 to predict reductions in distortion resulting from mixed-curvature embeddings for unseen pathway graphs. We create an 80-20 train-test split for this task using the results of our embedding experiment, that is, we train the model to predict the observed reductions in distortion. We then evaluate the model on the test set using scikit-learn. This returns the coefficient of determination, defined as

$$R^2 = \left(1 - \frac{u}{v}\right)$$

where u is the residual sum of squares, and v is the total sum of squares. The best possible score is 1.0, and the score can be negative for models that do worse than the mean value. Our model achieves an $R^2 = 0.37$ and is trained on the following features:

`Num Nodes, Num Edges, Density, Number of Self Loops, Node Connectivity, Avg Clustering Coeff, Diameter, Degree Assortativity, Is Bipartite`

Based on the R^2, the model does better than the mean at predicting reductions in distortion based on a small set of features easily generated within NetworkX (Hagberg et al., 2008).