Nonnegative Residual Matrix Factorization for Community Detection

Yulong Pei[1], Cong Liu[2], Chuanyang Zheng[3], and Long Cheng[4]

[1] Eindhoven University of Technology
[2] Shandong University of Technology
[3] Hong Kong Baptist University
[4] University College Dublin
y.pei.1@tue.nl, liucongchina@gmail.com
17251311@life.hkbu.edu.hk, long.cheng@ucd.ie

Abstract. Community detection is one of the most important and challenging problems in graph mining and social network analysis. Nonnegative Matrix Factorization (NMF) based methods have been proved to be effective in the task of community detection. However, real-world networks could be noisy and existing NMF based community detection methods are sensitive to the outliers and noise due to the utilization of the squared loss function to measure the quality of graph regularization and network reconstruction. In this paper, we propose a framework based on the nonnegative residual matrix factorization (NRMF) to overcome this limitation. In this method, a residual matrix, represented by the matrix reconstruction error, is explicitly introduced to capture the impact of outliers and noise. The residual matrix should be sparse intuitively so some sparse regularization can be used to model the sparsity. Specifically, three different types of sparse regularization, i.e., L_0, L_1 and $L_{2,1}$, have been studied. Multiplicative update rules and different thresholding operators are used to learn these lower-rank matrices. Extensive experiments on benchmark networks with and without known communities demonstrate that our framework is more robust so that it outperforms state-of-the-art NMF based approaches in community detection task.

Keywords: Nonnegative residual matrix factorization · Community Detection · L_p regularization.

1 Introduction

Network data is ubiquitous in our daily life, for example, social networks, road networks, and Internet networks. Analyzing these networks is of both theoretical and practical values. Recent years have witnessed numerous network analysis tasks. Among these tasks community detection is one of the most important and challenging problems. A community can be defined as a group of users that (1) interact with each other more frequently than with those outside the group and (2) are more similar to each other than to those outside the group [14].

The research on community detection is beneficial for a variety of real-world applications such as online marketing and recommendation systems.

A variety of approaches have been proposed to solve the problem of community detection in different types of networks, e.g., homogeneous networks, attributed networks, and heterogeneous networks. More details about community detection can be found in the survey paper [6]. Among these methods, Nonnegative Matrix Factorization (NMF) based methods have attracted increasing attention since it has been proved to be effective in detecting communities and has powerful interpretability in clustering. NMF based community detection approaches identify the hidden communities by decomposing the adjacency matrix into lower-rank matrices [21, 23]. From the perspective of statistics, NMF can be viewed as probabilistic [10] and Bayesian models [19] by factorizing the input data into distributions instead of real-value matrices. Nonnegative matrix tri-factorization (NMTF), which extends the standard NMF, factorizes the input matrix into three low-rank matrices so that it can provide more information about the interaction between lower-rank representations. Therefore, it has been utilized to community detection task to identify communities and learn community interaction simultaneously [26, 14]. However, real-world networks could be noisy. Existing NMF based community detection methods are sensitive to the outliers and noise due to the utilization of the squared loss function to measure the quality of graph regularization and network reconstruction [7]. The community detection performance may be degraded by the inevitable noise.

To deal with this issue, we propose a framework based on the nonnegative residual matrix factorization (NRMF) in this study. A residual matrix, represented by the matrix reconstruction error, is explicitly introduced to capture the impact of outliers and noise. The residual matrix should be sparse intuitively so some sparse regularization can be used to model the sparsity. Three different types of norms, i.e., L_0, L_1 and $L_{2,1}$, have been studied for the sparsity modeling. To optimize the NRMF framework with different types of regularization, multiplicative update rules are used to learn the lower-rank matrices and different thresholding operators are used to learn the residual matrix corresponding to different regularization types. To evaluate the performance of the proposed *NRMF*, we conduct community detection experiments on two types of real-world networks: networks with and without known communities.

Our contributions can be summarized as follows:

– We propose a nonnegative residual matrix factorization based framework *NRMF* to detect communities. A residual matrix, which is represented by the matrix reconstruction error, is explicitly introduced to model the impact of outliers and noise. The residual matrix is regularized to capture the sparsity.
– We explore three different regularization types to model the sparsity, i.e., L_0, L_1 and $L_{2,1}$ norms, on the residual matrix. We also propose the multiplicative update rules and different thresholding operators to optimize the objectives.
– We evaluate the effectiveness of *NRMF* using several real-world networks. The experimental results demonstrate that our method outperforms state-of-the-art NMF based methods in community detection.

2 Related Work

Communities are groups of vertices which probably share common properties and/or play similar roles within the graph [6]. Traditional community detection methods aim to partition nodes into different groups such that the number of edges between groups is minimal. For example, cut-based graph partition [9], k-core decomposition [5] and modularity maximization [12].

Recently, NMF-based clustering approaches have also been applied in community detection. NMF as well as NMTF techniques have been used for community detection in networks in [21] where it aims to factorize the adjacency matrix of the given network. Bayesian version of NMF has been proposed in [18] to identify overlapping communities. Since NMTF explicitly models data interactions through an extra latent factor, it provides better interpretability and then has been employed in community detection. NMTF with graph regularization has been used in [14]. REACT [16] employs two NMTF components for community detection and role discovery . Deep Autoencoder-like NMF [24] consists of encoder and decoder and it aims to capture the nonlinear information in networks. Local community detection problem has been studied in [8].

With the rapid development of deep learning, community preserving network embedding has attracted enormous attention recently. Many methods have reported promising results in community detection. For example, M-NMF [22] exploits the consensus relationship between the representations of nodes and community structure. Community Embedding (ComE) [3] is an embedding based method for joint node embedding and community detection. DNGE [15] uses Gaussian embedding to detect communities in dynamic networks.

Table 1: Summary of the notations.

Notation	Description
n	Number of nodes.
e	Number of edges.
c	Number of communities.
$A_{n \times n}$	Adjacency matrix of the given network.
$C_{n \times c}$	Community membership matrix.
$S_{n \times n}$	Residual matrix.
$M_{c \times c}$	Community interaction matrix.
$L_{n \times n}$	graph Laplacian matrix.
α	Trade-off parameter for the residual matrix.
β	Trade-off parameter for graph regularization.

3 Preliminary

We first summarize the notations used in this study in Table 1 and then introduce the backgrounds of the techniques we will use in this paper.

Data could be noisy and there may exist some entries of the data corrupted arbitrarily. To capture such noisy information in the framework of NMF, in [17,

7] a sparse error matrix S has been introduced to capture the sparse corruption. Thus, a robust factorization to approximate the input data matrix A is:

$$A \approx UV^T + S, \tag{1}$$

where S is supposed to be sparse. To guarantee the sparsity, [17] proposed to use L_0 norm to regularize S and [7] utilized L_1 norm as the sparseness constraint. Therefore, the objective function can be defined as

$$\min_{U,V,S} \|A - UV^T - S\|_F^2. \tag{2}$$

Then, we will introduce the formal definitions of different norms:
- L_0 **Norm**: L_0 norm is not really a norm and it is defined as the number of nonzero elements in the given matrix.
- L_1 **Norm**: L_1 norm is the sum of the absolute values of the columns. Given a matrix $A_{n \times n}$, formally

$$L_1(A) = \max_{1 \le j \le n} \sum_{i=1}^{n} |A_{ij}|. \tag{3}$$

L_1 Norm is robust to outliers because it considers the sparsity in rows.
- $L_{2,1}$ **Norm**: $L_{2,1}$ Norm combines L_1 and L_2 norms. Given a matrix $A_{n \times n}$, $L_{2,1}(A)$ is defined as

$$L_{2,1}(A) = \sum_{i=1}^{N} \Big(\sum_{j=1}^{N} |A_{ij}|^2 \Big)^{1/2}. \tag{4}$$

$L_{2,1}$ norm controls the capacity of A and also ensures A to be sparse in rows so it is robust to noise and outliers.

4   The Proposed Method

In this paper, motivated by the robust NMF [17, 7], we propose a general framework, which is less sensitive to noise and outliers, to detect communities. The framework can be formulated as:

$$\min_{C,M,S} \|A - CMC^T - S\|_F^2 + \alpha \cdot \|S\|_p + \beta \cdot Tr(C^T LC), \tag{5}$$

$$s.t.\ C \ge 0, M \ge 0, C^T C = I,$$

where $Tr(V)$ denotes the trance of the matrix V. L is the graph Laplacian and defined as $L = D - A$ where $D_{ii} = \sum_{l} A_{il}$. This framework is flexible to incorporate different types of sparse regularization and constraints. It is worth noting that:
- In this work we only consider undirected networks where the adjacency matrix A is symmetric. This framework can be extended to directed networks straightforwardly by changing the first term to $\|A - CMB^T - S\|_F^2$.

– Empirically we find that adding the interaction matrix M and orthogonal constraint would achieve worse performance so in the following discussion we will simplify the objective function as:

$$\min_{C,M,S} \|A - CC^T - S\|_F^2 + \alpha \cdot \|S\|_p + \beta \cdot Tr(C^T LC), \qquad (6)$$

$$s.t. \ C \geq 0, M \geq 0,$$

– In this work, we exploit three different norms, L_0, L_1 and $L_{2,1}$, to capture the reconstruction error, i.e., q is set to be 0, 1 and 2,1, respectively.

4.1 Optimization

The objective function in Eq. (6) is not convex for all parameters simultaneously. We use the multiplicative update rules to solve this problem due to its good compromise between speed and ease of implementation [11]. We will optimize the objective with respect to one variable while fixing the other variables.

The update rules for matrices C and M (since we will not consider M we can simply set $M = I$) are similar to the standard NMF, which is defined as:

$$C \leftarrow C \circ \frac{(A - S)CM + \beta AC}{CMC^T CM + \beta DC} \qquad (7)$$

$$M \leftarrow M \circ \frac{C(A - S)C}{C^T CMCC^T} \qquad (8)$$

where $\circ$ denotes the element-wise product.

The update rules for the residual matrix S is different from different norms.

L_0 **Regularized Residual**. For L_0 norm, a hard-thresholding update rule has been used in [17]. Formally, it is define as:

$$S_{ij} = \begin{cases} 0 & \text{if } |R_{ij}| \leq \alpha \\ R_{ij} & \text{otherwise} \end{cases} \qquad (9)$$

where $R_{ij} = (A - CC^T)_{ij}$.

L_1 **Regularized Residual**. For L_1 norm, a soft-thresholding update rule has been used in [7]. Formally, it is define as:

$$S_{ij} = \begin{cases} 0 & \text{if } |R_{ij}| \leq \frac{\alpha}{2} \\ R_{ij} - \frac{\alpha}{2} sign(R_{ij}) & \text{otherwise} \end{cases} \qquad (10)$$

where $R_{ij} = (A - CC^T)_{ij}$, $sign(R_{ij})$ is the sign function: if $R_{ij} > 0$, $sign(R_{ij}) = 1$; if $R_{ij} < 0$, $sign(R_{ij}) = -1$; otherwise $sign(R_{ij}) = 0$.

$L_{2,1}$ **Regularized Residual**. $L_{2,1}$ norm is superior to L_0 and L_1 norm because: (1) it can model the sparsity effectively, and (2) it has a simple and efficient solution to solve the optimization problem [13]. Based on [13], we have the update rule for $L_{2,1}$ regularized S as:

$$S \leftarrow S \circ \frac{A - CC^T}{S + \alpha DS}, \qquad (11)$$

where D is the diagonal matrix with the j-th diagonal element which is:

$$D_{jj} = \frac{1}{\|S_j\|_2} \qquad (12)$$

4.2   Computational complexity

Computational complexity. For simplicity, given two matrices $M_{n \times r}$ and $N_{r \times f}$, the computational complexity of the multiplication of M and N is $O(nrf)$. The complexity of updating rules in Algorithm 1 (Line 3 - 5) is $O(n^2c + nc^2 + n^3)$. To update S (Line 6 - 12), the complexity for L_0 and L_1 is $O(n^2 + nc^2)$ and for $L_{2,1}$ is $O(nc^2 + n^3)$. By taking the number of iteration i into consideration, the complexity is $O(i(n^2c + nc^2 + n^3))$.

5   Experiments

To validate the effectiveness of *NRMF* in community detection, we conduct experiments on two types of real-world networks from different domains, i.e., networks with and without known communities. To further exploit the influence of different regularization terms in sparsity modeling, we compare *NRMF* with L_0, L_1 and $L_{2,1}$ regularization.

Algorithm 1 Optimization Algorithm

Input: Adjacency matrix A, number of communities c, trade-off parameter α and β, regularization type q
Output: Community membership matrix C, community interaction matrix M and residual matrix S
 1: Initialize C, B, and S
 2: **while** not converge **do**
 3:     Calculate graph Laplacian L
 4:     Update C according to Eq. (7)
 5:     Update M according to Eq. (8)
 6:     **if** $q = 0$ **then**
 7:       Update S according to Eq. (9)
 8:     **else if** $q = 1$ **then**
 9:       Update S according to Eq. (10)
10:     **else if** $q = 2, 1$ **then**
11:       Update S according to Eq. (11)
12:     **end if**
13: **end while**

5.1   Datasets and Evaluation Metrics

We conduct experiments on two types of data: networks with and without known communities. For networks with known communities, we select Football, Email, and Wiki networks. There are 5, 42 and 19 communities in Football, Email and Wiki networks[5], respectively. For networks without known communities,

---
[5] These datasets are from `http://www-personal.umich.edu/~mejn/netdata/`, `http://snap.stanford.edu/data/index.html` and `https://linqs.soe.ucsc.edu/data`.

we select Jazz, Email-Univ and Hamsterster networks. The optimal number of communities for each network is inferred using Louvain algorithm [1]. A brief summary of these datasets is shown in Table 2. We employ purity and normalized mutual information (NMI) as the evaluation metrics for networks with known communities. Purity measures the extent to which each cluster contained data points from primarily one class. The purity of a clustering is obtained by the weighted sum of individual cluster purity values which defined as:

$$Purity = \frac{1}{N} \sum_{i=1}^{k} max_j |c_i \cap t_j|, \tag{13}$$

where N is number of objects, k is number of clusters, c_i is a cluster in C, and t_j is the classification which has the max count for cluster c_i. NMI evaluates the clustering quality based on information theory, and is defined by normalization on the mutual information between the cluster assignments and the pre-existing input labeling of the classes:

$$NMI(\mathcal{C}, \mathcal{D}) = \frac{2 * \mathcal{I}(\mathcal{C}, \mathcal{D})}{\mathcal{H}(\mathcal{C}) + \mathcal{H}(\mathcal{D})}, \tag{14}$$

where obtained cluster $\mathcal{C}$ and ground-truth cluster $\mathcal{D}$. The mutual information $\mathcal{I}(\mathcal{C}, \mathcal{D})$ is defined as $\mathcal{I}(\mathcal{C}, \mathcal{D}) = \mathcal{H}(\mathcal{C}) - \mathcal{H}(\mathcal{C}|\mathcal{D})$ and $\mathcal{H}(\cdot)$ is the entropy.

For networks without known communities we use modularity as the evaluation metric. Formally, modularity is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \tag{15}$$

where m is the number of edges, A is the adjacency matrix of the input graph, k_i is the degree of node i and $\delta(c_i, c_j)$ is 1 if node i and node j are in the same community and 0 otherwise.

Table 2: Summary of data sets used in the experiments where n, e and c denote the number of nodes, edges, and communities, respectively.

Data		n	e	c
Known Communities	Football	115	613	5
	Email	1005	25571	42
	Wiki	2405	17981	19
Unknown Communities	Jazz	198	2742	4
	Email-Univ	1133	5451	11
	Hamsterster	1858	12534	32

5.2 Baseline Methods

To demonstrate our model detects communities more effectively than other NMF based community detection approaches, we select several representative NMF based methods as our baselines:

Table 3: Community detection performance w.r.t. Purity.

	Football	Email	Wiki
NMF	0.9074	0.5834	0.2025
ONMF	0.8754	0.5734	0.3177
PNMF	0.9153	0.6002	0.2531
BNMF	0.9067	0.5587	0.3497
GNMF	0.9143	0.5675	0.2561
BigClam	0.9209	0.6358	0.2995
NSED	0.8994	0.6072	0.2773
NRMF-L_0	0.8914	0.5801	0.2880
NRMF-L_1	**0.9335**	0.6843	0.3774
NRMF-$L_{2,1}$	**0.9335**	**0.7021**	**0.4154**

NMF [21]: NMF is the basic matrix factorization framework. To make a fair comparison, the objective function for NMF is $\|A - CC^T\|_F^2$.

Orthogonal NMF (ONMF) [4]: ONMF is a variant of NMF by enforcing orthogonal constraints on the community membership matrix C, i.e., $C^T C = I$. In particular, since we add contraint to C, the objective is $\|A - CMC^T\|_F^2$.

Projected NMF (PNMF) [25]: PNMF directly projects the original network to a subspace by minimizing $\|A - CC^T A\|_F^2$.

Bayesian NMF (BNMF) [19]: BNMF is a bayesian NMF model. It models the matrix factorization in a Bayesian fashion.

Graph regularized NMF (GNMF) [2]: GNMF incorporates an affinity graph which is constructed to encode the geometrical information to NMF, and then seeks a matrix factorization which respects the graph structure.

BigClam [23]: a cluster affiliation model. It relaxes the graph fitting problem into a continuous optimization problem to find overlapping communities.

Nonnegative Symmetric Encoder-dedcoder (NSED) [20]: NSED is a nonnegative symmetric encoder-decoder approach proposed for community detection. It extracts the community membership by integrating a decoder and an encoder into a unified loss function.

For our framework, we compare three variants: L_0 regularized (NRMF-L_0), L_1 regularized (NRMF-L_1) and $L_{2,1}$ regularized NRMF (NRMF-$L_{2,1}$).

5.3   Experimental Results

The experimental results on networks with known communities are shown in Table 3 based on Purity and Table 4 based on NMI. Based on these results, it can be observed that: (1) NRMF-$L_{2,1}$ achieves the best performance in community detection based on both the purity and NMI metrics. It indicates $L_{2,1}$ regularization can better capture the sparsity of the residual matrix compared to other types of regularization such as L_0 and L_1. (2) NRMF methods with L_1 and $L_{2,1}$ sparse regularization outperform other NMF based approaches. It demonstrates that our proposed framework can effectively identify the communities because it explicitly takes noise into consideration when factorizing the input adjacency matrix. (3) NRMF methods with L_0 performs worse than some other NMF based

Table 4: Community detection performance w.r.t. NMI.

	Football	Email	Wiki
NMF	0.8973	0.6234	0.1876
ONMF	0.8843	0.6451	0.2108
PNMF	0.9035	0.6211	0.2404
BNMF	0.9142	0.6021	0.2411
GNMF	0.9038	0.5893	0.1976
BigClam	0.8963	0.5796	0.2722
NSED	0.9096	0.6845	0.2659
NRMF-L_0	0.8903	0.6746	0.2235
NRMF-L_1	0.9214	0.6746	0.2626
NRMF-$L_{2,1}$	**0.9233**	**0.7032**	**0.2737**

methods. It shows that L_0 norm is not a good choice for sparsity modeling. This may because L_0 norm only considers the number of nonzero elements but ignore the values of these elements. (4) It is interesting to observe that ONMF and GNMF achieve worse performance than other NMF based methods including the standard NMF. This may result from that real-world network data is very sparse and constraints on the latent representations may be difficult to achieve.

Table 5: Community detection performance w.r.t. Modularity.

	Jazz	Email-Univ	Hamsterster
NMF	0.4136	0.2876	0.2347
ONMF	0.4221	0.2751	0.2378
PNMF	0.4201	0.4015	0.1313
BNMF	0.4161	0.3963	0.1226
GNMF	0.1852	0.2747	0.2060
BigClam	0.4005	0.4024	0.2796
NSED	0.4231	0.3987	0.2553
NRMF-L_0	0.3843	0.3369	0.1892
NRMF-L_1	0.4355	0.4448	0.2856
NRMF-$L_{2,1}$	**0.4463**	**0.5011**	**0.2903**

The experimental results on networks without known communities are shown in Table 5 based on Modularity. From these results, we can draw the same conclusions to that on the networks with known communities. Since modularity is a measure only based on the network structure (ground-truth community labels may be related not only to structures but also semantics), good performance on modularity further demonstrate the effectiveness of our framework from the structural perspective.

5.4   Parameter Sensitivity

The *NRMF* involves two parameters shown in Eq. (5): α controls the trade-off of redisual modeling and β controls the graph regularization. In this section, we examine how the different choices of parameters affect the performance of *NRMF* in community detection. Specifically, we measure the NMI and Purity
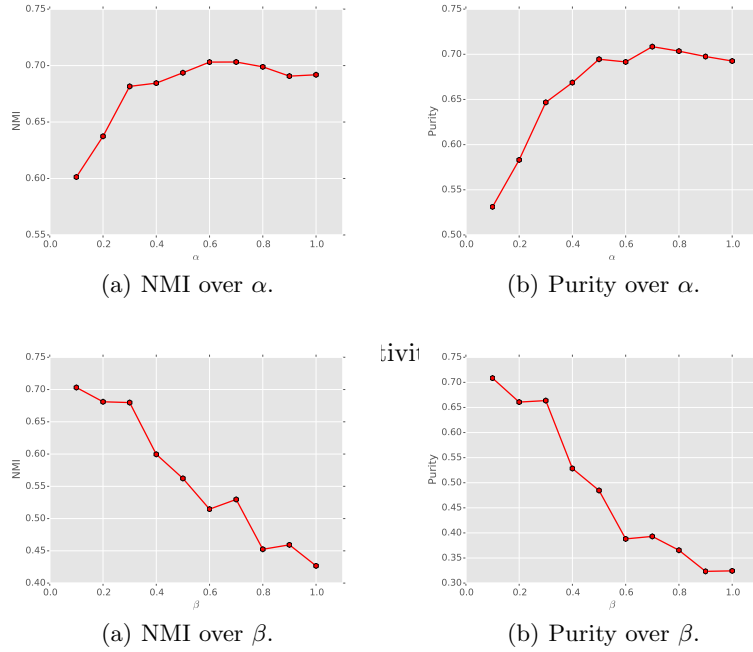
(a) NMI over α.   (b) Purity over α.

(a) NMI over β.   (b) Purity over β.

Fig. 2: Parameter sensitivity of β on Email data.

on Email data and Modularity on Email-Univ data. The results are shown in Fig. 1, 2 and 3. Note that only *NRMF* with $L_{2,1}$ is evaluated because it achieves the best performance in the experiments above.

It can be observed from these results that: (1) Large α is preferred in order to achieve better NMI and Purity in detecting communities. In specific, when α is 0.6 - 0.7, the performance is the best. (2) In contrast, smaller β brings better community detection performance. In the experiments, $\beta = 0.1$ is the best choice w.r.t. both NMI and Purity. (3) For the networks without ground-truth labels, the choices of α and β are different. In Email-Univ, smaller α and larger β give better performance. In this specific experiment, $\alpha = 0.1$ and $\beta = 0.8$ would be the best parameters.

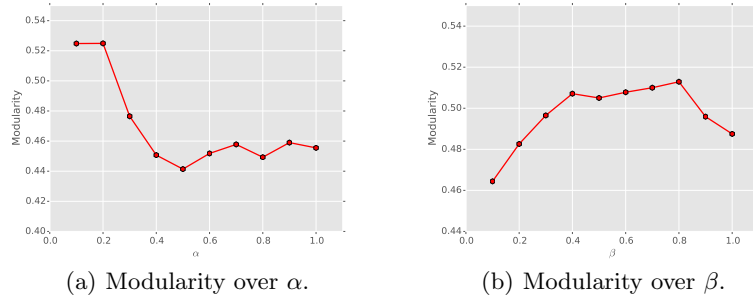(a) Modularity over α.   (b) Modularity over β.

Fig. 3: Parameter sensitivity results on Email-Univ data.

6   Conclusion

We proposed *NRMF*, a novel framework for community detection, that addresses the limitation of existing NMF based community detection approaches: being sensitive to noise and outliers. In *NRMF*, a residual matrix has been introduced explicitly to capture the noise. To model the sparsity of the *NRMF*, we exploit three types of regularization, i.e., L_0, L_1 and $L_{2,1}$. To optimize the objective w.r.t different regularization terms, different updating rules have been employed. Our experimental study demonstrated that *NRMF* effectively preserves community structures and captures noisy information, outperforming state-of-the-art NMF based methods in community detection. On the basis of *NRMF*, several new research lines can be pursued. For example, it is interesting to exploit more advanced optimization methods or extend it to attributed networks. We leave these extensions for future work.

References

1. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
2. Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1548–1560, 2010.
3. Sandro Cavallari, Vincent W Zheng, Hongyun Cai, Kevin Chen-Chuan Chang, and Erik Cambria. Learning community embedding with community detection and node embedding on graphs. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 377–386, 2017.
4. Chris Ding, Tao Li, Wei Peng, and Haesun Park. Orthogonal nonnegative matrix t-factorizations for clustering. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 126–135. ACM, 2006.
5. Sergey N Dorogovtsev, Alexander V Goltsev, and Jose Ferreira F Mendes. K-core organization of complex networks. *Physical review letters*, 96(4):040601, 2006.
6. Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.
7. Shudong Huang, Hongjun Wang, Tao Li, Tianrui Li, and Zenglin Xu. Robust graph regularized nonnegative matrix factorization for clustering. *Data Mining and Knowledge Discovery*, 32(2):483–503, 2018.
8. Dany Kamuhanda and Kun He. A nonnegative matrix factorization approach for multiple local community detection. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 642–649. IEEE, 2018.
9. George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1):359–392, 1998.
10. Hans Laurberg, Mads Græsbøll Christensen, Mark D Plumbley, Lars Kai Hansen, and Søren Holdt Jensen. Theorems on positive data: On the uniqueness of nmf. *Computational intelligence and neuroscience*, 2008, 2008.

11. Daniel D Lee and H Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
12. Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.
13. Feiping Nie, Heng Huang, Xiao Cai, and Chris H Ding. Efficient and robust feature selection via joint l2, 1-norms minimization. In *Advances in neural information processing systems*, pages 1813–1821, 2010.
14. Yulong Pei, Nilanjan Chakraborty, and Katia Sycara. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
15. Yulong Pei, Xin Du, George Fletcher, and Mykola Pechenizkiy. Dynamic network representation learning via gaussian embedding. In *NeurIPS 2019 Workshop on Graph Representation Learning*, 2019.
16. Yulong Pei, George Fletcher, and Mykola Pechenizkiy. Joint role and community detection in networks via l 2, 1 norm regularized nonnegative matrix tri-factorization. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 168–175, 2019.
17. Chong Peng, Zhao Kang, Yunhong Hu, Jie Cheng, and Qiang Cheng. Robust graph regularized nonnegative matrix factorization for clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 11(3):33, 2017.
18. Ioannis Psorakis, Stephen Roberts, Mark Ebden, and Ben Sheldon. Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E*, 83(6):066114, 2011.
19. Mikkel N Schmidt, Ole Winther, and Lars Kai Hansen. Bayesian non-negative matrix factorization. In *International Conference on Independent Component Analysis and Signal Separation*, pages 540–547. Springer, 2009.
20. Bing-Jie Sun, Huawei Shen, Jinhua Gao, Wentao Ouyang, and Xueqi Cheng. A non-negative symmetric encoder-decoder approach for community detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 597–606, 2017.
21. Fei Wang, Tao Li, Xin Wang, Shenghuo Zhu, and Chris Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521, 2011.
22. Xiao Wang, Peng Cui, Jing Wang, Jian Pei, Wenwu Zhu, and Shiqiang Yang. Community preserving network embedding. In *Thirty-first AAAI conference on artificial intelligence*, 2017.
23. Jaewon Yang and Jure Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.
24. Fanghua Ye, Chuan Chen, and Zibin Zheng. Deep autoencoder-like nonnegative matrix factorization for community detection. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1393–1402, 2018.
25. Zhijian Yuan and Erkki Oja. Projective nonnegative matrix factorization for image compression and feature extraction. In *Scandinavian Conference on Image Analysis*, pages 333–342. Springer, 2005.
26. Yu Zhang and Dit-Yan 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*, pages 606–614. ACM, 2012.