CS768 Project

Study on *Exact Representation with Symmetric Nonnegative Embeddings (NeurIPS 2023)*

Aryan Katiyar (22b1205), Aditya Agrawal (22b0955)
Kunal Chaudhari (22b1060), Navnit Kumar (22b0936)

November 16, 2025

Contents

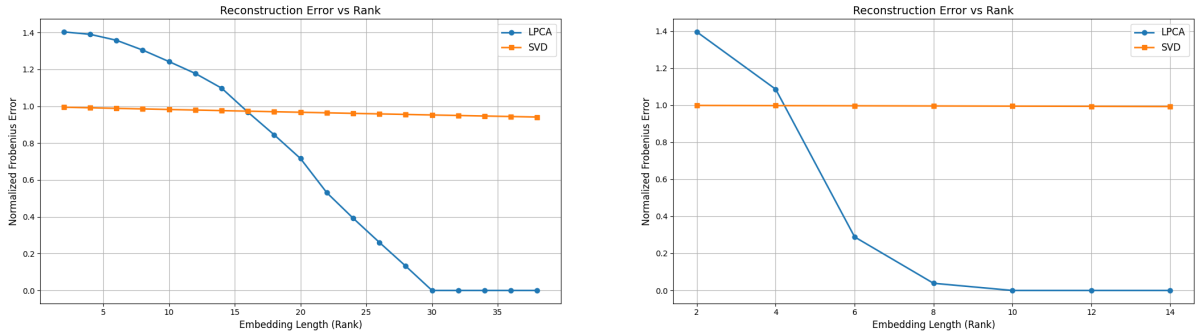
Code Repository: The code used in this work is also available at `https://github.com/cesiumstaar/Symmetric-Nonnegative-Embeddings-Empirical-Evaluation-and-Analysis`.

1 Introduction

We are primarily interested in finding low-dimensional embeddings for sparse graphs. which commonly arise in real-world networks such as social graphs.The paper introduces a novel approach for learning such embeddings that not only enable exact graph reconstruction but also capture structural properties like communities, thereby enhancing interpretability. We investigate the effectiveness and limitations of the $BB^\top$–$CC^\top$ model proposed in this paper. We compare this approach with existing methods such as SVD, BigClam, and LPCA. We also evaluate the impact of different link functions on the quality of the representation.

2 Related Work and Insights from Prior Literature

The work by Seshadhri(2020)[4] formally proves the limitations of dot-product-based models in representing triangle-rich networks. It establishes that graphs with high triangle density and predominantly low-degree nodes require a rank of at least $\Omega\left(\frac{n}{\log^2 n}\right)$ to achieve accurate low-rank positive semi-definite (PSD) approximations. These structural properties are characteristic of real-world social networks, where a user's friends are often mutual acquaintances, leading to many triangles, yet the graph remains sparse due to low average degree. This result highlights the inadequacy of simple linear models like SVD and underscores the necessity for more expressive nonlinear models, such as LPCA and DSNE[1] The NeurIPS 2020 paper *Node Embeddings and Exact Low-Rank Representations of Complex Networks* [1] counters the pessimistic view that sparse, triangle-rich graphs defy low-rank modeling by using a non-linear linking function. An important result of this paper is that for any graph with maximum degree c, there exist embeddings $X, Y \in \mathbb{R}^{n\times(2c+1)}$ such that the adjacency matrix A is exactly recovered as $A = \phi(XY^\top)$, where $\phi(x) = \max(0, \min(1, x))$ is applied entry-wise[2]. Our experiments support the main result of Seshadhri et al. (2020): dot-product low-rank



(a) G1 Frobenius Error

(b) G2 Frobenius Error

	Nodes	Edges	Triangles	Max deg	Avg deg	Arboricity
G1	2000	11967	4230	28	11.97	6
G2	3000	4000	1000	3	2.67	2

Figure 1: Frobenius error comparison and graph statistics for graph types G1 and G2.

models like SVD cannot accurately represent triangle-rich graphs. We evaluated the normalized Frobenius error for both SVD and LPCA on synthetic graphs with high triangle density for two different family of graphs[3]. LPCA, which minimizes a softplus loss over edge probabilities, significantly outperformed SVD as seen in Figure 1

BIGCLAM [5] models overlapping communities using non-negative node embeddings, with edge probabilities defined as $A_{ij} \sim 1 - \exp(-F_i^\top F_j)$. Scales well and optimizes via gradient ascent. However, since it

[1]Difference of Symmetric Nonnegative Embeddings, this is a made-up name used to refer the $BB^\top - CC^\top$ in our paper [2]. We will use this name futher as well to reference this method

[2]this bound is not tight in practice as stated in the paper[1]

[3]Refer to Experiment 1 of code for more details.

assumes homophily through dot-product affinity, it struggles in settings with heterophilic interactions, failing to capture cross-community edges effectively as seen in Figure 2[4]. The DSNE method takes motivaion

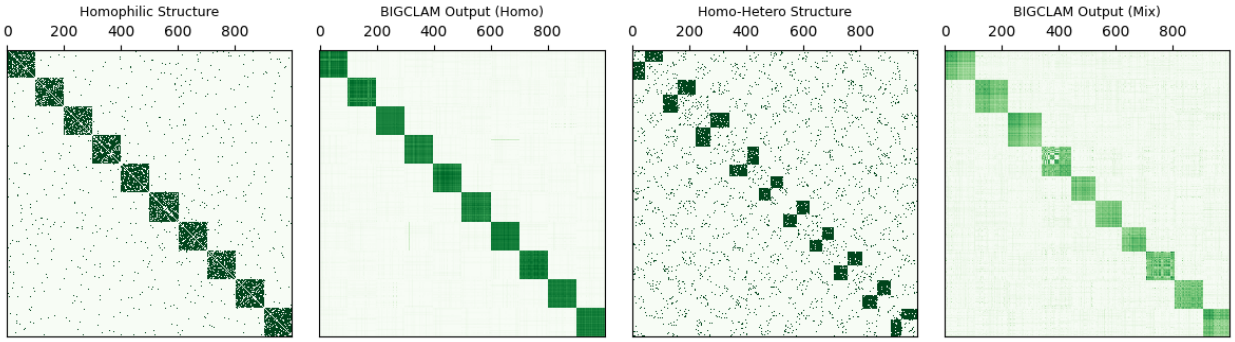


Figure 2: BigClam performance under different graph structures. Left: homophilic-only structure and its BigClam reconstruction. Right: homo-hetero mixed structure and corresponding BigClam output.

from the Attraction-Repulsion (AR) decomposition method[3], reconstructs a symmetric adjacency matrix Adj using a difference of two symmetric PSD matrices: $Adj \approx AA^\top - RR^\top$, where A captures "attractive" interactions and R captures "repulsive". Since the graphs we are dealing with dont contain self-loops, we can ignore the diagonal entries of the reconstructed adjacency matrix. The key idea behind this model is to perturb the diagonal of the original adjacency matrix such that the resulting matrix admits a lower-rank representation. This is achieved by minimizing its nuclear norm[5] using gradient-based optimization.Once we obtain the diagonally perturbed matrix with minimized nuclear norm, we perform an eigendecomposition to express it as a difference of two symmetric positive semidefinite matrices. Specifically, we split the matrix into two parts: a positive component $AA^\top$ and a negative component $RR^\top$, such that the reconstructed adjacency matrix is given by $AA^\top - RR^\top$. This is analogous to performing a symmetric SVD, but separating the attractive and repulsive components of the graph structure.

To evaluate its empirical effectiveness, we compared AR decomposition with standard SVD across varying target ranks k. Figure 3 shows the normalized Frobenius reconstruction error on a synthetic graph composed of 10 disjoint cliques of size 100.[6] As expected, once the rank hits 10 (the number of cliques), AR achieves a near-zero error, while SVD still lags slightly behind. However, when applied to a stochastic block model (SBM) with moderate inter-community noise ($p_{\text{in}} = 0.8$, $p_{\text{out}} = 0.05$)[7], the AR model shows little to no advantage over SVD (Figure 4). The reconstruction error curves nearly overlap across all ranks, revealing a key limitation.While the attraction-repulsion model is conceptually appealing, its empirical performance does not significantly surpass that of standard SVD. In dense structures like cliques and SBM, the leading singular values dominate, allowing SVD to already capture the essential information in the adjacency matrix. As a result, the AR decomposition yields only marginal improvement over SVD in such settings.

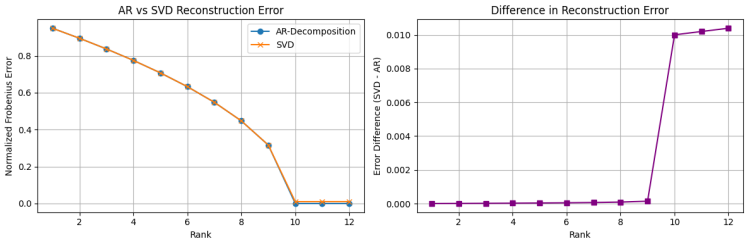


Figure 3: Comparison of AR and SVD on disconnected cliques.

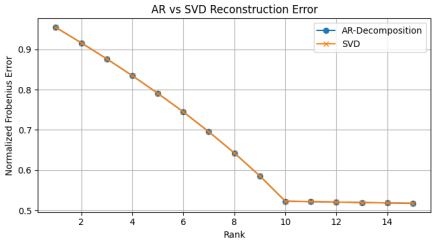


Figure 4: AR vs SVD on SBM

[4]Refer to experiment 2 of code for more details

[5]The nuclear norm of a matrix M, denoted $\|M\|_*$, is the sum of its singular values: $\|M\|_* = \sum_i \sigma_i(M)$. It serves as a convex proxy for matrix rank and is widely used in low-rank approximation problems.

[6]Refer to Experiment 3 of code

[7]Refer to Experiment 4 of code

3 Interpreting Embeddings in DSNE

The DSNE model factorizes the adjacency matrix using symmetric nonnegative embeddings B, C as:

$$A_{ij} = H\left((BB^\top)_{ij} - (CC^\top)_{ij}\right)$$

where $H(\cdot)$ denotes the Heaviside step function[8].In this formulation, $BB^\top$ models attraction (homophily), while $CC^\top$ captures repulsion (heterophily). The DSNE model predicts an edge between two nodes when the attractive force between them outweighs the repulsive force, this is analogous to the AR model.[3] To better understand these embeddings, we visualize the softmax-transformed rows of B and C produced on the homo-hetero graph in Figure 2. The class with maximum softmax value in B can be interpreted as the dominant community a node belongs to. Similarly, C's distribution highlights which repulsive dimension (if any) is most activated for a node.

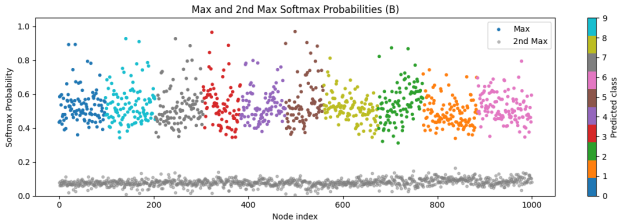


Figure 5: Top-2 softmax probabilities per node for B.

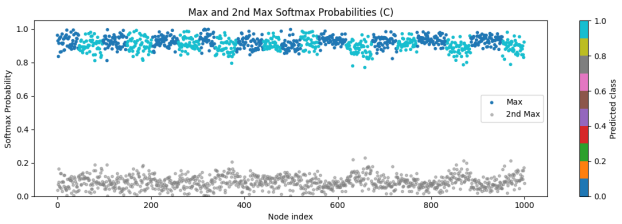


Figure 6: Top-2 softmax probabilities per node for C.

Limitations of DSNE Interpretability

We observe notable interpretability challenges in graphs with unbalanced structures. We first applied DSNE to a graph with communities of varying sizes (Experiment 6). As seen in Figures 7 and 8, nodes within larger blocks exhibit diffuse softmax activations across embedding dimensions, resulting in unclear or ambiguous community assignments.We then tested DSNE on a graph with uneven heterophily across communities (Experiment 7). Figures 9 and 10 show that interpretability goes for a toss. The adjacency matrices corresponding to these experiments are included in appendix 12 and 13. On varing both we get absolutely no interpretabilty, as seen in Exerpiment 8.

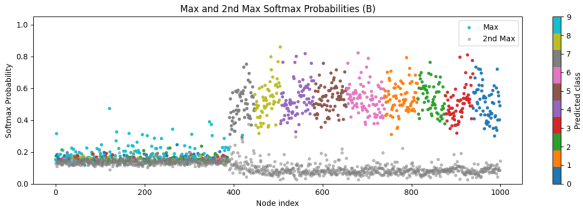


Figure 7: Softmax probabilities from B

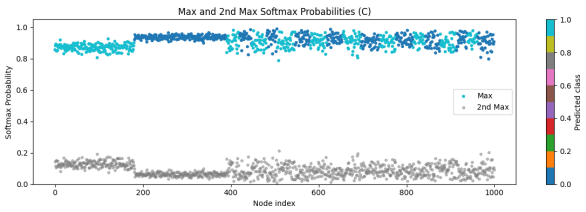


Figure 8: Softmax probabilities from C.

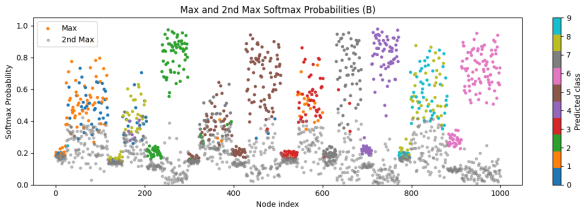


Figure 9: Softmax from B

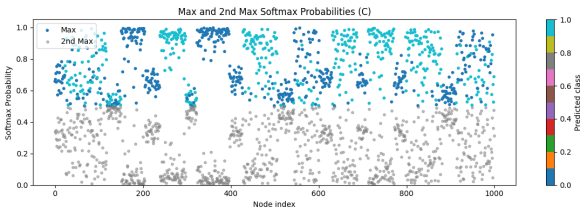


Figure 10: Softmax from C

[8]The Heaviside step function is defined as $H(x) = 1$ if $x > 0$, and $H(x) = 0$ otherwise. It serves as a hard threshold to determine whether node pair (i, j) is connected, based on whether the attractive score exceeds the repulsive score.

4 Abalation Experiments

Ablation of the Non-Linear Linking Function. We perform an ablation in which the non-linear linking function is removed. Instead, we train the embedding matrices using a margin-based objective formulated via hinge loss:

$$\text{score}_{ij} = \mathbf{b}_i^\top \mathbf{b}_j - \mathbf{c}_i^\top \mathbf{c}_j,$$

where $\mathbf{b}_i$ and $\mathbf{c}_i$ are the respective rows of the embedding matrices $\mathbf{B} \in \mathbb{R}^{n \times k_B}$ and $\mathbf{C} \in \mathbb{R}^{n \times k_C}$. We then optimized a hinge loss over these scores:

$$\mathcal{L}_{\text{hinge}} = \sum_{(i,j) \in \mathcal{E}} \max(0, 1 - \text{score}_{ij}) + \sum_{(i,j) \notin \mathcal{E}} \max(0, 1 + \text{score}_{ij}),$$

where $\mathcal{E}$ is the set of observed edges, and $(i, j) \notin \mathcal{E}$ denotes non-edges.
To prevent overfitting and control the scale of the embeddings, we added regularization as well.

Enhancing Interpretability via Softmax Constraints. We sought to enhance the interpretability of the learned embeddings by imposing a probabilistic constraint on the embeddings by applying a row-wise softmax transformation to the raw embedding vectors:

$$\mathbf{b}_i = \text{softmax}(\mathbf{b}_i^{\text{raw}}), \quad \mathbf{c}_i = \text{softmax}(\mathbf{c}_i^{\text{raw}}),$$

thereby ensuring that each embedding vector lies in the probability simplex. Since the embeddings are distributions now, visualization becomes easy.

Visualization of Embedding Results on a Homophilic- Heterophilic Graph

Following are the results obtained by applying the approaches to a synthetic graph with 1000 nodes and 10 homophilic communities and 2 heterophilic communities within a single large community.

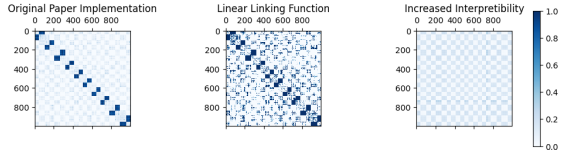


Figure 11: Visualizations of learned embeddings on the synthetic graph.

Ablation Study Conclusions

Our ablation study reveals several key insights:

- **Performance of New Method:** Replacing DSNE with a hinge loss results in good results. This is evident in Figure 11, where the hinge-based embeddings exhibit well-separated clusters. However, it has learnt some erratic patterns within the non-community sparse edges compared to smooth boundaries of DSNE. We haven't been able to provide theoretical guarantees on the ranks of k_b and k_c for which such a method works, but it appears highly competitive.
- **Frobenius Norm:** This directly questions whether frobenius norm is a good loss to consider for low rank decompositions. The best low rank decomposition we can get is an SVD according to the frobenius norm but as we saw SVD just simply ignores some of the communities totally, so frobenius norm may not be the best way to tackle the optimization problem

- **Sample Based Training:** A new problem faced while training sparse graphs with hinge loss is fitting the embeddings to be simply zero, due to the sparse graph. This is overcome by a simple negative, positive sampling strategy. This is not evident in the original method due to extremely high gradient penalty for total misclassifications
- **Interpretability Trade-off:** This fails miserably as is evident in Figure 11. It perhaps means that the system here has been overconstrained, and that the claim of the authors where the matrices represent affinity to a community is suspect at best

5 Conclusions and Possible Further Work

Through both theoretical and empirical investigations, we identified several critical issues with the claims and methodology in this paper. While the idea of interpretable embeddings sounds nice, it fails to hold any meaning on real-world graphs, beyond which this paper does not provide anything new

- **Questionable Contributions:** Our ablation experiments—varying community size and measuring arboricity suggests that the original contribution may be overstated. No example of theoretical bound on arboricity in the paper being tight, or the model being slightly mathematically interpretable, was conclusively provided in the original paper
- **Arboricity Bound Limitations:** One of the paper's central claims relates embedding dimensionality to the arboricity of the input graph. The proof uses extremely loose bounds, and for most instances, the graph was representable in linear arboricity itself. We tried to devise algorithms so that an α arboricity graph could be represented as a matrix such that the upper triangular matrix itself has indegree being less than alpha. This is a great theoretical direction to work in.
- **Constructing Hard Instances:** We explored whether there exist graphs that fundamentally require high-dimensional embeddings. We tried to theoretically explore what makes representing a graph hard. Many of the cliques or homophilic heterophilic graphs were easily representable in low dimension owing to their regular structure.
- **Limitations of Frobenius Error:** Throughout our study, we observed that Frobenius norm is a blunt tool for evaluating embedding quality. As we realized through the embeddings, for some other purposes the hinge based trained embeddings would perhaps be a better choice.

6 Contributions

- Aryan: Implemented the various embedding algorithms, plotted the nuclear norm minimization and prepared references and footnotes for the report
- Aditya: Designed graph generators, abalation code and came up with ideas to perform the abalation study, performed theoretical analysis
- Kunal: Conducted ablation tuning, plotted graphs and interpretibility analysis for varying sized communities
- Navnit: Compiled latex report, prepared well commented code and suggested and found other papers performing similar embeddings

Appendix

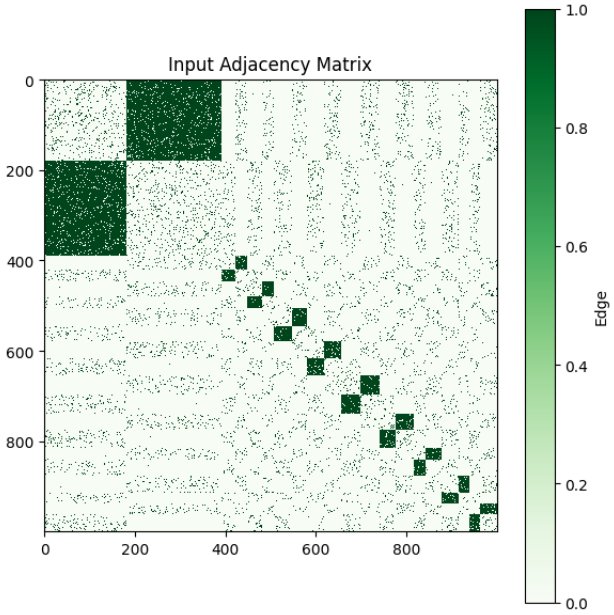


Figure 12: Adjacency matrix used for DSNE interpretability experiment with uneven homophilic community sizes.

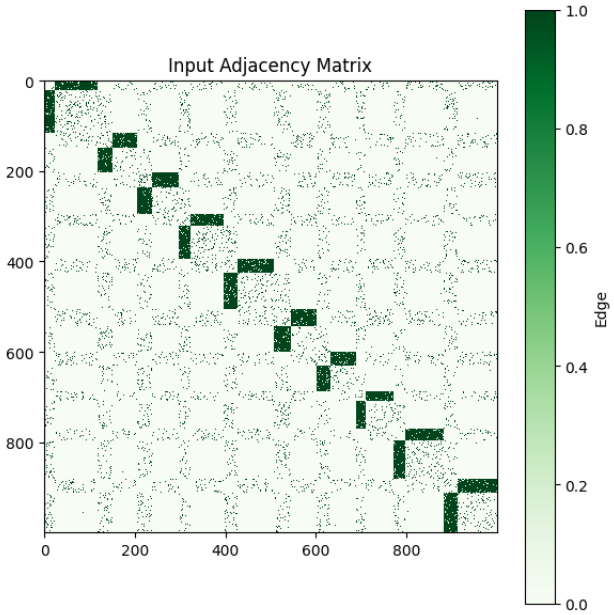


Figure 13: Adjacency matrix used for DSNE interpretability experiment with uneven heterophilic community sizes.

References

[1] Sudhanshu Chanpuriya, Cameron Musco, Charalampos E. Tsourakakis, and Konstantinos Sotiropoulos. Node embeddings and exact low-rank representations of complex networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 16925–16937, 2020.

[2] Sudhanshu Chanpuriya, Ryan A Rossi, Anup Rao, Tung Mai, Nedim Lipka, Zhao Song, and Cameron Musco. Exact representation of sparse networks with symmetric nonnegative embeddings. In *Advances in Neural Information Processing Systems*, 2023. NeurIPS.

[3] Alexander Peysakhovich, Anna Klimovskaia Susmelj, and Leon Bottou. Pseudo-euclidean attract-repel embeddings for undirected graphs, 2023. Preprint.

[4] C. Seshadhri, Aneesh Sharma, Andrew Stolman, and Ashish Goel. The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*, 117(11):5631–5637, 2020.

[5] Jaewon Yang and Jure Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.