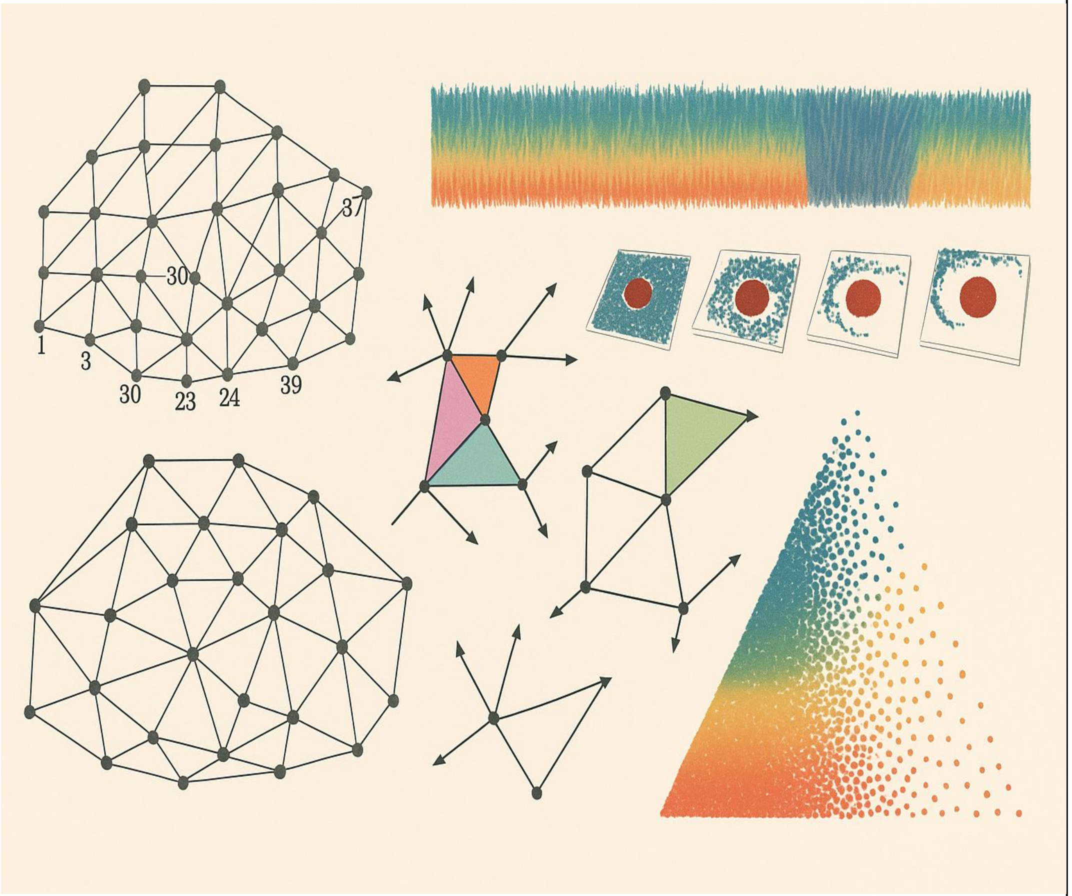
2025/6/10

Unveiling Latent Structural Collapse:

A Discrete Morse Perspective on SAE Representations for Gene Expression

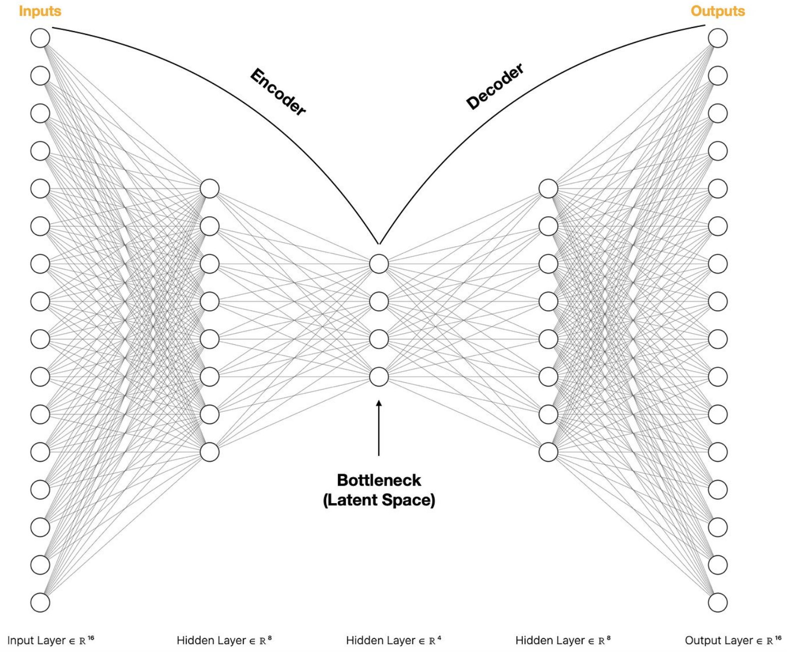
Kaijie Zhang

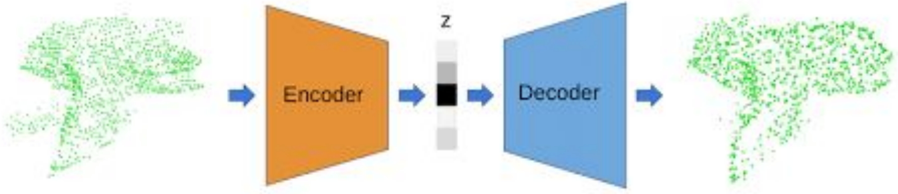


DSC 214: TDA

Motivation

- Autoencoders compress data nonlinearly
- Accuracy drops **non-linearly** with lower latent dims
- At some point: **sudden collapse** of meaningful structure
- We need to **detect** that collapse



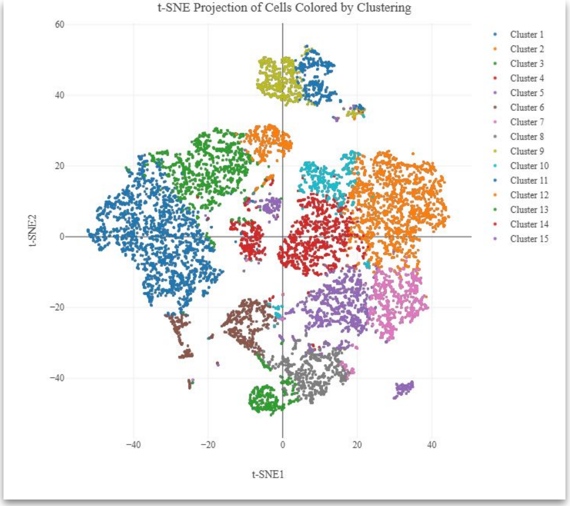


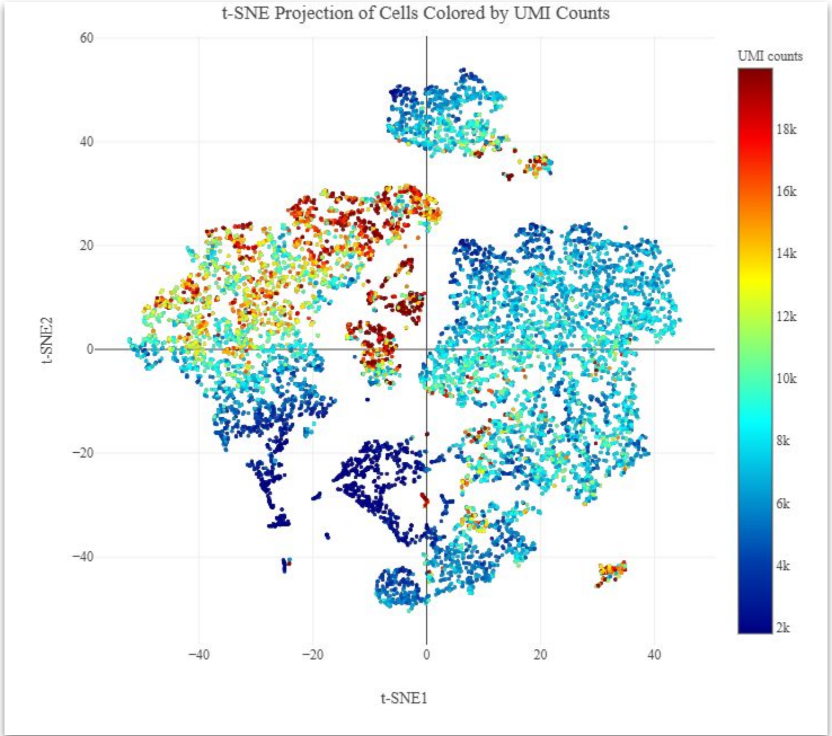
Dataset: PBMC Gene Expression

- Single-cell gene classification task
- High-dimensional, naturally structured data
- Structure matters for accurate classification
- Used in downstream BERT classifier

Label	Count	Percentage
1	1970	19.33%
2	1540	15.11%
3	967	9.49%
4	878	8.61%
5	753	7.39%
6	630	6.18%
7	606	5.94%
8	534	5.24%
9	468	4.59%
10	384	3.77%
11	353	3.46%
12	328	3.22%
13	278	2.73%
14	259	2.54%
15	246	2.41%
	Total	**10194 samples**

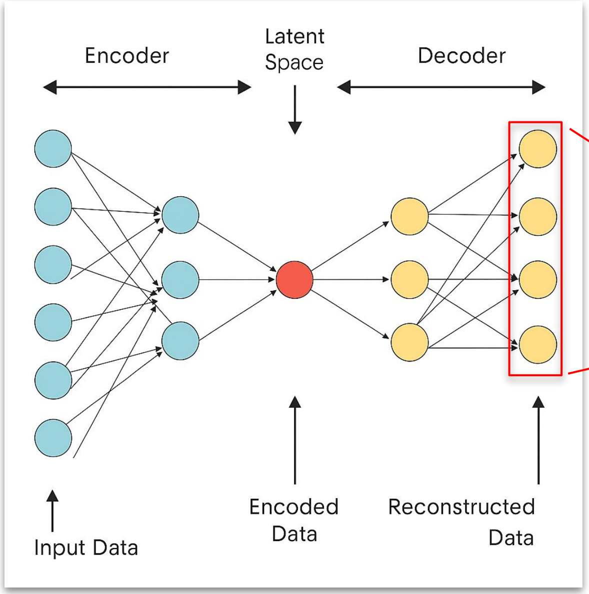
Table 1: Train Set Label Distribution



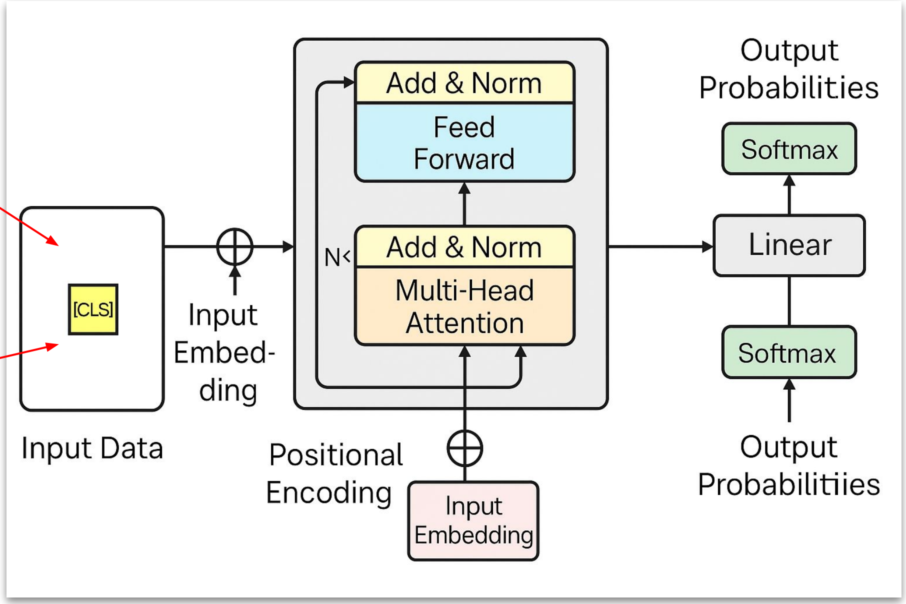


Pipeline Overview: SAE + BERT

- Use BERT-style classifier on latent features + Train SAE to compress gene vectors (→ latent dim 512/256/.../2)
- Track classification performance. But performance alone doesn't explain **why collapse happens**



Sparse Autoencoder

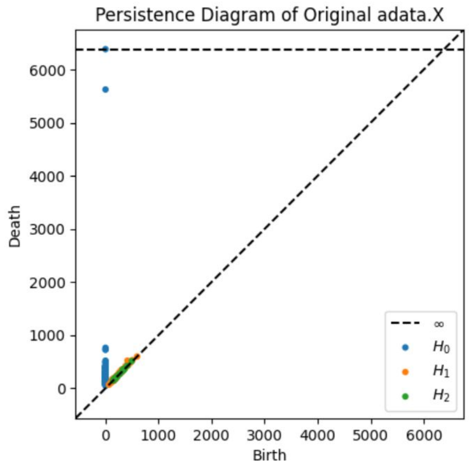


Bidirectional Encoder Representations from Transformers

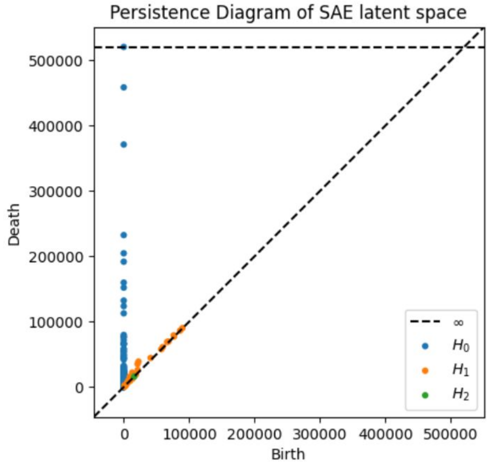
Persistent Diagram Analysis

- We tried using **persistent homology (H_0/H_1/H_2)**
- Misleading behavior:
  → Topological features appear stronger **after collapse**
- Conclusion: PH is sensitive to distortion X

Original
($H0=500$, $H1=218$, $H2=48$)



SAE -> dim=128



dim	H_0	H_1	H_2
2	1000	231	2
4	1000	344	34
8	1000	253	18
16	1000	102	2
32	1000	126	3
64	1000	154	5
128	1000	88	1
256	1000	46	2
512	1000	56	0

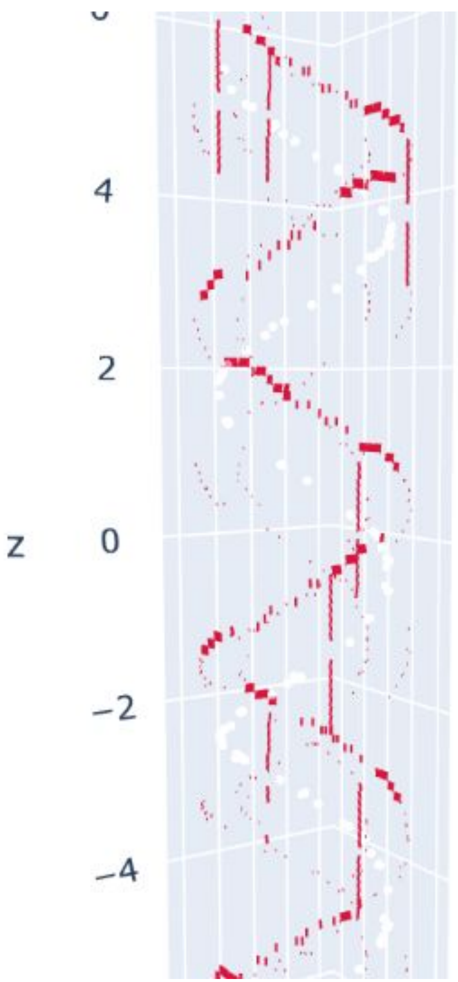
Table 1: Persistence counts at different scales for H_0, H_1, and H_2

Discrete Morse Skeleton

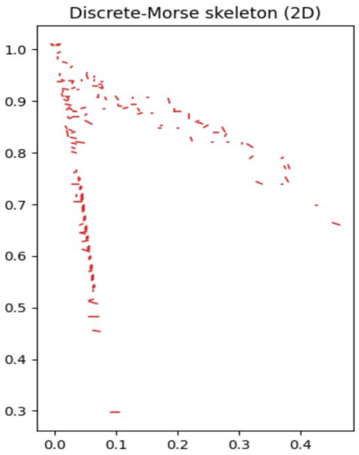
- Construct skeleton from PCA-lowered latent space (dim=2)
- Highlights 1-stable manifolds (i.e., **structural pathways**)
- Persistence simplification filters noise
- Better reflects **true connectivity** in compressed space

Skeleton captures structural collapse

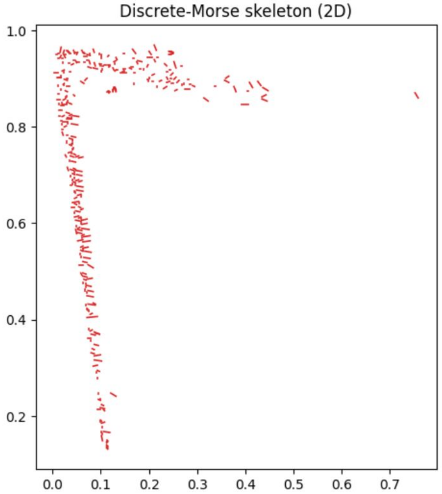
- At high dim (e.g. 512): smooth, coherent skeleton
- At low dim (e.g. 16): disconnected, fragmented skeleton
- Collapse visible as **topological fracture**
- Matches the accuracy cliff in classification



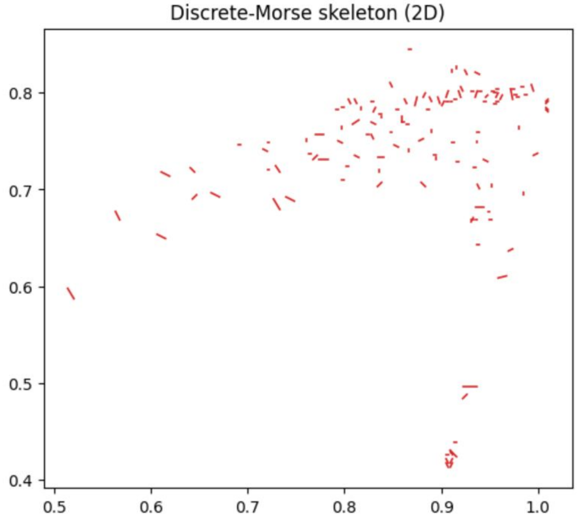
Dim = 128: 323 nodes, 162 edges

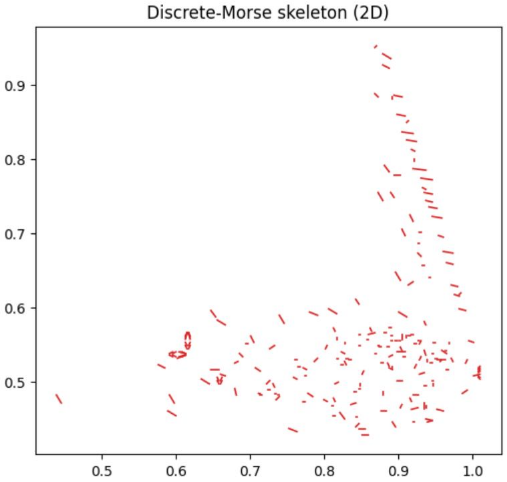


Dim = 64:618 nodes, 315 edges

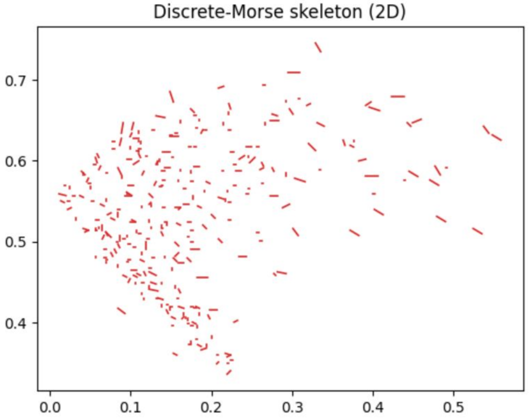


Dim = 32: 248 nodes, 125 edges





Dim = 16: 393 nodes, 202 edges



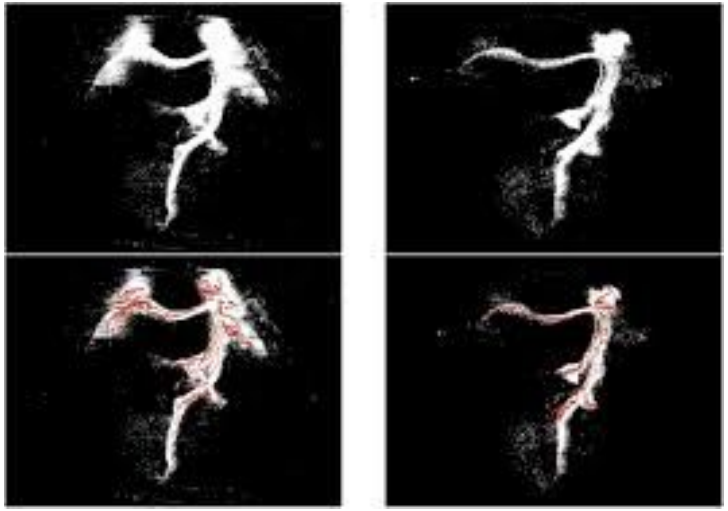
Dim = 8: 563 nodes, 286 edges

Dim	Train Loss	Val Acc	Test Acc
128	1.254	0.538	0.538
64	1.488	0.474	0.487
32	1.742	0.393	0.389
16	2.077	0.271	0.273
8	2.484	0.204	0.206

Performance of CellBERT with varying latent dimensions

Conclusion

- DMS provides interpretable signal of latent degradation
- Better than persistent diagrams in nonlinear AE settings
- Can serve as a **factor for structure loss**
- Future: Integrate DMS into training, adaptive SAE dim selection



Reference

[1] T. K. Dey, J. Wang, and Y. Wang. Graph Reconstruction by Discrete Morse Theory. *arXiv preprint arXiv:1803.05093*, 2018. Available at: https://arxiv.org/abs/1803.05093.

[2] M. Moor, M. Horn, B. Rieck, and K. Borgwardt. Topological Autoencoders. *arXiv preprint arXiv:1906.00722*, 2019. Available at: https://arxiv.org/abs/1906.00722.

[3] Y. Li, A. Potapenko, M. Raghu, and A. Achille. The Geometry of Concepts: Sparse Autoencoder Feature Structure. *arXiv preprint arXiv:2410.19750*, 2024. Available at: https://arxiv.org/abs/2410.19750.

[4] T. Gebhart, P. Schrater, and A. Hylton. Characterizing the Shape of Activation Space in Deep Neural Networks. *arXiv preprint arXiv:1901.09496*, 2019. Available at: https://arxiv.org/abs/1901.09496.

[5] L. Alessandri, F. Cordero, M. Beccuti, N. Licheri, M. Arigoni, M. Olivero, M. F. Di Renzo, A. Sapino, and R. Calogero. Sparsely-connected Autoencoder (SCA) for single cell RNAseq data mining. *npj Systems Biology and Applications*, **7**, Article number: 1, 2021. Available at: https://www.nature.com/articles/s41540-020-00162-6.

[6] 10x Genomics. PBMCs from a Healthy Donor 3 Gene Expression. *10x Genomics Dataset Resource*, 2021. Available at: https://www.10xgenomics.com/resources/datasets.

[7] W. Wang, F. Yang, Y. Fang, D. Tang, J. Huang, H. Lu, and J. Yao. scBERT: a Large-scale Pretrained Deep Language Model for Cell Type Annotation of Single-cell RNA-seq Data. *bioRxiv*, 2021. Available at: https://doi.org/10.1101/2021.12.05.471261.