

# Boosting nonnegative matrix factorization based community detection with graph attention auto-encoder

He, C., Zheng, Y., Fei, X., Li, H., Hu, Z. & Tang, Y.

Author post-print (accepted) deposited by Coventry University's Repository

**Original citation & hyperlink:**

He, C, Zheng, Y, Fei, X, Li, H, Hu, Z & Tang, Y 2021, 'Boosting nonnegative matrix factorization based community detection with graph attention auto-encoder', *IEEE Transactions on Big Data*, vol. (In-Press).

<https://dx.doi.org/10.1109/TBDATA.2021.3103213>

DOI 10.1109/TBDATA.2021.3103213

ESSN 2332-7790

Publisher: Institute of Electrical and Electronics Engineers

**© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.**

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the author's post-print version, incorporating any revisions agreed during the peer-review process. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

# Boosting Nonnegative Matrix Factorization Based Community Detection With Graph Attention Auto-encoder

Chaobo He, Yulong Zheng, Xiang Fei, Hanchao Li, and Yong Tang

**Abstract**—Community detection is of great help to understand the structures and functions of complex networks. It has become one of popular research topics in the field of complex networks analysis. Due to the simplicity, flexibility, effectiveness and better interpretability, Nonnegative Matrix Factorization (NMF)-based methods have been widely employed for community detection. However, most existing NMF-based community detection methods are linear and their performance is limited when facing networks with diversified structure information. In view of this, we propose a nonlinear NMF-based method named NMF-GAAE, which is composed of two main modules: NMF and Graph Attention Auto-Encoder (GAAE). This approach can boost the performance of NMF-based community detection methods by the aid of graph neural networks and deep clustering. More specifically, GAAE introduces an attention mechanism directed by NMF-based community detection to learn the node representations, while NMF can simultaneously factor these representations to uncover the community structure. We design a unified framework to jointly optimize GAAE and NMF modules, which is very beneficial to obtain better community detection results. We conduct extensive experiments on synthetic and real-world networks. The results show that our NMF-GAAE not only performs better than state-of-the-art NMF-based community detection methods, but also outperforms some network representation based baselines. More importantly, NMF-GAAE indeed can boost the performance of NMF-based community detection methods.

**Index Terms**—Community detection, nonnegative matrix factorization, graph attention auto-encoder, graph neural networks, deep clustering, complex networks

## I. INTRODUCTION

Community detection is a long-standing problem in the field of complex networks analysis. It aims to detect cohesive groups of nodes, where nodes in the same group connect to each other more densely than those in different groups [1], [2]. Effective methods for community detection not only can be served as the fundamental tools to characterize and understand

Manuscript received April 30, 2021; revised August \*\*, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62077045, Grant U1811263 and Grant 61772211, in part by the Humanity and Social Science Youth Foundation of Ministry of Education of China under Grant 19YJCZH049 and in part by the Natural Science Foundation of Guangdong Province of China under Grant 2019A1515011292. (Corresponding author: Yong Tang.)

Chaobo He is with the School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China (e-mail: hechaobo@foxmail.com).

Yulong Zheng and Yong Tang are with the School of Computer Science, South China Normal University, Guangzhou 510225, China (e-mail: 1457367033@qq.com; ytang@m.scnu.edu.cn).

Xiang Fei and Hanchao Li are with the Department of Computing, Coventry University, Coventry CV15FB, UK (e-mail: aa5861@coventry.ac.uk; lih30@uni.coventry.ac.uk).

the structures and functions of complex networks [3], but also having a great application value, such as helping to find groups with similar users in social networks [4], identify fraud groups in telecommunications networks [5], mine research teams in co-authorship networks [6] and discover functional units in protein-protein interaction networks [7]. Accordingly, community detection has attracted a lot of attention from both academia and industry. Moreover, related researchers cover a wide range of disciplines, including network science, computer science, sociology and bioinformatics.

In the past decades, various methods for community detection have been proposed, such as modularity optimization based methods [8], [9], stochastic block model based methods [10], game theoretic model based methods [11], label propagation based methods [12], deep learning based methods [13] and so on. It is worth mentioning that Nonnegative Matrix Factorization (NMF) based methods are also widely adopted. Compared with other types of methods, NMF-based methods are often more simple and effective, and more interpretable [14]. Particularly, they are competent in almost all community detection tasks, including overlapping community detection [15], [16], semi-supervised community detection [17], [18] and detecting communities from a variety of networks (e.g., undirected networks [19], [20], attributed networks [21], [22], signed networks [23], [24], multi-layer networks [25], [26], dynamic networks [27], [28] and large-scale networks [29], [30]). It can be argued that NMF-based methods are very versatile to deal with the problem of community detection. Until now, NMF-based community detection is still a hot topic and many methods are constantly presented in high-prestige journals and conferences.

NMF-based community detection has exhibited its unique advantages, but it may be less effective when encountering complex networks with complicated and diversified structures. This is due to the intrinsic linearity of NMF. Most NMF-based methods assume that the original networks could be directly reconstructed using a linear combination of community features. However, this linear reconstruction assumption is invalid for these networks mentioned above, because they often have various nonlinear features [31], [32]. To address this issue, a few methods have been proposed and can be roughly divided into the following two types:

- One is devising methods based on deep NMF [33], such as MvDGNMF [34] and DANMF [35]. These methods try to obtain better performance with the aid of deep NMF's expressive power enhanced by the multi-layer factoriza-

tion structure. However, due to the linear decomposition feature of each layer in deep NMF, deep NMF-based methods are essentially still linear, and their performance improvements are limited.

- The other type of methods is constructing nonlinear community detection model. A representative method is NAGC [36] which is based on symmetric NMF with positive unlabeled learning. This method introduces a nonlinear projection function to learn the relationship between the different community assignments of the topology and the attributes of networks. Although NAGC achieves better performance than some linear NMF-based methods, it cannot be generalized to networks with only structure information. Besides, it is very time consuming and can not deal with networks beyond a certain scale.

In summary, these two kinds of methods above both try to improve NMF-based community detection model by enhancing its expressive power. Considering that nonlinear models often have stronger expressive power to learn complicated features of data samples, constructing nonlinear community detection model will be a better strategy. However, relatively little attention has been dedicated to this aspect, and more effective methods still need to be explored.

In recent years, Deep Neural Networks (DNNs) have revolutionized many machine learning tasks, including image classification [37], speech recognition [38] and machine translation [39]. Due to the powerful nonlinear expressive power, DNNs have also been widely used to boost the performance of tradition machine learning models. Among them, deep clustering which combines clustering model (e.g., K-means and spectral clustering) with DNNs is very attractive and many effective methods have been proposed, such as DCN [40], DEC [41] and SC-EDAE [42]. Comparing with traditional clustering methods, deep clustering methods demonstrate great superiorities, especially in terms of the clustering performance. On the other hand, currently, as a special extension of DNNs to graph data, Graph Neural Networks (GNNs) have received increasing interest and many variants have been presented, such as famous Graph Convolutional Network (GCN) [43] and Graph Attention Network (GAT) [44]. Existing works show that GNNs are more suitable and powerful to conduct nonlinear feature learning on graph data than other general DNNs (e.g., auto-encoder and convolutional neural networks) [45], [46]. This makes them have great potential to improve many graph-related tasks, such as community detection focused in this paper. All of these above give us a good inspiration that using GNNs can be expected to better boost NMF-based community detection.

Motivated by the above-mentioned analyses, in this paper we propose a nonlinear NMF-based community detection method named NMFGAAE, which is powered by a specially designed variant of GNNs: graph attention auto-encoder (GAAE). To the best of our knowledge, it is the first time to utilize GNNs to boost the performance of NMF-based community detection. Our contributions are summarized as follows:

- NMFGAAE provides a unified framework to integrate

GAAE with NMF-based community detection model. Using this framework, GAAE and NMF-based community detection model can be optimized jointly, which enables GAAE to generate more beneficial node representations for NMF-based community detection.

- For the GAAE module in NMFGAAE, we design an attention mechanism which specially serve NMF-based community detection. Comparing with the attention mechanism used in GAT, our proposed attention mechanism can help NMFGAAE obtain better performance.
- We conduct extensive experiments to demonstrate the effectiveness of NMFGAAE on both synthetic and real-world networks. The results show that NMFGAAE not only outperforms state-of-the-art NMF-based community methods, but also outperforms some widely used network representation based methods, including DeepWalk+K-means and LINE+K-means.

The rest of this paper is organized as follows. We firstly briefly review the related work in Section II, and then respectively detail the proposed method NMFGAAE and report the experiment results in Section III and Section IV. Finally, we conclude this paper in Section V.

## II. RELATED WORK

In this section, we introduce some background knowledge and overview some work that relates to the features used in our method NMFGAAE: NMF-based community detection, GNNs and deep clustering.

### A. NMF-based community detection

NMF is a classical low-rank matrix factorization model proposed by Lee and Seung [47]. It can factor a nonnegative data sample matrix into the approximate product of two nonnegative matrices: the basis matrix and the coefficient matrix. Due to the nonnegativity constraint, every data sample can be represented as an additive linear combination of the basis features in the basis matrix, and the coefficients are from the corresponding vector in the coefficient matrix. Therefore, NMF is essentially a linear model.

In [48], Ding et al. proved that NMF has approximately equivalent relationships with K-means and spectral clustering models, which are both applicable to detect communities from complex networks. Besides, from the perspective of generative model, Psorakis et al. [14] demonstrated the powerful interpretability of NMF when used in community detection. These works provide a solid theoretical foundation to NMF-based community detection, and meanwhile stimulate the research interest in this field. Recently, various NMF-based community detection methods have been proposed, such as symmetric NMF-based methods [16], [19], [20], semi-supervised NMF-based methods [17], [18], semi-NMF-based methods [23], [24], joint NMF-based methods [25], [26], [27], [28], [49] and deep NMF-based methods [34], [35]. These methods construct the community detection model by extending the basic NMF, such as relaxing nonnegative constraint (semi-NMF), imposing graph regularization (graph regularized NMF), stacking multi-layer factorization structure (deep NMF) and so on. With these

strategies, they all have achieved good performance in some cases.

Although existing NMF-based community methods use various NMF variants, most of them are still within the category of linear methods due to the intrinsic linearity of NMF model. It is generally believed that linear community detection model may be less effective when facing networks with diversified structure information, which are considered to have various nonlinear features [31], [32]. In this paper, our proposed NMFGAAE is nonlinear due to the combination with a variant of GNNs: graph attention auto-encoder, which makes it fundamentally different from existing NMF-based community methods.

### B. GNNs

The concept of GNNs was formally proposed in [50], which extended existing neural networks to process graph data. Up to now, lots of GNNs have been presented, which mainly include three categories: graph recurrent neural networks (GRNNs), graph convolutional networks (GCNs) and graph auto-encoders (GAEs). Among them, GCNs have received more attention in recent years, which is largely attributed to the great success of convolution neural networks in computer vision domain. According to the difference in how to define graph convolutions, GCNs are further categorized into two types: spectral-based and spatial-based. Spectral-based GCNs define the graph convolutions as low-pass filters of graph signals, such as spectral CNN [51], ChebNet [52] and GCN [43]. Spatial-based GCNs define graph convolutions as the process of information propagation. The key idea behind them is to aggregate feature information from neighbor nodes via specific aggregation schemes, such as LSTM pooling used in GraphSage [53], attention mechanism used in GAT [44] and max-pooling used in GIN [54]. In general, spatial-based GCNs are more efficient, because spectral-based GCNs often need to conduct the costly Laplacian matrix eigen-decomposition. Besides, spatial-based GCNs are more flexible, because aggregation schemes, especially attention mechanism, allow them to only focus on task-relevant parts instead of the whole graph structure.

As another type of GNNs, GAEs is applied to learn node low-dimensional representations by an unsupervised training manner. They are composed of two parts: Encoder and Decoder. Encoder maps nodes into a latent feature space and decoder reconstructs graph structure from latent representations. In particular, the encoder part can adopt various GNNs flexibly, such as GCN encoder [55], [56] and GAT encoder [57]. Motivated by the unsupervised characteristic of community detection task here, and the effectiveness and flexibility of GAEs, we construct our method NMFGAAE based on the architecture of GAEs and specially design a spatial-based GCN with attention mechanism as its encoder.

### C. Deep clustering

Clustering with deep learning are called as deep clustering, which aims to improve the clustering performance via DNNs [58], [59]. In the early stage, most deep clustering methods

comprise two steps, such as StructAE [60] and SAE [61]. These methods firstly use DNNs to learn data representations and then feed these representations into the specified clustering model (e.g., K-means) directly. The drawback of two-step methods is that the learned representations may not be the best fit for the subsequent clustering model, and the clustering model is not beneficial to the representation learning. To achieve mutual benefit for these two steps, one-step deep clustering is presented. It provides a unified framework to jointly train representation learning model and clustering model. Representative methods include DCN [40], DEC [41] and SC-EDAE [42], which all demonstrate that one-step deep clustering can learn more clustering-friendly data representations, resulting in more significant increase of the clustering performance.

It should be noted that most existing deep clustering methods are designed for data with grid or sequence structure (e.g., image and text data). More recently, with the popularity of GNNs, a few deep clustering methods for graph data are proposed, such as SDCN [56], DAEGC [57] and EGAE-JOCAS [62]. These methods all belong to one-step deep clustering approaches, and meanwhile use GAEs to learn node representations. Our method NMFGAAE adopt the similar architecture with them, but it still has some obvious differences: (1) NMFGAAE selects NMF as the clustering model, but DAEGC and SDCN select self-supervised clustering model, and EGAE-JOCAS simultaneously use K-means and spectral clustering models. (2) NMFGAAE only exploits the structure information of the graph, but DAEGC, SDCN and EGAE-JOCAS all use the structure and attribute information. (3) NMFGAAE uses a different attention mechanism with DAEGC, whereas SDCN and EGAE-JOCAS do not use any attention mechanism. (4) More importantly, NMFGAAE is a goal-directed clustering method, which specially focuses on boosting NMF-based community detection.

## III. THE PROPOSED METHOD: NMFGAAE

In this section, we will first introduce the notations and preliminaries used in this paper, next give an overview about the proposed NMFGAAE framework, then detail its modules and finally present how to optimize NMFGAAE model.

### A. Notations and preliminaries

Throughout this paper, matrices are denoted by bold uppercase letters. For a given matrix  $\mathbf{Y}$ , its  $i$ -th row vector,  $(i, j)$ -th element, trace, transpose and Frobenius norm are denoted by  $\mathbf{Y}_i$ ,  $\mathbf{Y}_{ij}$ ,  $tr(\mathbf{Y})$ ,  $\mathbf{Y}^T$  and  $\|\mathbf{Y}\|_F$ , respectively. For clarity, we summarized all frequently used notations in Table I.

In this paper, we focus on non-overlapping community detection in undirected networks with only structure information. This is because most NMF-based community detection methods are suitable for this application scenario, and undirected networks with only structure information are more readily available in real-world.

Without loss of generality, we denote an undirected network as an undirected and unweighted graph  $G = (V, E)$ , where  $V = \{v_1, v_2, \dots, v_n\}$  is the nodes set and  $E = \{e_{ij} | v_i \in V \wedge v_j \in V\}$  is the edges set. To represent the structure information

TABLE I  
NOTATIONS AND DESCRIPTIONS

Notation	Description
$G$	The input network
$V$	Nodes set
$v_i$	The $i$ -th node
$E$	Edges set
$ E $	The number of edges
$e_{ij}$	The edge from $v_i$ to $v_j$
$n$	The number of nodes
$k$	The number of communities
$C$	Communities set
$C_i$	The $i$ -th community
$\mathbf{X}$	The node feature matrix
$m$	The dimension of node feature
$\mathbf{A}$	The adjacency matrix
$\mathbf{I}$	The identity matrix
$\mathbf{Z}$	The node latent representation matrix
$d$	The dimension of node latent representation
$\mathbf{U}$	The community membership matrix
$\mathbf{V}$	The community feature matrix
$\mathcal{N}_i$	The neighbor nodes set of $v_i$
$\mathbb{R}$	The real number set
$\mathbb{R}_+$	The nonnegative real number set

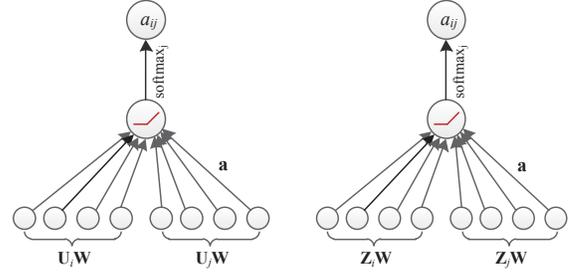
of  $G$ , we use a binary-valued adjacency matrix  $\mathbf{A} = [\mathbf{A}_{ij}]^{n \times n}$ , where  $\mathbf{A}_{ij} = 1$  if  $e_{ij} \in E$  and  $\mathbf{A}_{ij} = 0$  otherwise. Assuming that  $G$  comprises  $k$  non-overlapping communities, we denote the non-overlapping communities set as  $C = \{C_i | C_i \neq \emptyset, 1 \leq i \leq k\}$ , where  $C_i \cap C_j = \emptyset$  ( $i \neq j$ ). Based on the definitions above, the problem of community detection using NMFGAAE here is formally formulated as follows.

**Problem statement:** Given a network  $G = (V, E)$ , using NMFGAAE to obtain the community membership matrix  $\mathbf{U} = [\mathbf{U}_{ij}]^{n \times k} \in \mathbb{R}_+^{n \times k}$ , where  $\mathbf{U}_{ij}$  denotes the strength that  $v_i$  belongs to community  $C_j$ . For the given  $v_i$ , its non-overlapping community membership is determined by assigning it into the  $p$ -th community satisfying the condition that  $\mathbf{U}_{ip}$  is the maximum element in  $\mathbf{U}_i$ .

### B. An overview of NMFGAAE framework

The overall framework of NMFGAAE is shown in Fig. 1 and it consists of two main modules: GAAE (i.e., graph attention auto-encoder) and NMF. GAAE module learns node representation  $\mathbf{Z} \in \mathbb{R}^{n \times d}$  of the graph  $G$  by minimizing the reconstruction loss. It takes  $\mathbf{A}$  and a node feature matrix  $\mathbf{X} \in \mathbb{R}^{n \times m}$  as the input. Because we focus on networks with only structure information, the attribute features adopted in most GCNs are unavailable here. However,  $\mathbf{X}$  can be obtained by using any graph representation algorithm (e.g., DeepWalk [63], LINE [64] and DNGR [65]) or simply replaced with the identity matrix  $\mathbf{I}$ , which is suggested in [66]. Based on the learned  $\mathbf{Z}$ , NMF module iteratively factorizes it to obtain community membership matrix  $\mathbf{U} \in \mathbb{R}_+^{n \times k}$ .

GAAE is optimized by minimizing the sum of the GAAE loss and NMF loss. Through this unified framework, we can learn node representations and perform community detection simultaneously, so that GAAE and NMF modules can benefit from each other. In the following two subsections, we will detail GAAE and NMF modules, respectively.



(a) The attention mechanism directed by NMF-based community detection (b) GAT attention mechanism

Fig. 2. Comparison of attention mechanism used in NMFGAAE and GAT

### C. GAAE module

**Encoder.** As shown in Fig. 1, Encoder attempts to learn the node representation  $\mathbf{Z}$  via multiple graph convolutional layers with attention, which can adaptively measure the importance of neighbors to the target node. By borrowing from the attention mechanism used in GAT, we specially design an attention mechanism directed by NMF-based community detection. Specifically, we define the attention coefficient  $\delta_{ij}$  from  $v_j$  to  $v_i$  as

$$\delta_{ij} = a(\mathbf{U}_i \mathbf{W} || \mathbf{U}_j \mathbf{W}), \quad (1)$$

where  $a$  is the attention mechanism using a single-layer feedforward neural network on the concatenation ( $||$ ) of  $\mathbf{U}_i \mathbf{W}$  and  $\mathbf{U}_j \mathbf{W}$  with weight vector  $\mathbf{a} \in \mathbb{R}^{2k}$ , and  $\mathbf{W} \in \mathbb{R}^{k \times k}$  is a weight matrix shared by the linear transformation of each  $\mathbf{U}_i$ . Note that we use the community membership representation  $\mathbf{U}$  instead of node representation  $\mathbf{Z}$  to compute attention coefficients, which is obviously distinct from the strategy used in GAT (Fig. 2).

We believe that this attention mechanism directed by NMF-based community detection can well incorporate the community membership information into the attention coefficients: for a target node  $v_i$  and  $\forall v_j \in \mathcal{N}_i$ , if  $\mathbf{U}_i$  is more similar to  $\mathbf{U}_j$ , then  $v_i$  and  $v_j$  are more likely to be in the same community and the corresponding attention coefficient  $e_{ij}$  will be bigger, which means that  $v_j$  will contribute more to the representation of  $v_i$ , or vice versa. By this attention mechanism, it can be expected to learn the node representation  $\mathbf{Z}$  which is more friendly to the community detection. This will be confirmed in our experiments.

To make attention coefficients easily comparable, we normalize them across all neighbor nodes with a softmax function:

$$a_{ij} = \text{softmax}_j(\delta_{ij}) = \frac{\exp(\delta_{ij})}{\sum_{q \in \mathcal{N}_i} \exp(\delta_{iq})}. \quad (2)$$

Applying the LeakyReLU activation function, the coefficients can be finally expressed as:

$$a_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}[\mathbf{U}_i \mathbf{W} || \mathbf{U}_j \mathbf{W}]))}{\sum_{q \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\mathbf{a}[\mathbf{U}_i \mathbf{W} || \mathbf{U}_q \mathbf{W}]))}. \quad (3)$$

We stack  $L$  graph attention layers with nonlinear activation function ReLU, and obtain the layer-wise representation for

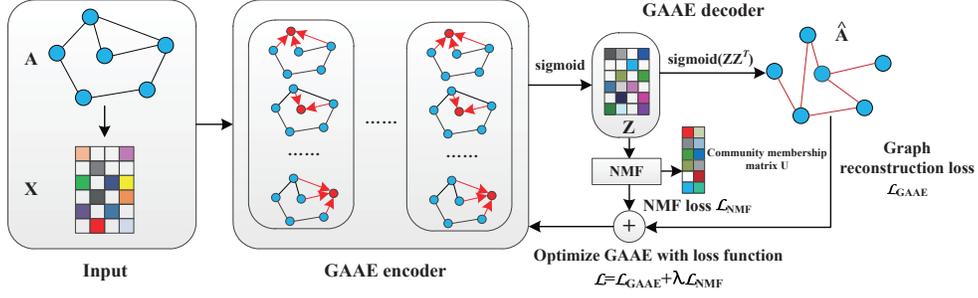


Fig. 1. The framework of NMFGAAE. Note that the input node feature  $\mathbf{X}$  is not the attribute feature matrix, and it can be obtained from  $\mathbf{A}$  by applying any network representation algorithm or simply replaced with the identity matrix  $\mathbf{I}$ .

every node via the following aggregation scheme:

$$\mathbf{Z}_i^{l+1} = \text{ReLU}\left(\sum_{j \in \mathcal{N}_i} a_{ij} \mathbf{W}^l \mathbf{Z}_j^l\right), \quad (4)$$

where  $\mathbf{Z}_i^l$  and  $\mathbf{W}^l$  denotes the representation of  $v_i$  and the linear transformation weight matrix at  $l$ -th layer, respectively. Note that we set  $\mathbf{Z}^0 = \mathbf{X}$ , and at the final layer we use sigmoid function to obtain the latent representation  $\mathbf{Z}$ :

$$\mathbf{Z} = \text{sigmoid}(\mathbf{Z}^{(L-1)}), \quad (5)$$

where  $\mathbf{Z}$  is nonnegative for  $0 \leq \mathbf{Z} \leq 1$ , and will be factorized to uncover community structures in the following NMF module.

Similar to GAT, we also introduce multi-head attention mechanism to independently execute the transformation of Eq. (4) and average their layer-wise latent representations as:

$$\mathbf{Z}_i^{l+1} = \text{ReLU}\left(\frac{1}{H} \sum_{h=1}^H \sum_{j \in \mathcal{N}_i} a_{ij}^h \mathbf{W}_h^l \mathbf{Z}_j^l\right), \quad (6)$$

where  $H$  is the number of attention heads,  $a_{ij}^h$  is the attention coefficient computed by the  $h$ -th attention mechanism ( $a^h$ ), and  $\mathbf{W}_h^l$  is the corresponding linear transformation weight matrix.

**Decoder.** After obtaining  $\mathbf{Z}$ , decoder uses it to reconstruct the structure of  $G$ . Here, we devise the following inner product decoder:

$$\hat{\mathbf{A}} = \text{sigmoid}(\mathbf{Z}\mathbf{Z}^T), \quad (7)$$

where  $\hat{\mathbf{A}} \in \mathbb{R}_+^{n \times n}$  denotes the reconstructed structure matrix of  $G$ . The reconstruction loss  $\mathcal{L}_{\text{DAAE}}$  is measured via the following binary cross entropy function:

$$\mathcal{L}_{\text{DAAE}} = -\frac{1}{n^2} \sum_{i,j} (\mathbf{A}_{ij} \log \hat{\mathbf{A}}_{ij} + (1 - \mathbf{A}_{ij}) \log(1 - \hat{\mathbf{A}}_{ij})). \quad (8)$$

#### D. NMF module

NMF module takes the output  $\mathbf{Z}$  of DAAE module as the input to detect communities from  $G$ . There are many NMF variants with better performance, such as graph regularized NMF [67] and orthogonal NMF [68], but we select the basic NMF model to factorize  $\mathbf{Z}$  to obtain the community membership matrix  $\mathbf{U} \in \mathbb{R}_+^{n \times k}$  and the community feature matrix  $\mathbf{V} \in \mathbb{R}_+^{k \times d}$ . In this way, we want to prove that even the combination of the basic NMF and GAAE also can achieve the

goal of boosting the performance of NMF-based community detection. The corresponding NMF loss  $\mathcal{L}_{\text{NMF}}$  is denoted as

$$\mathcal{L}_{\text{NMF}} = \|\mathbf{Z} - \mathbf{U}\mathbf{V}\|_F^2, \quad (9)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are both nonnegative.

#### E. Optimization

To train the whole NMFGAAE model, we merge  $\mathcal{L}_{\text{GAAE}}$  and  $\mathcal{L}_{\text{NMF}}$  into a unified loss function  $\mathcal{L}$

$$\mathcal{L} = \mathcal{L}_{\text{GAAE}} + \lambda \mathcal{L}_{\text{NMF}}, \quad (10)$$

where  $\lambda$  is a tradeoff parameter. By minimizing  $\mathcal{L}$ , it can not only optimize GAAE, but also can simultaneously optimize NMF. Essentially, this is a joint optimization process, which can make GAAE and NMF benefit from each other. Namely, GAAE can output more friendly node latent representation  $\mathbf{Z}$  for NMF community detection module, meanwhile NMF module also can well improve the accuracy of attention coefficients in GAAE by Eq. (3).

Obviously, minimizing  $\mathcal{L}$  is a non-convex constrained optimization problem, which is hard to obtain its close-form solution. In view of this, we adopt the following alternating optimization strategy to solve this problem.

**Optimize GAAE with NMF fixed.** When NMF related variables  $\mathbf{U}$  and  $\mathbf{V}$  are fixed as constants, we use  $\mathcal{L}$  to optimize GAAE and update its network parameters:  $\mathbf{W}$  and  $\mathbf{W}_h^l$  using the standard gradient decent:

$$\mathbf{W} = \mathbf{W} - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{W}}, \quad (11)$$

$$\mathbf{W}_h^l = \mathbf{W}_h^l - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{W}_h^l}, \quad (12)$$

where  $\alpha$  is the learning rate, and the gradients  $\frac{\partial \mathcal{L}}{\partial \mathbf{W}}$  and  $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_h^l}$  can be computed by the back-propagation algorithm widely used in DNNs.

**Optimize NMF with GAAE fixed.** When GAAE is fixed, minimizing  $\mathcal{L}$  is equivalent to:

$$\begin{aligned} \min \mathcal{L}_{\text{NMF}} &= \|\mathbf{Z} - \mathbf{U}\mathbf{V}\|_F^2, \\ \text{s.t. } &\mathbf{U} \geq 0, \mathbf{V} \geq 0. \end{aligned} \quad (13)$$

To solve this constrained optimization problem, we use the Lagrange multiplier method and define the following Lagrange

function:

$$\mathcal{F} = \|\mathbf{Z} - \mathbf{U}\mathbf{V}\|_F^2 + \text{tr}(\Phi\mathbf{U}^T) + \text{tr}(\Psi\mathbf{V}^T), \quad (14)$$

where  $\Phi$  and  $\Psi$  are the Lagrange multipliers. The derivatives of  $\mathcal{F}$  with respect to  $\mathbf{U}$  and  $\mathbf{V}$  are:

$$\frac{\partial \mathcal{F}}{\partial \mathbf{U}} = -\mathbf{Z}\mathbf{V}^T + \mathbf{U}\mathbf{V}\mathbf{V}^T + \Phi, \quad (15)$$

$$\frac{\partial \mathcal{F}}{\partial \mathbf{V}} = -\mathbf{U}^T\mathbf{Z} + \mathbf{U}^T\mathbf{U}\mathbf{V} + \Psi. \quad (16)$$

Using the Karush-Kuhn-Tucker (KKT) conditions that:  $\Phi_{ij}\mathbf{U}_{ij} = 0$  and  $\Psi_{ij}\mathbf{V}_{ij} = 0$ , we can derive the following multiplicative update rules for  $\mathbf{U}$  and  $\mathbf{V}$ :

$$\mathbf{U} = \mathbf{U} \odot \frac{\mathbf{Z}\mathbf{V}^T}{\mathbf{U}\mathbf{V}\mathbf{V}^T}, \quad (17)$$

$$\mathbf{V} = \mathbf{V} \odot \frac{\mathbf{U}^T\mathbf{Z}}{\mathbf{U}^T\mathbf{U}\mathbf{V}}, \quad (18)$$

where  $\odot$  and  $\frac{(\cdot)}{(\cdot)}$  are the element-wise product and division operators, respectively. Through iteratively execute Eq. (17) and Eq. (18) until  $\mathcal{L}_{\text{NMF}}$  converge, we can obtain the final  $\mathbf{U}$  to uncover community structure of  $G$ . It should be pointed out that: after each complete update for NMF module, the output  $\mathbf{U}$  and  $\mathbf{V}$  should be preserved as the initial values of the next update. Only by this way,  $\mathbf{U}$  can become better and better during the whole optimization process. Based on the analysis above, we summarize NMFGAAE community detection algorithm in Algorithm 1.

---

**Algorithm 1:** NMFGAAE community detection algorithm

---

**Input:**  $G = (V, E), k, \lambda;$

**Output:** Communities set  $C = \{C_1, C_2, \dots, C_k\};$

```

1 repeat
2    $\mathbf{Z} \leftarrow$  Optimize GAAE module via Eq. (11) and Eq. (12);
3    $\mathbf{U}, \mathbf{V} \leftarrow$  Optimize NMF module via iteratively executing
   Eq. (17) and Eq. (18);
4 until  $\mathcal{L}$  converges or exceed the maximum iterations;
5 for  $v_i \in V$  do
6    $q = \text{argmax}_p \mathbf{U}_{ip};$ 
7    $C_q = C_q^p \cup \{v_i\};$ 
8 return  $C;$ 

```

---

### F. Time complexity analysis

From Algorithm 1, we can observe that it is composed of two stages: iteratively optimizing GAAE and NMF, and extracting communities from  $\mathbf{U}$ . It is clear that the time complexity of the second stage is  $\mathcal{O}(nk)$ , and the first stage is the most time consuming. Assuming that the number of outer iterations is  $T$  and the number of inner iterations for NMF is  $t$ , the time complexity of the first stage can be expressed as  $\mathcal{O}(T(Hnmd + H|E|d + tndk))$ , where  $\mathcal{O}(T(Hnmd + H|E|d))$  and  $\mathcal{O}(Ttndk)$  are the time complexities of GAAE and NMF modules, respectively. By summing up and considering that  $m \ll n$ ,  $d \ll n$ ,  $k \ll n$  and  $H$  often takes a very small value, we can deduce that the overall time complexity of NMFGAAE is proportional to  $n$  and  $|E|$ . In real-world networks,  $n$  is often

far fewer than  $|E|$ , so the time cost of NMFGAAE will be largely dominated by  $|E|$ .

## IV. EXPERIMENTS

In this section, to validate the effectiveness of our proposed method NMFGAAE, we conduct extensive experiments by comparing it with state-of-the-art methods on several synthetic and real-world networks. All methods are implemented in Python 3.5 and all of experiments are conducted on a PC with 64bits Windows 7 system, 3.4G Intel Core i7-6700 CPU and 32GB RAM.

### A. Baselines

Our basic hypothesis in this paper is that NMFGAAE is a nonlinear NMF-based community detection method powered by graph attention auto-encoder, and is expected to achieve better performance than existing NMF-based community detection methods, almost all of which are linear. In view of this, we select eight representative NMF-based community detection methods as baselines. Besides, to further validate the superiority of NMFGAAE, we also compare it with three representative network representation based community detection methods. These baselines are detailed as follows.

The NMF-based community detection methods include:

- NMF. NMF is a module of NMFGAAE, and it can also be used to detect communities independently by replacing  $\mathbf{Z}$  in Eq. (9) with  $\mathbf{A}$ , i.e.,  $\|\mathbf{A} - \mathbf{U}\mathbf{V}\|_F^2$ .
- SNMF. SNMF is the symmetric NMF, which factorizes  $\mathbf{A}$  into the product of  $\mathbf{U}$  and  $\mathbf{U}^T$  with the objective function:  $\|\mathbf{A} - \mathbf{U}\mathbf{U}^T\|_F^2$ . It has been adopted for community detection in [19], [20].
- ONMF. ONMF is a variant of NMF by enforcing orthogonal constraints on community membership matrix  $\mathbf{U}$ , i.e.,  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ . In [48], ONMF is proven to be equivalent to K-means.
- HPNMF [69]. HPNMF is based on symmetric NMF with graph regularized constraint. Through this framework, it can integrate structure information with node homophily information to detect communities.
- DANMF [35]. DANMF is based on deep NMF. It consists of an encoder component and a decoder component, which enables it to learn the hierarchical mappings between the original network and the final community structure.
- M-NMF [70]. M-NMF is a modularized NMF based community detection method. It provides a unified framework to jointly optimize NMF-based representation learning model and modularity-based community detection model.
- NSED [49]. NSED is similar to DANMF and also comprises an encoder component and a decoder component, but every component is still based on shallow NMF with only one layer factorization structure.
- BigClam [29]. BigClam is a cluster affiliation model. It uncovers community structure by incorporating NMF into the probabilistic generative framework, and meanwhile provides an efficient optimization algorithm.

The network representation based methods include:

- DeepWalk [63]. DeepWalk is a random walks based network representation learning method. It learns node representations by using local information obtained from truncated random walks.
- LINE [64]. LINE is an edge modeling based method. Through modeling a joint probability distribution and a conditional probability distribution on connected nodes, it can learn the first-order and the second-order proximity preserving node representations.
- DNNGR [65]. DNNGR is a deep learning based method. It first constructs the high-dimensional positive pointwise mutual information matrix representations, and then applies the stacked denoising auto-encoder to obtain node representations.

Note that some NMF-based methods (e.g., BigClam) claim to detect overlapping communities, but they are also very straightforward to detect non-overlapping communities by following the steps of the second stage in Algorithm 1. NMF-based methods can directly obtain community detection results without seeking help from other clustering algorithms. For the network representation based methods, we utilize standard K-means algorithm to cluster the final node representations to obtain community detection results. For the Python implementations of SNMF, DANMF, M-NMF, NSED, BigClam and DeepWalk, we refer to the open-source Python framework for unsupervised learning on Graphs: Karate Club [71].

### B. Evaluation metrics

We select four widely used metrics to evaluate the performance of community detection. For the networks with ground-truth communities, we use accuracy (ACC), adjusted Rand index (ARI) and normalized mutual information (NMI). For the networks without ground-truth communities, we use modularity Q. Detailed definitions for these metrics can be found in [72], which is a comprehensive survey on metrics for community detection.

According to the definitions, we can obtain  $ACC \in [0, 1]$ ,  $ARI \in [-1, 1]$ ,  $NMI \in [0, 1]$  and  $Q \in [-1, 1]$ . For all these metrics, larger values indicate better performance.

### C. Parameter settings

For fair comparisons, the parameters of every baseline are set to be their optimal values. Specifically, for HPNMF, we set its regularization parameter  $\lambda$  to be 1. For DANMF, its regularization parameter  $\lambda$  and layer configuration are set to be 1 and  $n-256-128-k$ , respectively. For BigClam, we set its  $\ell_1$  regularization parameter  $\lambda$  to be 10. For DeepWalk, we set walks per vertex  $\gamma = 80$ , window size  $\omega = 10$  and walk length  $\ell = 40$ . For LINE, the number of negative samples  $K$  is set to be 5. For DNNGR, its layer configuration is set to be  $n-256-128-64$ . The size of node representations of all the network representation based methods is uniformly set to be 64. For NMFGAAE, we set  $\lambda$ ,  $H$ ,  $t$  and the layer configuration of graph encoder to be 1, 3, 50 and  $n-256-64$ , respectively. Note that the number of communities  $k$  for all methods is set to be the same value. Under the corresponding optimal settings for different methods, the experiments are repeated for 10 times and the average results are reported.

### D. Experiments on synthetic networks

1) *Datasets*. We utilize the well-known LFR tool [73] to generate benchmark synthetic networks with ground-truth communities. The setting of parameters of LFR networks is shown in Table II. By respectively varying  $n$  and  $mu$ , and fixing the other parameters, we generate 10 synthetic networks to test the performance of community detection. Because we only focus on non-overlapping community detection,  $on$  and  $om$  are both set to be 0.

TABLE II  
PARAMETERS OF LFR NETWORKS

Parameter	Explanation	Value
$n$	Number of nodes	To vary
$mu$	Mixing parameter	To vary
$on$	Number of overlapping nodes	0
$om$	Number of overlapping memberships	0
$d_{avg}$	The average degree	4
$d_{max}$	The maximum degree	15
$\lambda_1$	Exponent for node degree distribution	2
$\lambda_2$	Exponent for community size distribution	1
$C_{min}$	Minimum community size	50
$C_{max}$	Maximum community size	100

3) *Experiment results and analysis*. In our experiments, we first generate 5 synthetic networks by varying  $n$  from 1000 to 5000 with step size 1000 and fixing  $mu$  at 0.3. For all methods,  $k$  is set to be the number of ground-truth communities. The results on these networks are shown in Table III. We can observe that our proposed method NMFGAAE performs better than other methods in any case. Specifically, compared with the best NMF-based baseline: HPNMF, the average ACC, ARI and NMI scores of NMFGAAE respectively improve by 10.4%, 11% and 4.5%. Compared with the best network representation based baseline: LINE, the average ACC, ARI and NMI scores of NMFGAAE respectively improve by 7.2%, 14.4% and 5.7%.

We then generate another 5 synthetic networks by varying  $mu$  from 0.1 to 0.3 with step size 0.05 and fixing  $n$  at 3000. The results on these networks are shown in Fig. 3. As we can see, when  $mu$  increases, the performance of all methods tend to decrease. This is because the community structures become more and more obscure, which makes community detection more and more difficult. However, NMFGAAE still performs better than other methods at any  $mu$ . Even when  $mu = 0.3$  the scores of ACC, ARI and NMI of NMFGAAE are still competitive and satisfactory. These results show that NMFGAAE has better ability to detect communities from networks with more complicated structure.

### E. Experiments on real-world networks

1) *Datasets*. In this part, we select 6 real-world networks without ground-truth communities, including WebKB, Cora, Citeseer, Polblog, Blogcatalog and Pubmed. These networks are from LINQS data repository<sup>1</sup>. For the sake of simplicity, the number of communities  $k$  on each network is set to be the number of node labels. The basic information of real-world networks is shown in Table IV.

<sup>1</sup><https://linqs.soe.ucsc.edu/data>

TABLE III  
PERFORMANCE COMPARISON ON SYNTHETIC NETWORKS WITH DIFFERENT  $n$  (BOLD NUMBERS DENOTE THE BEST RESULTS).

$n$	Metrics	NMF	SNMF	ONMF	HPNMF	DANMF	M-NMF	NSED	BigClam	DeepWalk	LINE	DNGR	NMFGAAE
1000	ACC	0.51	0.5	0.49	0.54	0.55	0.23	0.27	0.12	0.25	0.44	0.57	<b>0.59</b>
	ARI	0.29	0.28	0.29	0.33	0.32	0.07	0.09	-0.01	0.08	0.21	0.31	<b>0.37</b>
	NMI	0.39	0.37	0.39	0.42	0.41	0.18	0.20	0.03	0.19	0.35	0.45	<b>0.46</b>
2000	ACC	0.48	0.51	0.51	0.54	0.50	0.24	0.25	0.13	0.14	0.55	0.51	<b>0.56</b>
	ARI	0.30	0.31	0.35	0.34	0.29	0.09	0.10	0.02	0.03	0.34	0.33	<b>0.36</b>
	NMI	0.47	0.47	0.50	0.52	0.47	0.28	0.30	0.14	0.19	0.52	0.51	<b>0.52</b>
3000	ACC	0.49	0.51	0.51	0.53	0.53	0.14	0.23	0.11	0.16	0.59	0.49	<b>0.61</b>
	ARI	0.31	0.31	0.32	0.34	0.33	0.04	0.10	0.02	0.04	0.36	0.31	<b>0.39</b>
	NMI	0.44	0.50	0.50	0.54	0.53	0.24	0.34	0.20	0.19	0.57	0.51	<b>0.58</b>
4000	ACC	0.45	0.48	0.52	0.54	0.52	0.14	0.26	0.11	0.11	0.60	0.52	<b>0.61</b>
	ARI	0.29	0.30	0.34	0.36	0.34	0.04	0.11	0.02	0.01	0.38	0.33	<b>0.40</b>
	NMI	0.41	0.52	0.55	0.58	0.56	0.26	0.38	0.22	0.18	0.59	0.51	<b>0.60</b>
5000	ACC	0.51	0.53	0.51	0.53	0.53	0.25	0.25	0.10	0.09	0.58	0.55	<b>0.59</b>
	ARI	0.31	0.34	0.32	0.35	0.34	0.11	0.11	0.02	0.01	0.38	0.31	<b>0.39</b>
	NMI	0.53	0.56	0.53	0.59	0.58	0.40	0.39	0.22	0.18	0.59	0.52	<b>0.61</b>

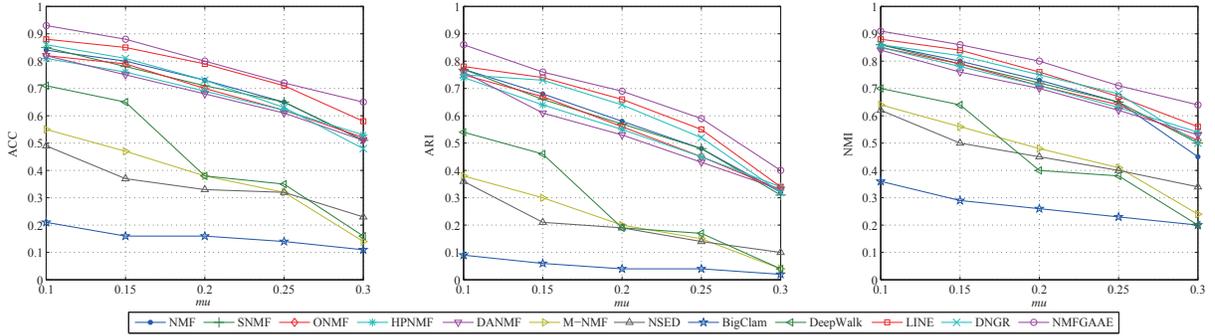


Fig. 3. Performance comparison on synthetic networks with different  $\mu$  in terms of ACC (Left), ARI (Middle) and NMI (Right).

TABLE V  
PERFORMANCE COMPARISON ON REAL-WORLD NETWORKS IN TERMS OF MODULARITY Q (BOLD NUMBERS DENOTE THE BEST RESULTS)

Networks	NMF	SNMF	ONMF	HPNMF	DANMF	M-NMF	NSED	BigClam	DeepWalk	LINE	DNGR	NMFGAAE
WebKB	0.57	0.63	0.64	0.68	0.35	0.62	0.66	0.30	0.61	0.45	0.58	<b>0.73</b>
Cora	0.53	0.52	0.62	0.71	0.71	0.70	0.69	0.47	0.65	0.61	0.71	<b>0.74</b>
Citeseer	0.59	0.61	0.58	0.65	0.65	0.63	0.59	0.58	0.49	0.67	0.65	<b>0.72</b>
Polblog	<b>0.43</b>	<b>0.43</b>	<b>0.43</b>	<b>0.43</b>	0.42	<b>0.43</b>	<b>0.43</b>	0.05	0.03	<b>0.43</b>	0.41	<b>0.43</b>
Blogcatalog	0.07	0.13	0.06	0.13	0.07	0.15	0.15	0.01	0.06	0.15	0.12	<b>0.17</b>
Pubmed	0.45	0.42	0.50	0.53	0.52	0.51	0.39	0.53	0.32	0.53	0.51	<b>0.55</b>

TABLE IV  
STATISTICS OF REAL-WORLD NETWORKS

Networks	$n$	$ E $	$k$
WebKB	877	1388	4
Cora	2708	5278	7
Citeseer	3327	4552	6
Polblog	1490	16715	2
Blogcatalog	10312	333983	39
Pubmed	19717	44324	3

2) *Experiment results and analysis.* We run every method on the real-world networks. Since these networks have no ground-truth communities, we select modularity Q metric to evaluate the performance. The results are shown in Table V. Although NMFGAAE achieves the same performance as most other methods on Polblog network, it performs the best on other networks and even exceed by a certain margin when comparing with NMF-based baselines. For example, on Citeseer network, the modularity Q scores of NMFGAAE are respectively 22%,

18%, 24.1%, 10.8%, 10.8%, 14.3%, 22% and 24.1% higher than those of NMF, SNMF, ONMF, HPNMF, DANMF, M-NMF, NSED and BigClam. Similar results can be found on WebKB, Cora and Pubmed networks. These results verify again that NMFGAAE indeed can boost the performance of NMF-based methods for community detection.

#### F. Attention mechanism comparison

In NMFGAAE, we adopt specially designed attention mechanism directed by NMF-based community detection (Fig. 2 (a)) instead of GAT mechanism (Fig. 2 (b)). Results on synthetic and real-world networks have demonstrated its effectiveness. To further validate its superiority, we conduct comparative experiments between these two attention mechanisms on real-world networks. For convenience, we call the method using GAT attention mechanism as NMFGAT. The experiments results are shown in Fig. 4.

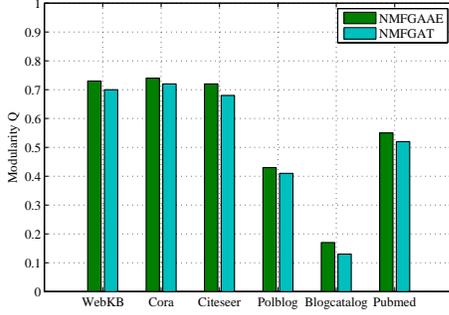


Fig. 4. NMFGAAE vs. NMFGAT

As we can see from Fig. 4, NMFGAAE performs better than NMFGAT on each network. The score of modularity  $Q$  of NMFGAAE increases by an average of 6.1%. This means that our proposed attention mechanism can better improve the performance of NMFGAAE than GAT attention mechanism.

### G. Joint optimization analysis

In NMFGAAE, we jointly optimize GAAE and NMF models, which can make GAAE and NMF benefit from each other. One of the most obvious intuitions is that the output  $\mathbf{Z}$  of GAAE will be more discriminative, because it is improved by NMF community detection model simultaneously. To verify this intuition visually, we respectively run NMFGAAE on real-world networks, and apply t-SNE tool [74] to visualize  $\mathbf{Z}$  in two-dimensional space. As a comparison, we run GAAE individually (i.e., it is not jointly optimized with NMF), and also visualize its output  $\mathbf{Z}$ . Note that here we train GAAE only using the graph reconstruction loss. Because the results of different networks have similar trends, here we just show two networks with fewer communities: WebKB and Polblog in Fig. 5 and Fig. 6, respectively.

From Fig. 5 and Fig. 6, we can observe that the visualization results of  $\mathbf{Z}$  in NMFGAAE both have shown clear boundaries of clusters (i.e., communities), but those in GAAE are not distinguishable: points with different color are extensively mixed together. Bad node representations will deteriorate the performance of community detection. By applying NMF to  $\mathbf{Z}$  of GAAE, the modularity  $Q$  scores on WebKB and Polblog are 0.55 and 0.36, respectively, which are both inferior to those of NMFGAAE as shown in Table V. These all fully demonstrate that optimizing GAAE and NMF jointly indeed can obtain better performance.

### H. Parameter $\lambda$ sensitivity analysis

There is one hyperparameter  $\lambda$  in the loss function of our NMFGAAE model. It is used to trade off GAAE reconstruction loss and NMF loss. To study its effects to the performance of NMFGAAE under different settings, we vary  $\lambda$  in the range of  $\{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2, 10^3\}$  and evaluate the corresponding performance on real-world networks. The results are shown in Fig. 7. As we can see, although there is degradation of performance as  $\lambda \geq 10$ , the degradation is mild and

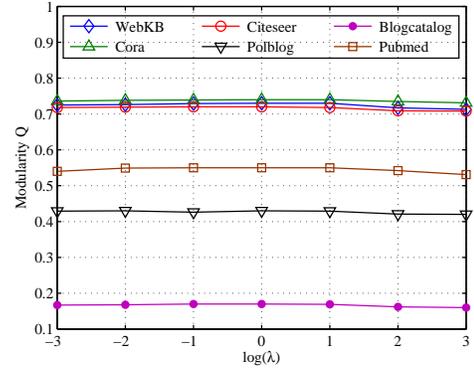


Fig. 7. The performance of NMFGAAE with different  $\lambda$

trivial. On the whole, when  $\lambda$  increases, NMFGAAE achieves consistent good performance. This shows that NMFGAAE is not sensitive to the exact value of  $\lambda$  and it can obtain satisfactory performance for a range of  $\lambda$ , e.g.,  $[10^{-3}, 10^1]$ . In all of our experiments, we equally treat the contributions of GAAE reconstruction loss and NMF loss, and thus set  $\lambda$  to be 1.

### I. Running efficiency analysis

As described in Algorithm 1, the results of community detection can be obtained by executing update rules iteratively. To investigate how many iterations NMFGAAE needs to take to obtain satisfactory results, we run NMFGAAE on each real-world network and explore how performance varies with the number of iterations. To better illustrate this problem, we select the best NMF-based baseline HPNMF as the competitor. The results are depicted in Fig. 8. As we can see, on every network NMFGAAE only needs to take about 10 iterations to obtain stable and satisfactory results, but HPNMF needs at least 30 iterations, and even more than 100 iterations on Blogcatalog and PubMed networks. Actually, most existing NMF-based methods for community detection need to take many iterations to get the best results on these networks. However, nonlinear NMFGAAE utilizes graph attention auto-encoder to significantly reduce the number of iterations.

To further analyze the running efficiency of NMFGAAE, we specially compare it with NMF-based baselines in term of the runtime on two larger networks: Blogcatalog and Pubmed. In the previous experiments, we have found that DANMF is very time consuming, due to its multiple factorization structure. Therefore, to better illustrate the comparison results, in this experiment we do not select DANMF as the competitor. We set the number of iterations for NMFGAAE on Blogcatalog and Pubmed networks to be 10, and that of NMF-based baselines to be 100. Experiment results are shown in Fig. 9.

As we can see from Fig. 9, in terms of runtime, on Blogcatalog network, NMFGAAE is inferior to NMF, SNMF, NSED and BigClam, but performs better than ONMF, HPNMF and M-NMF. On PubMed network, NMFGAAE performs the best. Overall, NMFGAAE is very efficient, because it requires fewer iterations. On Blogcatalog and PubMed networks, it

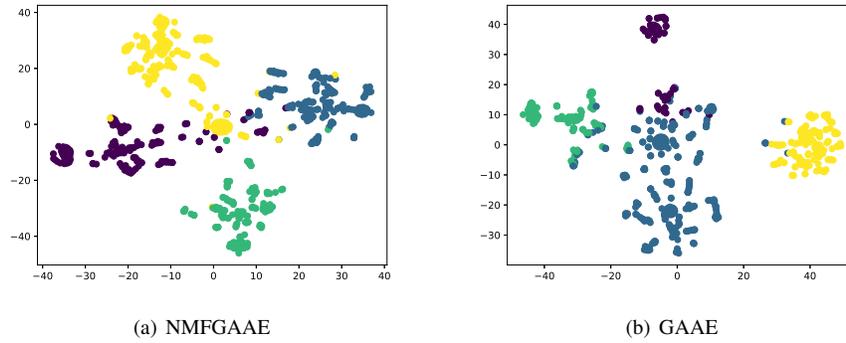


Fig. 5. 2D visualization of  $Z$  on WebKB

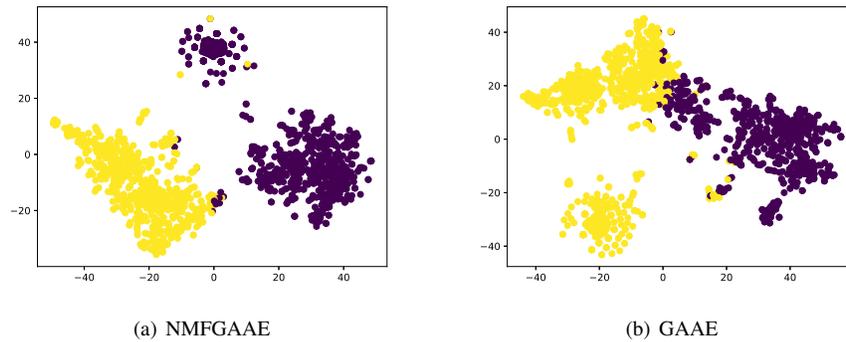


Fig. 6. 2D visualization of  $Z$  on Polblog

only takes about 100 and 800 seconds to converge to obtain satisfactory results, respectively. Besides, as we analyze in Section III.F, the time cost of NMFGAAE is largely dominated by the number of edges  $|E|$ . This also shows that NMFGAAE can be expected to run faster on networks with fewer edges. As we can see, NMFGAAE run faster on PubMed than on Blogcatalog, because the number of edges on Blogcatalog network is about 7 times more than that of PubMed network.

### V. CONCLUSIONS

NMF-based methods are widely applied to deal with the problem of community detection in complex networks. However, most of them are linear and still need to further boost the performance, especially when facing networks with diversified structure information. In this paper, we devise a nonlinear NMF-based method named NMFGAAE, which aims to combine NMF community detection and graph attention auto-encoder models to achieve better performance. We conduct extensive experiments on synthetic and real-world networks. The results show that NMFGAAE is effective and efficient, and consistently performs better than state-of-the-art NMF-based community detection methods. In the future, we think there are two related topics which are worth further study:

- In essence, NMFGAAE provides us a general framework to boost the performance of NMF-based community detection by incorporating GNNs, and its two main components NMF and GAAE can be replaced with other advanced NMF and GNN models, respectively. There

should be more combinations that can be expected to obtain better performance.

- NMF-based community detection in attributed networks is also popular, but NMFGAAE here focuses on networks with only structure information. Therefore, extending NMFGAAE to attributed networks is promising. This leads to the study of how to effectively fuse attributed information with structure information and how to learn the relevances of attributes to the communities.

### REFERENCES

- [1] S. Fortunato, D. Hric, Community detection in networks: a user guide, *Physics Reports* 659 (2016) 1-44.
- [2] M.A. Javed, M.S.Younis, S. Latif, J. Qadir, A. Baig, Community detection in networks: a multidisciplinary review, *Journal of Network and Computer Applications*, 108 (2018) 87-111.
- [3] M. Girvan, M.E.J. Newman, Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, 99 (12) (2002) 7821-7826.
- [4] G. Zhao, M. L. Lee, W. Hsu, W. Chen, H.J. Hu, Community-based user recommendation in uni-directional social networks, in: *Proceedings of the 22nd ACM international conference on Information and Knowledge Management*, 2013, pp. 189-198.
- [5] Y.C. Chang, K.T. Lai, S.C.T. Chou, M. Chen, Mining the networks of telecommunication fraud groups using social network analysis, in: *Proceedings of the 9th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2017, pp. 1128-1131.
- [6] A.P. Rodríguez, C.O. Gómez, F.M. Anegón. Detecting, identifying and visualizing research groups in co-authorship networks, *Scientometrics* 82 (2010) 307-319.
- [7] S.S. Bhowmick, B.S. Seah, Clustering and summarizing protein-protein interaction networks: A survey, *IEEE Transactions on Knowledge and Data Engineering*, 28 (3) (2016) 638-658.

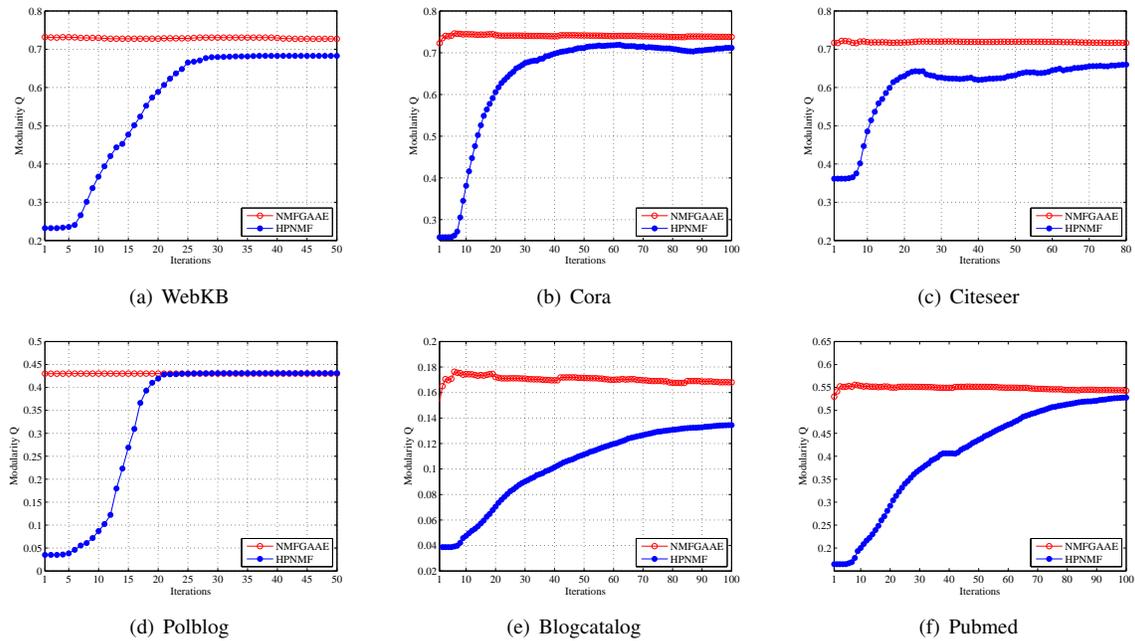


Fig. 8. The performance with respect to iterations on real-world networks

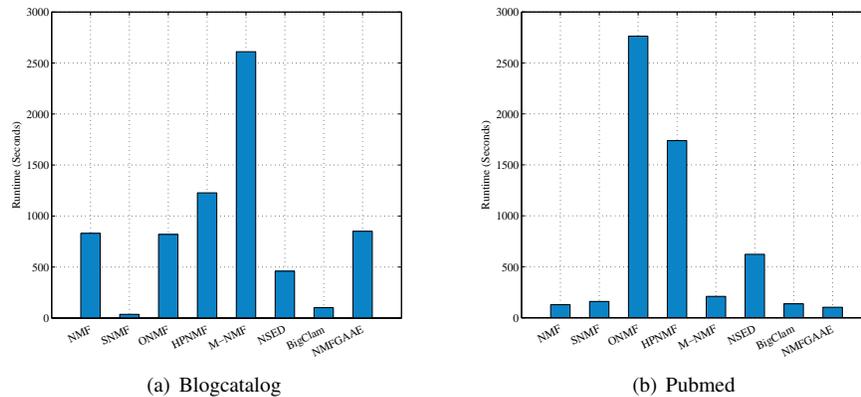


Fig. 9. Runtime comparison

- [8] M.E.J. Newman, Modularity and community structure in networks, *Proceedings of the National Academy of Sciences*, 103 (23) (2006) 8577-8582.
- [9] M.E.J. Newman, Equivalence between modularity optimization and maximum likelihood methods for community detection, *Physical Review E* 94 (2016) 052315.
- [10] E. Abbe, Community detection and stochastic block models: recent developments, *Journal of Machine Learning Research* 18 (2018) 1-86.
- [11] A. Jonnalagadda, L. Kuppasamy, A survey on game theoretic models for community detection in social networks, *Social Network Analysis and Mining*, 6 (1) (2016) 1-24.
- [12] S.E. Garza, S.E. Schaeffer, Community detection with the label propagation algorithm: a survey, *Physica A: Statistical Mechanics and its Applications* 534 (2019) 122058.
- [13] F.Z. Liu, S. Xue, J. Wu, C. Zhou, W.B. Hu, C. Paris, S. Nepal, J. Yang, P.S. Yu, Deep learning for community detection: progress, challenges and opportunities, in: *Proceedings of 29th International Joint Conference on Artificial Intelligence*, 2020, pp. 4981-4987.
- [14] I. Psorakis, S. Roberts, M. Ebdon, B. Sheldon, Overlapping community detection using Bayesian non-negative matrix factorization, *Physical Review E* 83 (2011) 066114.
- [15] Y. Zhang, D.Y. Yeung, Overlapping community detection via bounded nonnegative matrix tri-factorization, in: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2012, pp. 606-614.
- [16] F.H. Ye, C. Chen, Z.B. Zheng, R.H. Li, J.X. Yu, Discrete overlapping community detection with pseudo supervision, in: *Proceedings of the 19th IEEE International Conference on Data Mining*, 2019, pp. 708-717.
- [17] X. Liu, W. Wang, D. He, P. Jiao, D. Jin, C.V. Cannistraci, Semi-supervised community detection based on non-negative matrix factorization with node popularity, *Information Sciences*. 381 (2017) 304-321.
- [18] L. Yang, X.C. Cao, D. Jin, X. Wang, D. Meng, A unified semi-supervised community detection framework using latent space graph regularization, *IEEE Transactions on Cybernetics* 45 (11) (2015) 2585-2598.
- [19] F. Wang, T. Li, X. Wang, S.H. Zhu, C. Ding, Community discovery using nonnegative matrix factorization, *Data Mining and Knowledge Discovery* 22 (3) (2011) 493-521.
- [20] X. Luo, Z.G. Liu, L. Jin, Y. Zhou, M.C. Zhou, Symmetric nonnegative matrix factorization-based community detection models and their convergence analysis, *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2020.3041360, (2020).
- [21] Y. Li, C.F. Sha, X. Huang, Y.C. Zhang, Community detection in attributed graphs: An embedding approach, in: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018, pp. 338-345.
- [22] Z.C. Huang, Y.M. Ye, X.T. Li, F. Liu, H.J. Chen, Joint weighted non-negative matrix factorization for mining attributed graphs, in: *Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2017, pp. 368-380.

- [23] C.B. He, H. Liu, Y. Tang, S.Y. Liu, X. Fei, Q.W. Cheng, H.C. Li, Similarity preserving overlapping community detection in signed networks, *Future Generation Computer Systems* 116 (2021) 275-290.
- [24] J.Y. Shi, K.T. Mao, H. Yu, S.M. Yiu, Detecting drug communities and predicting comprehensive drug-drug interactions via balance regularized semi-nonnegative matrix factorization, *Journal of Cheminformatics* 11 (1) (2019) 1-16.
- [25] V. Gligorijević, Y. Panagakis, S. Zafeiriou, Non-negative matrix factorizations for multiplex network analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (4) (2019) 928-940.
- [26] X.K. Ma, D. Dong, Q. Wang, Community detection in multi-layer networks using joint nonnegative matrix factorization, *IEEE Transactions on Knowledge and Data Engineering* 31 (2) (2019) 273-286.
- [27] X.K. Ma, D. Dong, Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks, *IEEE Transactions on Knowledge and Data Engineering* 29 (5) (2017) 1045-1058.
- [28] X.K. Ma, B.H. Zhang, C.Z. Ma, Z.Y. Ma, Co-regularized nonnegative matrix factorization for evolving community detection in dynamic networks, *Information Sciences* 528 (2020) 265-279.
- [29] J. Yang, J. Leskovec, Overlapping community detection at scale: a nonnegative matrix factorization approach, in: *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, 2013, pp. 587-596.
- [30] C.B. He, X. Fei, H.C. Li, Y. Tang, H. Liu, S.Y. Liu, Improving NMF-based community discovery using distributed robust nonnegative matrix factorization with SimRank similarity measure, *The Journal of Supercomputing* 74 (10) (2018) 5601-5624.
- [31] L. Yang, X.H. Cao, D.X. He, C. Wang, X. Wang, W.X. Zhang, Modularity based community detection with deep learning, in: *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2016, pp. 2252-2258.
- [32] R. Ibrahim, D. Gleich, Nonlinear diffusion for community detection and semi-supervised learning, in: *Proceedings of 30th World Wide Web Conference*, 2019, pp. 739-750.
- [33] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, B.W. Schuller, A deep matrix factorization method for learning attribute representations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (3) (2017) 417-429.
- [34] J.Q. Li, G.X. Zhou, Y.N. Qiu, Y.J. Wang, Y. Zhang, S.L. Xie, Deep graph regularized non-negative matrix factorization for multi-view clustering, *Neurocomputing* 390 (2020) 108-116.
- [35] F.H. Ye, C. Chen, Z.B. Zheng, Deep autoencoder-like nonnegative matrix factorization for community detection, in: *Proceedings of 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 1393-1402.
- [36] S. Maekawa, K. Takeuchi, M. Onizuka, Non-linear Attributed Graph Clustering by Symmetric NMF with PU Learning, *arXiv:1810.00946* (2018).
- [37] W. Rawat, Z.H. Wang, Deep convolutional neural networks for image classification: a comprehensive review, *Neural Computation* 29 (9) (2017) 352-2449.
- [38] A.B. Nassif, I. Shahin, I. Attili, M. Azzeh, K. Shaalan, Speech recognition using deep neural networks: A systematic review, *IEEE Access* 7 (2019) 19143-19165.
- [39] J.J. Zhang, C.Q. Zong, Deep neural networks in machine translation: an overview, *IEEE Intelligent Systems* 30 (5) (2015) 16-25.
- [40] B. Yang, X. Fu, N.D. Sidiropoulos, M. Hong, Towards k-means-friendly spaces: simultaneous deep learning and clustering, in: *Proceedings of the 34th International Conference on Machine Learning*, 2017, pp.3861-3870.
- [41] J. Xie, Girshick, R. Girshick, A. Farhadi, Unsupervised deep embedding for clustering analysis, In: *Proceedings of the 33rd International Conference on Machine Learning*, 2016, pp. 478-487.
- [42] S. Affeldt, L. Labiod, M. Nadif, Spectral clustering via ensemble deep autoencoder learning, *Pattern Recognition* 108 (2020) 107522.
- [43] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: *Proceedings of the 5th International Conference on Learning Representations*, 2017, pp. 1-14.
- [44] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, in: *Proceedings of the 6th International Conference on Learning Representations*, 2018, pp. 1-12.
- [45] Z.H. Wu, S. Pan, F.W. Chen, G.D Long, C.Q Zhang, P.S. Yu, A comprehensive survey on graph neural networks, *IEEE Transactions on Neural Networks and Learning Systems* 32 (1) (2021) 4-24.
- [46] Z.W Zhang, P. Cui, W.W Zhu, Deep learning on graphs: A survey, *IEEE Transactions on Knowledge and Data Engineering*, doi: 10.1109/TKDE.2020.2981333, (2020).
- [47] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (1999) 788-791.
- [48] C. Ding, X. He, H.D. Simon, On the equivalence of nonnegative matrix factorization and spectral clustering, in: *Proceedings of the 5th SIAM International Conference on Data Mining*, 2005, pp. 606-610.
- [49] B.J. Sun, H.W. Shen, J.H. Gao, W.T. Ouyang, X.Q. Cheng, A non-negative symmetric encoder-decoder approach for community detection, in: *Proceedings of the 26th ACM Conference on Information and Knowledge Management*, 2017, pp. 597-606.
- [50] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, *IEEE Transactions on Neural Networks* 20 (1) (2009) 61-80.
- [51] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, Spectral networks and locally connected networks on graphs, in: *Proceedings of 2nd International Conference on Learning Representations*, 2014, pp. 1-14.
- [52] R. Levie, F. Monti, X. Bresson, and M. M. Bronstein, Cayleynets: Graph convolutional neural networks with complex rational spectral filters, *IEEE Transactions on Signal Processing* 67 (1) (2017) 97-109.
- [53] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: *Proceedings of 21st Conference on Neural Information Processing Systems*, 2017, pp. 1024-1034.
- [54] K.Y.L. Xu, W.H. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks, in: *Proceedings of 7th International Conference on Learning Representations*, 2019, pp. 1-17.
- [55] T.N. Kipf, M. Welling, Variational graph auto-encoders, *NIPS Workshop on Bayesian Deep Learning*, 2016, pp. 1-3.
- [56] D.Y. Bo, X. Wang, C. Shi, M.Q. Zhu, E. Lu, P. Cui, Structural deep clustering network, in: *Proceedings of 31st World Wide Web Conference*, 2020, pp. 1400-1410.
- [57] C. Wang, S.R. Pan, R.Q. Hu, G.D. Long, J. Jiang, C.Q. Zhang, Attributed graph clustering: A deep attentional embedding approach, in: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 3670-3676.
- [58] E. Min, X.F. Guo, Q. Liu, G. Zhang, J.J. Cui, J. Long, A survey of clustering with deep learning: from the perspective of network architecture, *IEEE Access* 6 (2018) 39501-39514.
- [59] E. Aljalbout, V. Golkov, Y. Siddiqui, and D. Cremers, Clustering with deep learning: Taxonomy and new methods, *arXiv:1801.07648* (2018).
- [60] X. Peng, J.S. Feng, S.J. Xiao, W.T. Yau, J.T. Zhou, S.F. Yang, Structured autoencoders for subspace clustering, *IEEE Transactions on Image Processing* 27(10) (2018) 5076-5086.
- [61] F. Tian, B. Gao, Q. Cui, E.H. Chen, T.Y. Liu, Learning deep representations for graph clustering, in: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014, pp. 1293-1299.
- [62] X.L. Li, H.Y. Zhang, R. Zhang, Embedding graph auto-encoder with Joint Clustering via Adjacency Sharing, *arXiv:2002.08643* (2020).
- [63] B. Perozzi, R. Al-Rfou, S. Skiena, DeepWalk: online learning of social representations, in: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 701-710.
- [64] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, Q. Mei, LINE: Large-scale information network embedding, in: *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1067-1077.
- [65] S. Cao, W. Lu, Q. Xu, Deep neural networks for learning graph representations, in: *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 2016, pp. 1145-1152.
- [66] A. Bojchevski, S. Günnemann, Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking, in: *Proceedings of the 6th International Conference on Learning Representations*, 2018, pp. 1-13.
- [67] D. Cai, X.F. He, J.W. Han, T.S. Huang, Graph regularized nonnegative matrix factorization for data representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (8) (2011) 1548-1560.
- [68] C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix tri-factorizations for clustering, in: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, pp. 126-135.
- [69] F.H. Ye, C. Chen, Z.Y. Wen, Z.B. Zheng, W.H. Chen, Y.U. Zhou, Homophily preserving community detection, *IEEE Transactions on Neural Networks and Learning Systems*, 31 (8) (2020) 2903-2915.
- [70] X. Wang, P. Cui, J. Wang, J. Pei, W.W. Zhu, S.Q. Yang, Community preserving network embedding, in: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 2017, pp. 203-209.
- [71] B. Rozemberczki, O. Kiss, R. Sarkar, Karate club: an API oriented open-source python framework for unsupervised learning on graphs, in: *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, 2020, pp. 3125-3132.

- [72] T. Chakraborty, A. Dalmia, A. Mukherjee, N. Ganguly, Metrics for community analysis: a survey, *ACM Computing Surveys* 50 (4) (2017) 54.
- [73] A. Lancichinetti, S. Fortunato, F. Radicchi, Benchmark graphs for testing community detection algorithms, *Physical Review E* 78(2) (2008) 046110.
- [74] L.V.D Maaten, G. Hinton, Visualizing data using t-sne, *Journal of Machine Learning Research* 9 (2008) 2579-2605.



**Chaobo He** received his Ph.D., M.S. and B.S. degree from South China Normal University, China in 2014, 2007 and 2004, respectively. He is currently a professor in Zhongkai University of Agriculture and Engineering, China, and is also a visiting scholar in School of Computer Science, South China Normal University, China. His research interests are machine learning and social computing. He has published over 30 papers on international journals and conferences.



**Yulong Zheng** received his BSc degree from School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, China, in 2020. He is currently a MSc student in South China Normal University, China. His research interests are machine learning and social computing.



**Xiang Fei** received a BSc and a PhD from Southeast University China in 1992 and 1999 respectively. After graduation, He worked, as a post-doctorial research fellow, on a number of projects including European IST Programs and EPSRC. He is currently working as a senior lecturer for the School of Computing, Electronics and Maths at Coventry University. His current research interests include machine learning and data mining in cyber-physical systems.



**Hanchao Li** received his BSc degree in Mathematics from University of Warwick in 2013, and MSc degree in Computing from Coventry University in 2015. He is currently a PhD research student in Coventry University and worked on music information retrieval, i.e., data mining in music subject area. His research interests are big data, data mining, machine learning and any mathematics-related researches. He has published several conference and journal papers.



**Yong Tang** received his BS and MSc degrees from Wuhan University in 1985 and 1990 respectively, and PhD degree from University of Science and Technology of China in 2001, all in computer science. He is now a Professor and Dean of the School of Computer Science at South China Normal University (SCNU). He serves as the director of services computing engineering research center of Guangdong province. He was vice dean of School of Information Science and Technology at Sun Yat-sen University, before he joined SCNU in 2009. He has published more than 200 papers and books. As a supervisor he has had more than 40 PhD students and Post Doc researchers since 2003 and more than 100 Master students since 1996. He is a Distinguished Member and the director of Technical Committee on Collaborative Computing of China Computer Federation. He has also served as general or program committee co-chair of more than 10 conferences.