Learning directed and hyperbolic gene embeddings

Aarthi Venkat [1*], Ferran Cardoso Rodriguez [2*], Joyce A. Chew [3], Michael Perlmutter [3], Christopher J. Tape [2], Smita Krishnaswamy [1]

[1]*Computational Biology and Bioinformatics Program, Yale University, New Haven, CT, USA.*
[2]*Cell Communication Lab, Department of Oncology, University College London Cancer Institute, London, UK.*
[3]*Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, USA*
*: These authors contributed equally to this work.
smita.krishnaswamy@yale.edu

***Abstract*—Molecular signals are transmitted in a hierarchical and directed manner, dictating a cell's response to both internal and external cues. Although these signaling networks are represented as directed graphs of genes, they are not consistently analyzed using methods that leverage the network's directed and hierarchical nature. Here, we present a Directed Scattering transform for Gene Embeddings (DSGE), which captures a directed and multiscale representation of the gene signaling network. We perform systematic comparison against existing undirected, directed, Euclidean, and hyperbolic approaches on the OmniPath gene network. DSGE outperforms existing methods on directed link prediction and performs among the best on standard link prediction. Our preliminary results suggest DSGE reveals key multiresolution and directed features for predicting and characterizing gene signaling relationships.**

***Index Terms*—directed graphs, hyperbolic embeddings, gene networks, representation learning**

I. Introduction

Cellular signaling involves overlapping, directed, and hierarchical signal transduction cascades [1] that govern a cell's response to its environment. This process can be represented as gene-gene networks. Understanding these networks across distinct biological settings is key, but methods to represent and infer gene-gene relationships rarely take into the account their directionality and hierarchical nature [2], [3].

For directed graphs, one possible approach for constructing low-dimensional embeddings while preserving directed structure [4]–[6] is to use the spectrum of a complex Hermitian matrix known as the magnetic Laplacian. Applied to a variety of data science tasks [7]–[10], recent works have used it to construct graph neural networks [11] and related versions of the scattering transform [12].

Concurrently, developments in hyperbolic geometry show potential for preserving latent hierarchies [13]–[19] . However, beyond some preliminary work [20], [21], little has been done to incorporate hyperbolic geometry into directed graph learning. Furthermore, there has been no work evaluating directed and hyperbolic approaches on biological networks.

Here, we describe an approach termed DSGE (Directed Scattering transform for Gene Embeddings), which learns a multiscale representation of directed biological networks via the magnetic Laplacian. We evaluate performance of *DSGE-Euc* (Euclidean representation) and *DSGE-Hyp* (hyperbolic representation) against baseline methods on the OmniPath directed gene signaling network [22] (4% reciprocal edges, Krachkardt hierarchy score (Khs) = 0.757) [18], [23], [24]. We additionally evaluate the top-performing methods on two smaller subgraphs, demonstrating the utility of directed scattering for capturing directedness and multiscale topology for biological graphs of various sizes.

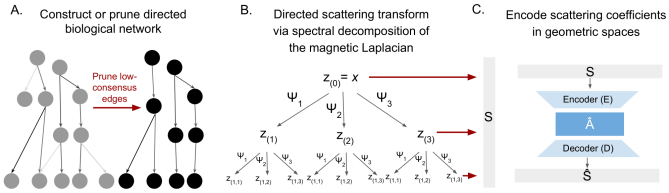

Fig. 1. **DSGE Schematic** A. Construction of directed biological graph B. Directed scattering transform. C. Low-dimensional embedding via Euclidean (DSGE-Euc) or hyperbolic (DSGE-Hyp) autoencoder layers.

II. Methodology

Our goal is to learn an embedding of a hierarchical, directed graph that preserves both directional and hierarchical information. Our method is based on a three-step approach described in each of the following subsections (Fig. 1).

A. Assembly of directed gene-gene graph

First, we define $G_O = (V_O, E_O)$ as the OmniPath signaling network, a weighted, directed gene-gene graph where V_O is a set of N_O vertices and $E_O \subseteq V_O \times V_O$ is a set of directed edges. G is derived from G_O by performing literature-based confidence edge-pruning using the edge-specific consensus score e_c as defined in [22], so that $G = (V, E)$ where $E = \{e \in E_O : e_c \geq \hat{c}\}$ for some user-defined threshold $\hat{c}$, and V is the subset of V_O with positive degree after the pruning. Additionally, we set $|V| = N$, and give all edges of E unit weight so as not to overemphasize well-studied genes.

B. Directed scattering transform via the magnetic Laplacian

We let A be the (asymmetric) adjacency matrix of G, let $A^{(s)} = \frac{1}{2}(A + A^T)$ be its symmetrized counterpart and let $D^{(s)}$ be the diagonal degree matrix corresponding to $A^{(s)}$, i.e. $D^{(s)}_{j,j} = \sum_{k=0}^{N-1} A^{(s)}_{j,k}$, and $D^{(s)}_{j,k} = 0$ if $j \neq k$. Then, we capture directional information via the phase matrix $\Theta^{(q)}$, where $\Theta^{(q)} = 2\pi q(A - A^T)$, where $q \geq 0$. This allows us to define the complex Hermitian adjacency matrix by $H^{(q)} = A^{(s)} \odot \exp(i\Theta^{(q)})$, where $i = \sqrt{-1}$ and $\odot$ denotes the Hadamard product (component-wise multiplication). Finally, we define the normalized magnetic Laplacian $L^{(q)}$, by $L^{(q)}_N = I - (D^{(s)})^{-1/2} H^{(q)} (D^{(s)})^{-1/2}$. It is shown in [11] that $L^{(q)}_N$ is positive semi-definite and therefore admits an orthonormal basis of eigenvectors $\mathbf{u}_k$, $0 \leq k \leq N-1$ such that $L^{(q)}\mathbf{u}_k = \lambda_k \mathbf{u}_k$, $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{N-1}$.

Next, following the lead of [12] (Section 7.3), for $t \geq 0$, we let H_t be the heat-kernel matrix defined by $H_t = \sum_{k=0}^{N-1} \exp(t\lambda_k)\mathbf{u}_k\mathbf{u}_k^*$ and for a fixed $J \geq 0$ we define a

	Method	Standard LP	Directed LP
Shallow	node2vec [26]	0.756	0.537
	PM [13]	0.787	0.525
	PM-D [13]	0.740	0.546
Undirected GNN	GAE [27]	0.874	0.602
	HGCN [16]	0.891	0.573
	GAE-D [27]	0.874	0.602
	HGCN-D [16]	0.886	0.599
Directed GNN	MagNet [11]	**0.947**	0.714
Scattering	UDS-AE [28]	0.937	0.581
Ours	DSGE-Euc	0.930	**0.718**
	DSGE-Hyp	0.901	0.716

TABLE I

TABLE 1. COMPARISON OF STANDARD AND DIRECTED LP AUROC. **BOLD** IS BEST-PERFORMING, UNDERLINED IS SECOND-BEST. PM: POINCARÉ MAP; GAE: GRAPH AUTOENCODER; HGCN: HYPERBOLIC GRAPH CONVOLUTIONAL NETWORK; UDS: UNDIRECTED SCATTERING

wavelet frame $\mathcal{W}_J = \{W_j\}_{j=0}^J \cup \{A_J\}$ where $W_0 = Id - H_1$, $A_J = H_{2^J}$, and $W_j = H_{2^{j-1}} - H_{2^j}$ for $1 \leq j \leq J$.

For a signal $\mathbf{x} : V \to \mathbb{R}$, we will, in a minor abuse of notation, identify $\mathbf{x}$ with the vector whose k-th entry $\mathbf{x}[k]$ is equal to the value of $\mathbf{x}$ at the k-th vertex (in some fixed ordering). We define the first- and second- order scattering coefficients of $\mathbf{x}$ by $S[j]\mathbf{x} = H_1 M W_j \mathbf{x}$ for $0 \leq j \leq J$ and $S[j_1, j_2]\mathbf{x} = H_1 M W_{j_2} M W_{j_1} \mathbf{x}$, $0 \leq j_1 \leq j_2$, where M is the entry-wise modulus (absolute value) operator, i.e. $M\mathbf{x}(v) = |\mathbf{x}(v)|$.[1] The zeroth-order scattering coefficient is defined simply by $H_1\mathbf{x}$. In our experiments, we will use N standard Gaussian random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_N$. We then let $S[v]$ denote the concatenation of all zeroth-, first-, and second-order scattering coefficients evaluated at the vertex v and let $S(G) = \{S[v] : v \in V\}$.

C. Low-dimensional encoding

The scattering representations $S[v]$ are typically redundant and unnecessarily high-dimensional. Therefore, our next step is to apply an autoencoder $D \circ E$ so that $\hat{S}(G) \approx D(E(S(G))$, i.e. so that $\|S(G) - \hat{S}(G)\|_2^2 = \sum_{v \in V} \|S[v] - \hat{S}[v]\|_2^2$ is as small as possible. We let d denote the dimension of $E(S[v])$.

In our experiments, we consider two versions of this model: one with a Euclidean encoder and decoder (DSGE-Euc), and one with a hyperbolic encoder and decoder (DSGE-Hyp). In both versions the encoder E and the deconder D are chosen to be multilayer perceptrons. The difference is that in the Euclidean version, the matrix multiplicaions are carried out via standard multiplication and addition operations whereas in the hyperbolic version they are carried out using using Möbius addition and multiplication (see, e.g., [15]). The inclusion of hyperbolic operations enables learning representations informed by hyperbolic in addition to directed geometry.

III. PRELIMINARY RESULTS

A. Evaluation of low-dimensional representations

We first evaluate our approaches for the full OmniPath graph. Table 1 compares the performance of DSGE-Euc and DSGE-Hyp versus baseline methods, including directed GNN MagNet [11] and hyperbolic GNN HGCN [16]. For standard link prediction (LP) (where negative edges are edges that do not exist in graph), DSGE-Euc and DSGE-Hyp perform comparably to MagNet and undirected scattering, and for directed link prediction (where negative edges are edges that are non-existent reciprocals to positive), they outperform all other methods. These results confirm that methods that account for directedness outperform undirected methods for node embeddings of directed graphs. This also highlights the utility of scattering for learning multiresolution representations.

[1]The use of H_1, rather than A_J, is based on the experimental setup in [12] and graph-residual convolutions in [25]

OmniPath Subgraph	Method	Standard LP	Directed LP
0.05 N = 2677 Khs = 0.973	MagNet	0.776	0.717
	UDS-AE	0.796	0.541
	DSGE-Euc	**0.837**	**0.756**
	DSGE-Hyp	0.816	0.749
0.25 N = 5771 Khs = 0.867	MagNet	0.876	0.703
	UDS-AE	0.878	0.565
	DSGE-Euc	**0.901**	**0.735**
	DSGE-Hyp	0.876	0.734

TABLE II

TABLE 2. COMPARISON OF STANDARD AND DIRECTED LP AUROC FOR TOP PERFORMING METHODS ON SUBGRAPHS.

B. Evaluation of representations for small signaling graphs

Often, gene relationships need to be characterized for small networks, such as those containing only genes related to a particular disease. To this end, we evaluate each method's ability to predict links from smaller graphs. Table 2 measures performance for downsampled OmniPath subgraphs with 5% (N=2677, Khs = 0.973) and 25% (N=5771, Khs=0.867) of the graph edges. For both graphs and both tasks, DSGE-Euc outperformed other top-performing methods. DSGE-Hyp performed second best for the 5% subsampled graph for both tasks, and on the directed LP task for the 25% subsampled graph. These results suggest that scattering is beneficial for learning the graph topology in smaller networks without over-smoothing across large neighborhoods (as GNNs can be susceptible to). We note that DSGE-Euc generally performs a bit better than DSGE-Hyp, perhaps because of the cumbersome nature of Riemannian optimization or perhaps because the OmniPath graph, which has a maximum depth of 6, is not deep enough to see the benefits of the hyperbolic approach. In future work, we intend to explore different optimization methods and deeper graphs.

IV. TRAINING DETAILS

All tasks were trained with an Adam optimizer (or RiemmanianAdam for hyperbolic methods) for 200 epochs with a patience of 50 epochs[2]. All methods learned representations with node dimension $d = 16$. The train/val/test split was 85/5/10. PM, GAE and HGCN refer to the implementations with the symmetrized adjacency matrix, and PM-D, GAE-D, and HGCN-D refer to the asymmetric matrices. For Node2Vec, we optimized for walk length and number of walks. For remaining methods, we optimized learning rate, bias, dropout, number of layers, activation, and weight decay. For MagNet, DSGE-Hyp, and DSGE-Euc, we also optimized for q, and for HGCN, Poincaré map, and DSGE-Hyp, we also optimized for c. For DSGE, we set $J = 15$. Results are reported as AUROC averaged over five runs.

[2]github.com/KrishnaswamyLab/Directed-Hierarchical-Gene-Networks

REFERENCES

[1] J. T. Hancock, "3 - the principles of cell signalling," in *On Growth, Form and Computers*, S. Kumar and P. J. Bentley, Eds. London: Academic Press, 2003, pp. 64–81.

[2] A. Pratapa, A. P. Jalihal, J. N. Law, A. Bharadwaj, and T. M. Murali, "Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data," *Nature Methods*, vol. 17, no. 2, pp. 147–154, Feb. 2020. [Online]. Available: https://doi.org/10.1038/s41592-019-0690-6

[3] E. Armingol, A. Officer, O. Harismendy, and N. E. Lewis, "Deciphering cell-cell interactions and communication from gene expression," *Nat. Rev. Genet.*, vol. 22, no. 2, pp. 71–88, Feb. 2021.

[4] E. Lieb and M. Loss, "Fluxes, laplacians and kasteleyn's theorem," Sep. 1992.

[5] A. Singer and H.-T. Wu, "Orientability and diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 44–58, Jul. 2011.

[6] ——, "Vector diffusion maps and the connection laplacian," *Commun. Pure Appl. Math.*, vol. 65, no. 8, pp. 1067–1144, Aug. 2012.

[7] S. Furutani, T. Shibahara, M. Akiyama, K. Hato, and M. Aida, "Graph signal processing for directed graphs based on the hermitian laplacian," in *Machine Learning and Knowledge Discovery in Databases*, 2020, pp. 447–463.

[8] A. Cloninger, "A note on markov normalized magnetic eigenmaps," *Applied and Computational Harmonic Analysis*, vol. 43, no. 2, pp. 370 – 380, 2017.

[9] M. Fanuel, C. M. Alaíz, Ángela Fernández, and J. A. Suykens, "Magnetic eigenmaps for the visualization of directed networks," *Applied and Computational Harmonic Analysis*, vol. 44, pp. 189–199, 2018.

[10] B. M. F. de Resende and L. d. F. Costa, "Characterization and comparison of large directed networks through the spectra of the magnetic laplacian," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 7, p. 073141, 2020.

[11] X. Zhang, Y. He, N. Brugnone, M. Perlmutter, and M. Hirn, "Magnet: A neural network for directed graphs," in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34. Curran Associates, Inc., 2021, pp. 27 003–27 015. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/e32084632d369461572832e6582aac36-Paper.pdf

[12] J. Chew, M. Hirn, S. Krishnaswamy, D. Needell, M. Perlmutter, H. Steach, S. Viswanath, and H.-T. Wu, "Geometric scattering on measure spaces," Aug. 2022.

[13] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/59dfa2df42d9e3d41f5b02bfc32229dd-Paper.pdf

[14] C. De Sa, A. Gu, C. Ré, and F. Sala, "Representation tradeoffs for hyperbolic embeddings," Apr. 2018.

[15] O.-E. Ganea, G. Bécigneul, and T. Hofmann, "Hyperbolic neural networks," May 2018.

[16] I. Chami, R. Ying, C. Ré, and J. Leskovec, "Hyperbolic graph convolutional neural networks," Oct. 2019.

[17] Q. Liu, M. Nickel, and D. Kiela, "Hyperbolic graph neural networks," Oct. 2019.

[18] I. Chami, A. Wolf, D.-C. Juan, F. Sala, S. Ravi, and C. Ré, "Low-dimensional hyperbolic knowledge graph embeddings," May 2020.

[19] Y. Bai, R. Ying, H. Ren, and J. Leskovec, "Modeling heterogeneous hierarchies with relation-specific hyperbolic cones," Oct. 2021.

[20] D. McDonald and S. He, "Hyperbolic embedding of attributed and directed networks," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–12, 2022.

[21] Z. Wu, Z. Di, and Y. Fan, "A hyperbolic embedding model for directed networks," Jun. 2019.

[22] D. Türei, A. Valdeolivas, L. Gul, N. Palacio-Escat, M. Klein, O. Ivanova, M. Ölbei, A. Gábor, F. Theis, D. Módos, T. Korcsmáros, and J. Saez-Rodriguez, "Integrated intra- and intercellular signaling knowledge for multicellular omics analysis," *Mol. Syst. Biol.*, vol. 17, no. 3, p. e9923, Mar. 2021.

[23] I. Balazevic, C. Allen, and T. Hospedales, "Multi-relational poincaré graph embeddings," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/f8b932c70d0b2e6bf071729a4fa68dfc-Paper.pdf

[24] D. Krackhardt, "Graph theoretical dimensions of informal organizations," *Computational organization theory, pages 107–130. Psychology Press.*

[25] F. Wenkel, Y. Min, M. Hirn, M. Perlmutter, and G. Wolf, "Overcoming oversmoothness in graph convolutional networks via hybrid scattering networks," *arXiv preprint arXiv:2201.08932*, 2022.

[26] A. Grover and J. Leskovec, "Node2vec: Scalable feature learning for networks," Jul. 2016.

[27] T. N. Kipf and M. Welling, "Variational graph Auto-Encoders," Nov. 2016.

[28] F. Gao, G. Wolf, and M. Hirn, "Geometric scattering for graph data analysis," Oct. 2018.