REVISITING SELF-ORGANIZING MAPS FOR DRUG-DISEASE ASSOCIATION PREDICTION: A GRAPH-BASED APPROACH

Anonymous authors
Paper under double-blind review

1 INTRODUCTION

Computational drug discovery demands robust methods to identify novel drug-disease interactions Chen et al. (2012). Given a drug-disease interaction matrix $Y \in \mathbb{R}^{n \times m}$, where $\mathcal{A}_{ij} = 1$ indicates an interaction, traditional models often lack interpretability and scalability Ching et al. (2018). We propose a hybrid framework that leverages Graph Neural Networks (GNNs) and Self-Organizing Maps (SOMs). GraphSAGE is employed for link prediction on a heterogeneous biomedical graph, while SOM projects high-dimensional embeddings onto a 2D lattice to reveal latent biological relationships.

2 METHODOLOGY

Our approach integrates heterogeneous graph learning with topological mapping in three phases (Algorithm 1).

We define a biomedical network $G = (V, E)$ where nodes (e.g., genes, proteins, diseases, drugs) are initialized with features:

$$h_v^0 = X_v. \tag{1}$$

GraphSAGE Hamilton et al. (2017) computes node embeddings by aggregating features from sampled neighbors:

$$h_v^{(k)} = \sigma\Big(W_k \cdot \text{MEAN}\big(\{h_u^{(k-1)} : u \in N(v)\}\big)\Big), \tag{2}$$

with W_k as trainable weights.

SOM Pasa et al. (2022) projects the embeddings onto a 2D hexagonal lattice. Neuron weights update via:

$$w_i(t+1) = w_i(t) + \alpha(t)\,\eta(i, i^*, t)\,(h_v - w_i(t)), \tag{3}$$

where $\alpha(t)$ is the decaying learning rate and $\eta(i, i^*, t)$ defines neighborhood influence.

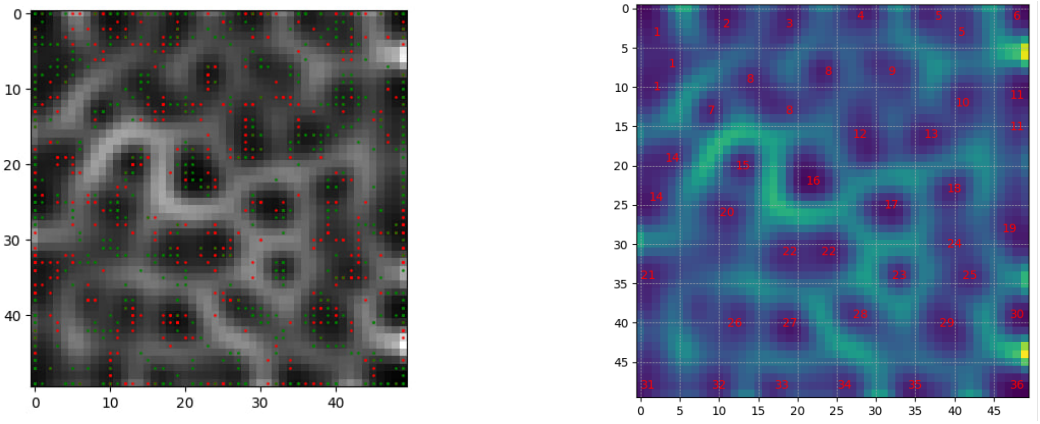
Algorithm 1 Biomedical Graph Analysis Pipeline

1: **Input:** Entities V, relationships E, features X
2: **Phase 1:** Build $G = (V, E)$; initialize $h_v^0 \leftarrow X_v$
3: **Phase 2:** For each epoch, sample $N(v)$ and update embeddings via GraphSAGE; optimize link prediction loss $L = \text{BCE}(S(u, v), y_{\text{true}})$
4: **Phase 3:** Initialize SOM grid; assign nodes to best-matching neurons; update weights using the SOM rule.
5: **Output:** A 2D map of biomedical relationships.

3 RESULTS AND EXPERIMENTS

Using the dataset from Liang et al. (2017), Figure 1 presents SOM of drug-disease interactions. In Figure 1a, red dots denote diseases and green dots represent drugs, clustered according to their

interaction profiles. This spatial organization reflects inherent biological relationships, thereby facilitating the identification of potential drug repurposing opportunities and novel therapeutic targets. Figure 1b illustrates the SOM trained on high-dimensional embeddings, with the background grayscale U-Matrix indicating similarity among neighboring nodes. Darker regions correspond to areas of high similarity, while lighter areas delineate cluster boundaries. This structured representation enhances biomarker discovery and disease classification, underscoring its utility in computational biology.



(a) Drug-Disease Associations.

(b) SOM Cluster Centers and Topology.

Figure 1: SOM Clustering Results

4 Conclusion

Our hybrid framework advances computational drug discovery by integrating interpretable graph embedding with Graph Neural Network and Self-Organizing Map methods.

Meaningfulness Statement

Our work leverages SOM to preserve topological relationships in high-dimensional biological data, ensuring smooth cluster transitions and biologically relevant proximities. This structured representation enhances interpretability, aiding disease classification and biomarker identification. Our findings demonstrate SOM's effectiveness in learning meaningful representations, improving biological data analysis, and supporting translational research in computational biology.

References

Xing Chen, Ming-Xi Liu, and Gui-Ying Yan. Drug–target interaction prediction by random walk on the heterogeneous network. *Molecular BioSystems*, 8(7):1970–1978, 2012.

Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of the royal society interface*, 15(141):20170387, 2018.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Xujun Liang, Pengfei Zhang, Lu Yan, Ying Fu, Fang Peng, Lingzhi Qu, Meiying Shao, Yongheng Chen, and Zhuchu Chen. Lrssl: predict and interpret drug–disease associations based on data integration using sparse subspace learning. *Bioinformatics*, 33(8):1187–1196, 2017.

Luca Pasa, Nicolò Navarin, and Alessandro Sperduti. Som-based aggregation for graph convolutional neural networks. *Neural Computing and Applications*, 34(1):5–24, 2022.