ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3315-6701-9/26/$31.00 ©2026 EU | DOI: 10.1109/ICASSP55912.2026.11460800

ENCODER-DECODER SYMMETRIC NONNEGATIVE MATRIX TRI-FACTORIZATION FOR GRAPH CLUSTERING

Amjad Seyedi and Nicolas Gillis
Department of Mathematics and Operational Research, University of Mons, Belgium

ABSTRACT

Symmetric nonnegative matrix tri-factorization (SNMTF) provides an interpretable, parts-based representation for graph data, where node memberships and cluster interactions jointly expose the community structure. Yet many SNMTF variants are decoder-only, reconstructing the graph without guaranteeing the latent variables are recoverable, which can yield drifting embeddings and unstable memberships. We introduce an encoder–decoder SNMTF (ED-SNMTF) that couples graph reconstruction with an encoder-consistency principle: the interaction matrix must be inferable by projecting the graph onto the learned membership space. This coupling induces implicit orthogonality: membership columns repel through the consistency term, so crisp clusters emerge without explicit orthogonality penalties or post hoc k-means. The optimization of the parameters of our model relies on GPU-accelerated computing with CUDA and is trained using the Adam optimizer, with a nonnegative projection step to guarantee feasibility and promote stable convergence. We evaluate ED-SNMTF on real-world networks and show that it outperforms both traditional baselines and recent deep methods.

***Index Terms*—** low-rank matrix factorization, encoder-decoder representation, graph clustering, Adam.

1. INTRODUCTION

Nonnegative Matrix Factorization (NMF) has emerged as a powerful method for part-based representation in data analysis [1–3]. Beyond its role in data analysis, it finds use in machine learning for reducing dimensionality and identifying meaningful features [1], in data mining for tasks such as community detection [4] and recommender systems [5], and in signal processing for separating mixed signals [6] and noise reduction [7]. Its strength lies in producing compact yet interpretable representations across these diverse applications. NMF approximates a nonnegative data matrix $\mathbf{X} \in \mathbb{R}_+^{m\times n}$ by the product of two nonnegative matrices, $\mathbf{W} \in \mathbb{R}_+^{m\times r}$ and $\mathbf{H} \in \mathbb{R}_+^{r\times n}$, where the factorization rank r is smaller than m and n. A standard formulation is to minimize the Frobenius norm of reconstruction error:

$$\min_{(\mathbf{W},\mathbf{H})\geq 0} \|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2, \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm. While basic NMF (1) provides an interpretable, part-based data representation, it does not inherently enforce clustering structure in the learned factors. To address this limitation, several variants have been proposed by introducing additional constraints or structures that encourage the learned representations to be well-separated, most notably orthogonal NMF [8], cluster NMF (also known as convex NMF) [9], and projective NMF (PNMF) [10], which reconstructs the input by its subspace projection as $\mathbf{X} \approx \mathbf{W}\mathbf{W}^\top\mathbf{X}$. These extensions help bridge the gap between NMF as a purely representational method and its use as a clustering technique, though challenges remain when the data exhibit complex relational or graph-structured dependencies, such as in graph clustering.

To address these challenges, various extensions of NMF have been proposed. Some enforce hard or soft orthogonality constraints to steer the factorization toward a clustering-friendly structure. Others, such as symmetric NMF (SNMF) and symmetric nonnegative matrix tri-factorization (SNMTF) [11], directly integrate clustering assumptions by factoring proximity matrices under the premise that similar nodes should be grouped, while dissimilar ones should be separated. These methods have proven more effective for graph clustering by preserving structural relationships in the data. SNMTF factorizes the adjacency (or similarity) matrix of an undirected or directed graph, $\mathbf{A} \in \mathbb{R}_+^{n\times n}$, using two factors, $\mathbf{W} \in \mathbb{R}_+^{n\times r}$ and $\mathbf{S} \in \mathbb{R}_+^{r\times r}$, as follows

$$\min_{(\mathbf{W},\mathbf{S})\geq 0} \|\mathbf{A} - \mathbf{W}\mathbf{S}\mathbf{W}^\top\|_F^2, \tag{2}$$

where $\mathbf{W}$ and $\mathbf{S}$ are cluster membership and interaction matrices, respectively. More specifically, $W_{[i,k]}$ quantifies the membership of node i in cluster k, and $S_{[k,l]}$ represents the interaction strength between clusters k and l.

Over the past decade, substantial efforts have been devoted to enhancing these models through the incorporation of additional constraints [12], regularizations [4, 13, 14], and refined similarity measures [15] that account for high-order relationships in graphs. For community detection, an encoder–decoder NMF [16] was proposed, where the encoder factorization $\mathbf{H} \approx \mathbf{W}^\top\mathbf{A}$ is combined with the standard decoder

The authors acknowledge the support by the European Union (ERC consolidator, eLinoR, no 101085607). Emails: {seyedamjad.seyedi,nicolas.gillis}@umons.ac.be.

factorization $\mathbf{A} \approx \mathbf{WH}$ as follows:

$$\min_{(\mathbf{W},\mathbf{H})\geq 0} \|\mathbf{A}-\mathbf{WH}\|_F^2 + \|\mathbf{H}-\mathbf{W}^\top\mathbf{A}\|_F^2.$$

It is closely related to PNMF [10], since substituting the encoder into the decoder leads to $\boldsymbol{A} \approx \mathbf{W}\mathbf{W}^\top\mathbf{A}$. More recently, following the development of deep MF [17–20], deep graph-based factorizations such as Deep SymNMF (DSNMF) [21] and Deep AsNMF (DAsNMF) [22] have been developed specifically for graph clustering. These deep extensions stack multiple layers of factorizations to capture hierarchical structures in the data. However, existing shallow and deep graph-based methods are predominantly decoder-only: they reconstruct $\mathbf{A}$ from latent variables but do not require that the same variables be recoverable from $\mathbf{A}$. Lacking this autoencoding constraint, the latent space may drift toward solutions that reconstruct well but encode poorly, yielding ambiguous memberships and less stable cluster assignments.

This work introduces Encoder–Decoder SNMTF (ED-SNMTF), a novel formulation that unifies graph reconstruction with latent recoverability by coupling the traditional SNMTF decoder with an encoder consistency term; see Fig. 1. In this autoencoder model, the encoder and decoder refine and verify each other through self-representation to ensure that the latent representation is robust and faithful to the data. We provide theoretical insights showing that ED-SNMTF naturally promotes orthogonality. This leads to a clear separation of nodes into distinct, non-overlapping communities for effective graph clustering. To solve the resulting nonconvex problem, we design an efficient GPU-accelerated optimization scheme that combines the Adam optimizer with nonnegativity projection, supported by a principled initialization strategy. Extensive experiments on eight benchmark networks show that ED-SNMTF consistently outperforms both shallow and deep state-of-the-art factorization methods in clustering quality.

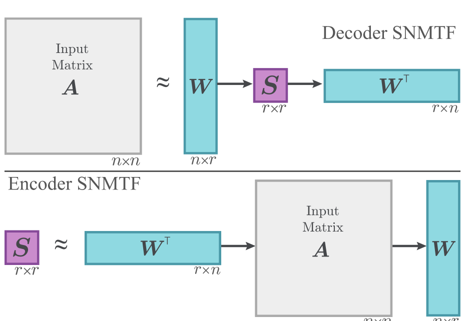

Fig. 1. Illustration of ED-SNMTF. The Decoder SNMTF (top) approximates the input $\mathbf{A}$ by factorizing it into $\mathbf{WSW}^\top$, where $\mathbf{W}$ and $\mathbf{S}$ are cluster membership and interaction matrices, resp. The Encoder SNMTF (bottom) recovers $\mathbf{S}$ by approximating it as $\mathbf{W}^\top\mathbf{AW}$.

2. PROPOSED MODEL: ED-SNMTF

In this section, we introduce ED-SNMTF, an extension of classical SNMTF designed to jointly address graph reconstruction and latent recoverability. Unlike decoder-only variants, ED-SNMTF enforces consistency between the encoding and decoding processes, yielding a more stable latent space and well-separated clusters. We begin by presenting the model formulation along with theoretical insights, followed by a detailed description of the optimization procedure.

2.1. Model Formulation

Extending a decoder-only SNMTF model into an encoder-decoder SNMTF enables both representation and reconstruction. The encoder maps the data into a structured latent space that captures essential patterns and relationships, ensuring the latent variables serve as a meaningful representation rather than just abstract factors. The decoder then reconstructs the original data, confirming that this representation retains enough information for accurate recovery. Together, these steps justify the encoder–decoder design by providing interpretable embeddings and validating them through reliable reconstruction. In the decoder SNMTF (2), node i is represented by a node membership vector $\mathbf{W}_{[i,:]} \in \mathbb{R}_+^r$, and each edge (similarity) $A_{i,j}$ is reconstructed via the cluster-interaction matrix $\mathbf{S} \in \mathbb{R}_+^{r\times r}$ as follows $A_{i,j} \approx \mathbf{W}_{[i,:]}\,\mathbf{S}\,\mathbf{W}_{[j,:]}^\top$. To ensure that the learned latent representation is consistent with the underlying graph structure, the encoder SNMTF projects the input adjacency matrix into a lower-dimensional cluster space, $\mathbf{S}$, as follows

$$S_{k,l} \approx \mathbf{W}_{[:,k]}^\top \mathbf{A}\,\mathbf{W}_{[:,l]} \iff \mathbf{S} \approx \mathbf{W}^\top\mathbf{AW},$$

where $\mathbf{W}_{[:,k]}$ and $\mathbf{W}_{[:,l]}$ are membership vectors for the kth and lth clusters, resp. The product $\mathbf{W}_{[:,k]}^\top\mathbf{A}$ is the weighted sum of the connections from cluster k to all nodes, and multiplying it by $\mathbf{W}_{[:,l]}$ extracts the total affinity between the two clusters. Enforcing consistency between the encoder and decoder factorizations yields ED-SNMTF:

$$\min_{(\mathbf{W},\mathbf{S})\geq 0} \left\|\mathbf{A}-\mathbf{WSW}^\top\right\|_F^2 + \lambda\left\|\mathbf{S}-\mathbf{W}^\top\mathbf{AW}\right\|_F^2, \quad (3)$$

with trade-off parameter $\lambda \geq 0$. Combining the decoder and encoder induces the self-consistency relation

$$\mathbf{W}^\top\mathbf{AW} \;\approx\; \mathbf{W}^\top\mathbf{W}\,\mathbf{S}\,\mathbf{W}^\top\mathbf{W},$$

that is, the encoded $\mathbf{S}$ should also be coherent under the (low-rank) projection induced by $\mathbf{W}$. When $\lambda = 0$, (3) reduces to SNMTF. When $\lambda \to \infty$, the optimal $\mathbf{S}$ satisfies $\mathbf{S} = \mathbf{W}^\top\mathbf{AW}$, and (3) becomes

$$\min_{\mathbf{W}\geq 0} \left\|\mathbf{A}-\mathbf{W}(\mathbf{W}^\top\mathbf{AW})\mathbf{W}^\top\right\|_F^2,$$

a self-consistent decoder that reconstructs $\mathbf{A}$ from its projection onto the cluster subspace, similarly as PNMF does in the non-symmetric case, where $\mathbf{X} \approx \mathbf{WW}^\top\mathbf{X}$.

2.2. Why ED-STNMF enforces orthogonality

In a noiseless scenario, $\mathbf{A} = \mathbf{W}\mathbf{S}\mathbf{W}^\top$ and $\mathbf{S} = \mathbf{W}^\top\mathbf{A}\mathbf{W}$. This means that $\mathbf{S} = \mathbf{W}^\top\mathbf{W}\mathbf{S}\mathbf{W}^\top\mathbf{W}$. By the lemma below, if $\mathbf{A}$ has rank r, this implies that $\mathbf{W}$ is orthogonal, that is, $\mathbf{W}^\top\mathbf{W} = \mathbf{I}_r$, where $\mathbf{I}_r$ is the r-by-r identity matrix. Hence, each node is associated with a single community; that is, ED-STNMF clusters the graph in r disjoint communities.

Lemma 1. *Let* $\mathbf{C} \in \mathbb{R}_+^{r\times r}$ *be symmetric and* $\mathbf{S} \in \mathbb{R}^{r\times r}$ *be such that* $\operatorname{rank}(\mathbf{S}) = r$ *and* $\mathbf{S} = \mathbf{C}\mathbf{S}\mathbf{C}$. *Then* $\mathbf{C} = \mathbf{I}_r$, *up to permutations.*

Proof. Since $\mathbf{S} = \mathbf{C}\mathbf{S}\mathbf{C}$, $\mathbf{S} = \mathbf{C}^q\mathbf{S}\mathbf{C}^q$ for any $q = 1, 2, \dots,$

$$\lambda_{\min}(\mathbf{S}) = \lambda_{\min}(\mathbf{C}^q\mathbf{S}\mathbf{C}^q) = \min_{\|x\|_2=1} x^\top \mathbf{C}^q\mathbf{S}\mathbf{C}^q x$$
$$= \min_{z=\mathbf{C}^q x, \|x\|_2=1} z^\top \mathbf{S} z \le \lambda_{\min}(\mathbf{C})^q \lambda_{\max}(\mathbf{S}),$$

where the inequality follows from taking the unit vector x as the eigenvector of $\mathbf{C}$ corresponding to the smallest eigenvalue of $\mathbf{C}$, so that $\|z\|_2 = \lambda_{\min}(\mathbf{C})^q$. Since $\operatorname{rank}(\mathbf{S}) = r$, $\lambda_{\min}(\mathbf{S}) > 0$, implying that $\lambda_{\min}(\mathbf{C}) \ge 1$ since the above inequality is true for all $q = \{1, 2, \dots\}$. Similarly, we can show that $\lambda_{\max}(\mathbf{S}) \ge \lambda_{\max}(\mathbf{C})^q \lambda_{\min}(\mathbf{S})$, implying that $\lambda_{\max}(\mathbf{C}) \le 1$. Hence $\lambda_{\min}(\mathbf{C}) = \lambda_{\max}(\mathbf{C}) = 1$ so that $\mathbf{C}$ is orthogonal. The only nonnegative orthogonal matrices are the permutation matrices [23]. □

When $\mathbf{C} = \mathbf{W}^\top\mathbf{W}$ for some $\mathbf{W}$ as in our setting, $\mathbf{C}$ can only be the identity matrix, as $\mathbf{C}_{i,i} = \|\mathbf{W}_{[:,i]}\|_2 = 1$ for all i.

2.3. Optimization

ED-SNMTF is a non-convex optimization problem. Unlike conventional NMF methods that typically employ standard optimization strategies including multiplicative updates, alternating minimization schemes, or coordinate descent algorithms, we propose to tackle the optimization challenge by integrating the Adaptive Moment Estimation (Adam) optimizer with nonnegativity constraints. We use Adam because its adaptive steps and momentum improve stability, reduce boundary stagnation, and yield consistently good results under comparable settings. This methodology establishes fresh pathways for efficient and interpretable NMF solutions in complex practical scenarios. Note that ED-SNMTF can also be applied to directed graphs when $\mathbf{A}$ is asymmetric. In fact, we did not explicitly constrain $\mathbf{S}$ to be symmetric. If $\mathbf{A}$ is symmetric (resp. asymmetric), the gradient direction is symmetric (resp. asymmetric), so the updates naturally preserve the structure. To initialize the membership matrix $\mathbf{W}$, we use the method from [12] which has been shown to be effective for SNMTF: it uses smoothed vertex component analysis (SVCA) [24] refined with an orthogonal nonnegative least squares (orthNNLS) step, which promotes column orthogonality of $\mathbf{W}$ (that is, disjoint clusters). This procedure yields a stable and meaningful starting point for the subsequent factorization. Also, we set $\mathbf{S}_0 = \mathbf{I}_r$, which corresponds to strong intra-cluster connections and disconnected clusters. The high-level optimization procedure for ED-SNMTF with the Adam optimizer is presented in Algorithm 1. Source code is publicly available at: https://github.com/amjadseyedi/ED-SNMTF.

Algorithm 1 ED-SNMTF via Adam optimizer

Input: input matrix $\mathbf{A}$, latent dimension r, learning rate η, encoding parameter λ;
Output: ED-SNMTF of $\mathbf{A} \approx \mathbf{W}\mathbf{S}\mathbf{W}^\top$ with $\mathbf{S} \approx \mathbf{W}^\top\mathbf{A}\mathbf{W}$,
Cluster assignments $\hat{c} \in \{1, \dots, r\}^n$;

```
Initialize W via SVCA and orthNNLS [12], S = I_r;
for t = 1 to T do
    Perform one Adam update step on (W, S);
    W ← max(W, 0), S ← max(S, 0);
end for
Obtain ĉ_i = arg max_k(W_[i,k]), for all i ∈ {1, ..., n};
```

3. EXPERIMENTS

We conduct experiments using 13 baseline methods that represent the current state-of-the-art in matrix factorization. Our benchmark encompasses traditional factorizations [1, 8, 10], specialized graph-based methods [4, 11, 13–16], and modern deep methods [19, 20, 22]. This selection provides comprehensive coverage of both conventional shallow methods and contemporary deep frameworks. Clustering quality is evaluated using Normalized Mutual Information (NMI), a standard metric that measures the consistency between predicted clusters and ground-truth classes, ranging from 0 to 1, where 1 indicates a perfect clustering. We implement our method in PyTorch with GPU acceleration on an RTX 4060 laptop GPU, using the Adam optimizer with a learning rate of 0.01.

Datasets. We use eight benchmark datasets. The web networks *Cornell* (195 nodes, 301 edges, 5 classes), *Texas* (187, 309, 5), *Washington* (230, 395, 5), and *Wisconsin* (197, 502, 5) model university webpages and hyperlinks. The *Email* network (1005, 25,571, 42) and *Wiki* network (2405, 16,523, 19) capture communication and social interactions. *CiteSeer* (3312, 4732, 6) and *Cora* (2708, 5429, 7) are citation graphs between research papers. Together, they span diverse scales, classes, and structures for robust evaluation.

Results. Table 1 provides the NMI. For all regularized methods, it reports the best performance using parameters from the predefined set specified in the corresponding papers. Matrix factorization techniques (NMF, PNMF, ONMF) exhibit inconsistent clustering effectiveness, with substantial variation depending on the specific datasets. Enhanced variants like RANMF and RAsNMF show improvements on certain data sets (Texas and Washington), yet fail to provide improvements on all data sets. Deep models (DANMF, SDNMF, DAsNMF) outperform their shallow counterparts, validating the effectiveness of hierarchical feature learning. DAsNMF emerges as the leading deep baseline, with the best or second-

Table 1. NMI scores of different methods. **Bold** indicates the best performance, underline the second best.

Method	Cornell	Texas	Wash.	Wisc.	Email	Wiki	Citeseer	Cora
[1] NMF	0.134	0.147	0.113	0.079	0.574	0.248	0.116	0.273
[10] PNMF	0.120	0.230	0.111	0.061	0.663	0.293	0.158	0.289
[8] ONMF	0.094	0.233	0.159	0.069	0.654	0.281	0.142	0.193
[11] SymNMF	0.163	0.155	0.113	0.064	0.493	0.243	0.126	0.260
[11] SNMTF-rnd	0.145	0.146	0.137	0.068	0.485	0.275	0.146	0.346
[11] SNMTF-SVD	0.155	0.162	0.139	0.073	0.489	0.258	0.146	0.346
[4] BigClam	0.043	0.068	0.063	0.073	0.565	0.254	0.089	0.092
[13] M-NMF	0.046	0.032	0.048	0.086	0.536	0.218	0.047	0.093
[14] RANMF-rnd	0.109	0.130	0.102	0.069	0.572	0.279	0.084	0.155
[14] RANMF-SVD	0.168	0.171	0.170	0.076	0.587	0.285	0.131	0.356
[15] RAsNMF	0.135	**0.305**	0.219	0.073	0.484	0.259	0.148	0.344
[16] NSED	0.075	0.075	0.034	0.068	0.670	0.257	0.146	0.183
[19] DANMF	0.098	0.134	0.082	0.089	0.674	0.289	0.110	0.326
[20] SDNMF	0.106	0.093	0.117	0.067	0.690	0.284	0.123	0.321
[22] DAsNMF	**0.211**	0.256	**0.221**	0.109	0.699	0.300	0.164	0.368
ED-SNMTF (λ=0)	0.132	0.282	0.131	0.086	0.698	**0.307**	0.137	**0.369**
ED-SNMTF-rnd	0.154	0.157	0.153	0.071	0.697	0.297	0.141	0.315
ED-SNMTF	0.197	0.299	**0.221**	**0.121**	**0.723**	**0.307**	**0.167**	**0.369**

best scores on multiple datasets. Nevertheless, these sophisticated models require extensive hyperparameter optimization and are sensitive to configuration choices.

ED-SNMTF delivers a robust performance. It provides the best performance on five datasets (Wisconsin, Email, Wiki, Citeseer, and Cora), ties for best on Washington, and second-best on Cornell and Texas. For the other datasets, performance gaps with leading methods are minimal, indicating consistent algorithmic behavior. Unlike complex deep alternatives, ED-SNMTF achieves competitive results through simplified parameter tuning, leveraging encoder regularization, clever initialization, and the Adam optimization algorithm. In fact, ED-SNMTF performs better than without regularization (ED-SNMTF with $\lambda = 0$), than with random initialization (ED-SNMTF-rnd, which is the average result over 5 random initializations), and than previous SNMTF algorithms (SNMTF-rnd and SNMTF-SVD).

Parameter analysis. We evaluated our model with different values of the regularization parameter λ, trying the values in the set $\{0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$; see Fig. 2. Our tests showed two clear patterns. For large datasets with lots of connections and rich information, like Email, Wiki, Citeseer, and Cora, ED-SNMTF works best for small λ values (close to 0). When λ is too large, it performs much worse because they do not need much regularization. In contrast, for smaller datasets (such as Cornell, Texas, Wisconsin, and Washington), which provide weaker and noisier signals due to limited sample size and sparse connections, ED-SNMTF performs better with larger values of λ. For Cornell, Texas, and Wisconsin, ED-SNMTF performs best when $\lambda = 1$, while it works better when $\lambda = 10$ for Washington. This happens because these datasets are small and can easily overfit, so they need more regularization to be stable. Based on our results, we recommend using $\lambda \in [0, 10^{-3}]$ for datasets with many connections (like Email, Wiki, Citeseer, Cora), $\lambda \in [10^{-1}, 1]$ for smaller datasets (like Cornell, Texas, Wisconsin), and λ as high as 10 for very sparse datasets (like Washington).

Convergence analysis. Fig. 3 shows the training dynamics of our method by tracking relative error reduction over successive epochs under varying step size configurations. The selection of the learning parameter emerges as a critical factor governing both convergence velocity and algorithmic stability. When employing conservative step sizes (η=10^{-4}), the optimization process converges slowly, demanding extensive computational cycles to achieve low error. In contrast, aggressive parameter settings (η=10^{-1}) introduce oscillatory behavior that undermines stability. However, intermediate configurations (η=10^{-3} and η=10^{-2}) provide good performance, delivering rapid error decay while maintaining algorithmic stability throughout the optimization trajectory.

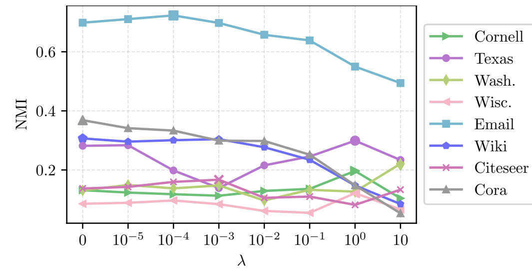

Fig. 2. Performance of the ED-SNMTF model on the Email dataset with varying values of the regularization parameter λ.

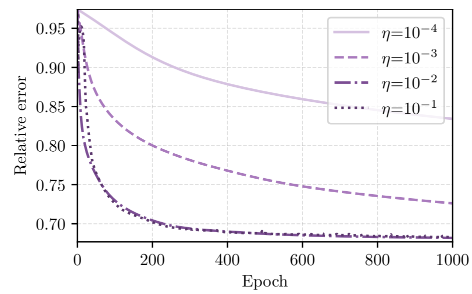

Fig. 3. Convergence behavior of the proposed ED-SNMTF model on the Email dataset ($\lambda = 10^{-4}$), optimized with Adam for different learning rates.

4. CONCLUSION

This work presented ED-SNMTF, an encoder–decoder formulation of symmetric NMTF that ensures consistency between reconstruction and latent recovery. By enforcing encoder-consistency, the model implicitly promotes orthogonality among membership vectors, enabling clear community structures to emerge without additional constraints or post-processing. Our GPU-accelerated implementation achieves stable training and consistently outperforms both classical factorization methods and recent deep approaches across diverse real-world networks. An interesting direction is to extend ED-SNMTF to nonlinear formulations that can capture more complex interaction patterns beyond linear projections. Incorporating deeper architectures could further enhance the model capacity to uncover hierarchical and multi-scale community structures.

5. REFERENCES

[1] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-ichi Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*, John Wiley & Sons, 2009.

[3] Nicolas Gillis, *Nonnegative Matrix Factorization*, SIAM, Philadelphia, PA, 2020.

[4] Jaewon Yang and Jure Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *ACM International Conference on Web Search and Data Mining*, 2013, pp. 587–596.

[5] Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu, "An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1273–1284, 2014.

[6] Paris Smaragdis and Judith C Brown, "Non-negative matrix factorization for polyphonic music transcription," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, pp. 177–180.

[7] Nasser Mohammadiha, Paris Smaragdis, and Arne Leijon, "Supervised and unsupervised speech enhancement using nonnegative matrix factorization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2140–2151, 2013.

[8] Chris Ding, Tao Li, Wei Peng, and Haesun Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 2006, pp. 126–135.

[9] Chris H.Q. Ding, Tao Li, and Michael I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[10] Zhijian Yuan and Erkki Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Scandinavian Conference on Image Analysis*. Springer, 2005, pp. 333–342.

[11] Fei Wang, Tao Li, Xin Wang, Shenghuo Zhu, and Chris Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, 2011.

[12] Alexandra Dache, Arnaud Vandaele, and Nicolas Gillis, "Orthogonal symmetric nonnegative matrix tri-factorization," in *IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2024.

[13] Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang, "Community preserving network embedding," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, p. 203–209.

[14] Ali Tosyali, Jinho Kim, Jeongsub Choi, and Myong K. Jeong, "Regularized asymmetric nonnegative matrix factorization for clustering in directed networks," *Pattern Recognition Letters*, vol. 125, pp. 750–757, 2019.

[15] Reyhaneh Abdollahi, Seyed Amjad Seyedi, and Mohamad Reza Noorimehr, "Asymmetric semi-nonnegative matrix factorization for directed graph clustering," in *2020 10th International Conference on Computer and Knowledge Engineering*, 2020, pp. 323–328.

[16] Bing-Jie Sun, Huawei Shen, Jinhua Gao, Wentao Ouyang, and Xueqi Cheng, "A non-negative symmetric encoder-decoder approach for community detection," in *Conference on Information and Knowledge Management*, 2017, pp. 597–606.

[17] George Trigeorgis, Konstantinos Bousmalis, Stefanos Zafeiriou, and Bjoern Schuller, "A deep semi-NMF model for learning hidden representations," in *International conference on machine learning*. PMLR, 2014, pp. 1692–1700.

[18] Pierre De Handschutter, Nicolas Gillis, and Xavier Siebert, "A survey on deep matrix factorizations," *Computer Science Review*, vol. 42, pp. 100423, 2021.

[19] Fanghua Ye, Chuan Chen, and Zibin Zheng, "Deep autoencoder-like nonnegative matrix factorization for community detection," in *International Conference on Information and Knowledge Management*, 2018, p. 1393–1402.

[20] Min Zhang and Zhiping Zhou, "Structural deep nonnegative matrix factorization for community detection," *Applied Soft Computing*, vol. 97, pp. 106846, 2020.

[21] Pierre De Handschutter, Nicolas Gillis, and Wivine Blekic, "Deep symmetric matrix factorization," in *2023 31st European Signal Processing Conference (EUSIPCO)*. IEEE, 2023, pp. 635–639.

[22] Akram Hajiveiseh, Seyed Amjad Seyedi, and Fardin Akhlaghian Tab, "Deep asymmetric nonnegative matrix factorization for graph clustering," *Pattern Recognition*, vol. 148, pp. 110179, 2024.

[23] Abraham Berman and Robert J Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, vol. 9, SIAM, 1994.

[24] Nicolas Nadisic, Nicolas Gillis, and Christophe Kervazo, "Smoothed separable nonnegative matrix factorization," *Linear Algebra and its Applications*, vol. 676, pp. 174–204, 2023.