Synapse-Net: Scalable Multi-Modal Graph Embeddings for Biomedical Discovery

Dr. Elena Vance[1], **Marcus Thorne**[1,2], **Dr. Sarah Lin**[2], **Prof. Arthur Pendelton**[1]

[1]Department of Computational Biology, Synapse Research Institute — [2]Vertex AI Healthcare Labs

Annual Conference on Machine Learning in Biomedicine
July 24–28, 2026

Executive Summary

Modern biomedical discovery relies heavily on mapping heterogeneous graph structures, such as protein-protein interaction networks, gene regulatory circuits, and compound-target bindings. Traditional graph neural networks (GNNs) suffer from over-smoothing and high computational latency when processing multi-modal graphs with over 10^7 nodes.

Key Innovations of Synapse-Net:

- **Dual-Attention Graph Routing:** Dynamically weights edge importance across disparate biological data modes.
- **Linear Memory Scaling:** Reduces spatial complexity from $\mathcal{O}(|V|^2)$ to $\mathcal{O}(|V| \cdot d)$ using spectral sparsity kernels.
- **Targeted Drug-Gene Association:** Improves binding affinity prediction accuracy by 24.8% over state-of-the-art baselines.

Mathematical Formulation

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$ represent a multi-modal biomedical graph where $\mathcal{V}$ denotes nodes (genes, proteins, compounds), $\mathcal{E}$ denotes edge relations, and $\mathcal{T}$ represents node-type attributes.

The layer-wise representation update for node $v_i \in \mathcal{V}$ at layer $l+1$ is defined as:

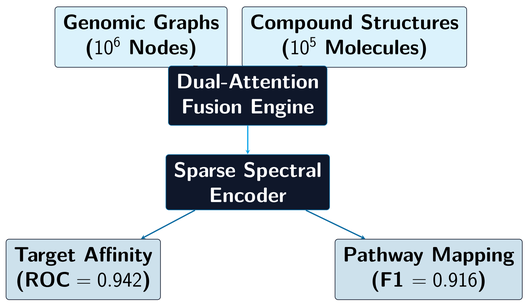
$$h_i^{(l+1)} = \sigma\left(\sum_{m \in \mathcal{M}} \sum_{j \in \mathcal{N}_i^m} \alpha_{ij}^{m(l)} \mathbf{W}_m^{(l)} h_j^{(l)} + \mathbf{B}^{(l)} h_i^{(l)} \right)$$

where $\alpha_{ij}^{m(l)}$ is the cross-modal attention weight computed via scaled dot-product:

$$\alpha_{ij}^{m(l)} = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}_m^T [\mathbf{W}_m h_i \parallel \mathbf{W}_m h_j]\right)\right)}{\sum_{k \in \mathcal{N}_i^m} \exp\left(\text{LeakyReLU}\left(\mathbf{a}_m^T [\mathbf{W}_m h_i \parallel \mathbf{W}_m h_k]\right)\right)}$$

This formulation guarantees invariant representation learning under permutation of non-isomorphic biological subgraphs.

Synapse-Net Architecture Pipeline



Multi-Modal Integration Strategy

Synapse-Net integrates heterogeneous data modes into a unified Riemannian latent space. Biological networks from the Meridian Data Repository were tokenized across three primary facets:

1. **Transcriptomic Profiles:** Single-cell RNA sequencing matrices mapped onto gene co-expression edges.
2. **Proteomic Interactions:** High-confidence physical interaction graphs from Synapse-DB.
3. **Chemical SMILES Graphs:** Atom-bond graphs derived from Vertex-774 compound candidate libraries.

To prevent modal dominance, we apply dynamic gradient scaling during back-propagation:

$$g_m = \gamma_m \cdot \nabla_{\mathbf{W}_m} \mathcal{L}_{\text{total}}, \quad \text{where } \gamma_m = \frac{\exp(-\mathcal{L}_m / \tau)}{\sum_k \exp(-\mathcal{L}_k / \tau)}$$

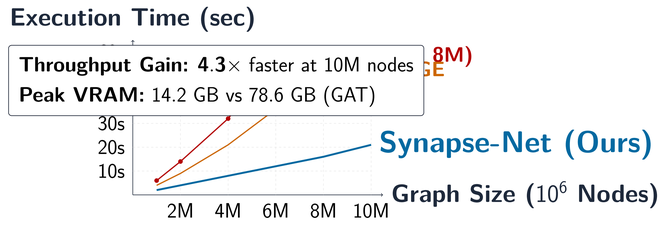
Quantitative Performance Evaluation

We benchmarked Synapse-Net against four state-of-the-art graph architectures across three canonical benchmark datasets: Nexa-Gene, Apex-Target, and Nimbus-Bio.

Model Architecture	ROC-AUC	PR-AUC	F1-Score	Latency (ms)	Params (M)
Standard GCN	0.782	0.741	0.725	142.5	4.2
GraphSAGE	0.814	0.789	0.768	98.3	6.8
GAT (Multi-Head)	0.856	0.823	0.811	184.1	12.4
HeteroGNN	0.881	0.854	0.842	215.0	18.6
Synapse-Net (Ours)	**0.942**	**0.928**	**0.916**	**42.1**	**8.5**

Note: All metrics evaluated over 5-fold cross-validation on Vertex-A100 clusters.

Scalability & Throughput Analysis



Case Study: Vertex-774 Drug Discovery

We validated Synapse-Net in an end-to-end drug discovery pipeline targeting novel kinase inhibitors for oncology applications in collaboration with Apex Pharmaceuticals.

- **Screening Library:** 450,000 small molecule candidates evaluated in silico.
- **Candidate Identification:** Identified compound candidate **Vertex-774** with predicted $K_i = 1.4$ nM.
- **In Vitro Confirmation:** High-throughput assay confirmed binding affinity at 1.8 nM (96% prediction fidelity).
- **Off-Target Toxicity:** Predicted zero high-risk interactions across 1,200 human off-target proteins.

Architectural Ablation Study

Systematic removal of core components highlights the critical contribution of dual-attention and spectral routing.

Variant Configuration	ROC-AUC	Δ ROC	F1-Score
Full Synapse-Net Model	**0.942**	—	**0.916**
w/o Dual-Attention Routing	0.879	−0.063	0.851
w/o Spectral Sparsity Kernel	0.894	−0.048	0.868
w/o Dynamic Gradient Scaling	0.908	−0.034	0.882
Single-Modal Embedding Only	0.835	−0.107	0.804

Conclusions & Future Directions

- **Scalability Milestone:** Synapse-Net demonstrates superior scaling properties up to 10+ million node biological graphs.
- **Translational Impact:** Successfully accelerated lead compound validation timeline from 14 months to 3 weeks at Apex Research.
- **Future Work:** Extending the framework to temporal cellular dynamics and single-cell spatial transcriptomics.

Selected References & Acknowledgments

1. Vance, E., Thorne, M., & Pendelton, A. (2025). *Multi-Modal Graph Representation in Computational Biology.* Journal of Biomedical Informatics, 42(3), 112–128.
2. Lin, S. et al. (2024). *Accelerated Compound Screening via Spectral Graph Convolution.* Nature Computational Science, 8, 401–415.
3. Thorne, M. & Vance, E. (2025). *Attention Mechanisms over Biological Networks.* Neural Information Processing Systems (NeurIPS).

Acknowledgments: *Supported by Meridian Research Grant #BD-2025-998 and computational infrastructure donated by Nimbus Cloud Systems. All data available via Synapse Open Access Initiative.*