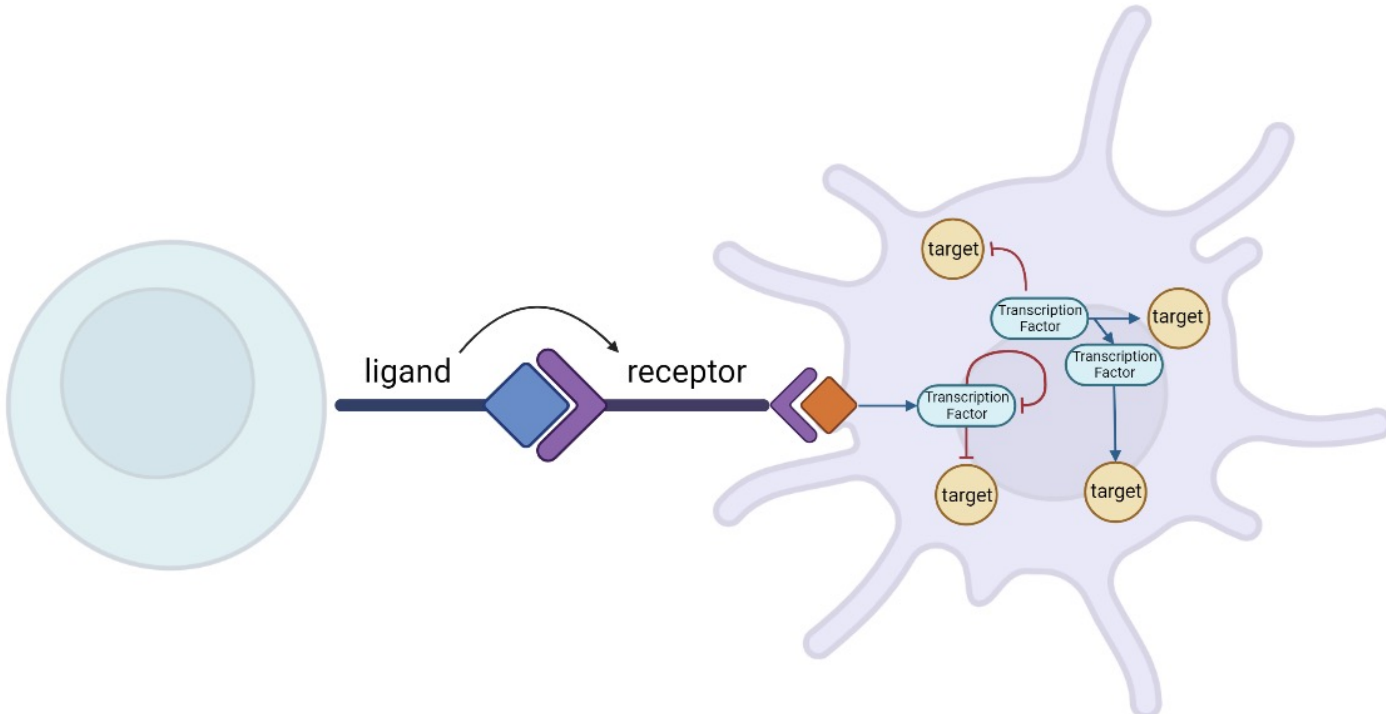
Directed Scattering for learning Euclidean & Hyperbolic Gene Embeddings

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

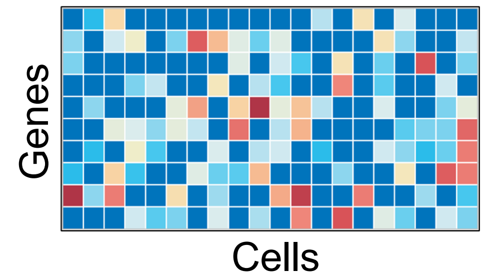
Genes "interact" by:

Ligands → Receptors → Transcription Factors→ Targets



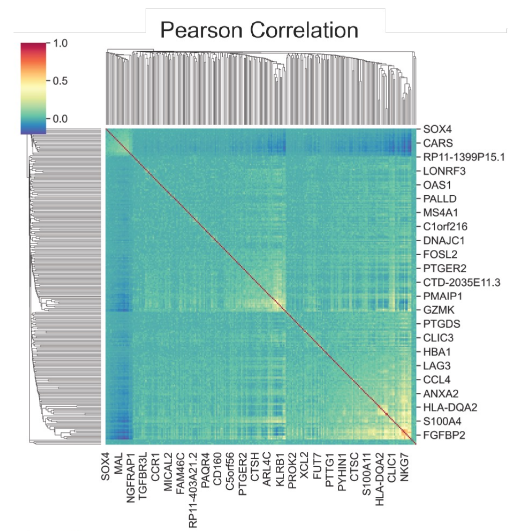
Yale

Gene representation is nontrivial



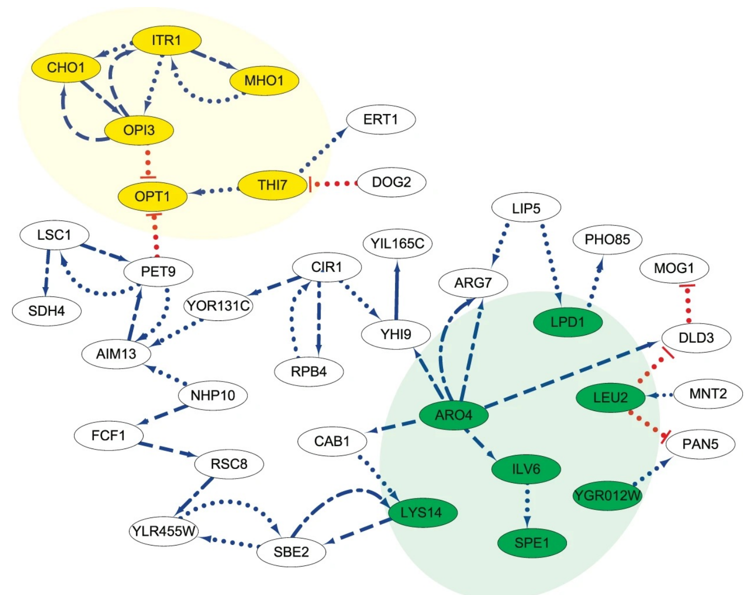
Gene signaling networks are hard to represent from data alone

- Metrics of coexpression are noisy and do not capture real interactions



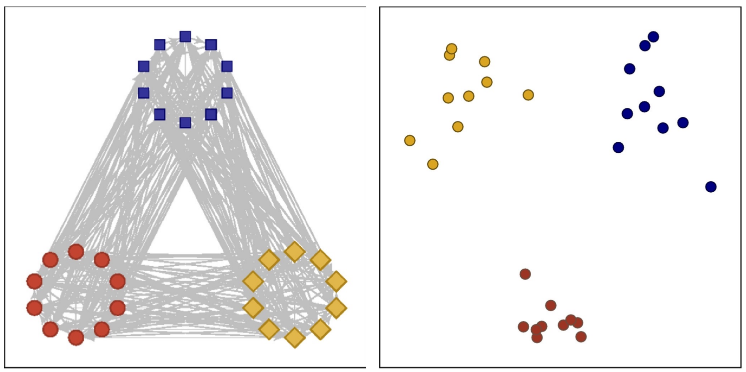
Gene representation is nontrivial

For representing prior gene signaling graphs, classical approaches don't account for directionality



Chen, C., Zhang, D., Hazbun, T.R. et al. Inferring Gene Regulatory Networks from a Population of Yeast Segregants. 2019

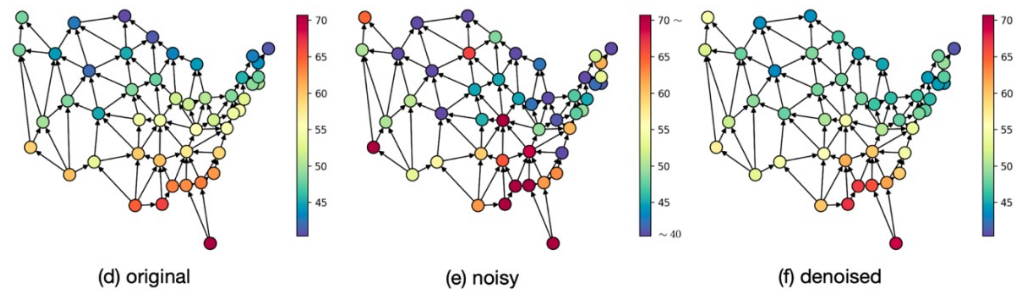
Magnetic Laplacian extends GSP to directed graphs



(a) Graph using the expert knowledge positions.

(b) First magnetic eigenmap ($\mathrm{Re}\,\phi_0^{(g)}$ vs. $\mathrm{Im}\,\phi_0^{(g)}$).

Visualization



Denoising

Magnetic Laplacian extends GSP to directed graphs

1. $A^{(s)} = \frac{1}{2}(A + A^T)$ **Symmetrized adjacency matrix**

2. $D^{(S)}_{j,j} = \sum_k A^{(S)}_{j,k}$ **Degree matrix of A(s)**

3. $\Theta^{(q)} = 2\pi q\ (A - A^T), q \geq 0$ **Phase matrix captures direction**

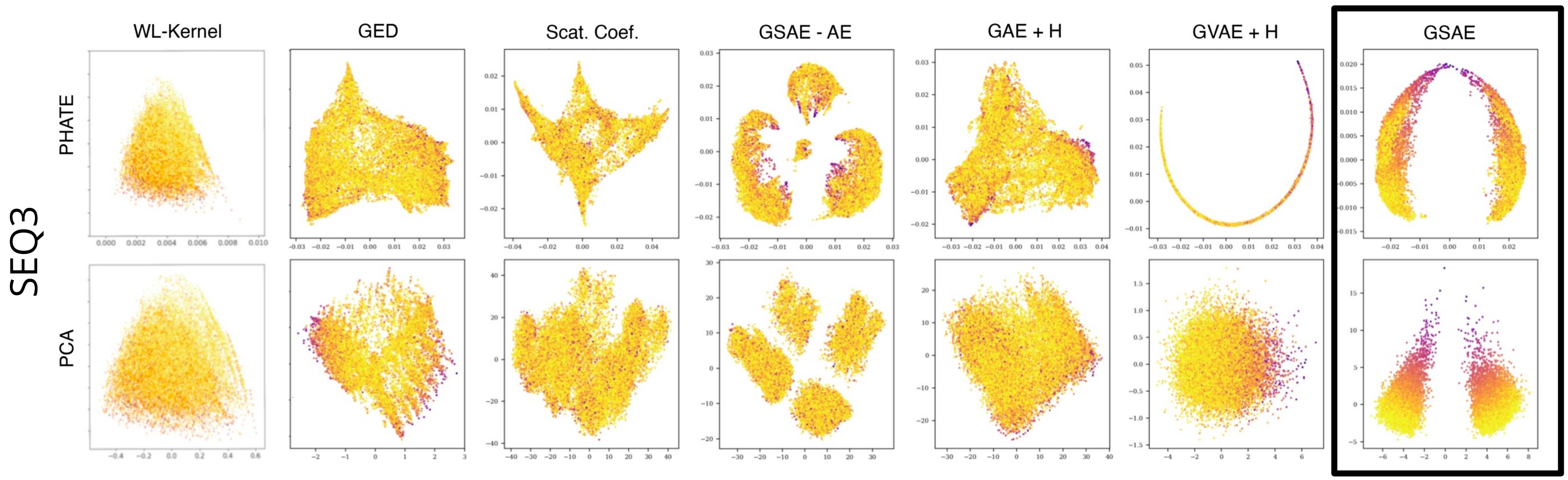
4. $H^{(q)} = A^{(s)} \odot \exp(i\ \Theta^{(q)})$ **Complex Hermitian adjacency matrix**

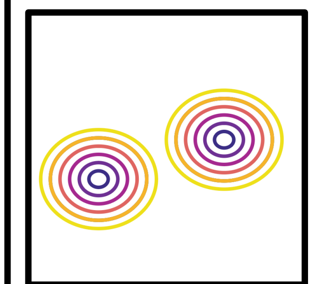
5. $L^{(q)}_N = I - (D^{(s)})^{-1/2} H^{(q)}\ (D^{(s)})^{-1/2}$ **Normalized magnetic Laplacian**

$L^{(q)}_N$ is positive semidefinite and admits an orthonormal basis of eigenvectors u_i with eigenvalues λ_i

Furutani et al. “Graph Signal Processing for Directed Graphs based on Hermitian Laplacian” 2020.

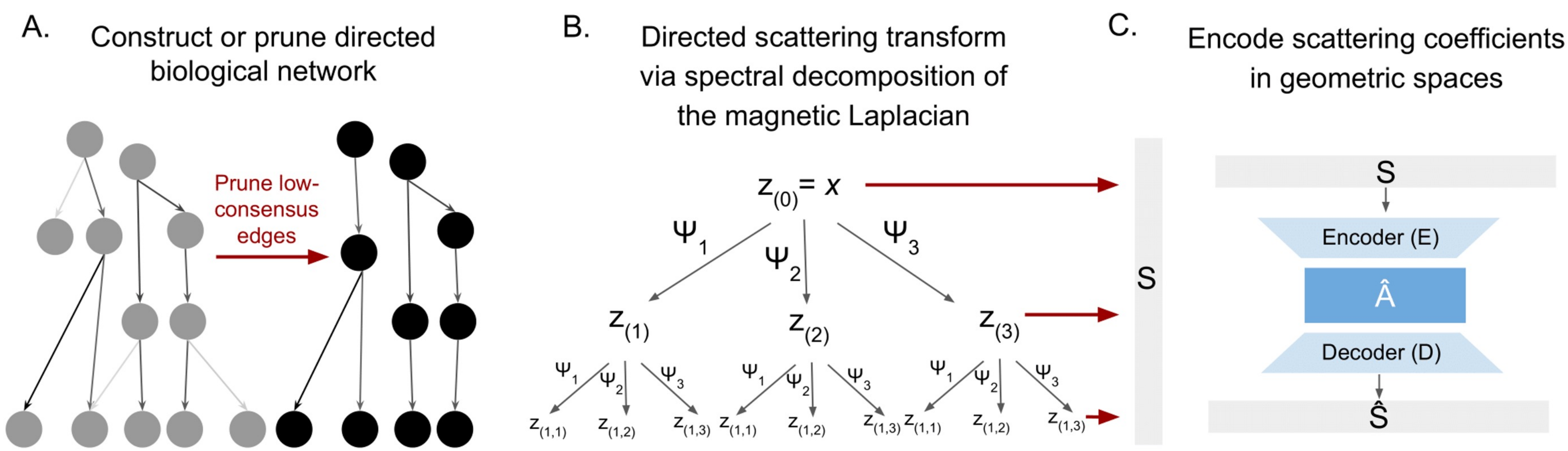
Geometric scattering for representation learning





Ground truth is two stable states

Directed Scattering for Gene Embeddings

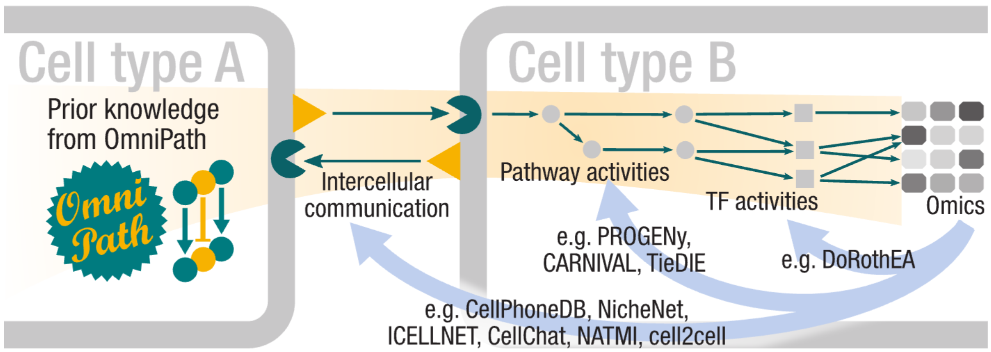


Assembly of directed gene-gene graph

OmniPath database

103,396 directed interactions
12,469 genes

Krackhardt hierarchy score = 0.757



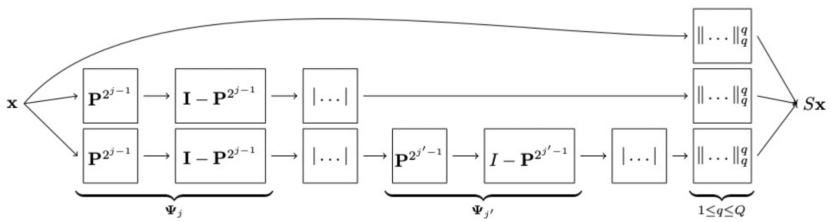
Intercellular (signaling) vs Intracellular (gene regulation)

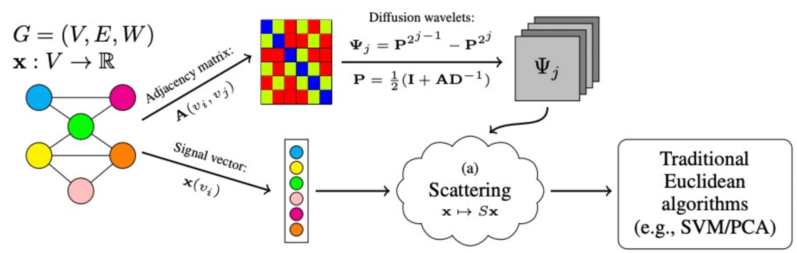
Turei et al. “Integrated intra- and intercellular signaling knowledge for multicellular omics analysis” 2021.

Directed Scattering Transform

Geometric Scattering for Graph Data Analysis

Feng Gao [1 2] Guy Wolf [* 3] Matthew Hirn [* 1 4]



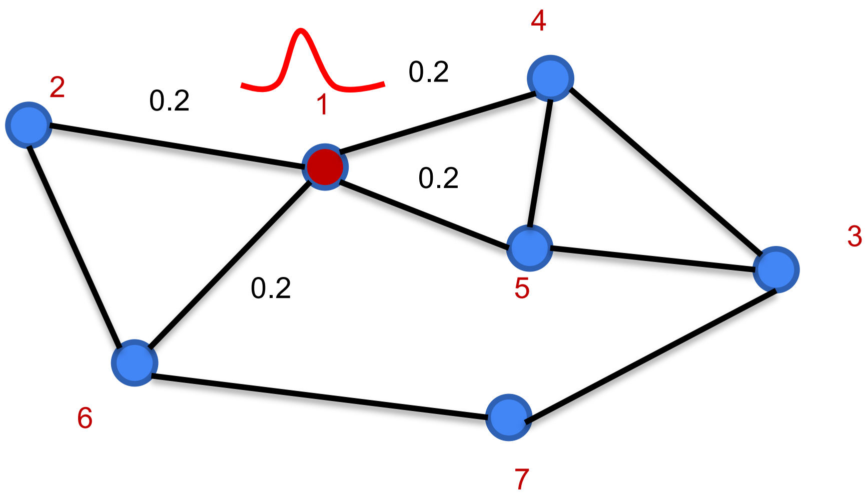


Yale

Diffusion Operator

.2	.2	0	.2	.2	.2	0
.333	.333	0	0	0	.333	0
0	0	.25	.25	.25	0	.25
.25	0	.25	.25	.25	0	0
.25	0	.25	.25	.25	0	0
.25	.25	0	0	0	.25	.25
0	0	.333	0	0	.333	.33

1	1	0	1	1	1	0
1	1	0	0	0	1	0
0	0	1	1	1	0	1
1	0	1	1	1	0	0
1	0	1	1	1	0	0
1	1	0	0	0	1	1
0	0	1	0	0	1	1

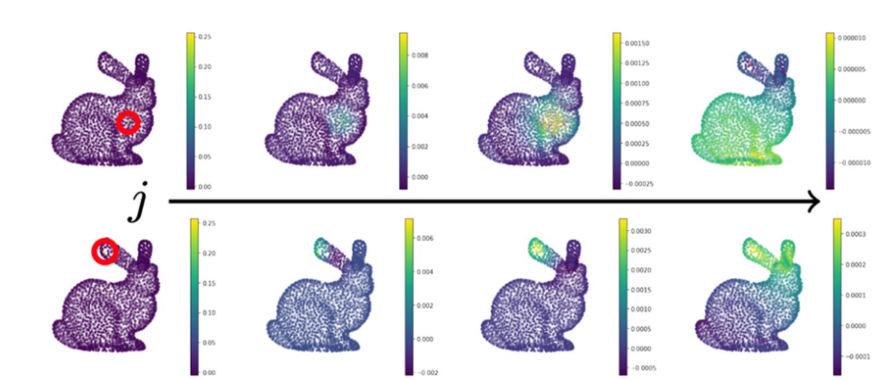


$$P = D^{-1}A$$

Diffusion Wavelets: Difference between scales of diffusion

Definition [Maggioni] *A diffusion wavelet transform is a difference between two scales of lazy diffusion on a graph*

$$\Psi_j = P^{2^j} - P^{2^{j-1}}$$

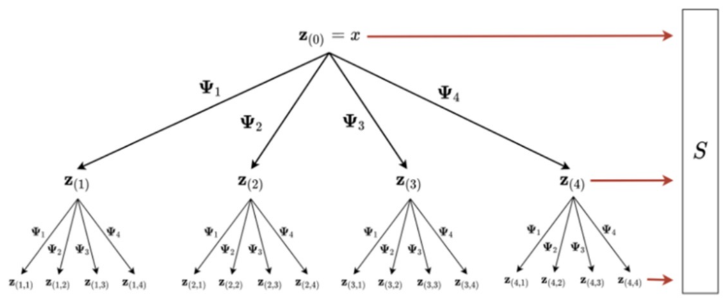


[

Geometric scattering transform

Definition [Gao et al.] *A* ***geometric scattering transform S*** *is an alternating cascade of wavelet coefficients and non-linear aggregations of a signal defined on the nodes of a graph*

$$\boldsymbol{U}_p\boldsymbol{x} := \boldsymbol{\Psi}_{j_m}|\boldsymbol{\Psi}_{j_{m-1}}\ldots|\boldsymbol{\Psi}_{j_2}|\boldsymbol{\Psi}_{j_1}\boldsymbol{x}||\ldots| \qquad \boldsymbol{S}_{p,q}\boldsymbol{x} := \sum_{i=1}^{n}|\boldsymbol{U}_p\boldsymbol{x}[v_i]|^q$$



Gao et al. “Geometric Scattering for Graph Data Analysis” 2018.

Generalized wavelets and scattering

Geometric Scattering on Measure Spaces

Joyce Chew, Matthew Hirn, Smita Krishnaswamy, Deanna Needell, Michael Perlmutter, Holly Steach, Siddharth Viswanath, Hau-Tieng Wu

October 17, 2022

Definition *Given a laplace beltrami operator.* $-\Delta$, *with eigenvalues* $0 = \lambda_0 \leq \lambda_1 \leq \lambda_2$ ... *and associated Eigenfunctions* $-\Delta\varphi_k = \lambda_k\ \varphi_k$, *the heat semigroup* H_t *is an operator that satisfies the heat equation.*

$H^t f = \sum_{k>0} g(\lambda_k)^t f(k)\varphi_k$ where $g(\lambda) = e^{-\lambda}$

Definition [Chew et al.] *A generalized wavelet transform is a difference between two scales of a heat semigroup*

$$\Psi_j = H^{2^{j-1}} - H^{2^{j-1}}$$

... scattering follows

Directed wavelets from a Magnetic Laplacian

$L_N^{(q)}$ has eigenvalues λ_k and eigenvectors u_k for $0 \leq k \leq N-1$

$H_k = \sum_{k-0}^{N-1} e^{t\lambda_k u_k u_k^*}$ $\quad$ $H_k = \sum_{k-0}^{N-1} e^{t\lambda_k u_k u_k^*}$

Directed wavelets are based on the heat kernel which is a matrix exponential of the magnetic Laplacian

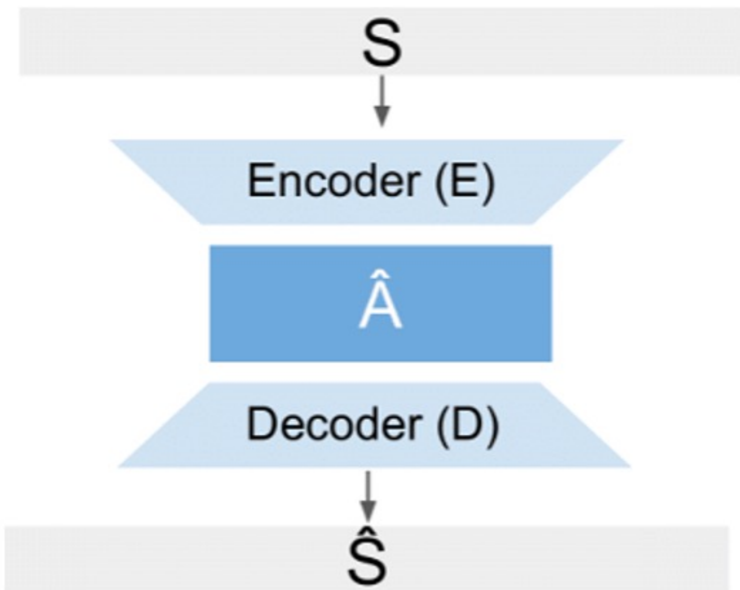
$$W_0 = I - H_1$$
$$W_j = H_{2^{j-1}} - H_{2^j}, 1 \leq j \leq J$$

Wavelets are differences between two scales of heat diffusion

$\mathfrak{W}_j = \{W_j\}, 0 \leq j \leq J \cup \ H_J$

The wavelet library consists of dyadic wavelet scales

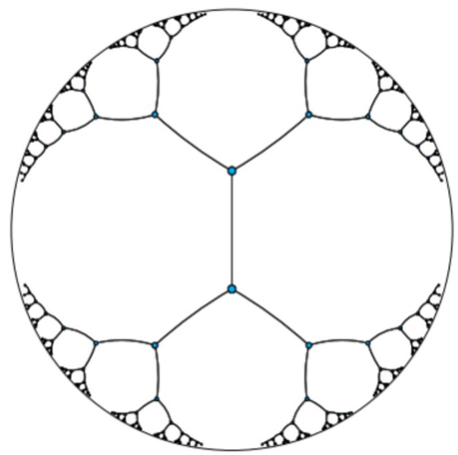
Learning low-dimensional encoding



$$\widehat{S(G)} \approx D(E(S(G)))$$

DSGE-Euc: Euclidean encoder and decoder
DSGE-Hyp: Hyperbolic encoder and decoder

Hyperbolic geometry for tree-like graphs



(b) Embedding of a tree in $\mathcal{B}^2$

Nickel and Kiela "Poincaré Embeddings for Learning Hierarchical Representations" 2017.

Hyperbolic operations for deep learning

Möbius addition. The *Möbius addition* of x and y in $\mathbb{D}_c^n$ is defined as

$$x \oplus_c y := \frac{(1 + 2c\langle x, y\rangle + c\|y\|^2)x + (1 - c\|x\|^2)y}{1 + 2c\langle x, y\rangle + c^2\|x\|^2\|y\|^2}. \tag{6}$$

Möbius scalar multiplication. For $c > 0$, the *Möbius scalar multiplication* of $x \in \mathbb{D}_c^n \setminus \{\mathbf{0}\}$ by $r \in \mathbb{R}$ is defined as

$$r \otimes_c x := (1/\sqrt{c}) \tanh(r \tanh^{-1}(\sqrt{c}\|x\|)) \frac{x}{\|x\|}, \tag{7}$$

Link prediction AUROC from OmniPath graph

	Method	Directed LP
Shallow	node2vec	0.537
	PM	0.525
	PM-D	0.546
Undirected GNN	GAE	0.602
	HGCN	0.573
	GAE-D	0.602
	HGCN-D	0.599
KGE	TransE	0.656
	TransE-edge	0.645
Scattering	UDS-AE	0.581
Directed GNN	MagNet	0.714
Ours	DSGE-Euc	**0.718**
	DSGE-Hyp	0.716

Drug–gene association prediction for Vitamin K3

Edges from DrugBank, ChEMBL, DrugCentral, and BindingDB

	Method	Precision@100	Recall@100
Shallow	node2vec	0.18	0.225
	PM	0.06	0.075
	PM-D	0.12	0.150
Undirected GNN	GAE	0.20	0.250
	HGCN	0.12	0.150
	GAE-D	0.20	0.250
	HGCN-D	0.16	0.200
KGE	TransE	0.12	0.150
	TransE-edge	0.11	0.138
Scattering	UDS-AE	0.18	0.225
Directed GNN	MagNet	0.14	0.175
Ours	DSGE-Euc	**0.21**	**0.263**
	DSGE-Hyp	<u>0.20</u>	<u>0.250</u>

Himmelstein et al. “Systematic integration of biomedical knowledge prioritizes drugs for repurposing” 2017.

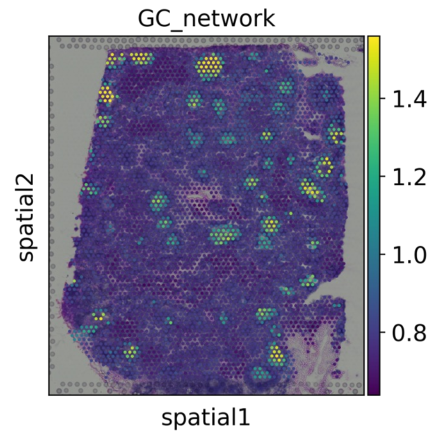
Disease–gene association prediction for Autistic Disorder

Edges from GWAS, DISEASES, DisGeNET, and DOAF

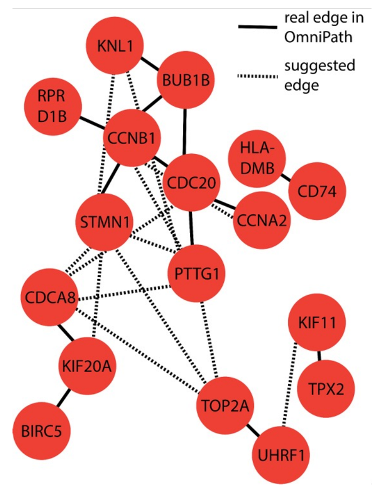
	Method	Precision@100	Recall@100
Shallow	node2vec	0.05	0.185
	PM	0.04	0.148
	PM-D	0.03	0.111
Undirected GNN	GAE	0.06	0.222
	HGCN	0.04	0.148
	GAE-D	0.06	0.222
	HGCN-D	0.03	0.111
KGE	TransE	0.05	0.185
	TransE-edge	0.02	0.074
Scattering	UDS-AE	0.04	0.148
Directed GNN	MagNet	0.08	0.296
Ours	DSGE-Euc	<u>0.11</u>	<u>0.407</u>
	DSGE-Hyp	**0.13**	**0.481**

Himmelstein et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing" 2017.

Integrate gene-gene correlation in space with DSGE representation to discover gene-gene signaling



Spatial gene expression of human lymph node (GC: Germinal centers)

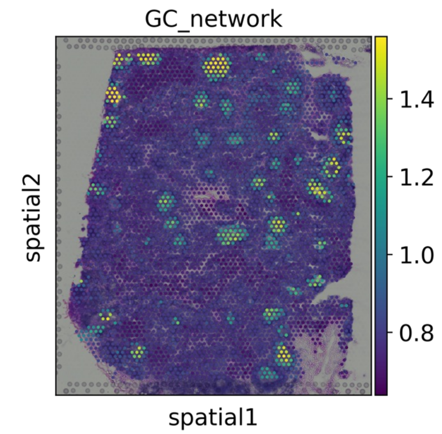


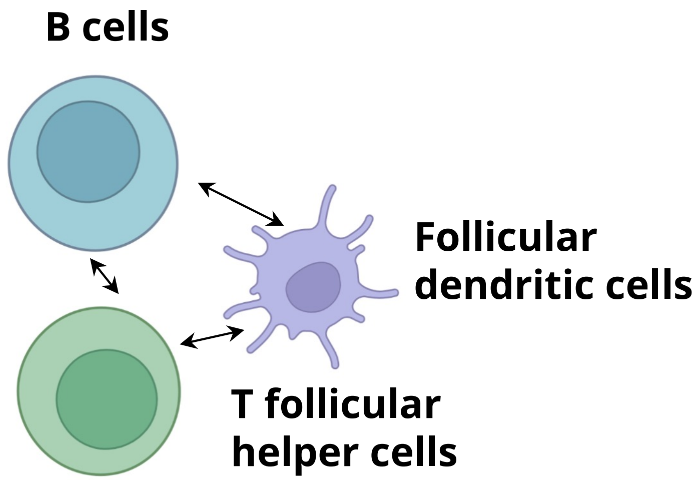
Inferred gene signaling network

Future work

1. Explore opportunities to infer directed cell-cell relationships based on gene-gene network

2. Further understand Gromov hyperbolicity of gene-gene networks





Krishnaswamy Lab at Yale University

Looking for students & postdocs!

- Machine learning methods incorporating signal processing, data geometry and topology
- Exploratory analysis and inference for biomedical datasets (genomics, fMRI, patient data)

Acknowledgments

Krishnaswamy Lab
Smita Krishnaswamy
Chen Liu
Danqi Liao
Dhananjay Bhaskar
Edward De Brouwer
Egbert Castro
Kevin Givechian
Dami Fasina
Scott Youlten
Xingzhi Sun
Holly Steach

Funding
Gruber Fellowship

Tape Lab
Ferran Cardoso Rodriguez
Christopher J. Tape

Perlmutter Lab
Michael Perlmutter
Joyce A. Chew

Thesis Committee
Smita Krishnaswamy
Rex Ying
Nikhil Joshi

Yale

GSP 2023

Thank you!