



Wei Zhang<sup>1</sup>, Shanshan Yu<sup>2,3,\*</sup>, Ling Wang<sup>1</sup>, Wei Guo<sup>1</sup> and Man-Fai Leung<sup>4</sup>

- <sup>1</sup> College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China; zw848438346@163.com (W.Z.); w1510223@163.com (L.W.); guowei123@email.swu.edu.cn (W.G.)
- <sup>2</sup> Training and Basic Education Management Office, Southwest University, Chongqing 400715, China
- <sup>3</sup> Key Laboratory of Cyber-Physical Fusion Intelligent Computing (South-Central Minzu University), State Ethnic Affairs Commission, Wuhan 430074, China
- <sup>4</sup> School of Computing and Information Science, Faculty of Science and Engineering, Anglia Ruskin University, Cambridge CB1 1PT, UK; man-fai.leung@aru.ac.uk
- \* Correspondence: yu33@swu.edu.cn

Abstract: Recently, community detection has emerged as a prominent research area in the analysis of complex network structures. Community detection models based on non-negative matrix factorization (NMF) are shallow and fail to fully discover the internal structure of complex networks. Thus, this article introduces a novel constrained symmetric non-negative matrix factorization with deep autoencoders (CSDNMF) as a solution to this issue. The model possesses the following advantages: (1) By integrating a deep autoencoder to discern the latent attributes bridging the original network and community assignments, it adeptly captures hierarchical information. (2) Introducing a graph regularizer facilitates a thorough comprehension of the community structure inherent within the target network. (3) By integrating a symmetry regularizer, the model's capacity to learn undirected networks is augmented, thereby facilitating the precise detection of symmetry within the target network. The proposed CSDNMF model exhibits superior performance in community detection when compared to state-of-the-art models, as demonstrated by eight experimental results conducted on real-world networks.

**Keywords:** community detection; undirected network; non-negative matrix factorization; deep learning; symmetry regularization

MSC: 68T30

# 1. Introduction

Networks are pervasive in the real world [1], and numerous intricate systems in nature can be effectively represented by networks, such as ecological networks [2] and social networks [3]. Consequently, network analysis has emerged as a paramount concern. The inherent community structure represents a fundamental aspect of networks, playing a pivotal role in network characterization. Community detection serves as the mechanism for investigating and identifying this essential feature. Community detection has emerged as a prominent research area in recent years [4]. It entails dividing networks into separate communities, marked by densely interconnected nodes within each community and fewer connections between different ones. The accurate detection of these communities can enhance our understanding of complex network structures. For instance, in social platforms, individuals with shared interests and close ties form cohesive communities, while those with no overlap are segregated into separate ones. Community detection is a research topic involving sociology, mathematics, physics, and other disciplines, with a wide range of applications, such as finding similar user groups in social networks and finding protein complexes in protein–protein interaction (PPI) networks.



Citation: Zhang, W.; Yu, S.; Wang, L.; Guo, W.; Leung, M.-F. Constrained Symmetric Non-Negative Matrix Factorization with Deep Autoencoders for Community Detection. *Mathematics* **2024**, *12*, 1554. https://doi.org/10.3390/ math12101554

Academic Editor: Iraklis Varlamis

Received: 8 April 2024 Revised: 6 May 2024 Accepted: 8 May 2024 Published: 16 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

Recently, significant efforts have been devoted to the study of community detection, leading to the establishment of a comprehensive model pyramid [5–7]. Numerous methodologies for community detection have been proposed and successfully applied to network structures, encompassing spectral-clustering-based methods [8], label-propagation-based methods [9], stochastic-block-model-based methods [10], and deep-learning-based methods [11]. It is important to note that non-negative matrix factorization (NMF) is widely utilized in community detection due to its commendable interpretability and scalability. The purpose of matrix decomposition is to decompose a given matrix into the product of two or more dimensionally reducing factor matrices [12]. NMF decomposition extracts the basic features of the network from the matrix, which is an effective way to identify hidden data [13]. The NMF-based methods approximately factorize the adjacency matrix A of a given network into two non-negative factor matrices,  $A \approx UV$  ( $U \ge 0, V \ge 0$ ). The matrix V symbolizes the community membership, with matrix U acting as the projection matrix between the original network and the community members' space. The adjacency matrix A is decomposed into a single factor in the SNMF model,  $A = VV^T$ , which can effectively improve the capture of symmetry information in undirected networks. The adjacency matrix A is decomposed by three factors in the orthogonal three-factor non-negative matrix decomposition model,  $A = VSV^{T}$ , where S provides more degrees of freedom for the decomposition. The CNMF model is often used in symbolic networks where the fundamental vectors are represented as convex combinations of data matrices,  $A_{\pm} = A_{\pm}UV^{T}$  [14]. The existing networks can be classified into six categories: topology networks, signed networks, attributed networks, multilayer networks, dynamic networks, and large-scale networks [15]. Although NMF-based models have achieved good results in some cases, most of the existing NMF-based models are shallow models, such as SNMF [16], GNMF [17], and CNMF [14]. However, we cannot completely capture all of the internal structure of the network through shallow models. Moreover, certain deep NMF models overlook symmetry, a fundamental characteristic of undirected networks. Deep NMF is a novel and effective feature extraction method that has been used in recent years, which can obtain a deeper representation by further recursive decomposition of the matrix results of the NMF algorithm. Deep NMF can discover the underlying hierarchical characteristics of the data, which prompts us to conduct further research on deep NMF [18]. The deep autoencoder, an extensively employed unsupervised learning algorithm in the field of deep learning [19], is capable of performing various tasks such as feature extraction, dimensionality reduction, and data reconstruction. In addition, it can also bridge the gap between low-level to highlevel networks for optimal community detection [20]. The deep autoencoder consists of two components: an encoder component and a decoder component [21]. By employing layer-by-layer encoding, a high-dimensional dataset is transformed into a low-dimensional encoding, which can then be reconstructed back to its original high-dimensional form using layer-by-layer decoding.

This process of dimensionality reduction and reconstruction from high-dimensional information can not only help us extract feature information but can also remove noise and reconstruct clean data representation in the process, and feature extraction is more efficient in low-dimensional information [22]. By integrating non-negative matrix factorization and the deep autoencoder, we can effectively decompose the mapping matrix *U* to enhance the learning of hidden information that represents the intricate network structure [23]. The internal structure of the target network is learned, and the learning capacity of the undirected network is bolstered by integrating graph regularization and symmetric regularization. Building upon the insights gained from the preceding discussion, we introduce a novel methodology termed constrained symmetric non-negative matrix factorization with deep autoencoders (CSDNMF) to tackle the community detection challenge. The structure of the proposed model is shown in Figure 1. Our key contributions can be summarized concisely as follows:

• The proposed model leverages a deep autoencoder to address the community detection challenge in undirected networks, proficiently capturing hierarchical information within the network. Additionally, it integrates graph regularization and symmetric regularization into the learning objective, thereby enhancing the model's representation learning capability.

- The convergence proof of the proposed CSDNMF model is provided. The mathematical convergence of the proposed model is proved by two steps: (a) The non-increasing nature of the objective function is established under the update rule through the utilization of an auxiliary function. (b) The update rule's finite solution satisfies the Karush–Kuhn–Tucker (KKT) optimality conditions.
- Extensive experiments are conducted to evaluate the efficacy of various NMF-based community detection models. The results indicate that CSDNMF outperforms stateof-the-art NMF-based community detection methods.



**Figure 1.** The structure of the CSDNMF model is depicted, with the upper left representing the encoder component, the upper right representing the decoder component, and the lower representing the symmetric regularization term.

The paper is structured as follows: Section 2 provides an introduction to the fundamental concepts; Section 3 presents the proposed CSDNMF model; Section 4 conducts an analysis of the designed algorithms; Section 5 showcases the experimental results; and finally, in Section 6, a comprehensive summary of the entire text is presented.

## 2. Related Works

In this section, we present the requisite symbols and foundational knowledge, followed by a concise overview of the community detection problem and the NMF model. Finally, we elucidate the application of the NMF model in addressing the community detection problem.

## 2.1. Notations

In this article, bold capital letters are employed to denote matrices. For a matrix A, its *i*-th row vector, *j*-th column vector, (i,j)-th element, trace, and Frobenius norm are, respectively,  $A_i, A_j, A_{ij}, \operatorname{tr}(A), ||A||_F^2$ . The undirected and unweighted networks are

denoted as G = (V, E), where  $V = \{v_1, v_2, \dots v_n\}$  denotes a set of *n* nodes and  $E = \{e_1, e_2, \dots e_m\}$  denotes a set of *m* edges. The topology of an unweighted undirected network *G* is represented by the adjacency matrix *A*, where  $a_{ij}$  equals 1 if an edge between nodes  $v_i$  and  $v_j$  exists; otherwise, it equals 0. In the case of an undirected network, the adjacency matrix *A* exhibits symmetry, with all diagonal elements being 0. We concentrate on the challenge of community detection within undirected networks.

# 2.2. Non-Negative Matrix Factorization (NMF)

The NMF algorithm is formally introduced by Lee and Seung in 1999 [24], providing a robust mathematical framework for reducing data dimensionality and extracting essential features. The goal of NMF is to represent a non-negative matrix *A* as the product of two non-negative matrices *U* and *V*, symbolized as  $A \approx UV$ . Here, matrix *A* embodies the original dataset, whereas matrices *U* and *V* depict the basis matrix and constant matrix, respectively. Through this decomposition, NMF facilitates the exploration of latent structures and features within the data while effectively eliminating noise and redundant information, thereby accomplishing dimensionality reduction and feature extraction. NMF finds extensive applications across diverse areas, encompassing image processing [25], text mining [7], speech processing [26], and so on. Through the application of NMF, we can uncover latent patterns, themes, and features within our data, facilitating a deeper comprehension and enhanced utilization of the data. The non-negative matrix factorization technique is a potent tool that facilitates the simplification of intricate data and extraction of valuable insights, thereby holding immense significance for data analysis and machine learning.

### 2.3. Community Detection with NMF

Community detection fundamentally resembles a clustering problem, with nodes in the complex network serving as the objects to be clustered. NMF inherently possesses clustering capabilities. By applying orthogonal constraints to the matrix V, the NMF model can be seen as an alternative formulation of the K-means clustering model. Additionally, if the non-negative matrix A demonstrates symmetry, it can be transformed into a symmetric decomposition form,  $A \approx VV^T$ , which aligns with the spectral clustering model. Both K-means and spectral clustering have demonstrated their efficacy in node clustering, thus making NMF a natural choice for community detection. In addition, a large part of the reason why the NMF model is applied to the community detection problem is that the generation ability of NMF can explain the community structure well. Usually, the adjacency matrix A of the community network is decomposed, that is,  $A \approx UV \ (U \ge 0, V \ge 0)$ , where U represents the community characteristic matrix, V represents the community indicator matrix, and  $V_i m$  represents the intensity of the *i* th node belonging to the *m* th community. NMF-based community detection models typically involve four essential stages [15]: feature matrix construction, NMF-based model construction, model solving, and community detection. The initial step entails the construction of the feature matrix by extracting features from network G and forming the corresponding matrix, denoted as A. In the subsequent stages, A is the executed factorization used to generate the community indicator matrix V. Finally, during the community extraction stage, regardless of whether the communities overlap, the final result is obtained based on the community indicator matrix V.

#### 3. Methods and Theories

## 3.1. Deep Non-Negative Matrix Factorization

The NMF model represents the adjacency matrix of a complex original network as the product of two matrices and employs a mapping matrix solely to capture the underlying information within the intricate network. The information capacity of a matrix is evidently limited, whereas the intricate nature of a complex network poses challenges in its representation through a single layer of mapping. Consequently, NMF can solely capture simplistic information within the network. If the basis matrix *U* can be further decomposed, the basis matrix U is decomposed into the characteristic matrix of p levels; each layer of abstraction can effectively capture inter-node similarities at various levels of granularity, thereby enabling more comprehensive information extraction from the network and enhancing the model's representation learning capability. The adjacency matrix A is specifically decomposed into p + 1 non-negative factor matrices in the following manner:

$$A \approx U_1 U_2 \cdots U_p V_p,\tag{1}$$

where  $V_P \in R_+^{k \times n}$ ,  $U_i \in R_+^{r_{i-1} \times r_i} (1 < i < p)$ , and we set  $n = r_0 \ge r_1 \ge \cdots r_{p-1} \ge r_p = k$ .

The formula presented in Equation (1) enables a hierarchical organization of abstract comprehension within the *p*-layer, which can be given by the following factorizations:

$$V_{p-1} \approx U_p V_p,$$
  
.....  

$$V_2 \approx U_3 \cdots U_p V_p,$$
  

$$V_1 \approx U_2 \cdots U_p V_p.$$
(2)

This intricate architecture yields more precise community detection outcomes. According to Equation (2), we deduce the following objective function:

$$\min_{U_i, V_p} \mathcal{L}_D = \|A - U_1 U_2 \cdots U_p V_p\|_F^2,$$
s.t.  $V_p \ge 0, U_i \ge 0, \forall i = 1, 2..., p.$ 
(3)

#### 3.2. Constrained Symmetric NMF with Deep Autoencoders (CSDNMF)

The deep autoencoder represents a prototypical unsupervised learning algorithm, wherein the fundamental concept involves multiplying the input data by a matrix to achieve dimensionality reduction. Subsequently, the resulting data is multiplied by the transpose of the preceding weight matrix to restore an approximate representation of the original data. If the dimensionality reduction–reconstruction process is represented by multi-layer matrices, then the dimensions of these matrices must be reduced layer by layer, and we can regard the dimensionality reduction-reconstruction process as a process of encoding and decoding. The high-dimensional data set is transformed into a lower-dimensional representation using layer-by-layer encoding and subsequently restored to its original high-dimensional form through layer-by-layer decoding. The closer the similarity between the pre-encoding and post-decoding data, the higher the quality of model training and the more meaningful the outcome of dimensionality reduction. The deep autoencoder consists of both an encoder and a decoder component, with Equation (3) specifically representing the decoder. To bolster the autoencoder's capacity for representation learning, we combine the encoder component into Equation (3), thereby constructing an NMF model rooted in the autoencoder paradigm [27]. The encoder is the inverse process of the decoder, encoding the multilayer mapping into the initial matrix A. The objective function for the encoder is deduced as follows:

$$\min_{U_i, V_p} \mathcal{L}_E = \left\| V_p - U_p^T \cdots U_2^T U_1^T A \right\|_{F'}^2$$
(4)  
s.t.  $V_p \ge 0, U_i \ge 0, \forall i = 1, 2 \cdots, p.$ 

In order to refine the precision of comprehending the internal geometry of the network, we incorporate a graph regularizer. Here,  $\lambda$  denotes the regularization parameter, tr(·)

computes the trace of the enclosed matrix, *L* denotes the graph Laplacian, and we define L = D - A (where *D* is a diagonal matrix comprising the row sums of *A*).

$$\lambda \min_{U_i, V_p} \mathcal{L}_{reg} = \lambda \operatorname{tr} \left( V_p L V_p^T \right).$$
(5)

The adjacency matrix of an undirected network *G* is inherently a symmetric square matrix. To effectively capture and characterize the symmetry inherent in the target network, we introduce a symmetric regularization term, which improves the representation learning ability of the model for undirected networks by establishing symmetric constraints between the lower-dimensional representation of the original matrix and its transpose. The symmetric regularization term limits the process of encoding and decoding to the category of undirected networks, denoted as:

$$\frac{1}{2}\mu\mathcal{L}_{\text{sym}} = \left\| U_1 U_2 \cdots U_p V_p - \left( U_1 U_2 \cdots U_p V_p \right)^T \right\|_F^2$$

By integrating an encoder, decoder, graph regularization term, and symmetric regularization term, the learning objective of CSDNMF can be formulated as follows:

$$\min_{U_i,V_p} \mathcal{L} = L_D + \mathcal{L}_E + \lambda \mathcal{L}_{\text{reg}} + \frac{1}{2} \mu \mathcal{L}_{\text{sym}}$$

$$= \left\| A - U_1 U_2 \cdots U_p V_p \right\|_F^2 + \left\| V_p - U_p^T \cdots U_2^T U_1^T A \right\|_F^2$$

$$+ \lambda \operatorname{tr} \left( V_p L V_p^T \right) + \frac{1}{2} \mu \left\| U_1 U_2 \cdots U_p V_p - \left( U_1 U_2 \cdots U_p V_p \right)^T \right\|_F^2.$$
(6)

# 4. Optimization

In order to expedite the approximation of the factor matrix in the proposed model, we employ a pre-training approach to acquire the initial estimation of factor matrices  $U_i$  and  $V_i$ . Utilizing pre-training techniques proves highly effective for gradually revealing lowdimensional nonlinear structures within high-dimensional data. This process significantly diminishes the training duration of the proposed model, thus enhancing its efficiency and scalability. Following the pre-training phase, global fine-tuning further refines the model. The efficacy of pre-training has been previously evidenced, particularly in the domain of deep autoencoder networks. Detailed explanations of the update rules for the model will be presented in this section.

#### 4.1. Optimization of Update Rules for $U_i$ $(1 \le i \le p)$

The update rules of  $U_i$  in fine-tuning are derived by considering  $U_i$  as the object. Meanwhile, the objective function of CSDNMF is rewritten while keeping  $U_i$  unchanged:

$$\min_{U_{i}} \mathcal{L}(U_{i}) = \left\| A - \omega_{i-1} U_{i} \omega_{i+1} V_{p} \right\|_{F}^{2} + \left\| V_{p} - \omega_{i+1}^{T} U_{i}^{T} \omega_{i-1}^{T} A \right\|_{F}^{2} 
+ \frac{1}{2} \mu \| \omega_{i-1} U_{i} \omega_{i+1} V_{p} - (\omega_{i-1} U_{i} \omega_{i+1} V_{p})^{T} \|_{F}^{2},$$
s.t.  $U_{i} \ge 0,$ 
(7)

where  $\omega_{i-1} = U_1 U_2 \cdots U_{i-1}, \omega_{i+1} = U_{i+1} \cdots U_{p-1} U_p$ . When  $U_i = U_1, \omega_{i-1} = 0$ . When  $U_i = U_p, \omega_{i+1} = 0$ .

The Lagrangian multiplier matrix  $\Theta_i$  is introduced to enforce non-negative constraints on  $U_i$ , thereby enabling the solution of this equation. The significance of the Lagrange multiplier matrix is to effectively incorporate non-negative constraints into the optimization process, so as to ensure that the elements in the factor matrix in the whole process are non-negative. The resulting function can be expressed as follows:

$$\min_{U_{i},\Theta_{i}} \mathcal{L}(U_{i},\Theta_{i}) = \|A - \omega_{i-1}U_{i}\omega_{i+1}V_{p}\|_{F}^{2} + \|V_{p} - \omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}A\|_{F}^{2} 
+ \frac{1}{2}\mu\|\omega_{i-1}U_{i}\omega_{i+1}V_{p} - (\omega_{i-1}U_{i}\omega_{i+1}V_{p})^{T}\|_{F}^{2} - \operatorname{tr}\left(\Theta_{i}U_{i}^{T}\right).$$
(8)

Based on the universally accepted properties  $||X||_F^2 = tr(XX^T)$  and tr(AB) = tr(BA), we can derive the following result:

$$\min_{U_{i}} \mathcal{L}(U_{i}, \Theta_{i}) = \operatorname{tr}(AA^{T} + V_{p}V_{p}^{T} - 4AV_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T} + \omega_{i-1}U_{i}\omega_{i+1}V_{p} V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T} 
+ \mu\omega_{i-1}U_{i}\omega_{i+1}V_{p} V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T} + \omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1} 
- \mu\omega_{i-1}U_{i}\omega_{i+1}V_{p}\omega_{i-1}U_{i}\omega_{i+1}V_{p} - \Theta_{i}U_{i}^{T}).$$
(9)

According to Lagrange's theorem, in order to determine the minimum value of this expression, it is necessary for the partial derivative of  $\mathcal{L}(U_i, \Theta_i)$  with respect to  $U_i$  to be equal to 0. Consequently, we can deduce the subsequent expression.

$$\Theta_{i} = 2\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} - 4\omega_{i-1}^{T}AV_{p}^{T}\omega_{i+1}^{T} + 2\mu\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + 2\omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T} - 2\mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T} = 0.$$
(10)

Then, in conjunction with the complementary relaxation condition of the KKT condition, the equation is constrained in the iterative process to ensure the satisfaction of the KKT condition throughout; particularly by ensuring that the KKT multiplier is 0 at the extreme value, we can deduce the subsequent equation.

$$\Theta_{i} \odot U_{i} = (-4\omega_{i-1}^{T}AV_{p}^{T}\omega_{i+1}^{T} + 2\mathcal{M}_{i} - 2\mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}) \odot U_{i} = 0, \quad (11)$$

where  $\mathcal{M}_i = \omega_{i-1}^T \omega_{i-1} U_i \omega_{i+1} V_p V_p^T \omega_{i+1}^T + \mu \omega_{i-1} U_i \omega_{i+1} V_p V_p^T \omega_{i+1}^T + \omega_{i-1}^T A A^T \omega_{i-1} U_i \omega_{i+1} \omega_{i+1}^T$ . The update rule for  $U_i$  can be derived by transforming Equation (11).

$$U_i \leftarrow U_i \odot \frac{2\omega_{i-1}^T A V_p^T \omega_{i+1}^T + \mu \omega_{i-1}^T V_p^T \omega_{i+1}^T U_i^T \omega_{i-1}^T V_p^T \omega_{i+1}^T}{\mathcal{M}_i}.$$
(12)

4.2. Optimization of Update Rules for  $V_i$   $(1 \le i \le p)$ 

While maintaining the constancy of  $V_i$ , the objective function of CSDNMF is reformulated as:

$$\min_{V_i} \mathcal{L}(V_i) = \|A - \omega_i V_i\|_F^2 + \|V_i - \omega_i^T A\|_F^2 + \frac{1}{2}\mu \|\omega_i V_i - (\omega_i V_i)^T\|_{F'}^2$$
(13)

where  $\omega_i = U_1 \cdots U_i$ .

By introducing the Lagrangian multiplier matrix  $\Theta_i$  to enforce non-negative constraints on  $V_i$ , we derive the following function:

$$\min_{V_i,\Theta_i} \mathcal{L}(V_i,\Theta_i) = \|A - \omega_i V_i\|_F^2 + \|V_i - \omega_i^T A\|_F^2 + \frac{1}{2}\mu \|\omega_i V_i - (\omega_i V_i)^T\|_F^2 - \operatorname{tr}(\Theta_i V_i^T).$$
(14)

The expression of Optimization problem (14) can be further refined as follows:

$$\min_{V_i} \mathcal{L}(V_i, \Theta_i) = \operatorname{tr}(AA^T + V_i V_i^T - 4A V_i^T \omega_i^T + \omega_i V_i V_i^T \omega_i^T + \omega_i^T AA^T \omega_i + \mu \omega_i V_i V_i^T \omega_i^T - .\mu \omega_i V_i \omega_i V_i - \Theta_i V_i^T) = 0.$$
(15)

According to Lagrange's extremum theorem and the Karush–Kuhn–Tucker (KKT) condition, we derive the following equation:

$$\Theta_i \odot V_i = (2V_i + 2\omega_i^T \omega_i V_i + 2\mu \omega_i^T \omega_i V_i - 2\mu \omega_i^T V_i^T \omega_i^T - 4\omega_i^T A)) \odot V_i = 0.$$
(16)

The update rule of  $V_i$  is derived by transforming Equation (16).

$$V_i \leftarrow V_i \odot \frac{2\omega_i^T A + \mu \omega_i^T V_i^T \omega_i^T}{V_i + \omega_i^T \omega_i V_i + \mu \omega_i^T \omega_i V_i}.$$
(17)

# 4.3. Optimization of Update Rules for $V_p$

While maintaining the constancy of  $V_p$ , the objective function of CSDNMF is reformulated as:

$$\min_{V_p} (V_p) = \|A - \omega_p V_p\|_F^2 + \|V_p - \omega_p^T A\|_F^2 + \lambda \operatorname{tr} (V_p L V_p^T) + \frac{1}{2} \mu \|\omega_p V_p - (\omega_p V_p)^T\|_{F'}^2 \quad (18)$$
s.t.  $V_p \ge 0$ ,

where  $\omega_p = U_1 U_2 \cdots U_p$ .

By introducing the Lagrangian multiplier matrix  $\Theta_i$  to enforce non-negative constraints on  $V_p$ , we derive the following function:

$$\min_{V_p,\Theta_p} \mathcal{L}(V_p,\Theta_p) = \left\|A - \omega_p V_p\right\|_F^2 + \left\|V_p - \omega_p^T A\right\|_F^2 + \lambda \operatorname{tr}\left(V_p L V_p^T\right) \\ + \frac{1}{2} \mu \|\omega_p V_p - (\omega_p V_p)^T\|_F^2 - \operatorname{tr}\left(\Theta_p V_p^T\right).$$
(19)

According to the deduction process of updating rules for  $U_i$  and  $V_i$ , Optimization problem (19) can be reformulated as follows:

$$\min_{V_p,\Theta_p} \mathcal{L}(V_p,\Theta_p) = \operatorname{tr}(AA^T + V_pV_p^T - 4AV_p^T\omega_p^T + \omega_pV_pV_p^T\omega_p^T + \omega_p^TAA^T\omega_p + \mu\omega_pV_pV_p^T\omega_p^T - \mu V_p^T\omega_p^TV_p^T\omega_p^T + \lambda V_pLV_p^T - \Theta_pV_p^T).$$
(20)

Moreover, the subsequent equation can be derived as follows:

$$\Theta_p \odot V_p = (2V_p + 2\omega_p^T \omega_p V_p + 2\mu \omega_p^T \omega_p V_p + 2\lambda V_p L - 2\mu \omega_p^T V_p^T \omega_p^T - 4\omega_p^T A) \odot V_i = 0.$$
(21)

Substitute L = D - A into Equation (21):

$$\Theta_p \odot V_p = (2V_p + 2\omega_p^T \omega_p V_p + 2\mu \omega_p^T \omega_p V_p + 2\lambda V_p D - 2\lambda V_p A - 2\mu \omega_p^T V_p^T \omega_p^T - 4\omega_p^T A) \odot V_i = 0.$$
(22)

The update rule of  $V_p$ , as transformed by Equation (22), is as follows:

$$V_p \leftarrow V_p \odot \frac{2\omega_p^T A + \mu \omega_p^T V_p^T \omega_p^T + \lambda V_p A}{V_p + \omega_p^T \omega_p V_p + \mu \omega_p^T \omega_p V_p + \lambda V_p D}.$$
(23)

### 4.4. Convergence Analysis

In this section, we establish the convergence of the proposed CSDNMF model using the update rules provided in Update rule (12) and Update rule (23). Since the update of  $V_i$ does not impact the objective function outlined in Optimization problem (6), we only need to consider the influence of  $U_i$  and  $V_p$ . Given the non-negative constraint, the objective function's lower limit in Optimization problem (6) is 0. Hence, the convergence analysis of the model can be simplified to proving the following two points: (a) The objective function in Optimization problem (6) demonstrates non-increasing behavior under the update rules outlined in (12) and (23). (b) The finite solutions yielded by the update rules in (12) and (23) adhere to the KKT optimality condition.

(a) The objective function in Equation (6) demonstrates non-increasing behavior under the update rules outlined in (12) and (23).

The non-incrementality of the objective function under the update rule will be demonstrated by employing the auxiliary function method, which is defined as follows:

**Definition 1.** The function M(u, u') is considered as an auxiliary function of F(u) when the following conditions are met.

$$M(u,u') \ge F(u)$$
 and  $M(u,u) = F(u)$ .

**Lemma 1.** Depending on the following update rule for the auxiliary function M(u, u'), F(u) is non-increasing:

$$u^{t+1} = \operatorname{argmin} M(u, u^t) \tag{24}$$

The detailed proof of Lemma 1can be found in reference [28]; thus, we only need to appropriately define the auxiliary function to ensure that our update formula satisifies Equation (24).

Firstly, we establish the non-increasing property of the objective function in Optimization problem (6) under the update rule of  $U_i$ . The  $U_i$ -related component of the objective function is denoted by  $F(u_{ik})$ :

$$F(u_{jk}) = \|A - \omega_{i-1}U_{i}\omega_{i+1}V_{p}\|_{F}^{2} + \|V_{p} - \omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}A\|_{F}^{2} + \frac{1}{2}\mu\|\omega_{i-1}U_{i}\omega_{i+1}V_{p} - (\omega_{i-1}U_{i}\omega_{i+1}V_{p})^{T}\|_{F}^{2}.$$
(25)

Firstly, we judiciously select the auxiliary function of  $F(u_{jk})$  in the following manner:

$$\frac{M(u_{jk}, u_{jk}^{t}) = F(u_{jk}^{t}) + F'(u_{jk}^{t})(u_{jk} - u_{jk}^{t}) + ((1 + \mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})}{2u_{jk}^{t}} \times (u_{jk} - u_{jk}^{t})^{2}.$$
(26)

**Proof.** M(u, u) = F(u) is evidently apparent, so we need to demonstrate  $M(u, u') \ge F(u)$ .  $\Box$ 

We expand the function  $F(u_{jk})$  at point  $u_{jk}$  by employing a second-order Taylor series expansion:

$$F(u_{jk}) = F(u_{jk}^{t}) + F'(u_{jk}^{t})(u_{jk} - u_{jk}^{t}) + \frac{1}{2}((1+\mu)\omega_{i-1}^{T}\omega_{i-1}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T}.$$
  
+  $\omega_{i-1}^{T}AA^{T}\omega_{i-1}\omega_{i+1}\omega_{i+1}^{T} - \mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T})(u_{jk} - u_{jk}^{t})^{2}.$  (27)

The comparison of Equation (26) with Equation (27) reveals that the condition  $M(u, u') \ge F(u)$  can be expressed equivalently as follows:

$$\frac{((1+\mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T}+\omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})}{u_{jk}^{t}} \geq ((1+\mu)\omega_{i-1}^{T}\omega_{i-1}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T}+\omega_{i-1}^{T}AA^{T}\omega_{i-1}\omega_{i+1}\omega_{i+1}^{T}-\mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}W_{p}^{T}\omega_{i+1}^{T}).$$
(28)

Given that  $u_{ij}^t > 0$ , Inequality (28) can be equivalently expressed as the following equation:

$$((1+\mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})$$

$$\geq ((1+\mu)\omega_{i-1}^{T}\omega_{i-1}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}\omega_{i+1}\omega_{i+1}^{T} - \mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}W_{p}^{T}\omega_{i+1}^{T})u_{jk}^{t}.$$

$$(29)$$

The following formula can be readily derived:

$$\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} = \sum_{j,k}\omega_{i-1}^{T}\omega_{i-1}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T}u_{jk}^{t} > \omega_{i-1}^{T}\omega_{i-1}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T}u_{jk}^{t}.$$
 (30)

Similarly, the following formula can be obtained:

$$\omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T} > \omega_{i-1}^{T}AA^{T}\omega_{i-1}\omega_{i+1}\omega_{i+1}^{T}u_{jk}^{t}.$$
(31)

The condition of  $M(u, u') \ge F(u)$  is proven by combining Inequality (30) and Inequality (31), thereby establishing  $M(u_{jk}, u_{jk}^t)$  as an auxiliary function of  $F(u_{jk})$ . The subsequent step involves demonstrating the essential equivalence between the updates presented in Update rule (12) and Update rule (23), and those outlined in Equation (24).

The auxiliary function in Equation (26) is substituted into Equation (24).

$$\begin{split} u_{jk}^{t+1} &= \operatorname{argmin} M(u_{jk}, u_{jk}^{t}) \\ \Rightarrow M'(u_{jk}, u_{jk}^{t}) &= F'(u_{jk}^{t}) \\ &+ \frac{((1+\mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})}{u_{jk}^{t}} \times (u_{jk} - u_{jk}^{t}) = 0 \\ \Rightarrow u_{jk} &= u_{jk}^{t} - u_{jk}^{t} \times \frac{F'(u_{jk}^{t})}{((1+\mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})} \\ \Rightarrow u_{jk}^{t+1} &= u_{jk}^{t} \times \frac{2\omega_{i-1}^{T}AV_{p}^{T}\omega_{i+1}^{T} + \mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}}{((1+\mu)\omega_{i-1}^{T}\omega_{i-1}U_{i}\omega_{i+1}V_{p}V_{p}^{T}\omega_{i+1}^{T} + \omega_{i-1}^{T}AA^{T}\omega_{i-1}U_{i}\omega_{i+1}\omega_{i+1}^{T})}. \end{split}$$
(32)

The update rule depicted in Equation (32) mirrors that of Update rule (12), thereby guaranteeing the non-increasing characteristic of the objective function according to the update rule outlined in (12). Similarly, we can affirm the non-increasing property of the objective function under the update rule delineated in (23) through a comparable methodology.

(b) The finite solutions yielded by the update rules in (12) and (23) adhere to the KKT optimality condition.

The convergence condition enables us to obtain  $U_i^{\infty} = U_i^{t+1} = U_i^t = U_i$ , where *t* represents the *t*-th iteration. The expansion of the update rule in Equation (12) can be derived as follows:

$$(2\mathcal{M}_{i} - 4\omega_{i-1}^{T}AV_{p}^{T}\omega_{i+1}^{T} - 2\mu\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}U_{i}^{T}\omega_{i-1}^{T}V_{p}^{T}\omega_{i+1}^{T}) \odot U_{i} = 0.$$
(33)

The equivalence between Equations (11) and (33) is evident, thereby satisfying the complementary slackness condition of KKT optimality. The update rule for  $V_p$  can be demonstrated using the same methodology. The convergence of the objective function under the update rule in (12) and (23) is guaranteed, as evidenced by the content presented in this section.

#### 4.5. Algorithm Analysis

. . .

The proposed CSDNMF model in this paper comprises two components:

• Pre-training process: The pre-training process involves layer-by-layer training, commencing from the first layer and advancing up to the *p*-th layer, in alignment with the objective function. This ensures that each layer's loss function is trained to minimize to the fullest extent possible. The time complexity of the pre-training process is  $O\left(pt_p(n^2r + nr^2)\right)$ . Here, *p* denotes the number of layers in the layer configuration,  $t_p$  represents the number of pre-training iterations, and *r* denotes the maximum value within the layer configuration.

• Fine-tuning process: The update rules mandate the iterative optimization of  $U_i$ ,  $V_i$ , and  $V_p$  for each layer until the model converges. The time complexity of the fine-tuning process is  $O(ptf(n^2r + nr^2))$ , where  $t_f$  represents the number of fine-tuning iterations. The time complexity of the CSDNMF model can be summarized as

$$O\left(p\left(t_f + t_p\right)\left(n^2r + nr^2\right)\right)$$

The procedure of the CSDNMF model is illustrated in Algorithm 1.

```
Require: The adjacency matrix A
  layer configuration L
  number of communities k
  graph regular parameters \lambda
  symmetric regular parameters \mu
Ensure: mapping matrix U<sub>i</sub>
  feature matrix V_i
  community membership matrix V_p
  /*Pre-training*/
  U_1, V_1 = \text{Pro-training}(A, l_1)
  for i = 2 : m do
      U_i, V_i = \text{Pro-training}(A, l_{i-1})
  end for
  /*Fine-tuning*/
  while not converged do
  for i = 2 : m do
      update U_i according to Update rule (12)
      update V_i according to Update rule (17)
      update V_p according to Update rule (23)
  end for
  endwhile
  return U_i, V_i, V_p
```

# 5. Experiments

### 5.1. Evaluation Metric

In order to obtain precise experimental results, we utilize four evaluation metrics that have been empirically validated and extensively utilized as the assessment methodologies for our experimental outcomes. The four evaluation metrics are Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), Accuracy (ACC), and F-score.

# 5.2. Dataset

In our experiment, seven data sets obtained in practical applications are used to evaluate the performance of each model, and the data sets used are all public and widely adopted data sets. In order to verify the performance of each model under different data sets, we select data sets with large spans in the number of nodes. The smallest dataset is Email, which has 1005 nodes, and the largest is Facebook, which has 22,470 nodes. The selection of data sets in this way can verify, in the experimental results, whether the proposed model has universal significance in data sets of different sizes. The particulars of each dataset are delineated in Table 1.

Dataset	Node	Communities	Layer Configuration
Email	1005	42	1005-256-128-42
Wiki	2405	19	2405-256-128-19
Core	2708	7	2708-256-64-7
Citeseer	3312	6	3312-256-64-6
Facebook	22,470	4	22,470-512-64-4
Lastfm	7624	18	7625-256-128-18
Pubmed	19,717	3	19,717-512-64-3

Table 1. Dataset and layer configuration.

# 5.3. Comparison Models

To showcase the superiority of the proposed CSDNMF model, we have chosen seven prominent NMF-based community detection models for comparative analysis:

- NMF: The NMF model, based on non-negative matrix factorization, serves as a fundamental approach for community detection [15].
- SNMF: The SNMF algorithm employs a distinctive factor matrix to characterize the symmetry of undirected networks [16].
- ONMF: The ONMF imposes an orthogonal constraint on the mapping matrix *U* [29].
- MNMF: MNMF introduces a novel approach called Modularized NMF, which integrates the community structure into network embedding [30].
- GNMF: The GNMF model, which integrates NMF with graph regularization, is proposed for addressing community detection problems [17].
- DNMF: The DNMF model, a deep non-negative matrix factorization approach, exclusively incorporates the decoder component [27].
- DANMF: The DANMF model is a deep autoencoder NMF, comprising an encoder and a decoder [27].

## 5.4. Experiments Results

The results of four evaluation metrics across various datasets and models are presented in Tables 2–8. The optimal outcomes are emphasized in the color blue. Repetitively execute ten iterations on each dataset and compute the average outcome. The table clearly demonstrates that the CSDNMF model proposed in this paper outperforms other models in the majority of cases. The CSDNMF model only ranks second in the F-score metric in the Wiki dataset, the ARI metric in the Citeser dataset, and the ACC metric in the Lastfm dataset. The results demonstrate that the integration of a deep autoencoder with the model outperforms the shallow model, and further, incorporating symmetric regularization into the deep model surpasses other deep models in terms of performance on undirected networks.

Table 2. Performance evaluation of each model on the Email dataset.

		En	nail	
-	NMI	ARI	F-Score	ACC
NMF	0.6694	0.4298	0.4953	0.6597
SNMF	0.6880	0.4223	0.5358	0.6897
ONMF	0.6964	0.4918	0.5611	0.6896
MNMF	0.2366	0.0126	0.0295	0.1352
GNMF	0.6631	0.4362	0.5055	0.6152
DNMF	0.6744	0.4713	0.5609	0.6612
DANMF	0.6863	0.4919	0.5797	0.6736
CSDNMF	0.7014	0.5146	0.5937	0.6973

The optimal outcomes are emphasized in the color blue.

	Wiki			
-	NMI	ARI	F-Score	ACC
NMF	0.2651	0.1294	0.2041	0.4252
SNMF	0.2912	0.1223	0.2045	0.4397
ONMF	0.2812	0.1273	0.2056	0.4478
MNMF	0.0462	0.0215	0.0261	0.1063
GNMF	0.2722	0.1183	0.2152	0.4135
DNMF	0.2863	0.1121	0.1904	0.4083
DANMF	0.3136	0.1352	0.2285	0.4525
CSDNMF	0.3224	0.1394	0.2231	0.4567

Table 3. Performance evaluation of each model on the Wiki dataset.

The optimal outcomes are emphasized in the color blue.

# Table 4. Performance evaluation of each model on the Core dataset.

		C	ore	
_	NMI	ARI	F-Score	ACC
NMF	0.3142	0.1873	0.3411	0.5298
SNMF	0.3309	0.2305	0.3823	0.5462
ONMF	0.2046	0.1410	0.2877	0.4494
MNMF	0.0016	0.0012	0.0147	0.2044
GNMF	0.3362	0.1937	0.3655	0.5389
DNMF	0.3403	0.2466	0.3854	0.5016
DANMF	0.3484	0.2381	0.3931	0.5374
CSDNMF	0.3746	0.2538	0.4039	0.5545

The optimal outcomes are emphasized in the color blue.

 Table 5. Performance evaluation of each model on the Citesser dataset.

	Citesser			
-	NMI	ARI	F-Score	ACC
NMF	0.1172	0.0446	0.1436	0.3322
SNMF	0.1156	0.0656	0.1329	0.3491
ONMF	0.1085	0.0833	0.1337	0.3551
MNMF	0.0021	0.0013	0.0069	0.1065
GNMF	0.1362	0.1006	0.1497	0.3521
DNMF	0.1325	0.0962	0.1566	0.3427
DANMF	0.1269	0.1124	0.1437	0.3516
CSDNMF	0.1524	0.0956	0.1652	0.3626

The optimal outcomes are emphasized in the color blue.

Table 6. Performance evaluation of each model on the Lastfm dataset.

	Lastfm			
-	NMI	ARI	F-Score	ACC
NMF	0.5043	0.3781	0.4345	0.7176
SNMF	0.5144	0.4183	0.4977	0.7263
ONMF	0.5297	0.4016	0.4861	0.7056
MNMF	0.1296	0.0643	0.1016	0.3642
GNMF	0.5291	0.4291	0.4762	0.7651
DNMF	0.5811	0.4771	0.5783	0.7162
DANMF	0.5874	0.4962	0.6129	0.7439
CSDNMF	0.6078	0.5816	0.6469	0.7555

The optimal outcomes are emphasized in the color blue.

	Pumbed			
-	NMI	ARI	F-Score	ACC
NMF	0.1494	0.0939	0.5767	0.4314
SNMF	0.1647	0.1055	0.5437	0.4837
ONMF	0.1566	0.1647	0.5147	0.5171
MNMF	0.0006	0.0015	0.0167	0.2644
GNMF	0.1637	0.1129	0.6065	0.4933
DNMF	0.1752	0.1327	0.5261	0.5579
DANMF	0.2173	0.2331	0.6267	0.5916
CSDNMF	0.2464	0.2473	0.6684	0.6077

Table 7. Performance evaluation of each model on the Pumbed dataset.

The optimal outcomes are emphasized in the color blue.

Table 8. Performance evaluation of each model on the Facebook dataset.

	Facebook			
-	NMI	ARI	F-Score	ACC
NMF	0.0601	0.0280	0.3363	0.3919
SNMF	0.0682	0.0411	0.3677	0.4006
ONMF	0.0657	0.0395	0.3535	0.4155
MNMF	0.0014	0.0009	0.1065	0.1273
GNMF	0.0711	0.0306	0.3587	0.4236
DNMF	0.0944	0.0674	0.4421	0.4697
DANMF	0.1065	0.0811	0.4758	0.4966
CSDNMF	0.1263	0.0984	0.5271	0.5563

The optimal outcomes are emphasized in the color blue.

## 5.5. Parameter Sensitivity

The parameters  $\lambda$  and  $\mu$  are used to adjust the contribution of graph regularization and symmetric regularization, respectively, and the range of both parameters is  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1\}$ . We conducted experiments to assess the sensitivity of parameter values on the Email and Lastfm datasets, respectively. When conducting parameter testing, the other parameter should be set to a value of 1. The experimental results are presented in Figures 2–5. The results demonstrate the stability of CSDNMF across various values of  $\lambda$  and  $\mu$  while emphasizing that appropriate parameter selection can enhance model performance and robustness. In this experiment, both  $\lambda$  and  $\mu$  were set to 1.



**Figure 2.** The influence of graph regularization  $\lambda$  on the Email dataset.



**Figure 3.** The influence of symmetric regularization  $\mu$  on the Email dataset.



**Figure 4.** The influence of graph regularization  $\lambda$  on the Lastfm dataset.



**Figure 5.** The influence of symmetric regularization  $\mu$  on the Lastfm dataset.

# 5.6. Wilcoxon Signed Rank Test

To rigorously compare the performance of the proposed CSDNMF model in this paper with other state-of-the-art models, we conduct a widely recognized verification experiment known as the Wilcoxon Signed Rank Test [31]. The results from the Wilcoxon Signed Rank Test experiments, conducted at a specified significance level, are outlined in Table 9. The *p*-value stands as the key indicator denoting the significance level. Analyzing the data from Table 9, we confidently affirm that the CSDNMF model demonstrates significantly

greater accuracy in community detection compared to other models, with a confidence level of 95.

Table 9. Wilcoxon Signed Rank Test.

Proposed Model	<b>Comparison Model</b>	<i>p</i> -Value
CSDNMF	NMF	$6.25 \times 10^{-6}$
CSDNMF	SNMF	$2.02 \times 10^{-7}$
CSDNMF	ONMF	$1.31  imes 10^{-6}$
CSDNMF	GNMF	$4.12 imes10^{-7}$
CSDNMF	MNMF	$7.35  imes 10^{-7}$
CSDNMF	DNMF	$2.80 imes10^{-4}$
CSDNMF	DANMF	$1.63 imes10^{-4}$

#### 6. Conclusions

The utilization of NMF-based models in community detection problems is driven by their advantageous traits. However, the linearity of NMF poses limitations, particularly when dealing with nonlinear complex networks. To overcome this constraint, we propose the novel CSDNMF model, which integrates deep learning techniques to outperform shallow models. Our proposed CSDNMF model leverages the deep learning capabilities of deep autoencoders, the effective learning abilities of graph regularization for capturing network internal structures, and the representation power of undirected networks through symmetric regularization. Through extensive experiments conducted on seven datasets and employing four evaluation metrics, we demonstrate the model's superiority over other NMF-based models. Additionally, we showcase its robustness to regularization parameters.

Author Contributions: Conceptualization, W.Z., S.Y. and L.W.; Methodology, W.Z., S.Y., L.W. and W.G.; Software, W.Z., S.Y., L.W. and W.G.; Validation, W.Z., L.W., W.G. and M.-F.L.; Formal analysis, L.W., W.G. and M.-F.L.; Investigation, S.Y. and L.W.; Resources, L.W., W.G. and M.-F.L.; Data curation, L.W., W.G. and M.-F.L.; Writing—original draft, W.Z.; Writing—review and editing, S.Y. and M.-F.L.; Visualization, W.Z., S.Y., L.W. and W.G.; Supervision, S.Y., L.W. and W.G.; Project administration, M.-F.L.; Funding acquisition, M.-F.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Open Fund of the Key Laboratory of Cyber-Physical Fusion Intelligent Computing (South-Central Minzu University), State Ethnic Affairs Commission under Grant CPFIC202303.

**Data Availability Statement:** The data used to support the findings of the study are available from the first author upon request. The author's email address is zw848438346@163.com.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

- Wang, Y.; Zou, L.; Ma, L.; Zhao, Z.; Guo, J. A survey on control for Takagi-Sugeno fuzzy systems subject to engineering-oriented complexities. *Syst. Sci. Control Eng.* 2021, 9, 334–349. [CrossRef]
- 2. Cui, L.; Wang, J.; Sun, L.; Lv, C. Construction and optimization of green space ecological networks in urban fringe areas: A case study with the urban fringe area of Tongzhou district in Beijing. *J. Clean. Prod.* **2020**, *276*, 124266. [CrossRef]
- Sanchez-Vega, F.; Mina, M.; Armenia, J.; Chatila, W.K.; Luna, A.; La, K.; Dimitriadoy, S.; Liu, D.L.; Kantheti, H.S.; Heins, Z.; et al. Abstract 3302: The molecular landscape of oncogenic signaling pathways in The Cancer Genome Atlas. *Cancer Res.* 2018, 78, 3302. [CrossRef]
- 4. Hu, J.; Zhang, H.; Liu, H.; Yu, X. A survey on sliding mode control for networked control systems. *Int. J. Syst. Sci.* 2021, 52, 1129–1147. [CrossRef]
- 5. Fortunato, S. Community detection in graphs. *Phys. Rep.* 2010, 486, 75–174. [CrossRef]
- Das, S.; Biswas, A. Deployment of Information Diffusion for Community Detection in Online Social Networks: A Comprehensive Review. *IEEE Trans. Comput. Soc. Syst.* 2021, *8*, 1083–1107. [CrossRef]

- Barman, P.C.; Iqbal, N.; Lee, S.Y. Nonnegative Matrix Factorization Based Text Mining: Feature Extraction and Classification. In Proceedings of the 13th International Conference, ICONIP 2006, Hong Kong, China, 3–6 October 2006; pp. 703–712.
- 8. Tang, F.; Wang, C.; Su, J.; Wang, Y. Spectral clustering-based community detection using graph distance and node attributes. *Comput. Stat.* **2020**, *35*, 69–94. [CrossRef]
- Xie, J.; Szymanski, B.K. Community detection using a neighborhood strength driven Label Propagation Algorithm. In Proceedings of the 2011 IEEE Network Science Workshop, West Point, NY, USA, 22–24 June 2011; pp. 188–195. [CrossRef]
- Lyzinski, V.; Tang, M.; Athreya, A.; Park, Y.; Priebe, C.E. Community Detection and Classification in Hierarchical Stochastic Blockmodels. *IEEE Trans. Netw. Sci. Eng.* 2017, 4, 13–26. [CrossRef]
- 11. Wu, L.; Zhang, Q.; Chen, C.H.; Guo, K.; Wang, D. Deep Learning Techniques for Community Detection in Social Networks. *IEEE Access* 2020, *8*, 96016–96026. [CrossRef]
- Peng, C.; Hou, X.; Chen, Y.; Kang, Z.; Chen, C.; Cheng, Q. Global and local similarity learning in multi-kernel space for nonnegative matrix factorization. *Knowl.-Based Syst.* 2023, 279, 110946. [CrossRef]
- 13. Fathi Hafshejani, S.; Moaberfard, Z. Initialization for nonnegative matrix factorization: A comprehensive review. *Int. J. Data Sci. Anal.* **2023**, *16*, 119–134. [CrossRef]
- 14. Yan, C.; Chang, Z. Modularized convex nonnegative matrix factorization for community detection in signed and unsigned networks. *Phys. A Stat. Mech. Its Appl.* **2020**, *539*, 122904. [CrossRef]
- 15. He, C.; Fei, X.; Cheng, Q.; Li, H.; Hu, Z.; Tang, Y. A Survey of Community Detection in Complex Networks Using Nonnegative Matrix Factorization. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 440–457. [CrossRef]
- 16. Luo, X.; Liu, Z.; Jin, L.; Zhou, Y.; Zhou, M. Symmetric Nonnegative Matrix Factorization-Based Community Detection Models and Their Convergence Analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 1203–1215. [CrossRef] [PubMed]
- 17. Deng, P.; Li, T.; Wang, H.; Wang, D.; Horng, S.J.; Liu, R. Graph Regularized Sparse NonNegative Matrix Factorization for Clustering. *IEEE Trans. Comput. Soc. Syst.* 2023, 10, 910–921. [CrossRef]
- 18. Chen, W.-S.; Zeng, Q.; Pan, B. A survey of deep nonnegative matrix factorization. *Neurocomputing* **2022**, 491, 305–320. [CrossRef]
- Tang, J.; Deng, C.; Huang, G.B.; Hou, J. A fast learning algorithm for multi-layer extreme learning machine. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 175–178. [CrossRef]
- 20. Al-sharoa, E.; Rahahleh, B. Community detection in networks through a deep robust auto-encoder nonnegative matrix factorization. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105657. [CrossRef]
- 21. Lv, L.; Bardou, D.; Liu, Y.; Hu, P. Deep Autoencoder-like nonnegative matrix factorization with graph regularized for link prediction in dynamic networks. *Appl. Soft Comput.* **2023**, *148*, 110832. [CrossRef]
- Li, Y.; Chen, J.; Chen, C.; Yang, L.; Zheng, Z. Contrastive Deep Nonnegative Matrix Factorization For Community Detection. In Proceedings of the ICASSP 2024—2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Republic of Korea, 14–19 April 2024; pp. 6725–6729. [CrossRef]
- Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 2013, 35, 1798–1828. [CrossRef] [PubMed]
- 24. Lee, D.; Seung, H. Learning the parts of objects by nonnegative matrix factorization. Nature 1999, 401, 788–791. [CrossRef]
- Luong, T.X.; Kim, B.-K.; Lee, S.-Y. Color image processing based on Nonnegative Matrix Factorization with Convolutional Neural Network. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 2130–2135. [CrossRef]
- 26. Nie, S.; Liang, S.; Liu, W.; Zhang, X.; Tao, J. Deep Learning Based Speech Separation via NMF-Style Reconstructions. *IEEE/ACM Trans. Audio Speech Lang. Process.* 2018, 26, 2043–2055. [CrossRef]
- Ye, F.; Chen, C.; Zheng, Z. Deep Autoencoder-like Nonnegative Matrix Factorization for Community Detection. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; Association for Computing Machinery: New York, NY, USA, 2018. [CrossRef]
- 28. Lee, D.; Seung, H.S. Algorithms for Nonnegative Matrix Factorization. In *Advances in Neural Information Processing Systems*; Leen, T., Dietterich, T., Tresp, V., Eds.; MIT Press: Cambridge, MA, USA, 2000; Volume 13.
- Choi, S. Algorithms for orthogonal nonnegative matrix factorization. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 1828–1832. [CrossRef]
- 30. Ghadirian, M.; Bigdeli, N. Hybrid adaptive modularized tri-factor nonnegative matrix factorization for community detection in complex networks. *Sci. Iran.* 2023, *30*, 1068–1084. [CrossRef]
- 31. Taheri, S.M.; Hesamian, G. A generalization of the Wilcoxon signed-rank test and its applications. *Stat. Pap.* **2013**, *54*, 457–470. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.