Front. Comput. Sci., 2021, 15(4): 154324
https://doi.org/10.1007/s11704-020-9203-0

RESEARCH ARTICLE

Semi-supervised community detection on attributed networks using non-negative matrix tri-factorization with node popularity

Di JIN[1], Jing HE[1], Bianfang CHAI (✉)[2], Dongxiao HE[1]

1 College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
2 Department of Information Engineering, Hebei GEO University, Shijiazhuang 050031, China

Abstract The World Wide Web generates more and more data with links and node contents, which are always modeled as attributed networks. The identification of network communities plays an important role for people to understand and utilize the semantic functions of the data. A few methods based on non-negative matrix factorization (NMF) have been proposed to detect community structure with semantic information in attributed networks. However, previous methods have not modeled some key factors (which affect the link generating process together), including prior information, the heterogeneity of node degree, as well as the interactions among communities. The three factors have been demonstrated to primarily affect the results. In this paper, we propose a semi-supervised community detection method on attributed networks by simultaneously considering these three factors. First, a semi-supervised non-negative matrix tri-factorization model with node popularity (i.e., PSSNMTF) is designed to detect communities on the topology of the network. And then node contents are integrated into the PSSNMTF model to find the semantic communities more accurately, namely PSSNMTFC. Parameters of the PSSNMTFC model is estimated by using the gradient descent method. Experiments on some real and artificial networks illustrate that our new method is superior over some related state-of-the-art methods in terms of accuracy.

Keywords community detection, non-negative matrix trifactorization, node popularity, attributed networks

1 Introduction

With the development of Internet, online social networks generate more and more data with both links and semantic contents, such as user blogs, research papers, etc. These datasets are always modeled as attributed networks [1], where links form the topology of a graph and contents are modeled as attributes of nodes in the graph. It is of great significance to detect the semantic communities of these networks. For example, in a paper citation network, each node represents a paper, and the papers contain hyperlinks from one to another, and each paper has its contents. Identifying communities of these papers according to the links and contents among them helps researchers understand the fields of the current frontiers. Thus, how to integrate links and contents in attributed networks to identify more accurate semantic community structures is a challenging and meaningful problem.

In the past few decades, some methods have been proposed to detect semantic communities in attributed networks [2–9]. Based on the data types used for this task, they are mainly classified into four categories, i.e., topological-based methods [10–14], attributed-based methods [15,16], ensemble methods [17–20] and model-based methods [21–24]. The first type transforms community detection on the attributed network into graph clustering on a new reconstructed network (where nodes' attributes are modeled as topological information). The second type transforms community detection on the attributed network into a traditional vector data clustering task (where links and contents are merged to compute similarities or dissimilarities between all pair of nodes). Ensemble methods combine the results of different clustering. Model-based methods jointly model links and contents by some statistical models such as NMF [25] and probabilistic model [26]. By doing so, they can make full use of links and contents and formulate the clustering problem as an optimization process.

Topological-based community detection is mainly based on some early approaches. While a large number of approaches in this realm have been provided for network community detection in recent years [7,8], NMF based methods have attracted many interests due to its good performance and strong interpretability. They are able to cluster data with different distributions and detect non-overlapping and overlapping communities. There are also some variants introducing different factors to improve the performance of community detection on networks. On the other hand, semi-supervised community detection methods integrate priors to improve the performance of community detection [27–29]. For example, Liu et al. [13] introduced node popularity to a semi-supervised NMF model (PSSNMF) for community detection, which utilized the heterogeneity of node degree and the prior constraints at the same time. The literature [30] incorporated the community structure into network embedding, which jointly optimized NMF-based representation learning model and modularity-based community detection model in

Received June 4, 2019; accepted April 2, 2020

E-mail: chaibianfang@163.com

a unified framework. Generally speaking, the existing community detection methods based on NMF were more efficient than unsupervised NMF-based methods, since they took one or two of the three factors into considerations, i.e., priors, node heterogeneity and interactions among communities. But it is more reasonable that all of these three factors should be considered in the topological-based community detection methods together. In addition, these NMF-based methods just utilize the topology of a network, and thus are not able to detect semantic communities directly on attributed networks As the best of our knowledge, there is no research that extends the above community detection methods from topological networks to attributed networks. Thus, it is necessary to design a new community detection model based on links and contents on attributed networks, and the model for the topology should also consider the three important factors (i.e., prior information, degree heterogeneity and interactions among communities) simultaneously

Model-based methods for community detection on attributed networks are believed to have a good performance due to owning the solid theoretical foundation (compared with other type of methods). For example, Zhu et al. [25] formalized community detection as an optimization problem based on the factorization of link matrix and content matrix. But this model did not consider the interactions among different communities. In addition, it ignored the factors of node degrees and prior information based on the topology. Lately, Wang et al. [31] proposed a model based on NMF for detecting semantic communities on attributed networks by integrating priors and attributed network information. As we all know, this is the only semi-supervised community detection model on attributed networks. It does not consider the heterogeneity of node degrees and the interactions among communities (which are demonstrated to be important for community detection).

Based on the above discussions, it is concluded that these three important factors, including node priors, node heterogeneity and interactions among communities, should be considered together for semantic community detection via integrating links and contents on attributed networks. In this paper, we propose a Semi-Supervised community detection Non-negative matrix Tri-Factorization (NMTF) model with node Popularity for attributed networks based on links and node contents, namely PSSNMTFC. It can not only combine the links and contents of nodes seamlessly based on the NMTF model to detect semantic community, but also utilize the pairwise constraints, the heterogeneity of node degrees and the interactions among communities together to enhance the performance of community detection.

The contributions of this work are as follows:

• We propose a semi-supervised community detection model for attributed networks. By combining links and node contents based on the NMTF model, the semantic communities are detected more accurately, and the interactions among communities can also be inferred. By utilizing a few priors and considering the node popularity, the proposed model owns a better performance, especially on networks with degree heterogeneity of the nodes or the network with fuzzy community structure.

• Parameter estimating of our PSSNMTFC model is inferred using gradient descent, leading to an efficient algorithm.

• The experiments on several networks demonstrate that the algorithm based on our PSSNMFC model derive network communities with a higher accuracy

The remainder of the paper is organized as follows. A brief review of the related works on community detection based on NMF is given in Section 2. Section 3 provides a detailed description of our PSSNMTFC model which integrates the topological information and node contents based on NMTF with node popularity. Section 4 shows in details the experiment results of our model on artificial and real networks. Section 5 concludes the contributions of this paper and looks forward to the future work.

2 Related work

Our work is to design a semi-supervised community detection method for attributed network. It is related on unsupervised community detection methods and semi-supervised community detection methods based on the topology of a network, and is also associated with community detection methods based on links and contents of attributed networks.

A large number of approaches have been provided for community detection on a network [7,8], such as modularity-based methods, statistical inference methods, NMF based methods [32], network embedding based methods [33], etc. Modularity-based methods detect communities by improving the optimization of modularity measure. Statistical inference methods identify different types of structures based on flexible generating model, such as the stochastic block model. NMF based methods captain node memberships or community labels by factorizing link matrix. Recently, many network embedding methods are used to detect community. Some methods learn an effective low dimensional vectors of nodes by preserving the network structure, and then clustering algorithms use the embedding vectors to captain communities. Some methods incorporated the community structure into network embedding, and jointly optimized representation learning model and modularity based community detection model in a unified framework. The results of these methods are always inaccuracy, especially on a network with unclear community structure.

Recently, researchers have proposed many semi-supervised community detection approaches to improve the performance of community detection by labeling a few priors [27–29]. There are mainly two kinds of priors: the individual labels and the pairwise constraints. Compared with the first kind of priors, the latter is easier to get. For example, in a paper citation network, it is easier to know whether two papers belong to the same topic. On the contrary, it is difficult to label the category of a paper. These methods are mainly based on modularity [34] or NMF [11–13]. For example, Eaton and Mansbach [34] developed a semi-supervised automated community detection model, which incorporated background knowledge in the forms of individual labels and pairwise constraints to guide the process of community detection. This model was demonstrated to be equivalence to modularity which existed the resolution limit problem. Yang et al. [11] integrated the must-link constraints with network topology to obtain a unified semi-supervised community detection framework based on NMF. Shi et al. [12] built a nonnegative symmetric matrix factorization (PCSNMF) model

with the pairwise constraints to enhance the community detection. Liu et al. [13] developed the PSSNMF model for semi-supervised community detection, which considered the heterogeneity of node degree and the prior constraints at the same time. Different from other NMF-based methods, the PSSNMF model preserved that the Euclidean distance of two nodes with degree heterogeneity in the same community was small. However, the distance based on other NMF models was larger Thus, the PSSNMF model was able to use the priors more accurately on this kind of nodes, which had better accuracy performance than other semi-supervised community detection methods. The PSSNMF model used two matrices factorization and did not consider the interactions among communities. The GNMTF model [35] incorporated the graph structure as a regularization term into the objective function of the symmetric three factor matrices factorization. It explicitly modeled the node memberships and the interactions among communities. But the node degree heterogeneity was not considered. In a word, the existed topological community detection methods mainly considered some of the three factors, i.e., the priors, degree heterogeneity and the interactions among communities. But these methods are not able to directly use to detect semantic communities on attributed networks.

Some classical model-based methods for community detection on attributed networks are provided. For example, Yang et al. [26] proposed a discriminative model for combining the link and content analysis for community detection. It designed a probabilistic model to generate directed links by considering the popularity and productivity of nodes, and inferenced node memberships by maximizing the likelihoods of generating links. Pei et al. [36] proposed a nonnegative matrix tri-factorization (NMTF) clustering framework to combine three types of graph regularization in a social network which employ the user relations, user-words and message-words together. This method modeled the interactions among communities and utilized three kinds of local similarities to improve the performance of user clustering. Zhu et al. [37] performed matrix factorization on the term-document matrix and the adjacency matrix of citation networks. The two factorizations shared a common base and the discovered latent factors represented the memberships of nodes based on both contents and links of citation networks. Wang et al. [38] defined two different matrices, i.e., the community membership matrix based on network topology and community attribute matrix based on node attribute to identify network community structure and semantic annotation. Although some methods [25,36,38] also used NMF-based model to detect communities on attributed networks, they are different from our model. Our model for links not only considers the priors and node degree heterogeneity factors, but also interactions among communities. While the existed methods did not consider them at the same time.

3 The method

In this section, we first propose a semi-supervised NMTF model with node popularity (PSSNMTF) for community detection according to the network topology. The PSSNMTF model is then extended to combine links and node contents of attributed networks for community detection. The main framework is shown in Fig. 1. Finally, a gradient descent method is inferred to update the model parameters and optimizes the objective function of the proposed model.

Assume a given attribute network, represented as $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{C})$, and our method aims to detect semantic communities Table 1 gives the principal notations used in this paper.

Table 1 Attributed networks: notations

Notation	Interpretation
$\boldsymbol{V}$	The set of nodes, $V = \{v_1, v_2, \ldots, v_n\}$, n is the number of nodes
$\boldsymbol{E}$	The set of edges
$\boldsymbol{C}$	The content information of attributed graph $\boldsymbol{G}$
$\boldsymbol{A}$	Adjacency matrix its element equals to 1 if there is an edge from v_i to v_j, and 0 otherwise
$\boldsymbol{X}$	Represents the membership matrix of nodes and communities
$\boldsymbol{U}$	The relationship matrix between communities
$\boldsymbol{M}$	Must link matrix
$\boldsymbol{W}$	Node popularity matrix
$\boldsymbol{V}$	Attribute matrix in k-dimensional potential space

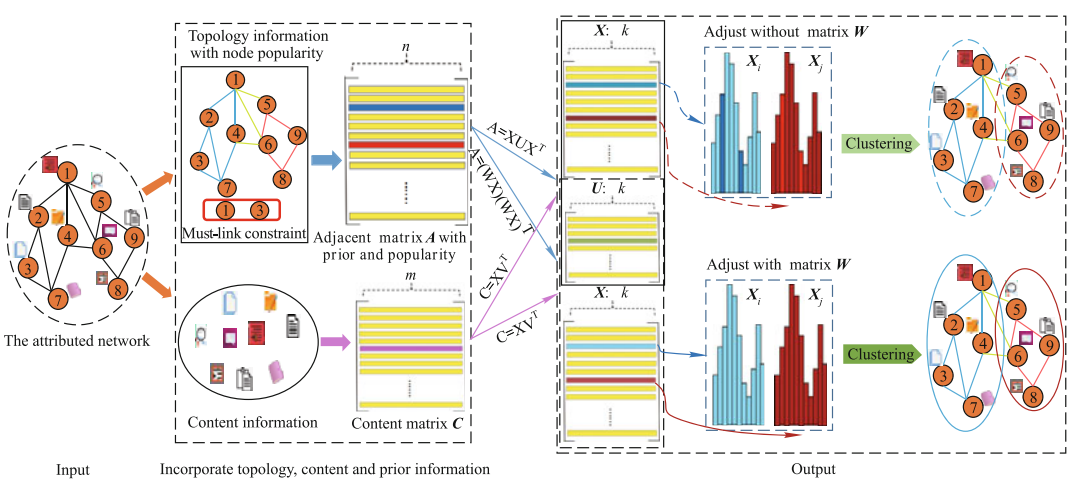


Fig. 1 Proposed semi-supervised framework which incorporates the topology and content information, as well as prior information for community detection on attributed networks. After dimensional reduction in the same latent space, we can obtain the membership matrix $\boldsymbol{X}$ of nodes and the relationship matrix $\boldsymbol{U}$ between communities. Furthermore, by introducing the node popularity matrix $\boldsymbol{W}$ to adjust the model (based on degree heterogeneity), the clearer community structure will be obtained on attributed networks

3.1 The NMTF model

The NMF is a powerful clustering technology for expressing data, and has obvious advantages in community detection. The NMF [39,40] and SNMF [41,13] models are able to learning a high-quality and low dimensional features of nodes in a network, and then communities are detected by clustering nodes based on these new features. However, both NMF and SNMF do not consider the interactions among communities. While the NMTF model explicitly models the interactions among communities, which generates networks that are more similar to real networks. Thus, here the NMTF model is used to uncover underlying communities based on the network topology as follows:

$$\min_{X\geqslant 0, U\geqslant 0} \mathcal{F}(\boldsymbol{X},\boldsymbol{U}) = \|\boldsymbol{A}-\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2. \tag{1}$$

The NMTF model generates links between two nodes of any two clusters by terms $\boldsymbol{X}$ and $\boldsymbol{U}$. $\boldsymbol{X}\in\mathbb{R}_+^{n\times k}$ is a $n\times k$ matrix, and its element x_{iz} represents the possibility that node v_i belongs to community z. $U\in R_+^{k\times k}$ represents the interactions among communities. And $\|\cdot\|_F$ denotes the Frobenius norm. Since the adjacency matrix $\boldsymbol{A}$ is positive, the nonnegative constraints are also added to matrix $\boldsymbol{X}$ and $\boldsymbol{U}$ simultaneously. η is a parameter to control the proportion of the Frobenius norm in optimization

3.2 The Semi-Supervised NMTF model

In order to improve the performance of the NMTF model, we combine the must-link constraints with the NMTF model to infer the community structure on attributed networks. Here, we denote the must-link constraints set as C_{ml}, and each element $(v_i,v_j)\in C_{ml}$ indicates that two nodes v_i and v_j belong to the same community. C_{ml} is formalized as a must-link matrix $\boldsymbol{M}$ in the following.

$$(\boldsymbol{M})_{ij} = \begin{cases} 1, & \text{if } i=j,\\ \varepsilon, & \text{if } (v_i,v_j)\in C_{ml},\\ 0, & \text{others.}\end{cases}$$

Intuitively, if two nodes v_i and v_j belong to the same community, their distance in the latent low-dimensional space should be small. i.e., the distance based on vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ should be small in the latent space. The Euclidean distance is usually used to measure the distance between two vectors. Combining the must-link constraints with the distance constants in the latent space, a graph regularization is formulated as:

$$\begin{aligned}\boldsymbol{D}(\boldsymbol{X}) &= \sum_{ij}\|\boldsymbol{X}_i-\boldsymbol{X}_j\|^2\boldsymbol{M}_{ij}\\ &= \sum_{ij}\sum_{z}(\boldsymbol{x}_{iz}-\boldsymbol{x}_{jz})^2\boldsymbol{M}_{ij}\\ &= 2\sum_{i}\sum_{z}\boldsymbol{x}_{iz}^2\boldsymbol{Q}_{ii} - 2\sum_{ij}\sum_{z}\boldsymbol{x}_{iz}\boldsymbol{x}_{jz}\boldsymbol{M}_{ij}\\ &= 2\boldsymbol{Tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) - 2\boldsymbol{Tr}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}),\end{aligned} \tag{2}$$

where $\boldsymbol{Q}$ is a diagonal matrix ($\boldsymbol{Q}_{ii}=\sum_j\boldsymbol{M}_{ij}$ and $\boldsymbol{Q}_{ij}=0$ if $(i\neq j)$, and $Tr(\cdot)$ denotes the trace of the matrix.

Incorporating the graph regularization term with prior information into the NMTF model, a semi-supervised NMTF model (SSNMTF) for community detection are built as:

$$\begin{aligned}O(\boldsymbol{X},\boldsymbol{U}) &= \|\boldsymbol{A}-\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2 + \frac{\lambda}{2}D(\boldsymbol{X})\\ &= \|\boldsymbol{A}-\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2\\ &\quad +\lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) - \lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}).\end{aligned} \tag{3}$$

where λ is a balance constant parameter and its value will be analyzed in details in Section 4.

However, there is a serious flaw in the above model of Eq. (3), which cannot minimize the graph regularization term that embeds the must-link prior information due to the degree heterogeneity of any two nodes with a must-link constraint., Take a network which is divided into four communities as an example We select two sampled nodes with heterogeneous degrees which belong to the four communities with different propensities, i.e., (0.2, 0.4, 0.6, 0.2) and (0.1, 0.2, 0.3, 0.1) respectively. According to the propensities, the two nodes should be both assigned to the third community with the maximal propensity. But their Euclidean distance in low dimensional space is 0.55, which is very large. If a must-link constraint between the two nodes is used on the two nodes, the distance on them according to their regularization term is large. When the objective function Eq.(3) is optimized, this regularization term will be minimized and then wrong memberships of the two nodes are captained. The two nodes may be assigned to different communities although they have a must-link constraint.

3.3 The SSNMTF with Popularity model

In order to solve the serious flaw mentioned above and avoid affecting the results of community detection we adjust our model by introducing the popularity parameters of nodes. By defining the popularity of nodes v_i as w_i, the popularity vector of all nodes is denoted as $\boldsymbol{w}=(w_1,w_2,\ldots,w_n)$. Then, the SSNMTF with popularity model (PSSNMTF) is modified as follows:

$$\begin{aligned}O(\boldsymbol{X},\boldsymbol{U},\boldsymbol{W}) &= \sum_{ij}\left(\boldsymbol{A}_{ij}-\sum_z w_i x_{iz}u_z w_j x_{jz}\right)^2\\ &\quad +\frac{\lambda}{2}\sum_{ij}\sum_z(x_{iz}-x_{jz})^2M_{ij} + \eta\sum_z u_{iz}u_{zj}\\ &= \sum_{ij}\left(A_{ij}-w_iw_j\sum_z x_{iz}u_zx_{jz}\right)^2 + \eta\sum_z u_{iz}u_{zj}\\ &\quad +\frac{\lambda}{2}\sum_{ij}\sum_z(x_{iz}^2-2x_{iz}x_{jz}+x_{jz}^2)M_{ij}\\ &= \sum_{ij}\left(A_{ij}-w_iw_j\sum_z x_{iz}u_zx_{jz}\right)^2 + \eta\sum_z u_{iz}u_{zj}\\ &\quad +\lambda\sum_i\sum_z x_{iz}^2Q_{ii} - \lambda\sum_{ij}\sum_z x_{iz}x_{jz}M_{ij}\\ &= \|\boldsymbol{A}-(\boldsymbol{W}\boldsymbol{X})\boldsymbol{U}(\boldsymbol{W}\boldsymbol{X})^{\mathrm{T}}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2\\ &\quad +\lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) - \lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X})\\ &\quad \|\boldsymbol{X}_i\|_1 = 1,\quad i=1,2,\ldots,n.\end{aligned} \tag{4}$$

The popularity of nodes is mainly used to solve the increase of Euclidean distance between nodes caused by degree heterogeneity. The normalization of $\boldsymbol{X}$ rows can then be taken as a

special case of node popularity. Considering the example in Section 2.2 again, the former Euclidean distance between two nodes, whose propensities x_i, x_j are (0.2, 0.4, 0.6, 0.2) and (0.1, 0.2, 0.3, 0.1), is 0.55. After normalizing the rows of $\boldsymbol{X}$, i.e., the sum of rows of matrix $\boldsymbol{X}$ equals to 1, the two nodes propensities both become (1/7, 2/7, 3/71/7), and the Euclidean distance between the two nodes turns to zero. The row normalizing of $\boldsymbol{X}$ is equivalent to set $w_i = 1.4$ and $w_j = 0.7$. Thus, the normalization of $\boldsymbol{X}$ is a special case of node popularity $\boldsymbol{w}$. This illustrates that the popularity of nodes plays an important role in minimizing the distances between. But the normalization of $\boldsymbol{X}$ is not able to model the degree heterogeneity. Therefore, not only node popularity but also the normalization of $\boldsymbol{X}$ are modeled in Eq. (4).

Based on the above analysis, we have combined the must-link constraint information with topological information and introduced the node popularity matrix $\boldsymbol{X}$ to adjust the degree heterogeneity. Consequently, our objective function can be summarized as:

$$\begin{aligned}\min_{\boldsymbol{X},\boldsymbol{U},\boldsymbol{W}} \boldsymbol{O}(\boldsymbol{X},\boldsymbol{U},\boldsymbol{W}) = &\|\boldsymbol{A} - (\boldsymbol{W}\boldsymbol{X})\boldsymbol{U}(\boldsymbol{W}\boldsymbol{X})^{\mathrm{T}}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2 \\ &+\lambda\boldsymbol{T}'br(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) - \lambda\boldsymbol{T}\boldsymbol{r}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}), \\ \text{s.t.} \quad &\boldsymbol{X} \geqslant 0, \boldsymbol{U} \geqslant 0, \boldsymbol{W} \geqslant 0, \\ &\|\boldsymbol{X}_i\|_1 = 1, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{5}$$

3.4 Jointing the PSSNMTF model with content model

In attribute networks, nodes not only connect with each other, but also contain rich semantic contents. Two nodes with similar contents may belong to the same community. Taking the content information of nodes into account. In order to combine the topological information between nodes with the content information on nodes, we want to use the same potential space to approximate the potential space of connection between nodes. Using the bag-of-words approach, the content matrix of the attributed network can be denote as $\boldsymbol{C}$, which is an $n \times m$ matrix, where m is the number of keyword features in a document. Similar to the latent semantic indexing (LSI), the k-dimensional latent space of words is expressed by an $m \times k$ matrix $\boldsymbol{V}$. Thus, to identify the semantic representation of the node, we consider the approximation of matrix $\boldsymbol{C}$ by $\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}$, defined as follow:

$$\begin{aligned}\min_{\boldsymbol{X},\boldsymbol{V}} &\|\boldsymbol{C} - \boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\|_F^2 + \beta\|\boldsymbol{V}\|_F^2, \\ \text{s.t.} \quad &\boldsymbol{X} \geqslant 0, \end{aligned} \tag{6}$$

where β is a small positive number, and $\beta\|\boldsymbol{V}\|_F^2$ serves as a regularization term to improve the robustness.

There are many ways to employ both links and contents in attributed networks. Our model combines them into a single, consistent, and compact feature representation in a low dimensional space, which is formalized in the following, noted as PSSNMTFC:

$$\begin{aligned}\min_{\boldsymbol{X},\boldsymbol{U},\boldsymbol{W},\boldsymbol{V}} \boldsymbol{O}(\boldsymbol{X},\boldsymbol{U},\boldsymbol{W},\boldsymbol{V}) \overset{\text{def}}{=} &\|\boldsymbol{A} - (\boldsymbol{W}\boldsymbol{X})\boldsymbol{U}(\boldsymbol{W}\boldsymbol{X})^{\mathrm{T}}\|_F^2 \\ &+\lambda\boldsymbol{T}\boldsymbol{r}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) - \lambda\boldsymbol{T}\boldsymbol{r}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}) \\ &+\alpha\|\boldsymbol{C} - \boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\|_F^2 + \beta\|\boldsymbol{V}\|_F^2 + \eta\|\boldsymbol{U}\|_F^2 \\ \text{s.t.} \quad &\boldsymbol{X} \geqslant 0, \boldsymbol{U} \geqslant 0, \boldsymbol{W} \geqslant 0, V \geqslant 0, \\ &\|\boldsymbol{X}_i\|_1 = 1, \quad i = 1, 2, \ldots, n. \end{aligned} \tag{7}$$

3.5 Optimization

Since the objective function in Eq. (7) is not convex, it is impractical to obtain the optimal solution. In this paper, we use the gradient descent method to optimize our objective function (7) and obtain the global minimum.

Here, let the $\boldsymbol{\Psi}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ be the Lagrange multipliers for constraints $\boldsymbol{X} \geqslant 0, \boldsymbol{U} \geqslant 0, \boldsymbol{W} \geqslant 0$, and $\boldsymbol{V} \geqslant 0$, respectively. We then define the Lagrange function $\mathcal{L}$ as:

$$\begin{aligned}\mathcal{L} = &\|\boldsymbol{A} - (\boldsymbol{W}\boldsymbol{X})\boldsymbol{U}(\boldsymbol{W}\boldsymbol{X})^{\mathrm{T}}\|_F^2 + \lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{X}) \\ &-\lambda Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{X}) + \eta\|\boldsymbol{U}\|_F^2 + \alpha\|\boldsymbol{C} - \boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\|_F^2 + \beta\|\boldsymbol{V}\|_F^2 \\ &+Tr(\boldsymbol{\Psi}\boldsymbol{X}^{\mathrm{T}}) + Tr(\boldsymbol{\Omega}\boldsymbol{U}^{\mathrm{T}}) + Tr(\boldsymbol{\Phi}\boldsymbol{W}^{\mathrm{T}}) + Tr(\boldsymbol{\Theta}\boldsymbol{V}^{\mathrm{T}}). \end{aligned} \tag{8}$$

Correspondingly, the partial derivatives of $\mathcal{L}$ with respect to $\boldsymbol{X}$, $\boldsymbol{U}$, $\boldsymbol{W}$ and $\boldsymbol{V}$ are as follow:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\boldsymbol{X}} = &-2\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} - 2\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} + 2\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} \\ &+2\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} - 2\alpha\boldsymbol{C}\boldsymbol{V} + 2\alpha\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} \\ &+2\lambda\boldsymbol{Q}\boldsymbol{X} - 2\lambda\boldsymbol{M}\boldsymbol{X} + \boldsymbol{\Psi}, \end{aligned} \tag{9}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{U}} = -2\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X} + 2\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X} + 2\eta\boldsymbol{U} + \boldsymbol{\Omega}, \tag{10}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\boldsymbol{W}} = &2\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}\boldsymbol{W} + 2\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}} \\ &-2\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} - 2\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{\Phi}, \end{aligned} \tag{11}$$

$$\frac{\partial\mathcal{L}}{\partial\boldsymbol{V}} = -2\alpha\boldsymbol{C}^{\mathrm{T}}\boldsymbol{X} + 2\alpha\boldsymbol{V}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + 2\beta\boldsymbol{V} + \boldsymbol{\Theta}. \tag{12}$$

Using the KKT conditions ($\boldsymbol{\Psi}_{ik}x_{ik} = 0, \boldsymbol{\Omega}_{iz}u_{iz} = 0, \boldsymbol{\Phi}_{ii}W_{ii} = 0, \boldsymbol{\Theta}_{ik}v_{ik} = 0$), we obtain the following update rules:

$$x_{ik} \leftarrow x_{ik} \cdot \frac{(\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} + \boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} + \alpha\boldsymbol{C}\boldsymbol{V} + \lambda\boldsymbol{M}\boldsymbol{X})_{ik}}{(\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} + \lambda\boldsymbol{Q}\boldsymbol{X} + \alpha\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{ik}}, \tag{13}$$

$$u_{iz} \leftarrow u_{iz} \cdot \frac{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X})_{iz}}{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X} + \eta\boldsymbol{U})_{iz}}, \tag{14}$$

$$w_{ii} \leftarrow w_{ii} \cdot \frac{(\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}} + \boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}})_{ii}}{(\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}\boldsymbol{W} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}})_{ii}}, \tag{15}$$

$$v_{ik} \leftarrow v_{ik} \cdot \frac{(\alpha\boldsymbol{C}^{\mathrm{T}}\boldsymbol{X})_{ik}}{(\alpha\boldsymbol{V}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} + \beta\boldsymbol{V})_{ik}}. \tag{16}$$

However, the above update rules on $\boldsymbol{X}$ do not take into account the constraint of $\|\boldsymbol{X}_i\|_1 = 1, i = 1, 2, \ldots, n$. That is, the sum of each row of $\boldsymbol{X}$ should equal to 1. So we define the loss function for this constraint as: $S = \gamma\sum_i(\sum_z x_{iz} - 1)^2$.

The partial derivatives on $\mathcal{S}$ with x_{iz} is:

$$\frac{\partial \mathcal{S}}{\partial x_{iz}} = 2\gamma\left(\sum_z x_{iz} - 1\right) = 2\gamma\sum_z x_{iz} - 2\gamma = 2\gamma H - 2\gamma \boldsymbol{E}.$$

Here, both $\boldsymbol{H}$ and $\boldsymbol{E}$ are $n \times k$ matrices, and the element H_{ij} of $\boldsymbol{H}$ equals $\sum_z x_{iz}$, $j = 1, 2, \ldots, k$ All the element of matrix $\boldsymbol{E}$ are 1.

Then the update rule about $\boldsymbol{X}$ becomes:

$$x_{ik} \leftarrow x_{ik} \cdot \frac{(\boldsymbol{W}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} + \boldsymbol{W}\boldsymbol{A}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} + \boldsymbol{L})_{ik}}{(\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U}^{\mathrm{T}} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{U}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{W}\boldsymbol{X}\boldsymbol{U} + \boldsymbol{T})_{ik}}, \tag{17}$$

where $\boldsymbol{L} = \alpha\boldsymbol{C}\boldsymbol{V} + \lambda\boldsymbol{M}\boldsymbol{X} + \gamma\boldsymbol{E}, \boldsymbol{T} = \lambda\boldsymbol{Q}\boldsymbol{X} + \alpha\boldsymbol{X}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} + \gamma\boldsymbol{H}$. Finally, the whole algorithm that estimates parameters of the PSSNMTFC model is described in Algorithm 1.

3.6 Complexity analysis

The time complexity of our algorithm PSSNMTFC is composed of two parts. The first is the computational costs for the updating of $\boldsymbol{X}$, $\boldsymbol{U}$ and $\boldsymbol{W}$ in Eq. (5). According to the multiplication law and the multiplication rule of diagonal matrix, we have that the time complexity of updating of $\boldsymbol{X}$, $\boldsymbol{U}$ and $\boldsymbol{W}$ once needs $O(n^2+n^2k+nk^2+nk)$, $O(k^2)$ and $O(n^2k+nk^2+nk)$, respectively. Thus the time complexity of the first part is $O(tn^2k + tnk^2)$, where t, n, k denote the number of iterations, the number of nodes, and the number of communities, respectively. The second part is the computational costs of updating $\boldsymbol{V}$, which is $O(tnk^2)$. Consequently, the total time complexity of the algorithm is $O(tn^2k + tnk^2)$.

However, we all know that most of the networks in real life are sparse, i.e., $m << n^2$, so considering the sparsely of the real networks, our algorithm's time complexity can be further reduced to $O(tmk + tnk^2)$, where m is the number of edges. Moreover, the number of communities k is often much smaller than m, leading to that the algorithm's time complexity approximates linearity.

4 Experiments

In this section, we will evaluate the proposed approach PSSNMTFC on both artificial and real networks. First, we describe the datasets, evaluation measures and the baseline algorithms. Then, we compare our method with some baselines on artificial and real networks. Finally, the involved hyper-parameters are analyzed in detail.

Algorithm 1 The procedure of PSSNMTFC

Input: adjacency matrix $\boldsymbol{A}$, content matrix $\boldsymbol{C}$, and the prior information

Output: the community label l_i of each node v_i ($i = 1, 2, \ldots, n$)

1 Obtain the must-link matrix $\boldsymbol{M}$ according to the definition of must-link constraints;
2 Compute the diagonal matrix $\boldsymbol{Q}$;
3 Initialize the matrix $\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{V}$ and $\boldsymbol{W}$ randomly;
4 For $t = 1$: *iter* do
  Updating $\boldsymbol{X}, \boldsymbol{U}, \boldsymbol{W}, \boldsymbol{V}$ according to Eqs. (17), (14)–(16);
5 End For
6 $(v_i, l_i) = \arg\max_{j \leqslant k} x_{ij}$.

4.1 Datasets description

4.1.1 Artificial networks

First, we describe the features of synthetic networks that are used to test the algorithms. The synthetic network includes link and attribute information together. The links are generated based on the SBM model [42]. And attributes are generated according to Gauss model, whose central nodes are all generated in [–10,10] and covariance of off-diagonal elements are randomly generated between [–1,1], and the diagonal elements are a random number generated by the sum of off-diagonal elements $[0, 20*\sqrt{m}]$ (m is the dimension of attributes). According to this rule, several artificial networks are generated, shown in Table 2.

4.1.2 Real-world networks

Here, we introduce eight real attributed networks, including the Cornell, Texas, Washington, Wisconsin, Cora, Citeseer, DBLP, Pubmed, and a real large scale network, i.e., Flickr. The Cornell, Texas, Washington and Wisconsin are four WebKB subnetworks gathered from four different universities, respectively. Moreover, these four datasets represent the link relationships between webpages. Each sub-network is divided into five communities. There are total 877 webpages with 1,608 edges. Each webpage is annotated by 1,703-dimensional binary-valued word attributes. The Cora, Citeseer and DBLP datasets are paper citation datasets and nodes represent papers and edges indicate that one paper is cited by another paper. Pubmed network consists of 19,717 scientific publications with three classes from Pubmed database and 44,338 links between publications. Flickr is an image hosting and video hosting website, web services suite and online community. It includes 80513 users. The relationship between users and friends is represented by edges, with a total of 5899882 edges. Some specifically features of these real networks are shown in Table 3.

4.2 Evaluation measures

We use normalized mutual information (NMI) and accuracy (AC) to measure the performance of different algorithms. The NMI Index [43] is a commonly used index to measure the accuracy of community detection. It indicates the similarity between the actual community partition and the partition obtained by the

Table 2 Description of artificial networks

Networks	Nodes	Edges	Attributes	Communities
artificial network1	220	4,356	4	5
artificial network2	500	25,691	4	5

Table 3 Description of real-world networks

Datasets	Nodes	Edges	Attributes	Communities
Texas	187	328	1,703	5
Cornell	195	304	1,703	5
Washington	230	446	1,703	5
Wisconsin	265	530	1,703	5
Cora	2,708	5,429	1,433	7
Citeseer	3,312	4,732	3,706	6
DBLP	6,936	12,353	500	5
Pubmed	19,717	44,338	500	3
Flickr	80,513	5,899,882		195

proposed algorithm. Assume a and b are the sets of the ground-truth labels and the detected labels respectively. The value of NMI is formulated as:

$$NMI(a,b) = \frac{-2\sum_{i=1}^{c_a}\sum_{j=1}^{c_b} N_{ij}\log\left(\frac{N_{ij}n}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{c_a} N_{i.}\log\left(\frac{N_{i.}}{n}\right)+\sum_{j=1}^{c_b} N_{.j}\log\left(\frac{N_{.j}}{n}\right)},$$

where c_a is the real number of communities, and n denotes the number of nodes, and c_b denotes the number of the derived communities, and the entities N_{ij} of the matrix N denotes the number of nodes belonging to group i in set a, which is also treated as the size of group j in set b. If the partition obtained by the running algorithm perfectly matches the actual community partition, the value of NMI equals to 1, and 0 otherwise.

AC [43] is another measurement used to evaluate the performance of community detection algorithms, where a is the ground-truth labels of the nodes and b is the labels of the nodes derived from the algorithm. AC is defined as the accuracy rate of the tested algorithm as follows:

$$AC(a,b) = 1 - \frac{\sum_{i=1}^{|a|}\mathrm{II}(a_i \neq b_i)}{|a|},$$

where a_i is the label of node i, and $\mathrm{II}(x)$ is 1 if x is true and 0 otherwise.

4.3 Baselines

In order to evaluate the performance of our PSSNMTFC algorithm and study the influence of prior information, content information and the node popularity for community detection, we compare it with five types of methods:

• **Topology-based methods** UNMF [44] is a standard community detection method which is based on the unsupervised symmetric NMF, and considers only the link information. Some network embedding methods are also used to tested, including DeepWalk [45], LINE [46], and Node2Vec [47].

• **Content-based method** SMR [38] is a method which only use content information to cluster objects.

• **Methods using both network topology and prior information** GNMF [11] utilizes the prior information and topology information together to obtain a unified semi-supervised community detection framework based on NMF. The PSSNMF [13] algorithm is a semi-supervised NMF algorithm with node popularity, which combines the must-links prior and node popularity to guide the process of community detection. The SSNMF is a semi-supervised algorithm based on NMF without considering node popularity[13]. The FSSNMF [48] is a semi-supervised NMF framework, which embeds the must-links constraints and cannot-links constraints into the adjacency matrix, and modifies the topological structure to make the community structure of the network clearer.

• **Methods using both network topology and node contents** The SCI [38] method is based on NMF, which contains two set of parameters, i.e., the community membership matrix and community attribute matrix without prior information.

• **Methods using network topology, node contents and prior together** WSCDSM [31] integrates network topology and node content with the prior information.

The parameters involved in all algorithms are set to 1.

4.4 Experiment results

In the artificial datasets, we compare our algorithm with these five types of algorithms and measure by NMI and AC. As shown in Table 3, Table 4 and Fig. 2, our algorithm PSSNMTFC is overall superior to the baselines.

To be specific, Table 3 shows the comparison results between our PSSNMTFC and five types of algorithms on two artificial networks, and the priori ratio is set as 2%. Results in term of NMI and AC illustrate that the performance of our algorithm is superior to other algorithms. In order to test how the prior affects the performance, the third type of methods are further compared with our PSSNMTFC in Table 4 and Fig. 2. Results show our algorithm is also superior to other algorithms with different prior ratios, which illustrates that our algorithm is effective.

Here we further compare these methods on real-world networks. First, we compare our algorithm PSSNMTFC with other algorithms on two attributed networks, Washington and Cora, and results are shown in Table 5. As shown, our algorithm has better performance compared with other types of algorithms.

Tables 6 and 7 show the results of our algorithm compared with other algorithms at different prior ratios in term of NMI on

Table 4 Comparison results on artificial networks in term of NMI and AC

Methods	Information used	Artificial network1		Artificial network2	
		NMI	AC	NMI	AC
UNMF	Links only	0.1854	0.3455	0.2922	0.4400
SMR	Content only	0.1264	0.2541	0.1132	0.2256
SCI	Links+content	0.2556	0.2631	0.2263	**0.5432**
GNMF	Links+prior	0.2052	0.3300	0.3465	0.3813
WSCDSM	Links+content	**0.3540**	0.3320	0.2110	0.3210
PSSNMTFC	+prior	0.2442	**0.4023**	**0.4116**	0.4896

Table 5 Experiment results on artificial networks with prior information ranging from 1% to 30% in term of NMI

Prior	Artificial network1				Artificial network2			
	PSSNMTFC	PSSNMF	SSNMF	FSSNMF	PSSNMTFC	PSSNMF	SSNMF	FSSNMF
1%	0.2299	0.2448	**0.3141**	0.2609	0.2955	**0.3415**	0.1463	0.2221
4%	0.5636	**0.6652**	0.3218	0.3444	0.7196	**0.7215**	0.3493	0.4752
6%	**0.7465**	0.6494	0.4912	0.5931	0.7235	**0.7622**	0.3556	0.5802
8%	0.8105	**0.8390**	0.4579	0.6926	**0.8734**	0.8693	0.3610	0.6127
10%	**0.8709**	0.8425	0.4821	0.8042	**0.9179**	0.8792	0.3989	0.6479
20%	**0.8768**	0.8416	0.4952	0.8001	**0.9353**	0.8940	0.3867	0.8561
30%	**0.9101**	0.8691	0.5200	0..8455	**0.9567**	0.9114	0.4654	0.8782

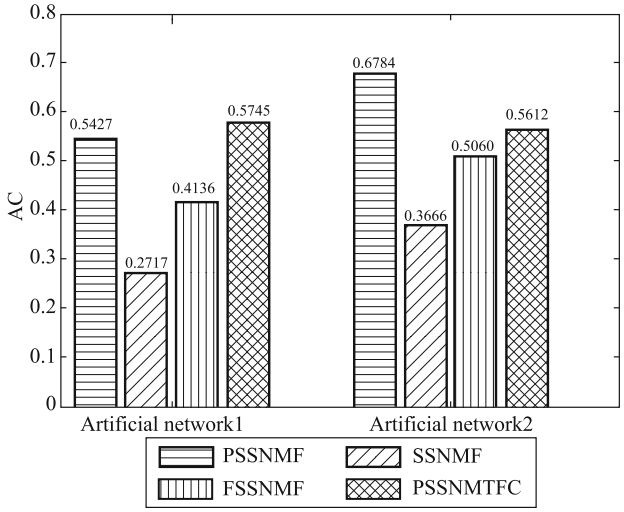


Fig. 2 Comparison results with prior 5% in term of AC

Table 6 Comparison results on Washington and Cora in terms of NMI

Information used	method	Washington		Cora	
links only	UNMF	0.0535		0.0963	
content only	SMR	0.0632		0.0078	
links+content	SCI	0.1257		0.1780	
links+prior	GNMF	2%	8%	2%	8%
		0.0225	0.6552	0.3256	0.8555
links+content	WSCDSM	**0.2564**	0.5552	0.5254	0.8083
+prior	PSSNMTFC	0.2225	**0.7072**	**0.8556**	**0.8999**

attributed networks. As shown in Table 6, our PSSNMTFC performs better than the other two methods on the small attributed networks WebKB. Table 7 also illustrates that our method PSSNMTFC not only performs better than the other methods, but also achieves a high accuracy with the prior of 10%. For example, the accuracy achieves 0.8906 when the prior is only 2% for Cora. When the prior ranges from 2% to 10%, the accuracy of our algorithm PSSNMTFC is improved by 6.21% and 9.77% respectively for the Cora and Citeseer

For larger networks, Citeseer, DBLP and PubMed, our algorithm PSSNMTFC still has advantages. We further use AC to validate the superiority of our algorithm on the three networks, and the results are shown in Fig. 3. It further proves that our method is more suitable on larger datasets with few prior information, such as 5% prior information. The reason is that our algorithm considers the node popularity, node contents and priors at the same time

Moreover, our algorithm PSSNMTF is compared with other embedding methods on a larger data set (Flickr), and results in Table 8 shows the superiority of our algorithm.

4.5 Parameter analysis and convergence analysis

In this paper, our proposed model involves several parameters, including the must-link matrix parameter ε, the balancing parameters α, β, λ, η and the Lagrange multiplier γ. For the parameter ε, we set it to 2 on all the experiments in this paper, just as did in most other semi-supervised community detection papers [38].

The function of the Lagrange multiplier γ (introduced in the optimization process) is to mitigate the constraints of $\|\boldsymbol{X}_i\|_1 = 1, i = 1, 2, \ldots, n$. It is only related to the updating rule of matrix X. As shown in the updating rule of formula (17), the correlation term of γ in molecule of formula (17) is an independent matrix, which is independent of any information on attributed networks, while the correlation term of γ in denominator of formula (17) is closely related to matrix $\boldsymbol{X}$. For the value of γ, if we set the same value for all network datasets, it may bring some fluctuations on $\boldsymbol{X}$, just because there is a huge difference between the denominator and the molecule in formula (17). Therefore, in order to alleviate this problem, we need to allocate the size of the parameters according to the scale of the attribute networks. Therefore, a relatively large value is set for

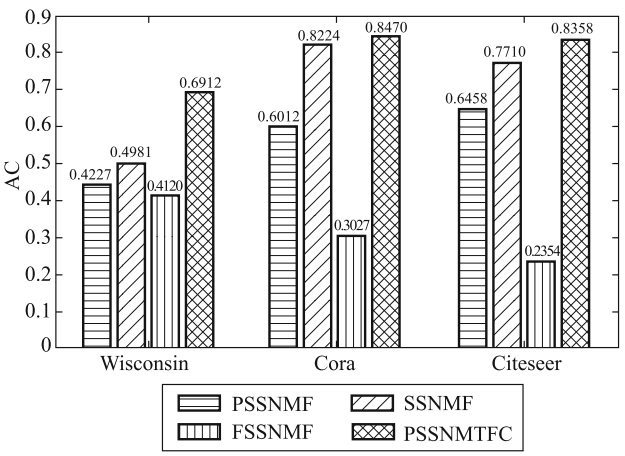


Fig. 3 Comparison results with prior 5% in term of AC

Table 7 Experiment results on WebKB with prior information ranging from 4% to 20% in term of NMI

Prior	PSSNMTFC	PSSNMF	FSSNMF	PSSNMTFC	PSSNMF	FSSNMF
	Texas			Cornell		
4%	0.2952	**0.3011**	0.1653	0.3809	**0.4016**	0.2390
8%	0.7735	**0.8053**	0.5210	**0.8562**	0.7216	0.5129
10%	**0.8210**	0.8002	0.6661	**0.9006**	0.8174	0.5793
15%	**0.8438**	0.8416	0.6735	**0.8244**	0.8094	0.6090
20%	**0.8677**	0.8599	0.7053	**0.8850**	0.8710	0.6125
	Washington			Wisconsin		
4%	0.2533	0.2116	**0.3281**	**0.5924**	0.5621	0.3137
8%	**0.7072**	0.6871	0.4138	**0.8530**	0.7363	0.3977
10%	0.7003	**0.9279**	0.4498	**0.9370**	0.8858	0.6271
15%	**0.8546**	0.8249	0.4400	**0.9088**	0.8216	0.6510
20%	0.8604	**0.9566**	0.4683	**0.9535**	0.9760	0.6932

Table 8 Experiment results on Cora, Citeseer, DBLP, Pubmed with prior information ranging from 2% to 10% (NMI.)

Prior	PSSNMTFC	PSSNMF	FSSNMF	PSSNMTFC	PSSNMF	FSSNMF
	Texas			Cornell		
2%	**0.8906**	0.5361	0.6631	**0.8611**	0.8315	0.3033
4%	**0.9399**	0.8066	0.6699	**0.9265**	0.9177	0.3359
6%	**0.9401**	0.9362	0.7615	**0.9411**	0.9224	0.3382
8%	**0.9670**	0.9430	0.7661	**0.9524**	0.9300	0.3679
10%	**0.9527**	0.9362	0.7745	**0.9588**	0.9568	0.4020
	DBLP			Pubmed		
2%	0.8378	0.8383	**0.8879**	**0.7753**	0.7331	0.6653
4%	0.8617	**0.8930**	0.8889	**0.7916**	0.7336	0.7034
6%	0.8882	**0.9028**	0.8890	**0.8882**	0.8836	0.8567
8%	**0.9347**	0.9053	0.8902	**0.9254**	0.8836	0.8974
10%	**0.9651**	0.9264	0.8941	**0.9321**	0.9041	0.9241

Table 9 results compared with some embedding methods on Flickr dataset (NMI, AC)

	DeepWalk	LINE	Node2Vec	PSSNMTF
NMI	0.3257	0.4628	0.3357	**0.4965**
AC	0.4673	0.5328	0.4732	**0.5573**

small data sets and a relatively small value is set for large data sets. Specifically, for the 4 small WebKB network datasets, we set a larger value of 10; and for the two larger network datasets (Cora and Citeseer), we set a smaller value of 0.5.

As for the balancing parameter α, β, λ and η, we analyse their effects on the Wisconsin dataset with the proposed algorithm PSSNMTFC. From 1 to 10, we search for the best parameter by binary search as shown in Fig. 4. For each balance parameter, the optimal parameters of the objective function are set as $\alpha = 1, \beta = 5, \lambda = 1, \eta = 10$.

The convergence of the algorithm is proved on the Wisconsin dataset, shown in Fig. 5. As the iterations increase, our algorithm is convergent, which is due the convergence of our loss function.

5 Conclusion and future work

In this paper, we develop a novel method (i.e., PSSNMTFC) for semantic community detection in attribute networks. We find semantic communities by incorporating the content and link information as well as the prior information altogether, by giving a new semi-supervised model based on the NMTF model. At the same time, we further consider the node popularity hidden in networks to better utilize the prior information to enhance this model. The extensive experiments illustrate that the new algorithm owns a higher accuracy for community detection compared with some state-of-the-art baselines on networks with different scales.

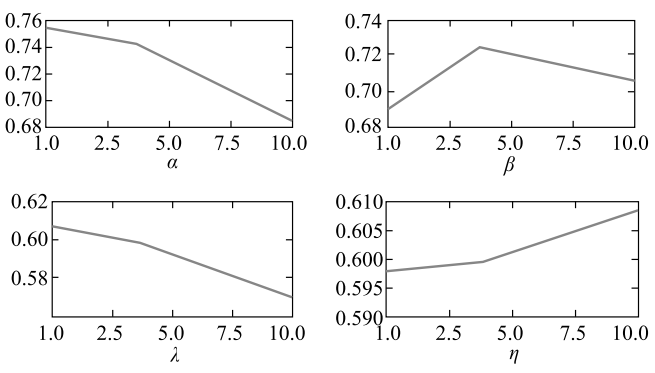


Fig. 4 NMI with different values for parameters α, β, λ and η

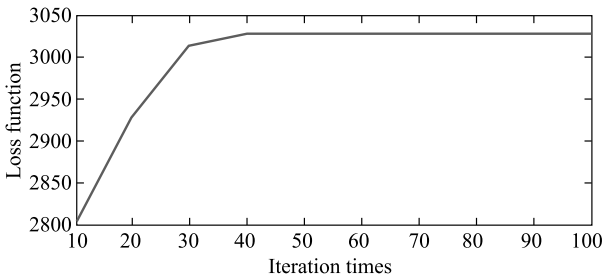


Fig. 5 Convergence curve that loss function changes with iteration times in our algorithm

In order to maintain the original geometric structure of the attribute networks, in this work we only incorporate the must-link prior information into the model while do not employ the cannot-link constraints. We plan to further improve the effectiveness of the algorithm by introducing the cannot-link constraint to our model in the future.

Acknowledgements This work was partly supported by the National Natural Science Foundation of China (Grant Nos. 61876128, 61772361) and the National Science Foundation of Hebei (F2019403070) and the science and technology research project for universities of Hebei (ZD2020175).

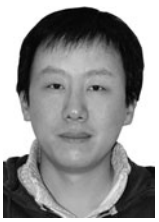
References

1. Yang J, McAuley J, Leskovec J. Community detection in networks with node attributes. In: Proceedings of IEEE International Conference on Data Mining. 2013, 1151–1156
2. Peel L, Larremore D B, Clauset A. The ground truth about metadata and community detection in networks. Science Advances, 2016, 3(5): e1602548
3. Newman M E J, Clauset A. Structure and inference in annotated networks. Nature Communications, 2016, 7: 11863
4. Bothorel C, Cruz J D, Magnani M, Micenkova B. Clustering attributed graphs: models, measures and methods. Network Science, 2015, 3(3): 408–444
5. Moayedikia A. Multi-objective community detection algorithm with node importance analysis in attributed networks. Applied Soft Computing, 2018, 67: 434-451
6. Atzmüller M. Subgroup and community analytics on attributed graphs. In: Proceedings of CEUR Workshop. 2015
7. Boden B. Combined Clustering of Graph and Attribute Data. Rwth Aachen, 2012, 13–18

8. Günnemann S, Boden B, Färber I, Seidl T. Efficient mining of combined subspace and subgraph clusters in graphs with feature vectors. In: Proceedings of Pacific-asia Conference on Knowledge Discovery and Data Mining. 2013, 261–275
9. Günnemann S, Färber I, Boden B, Seidl T. Subspace clustering meets dense subgraph mining: a synthesis of two paradigms. In: Proceedings of 2010 IEEE International Conference on Data Mining. 2010, 845–850
10. Chai B F, Wang J L, Xu J W, Li W B. Active semi-supervised community detection method based on link model. Journal of Computer Applications, 2017, 37(11): 3090–3094
11. Yang L, Cao X, Jin D, Wang X, Meng D. A unified semi-supervised community detection framework using latent space graph regularization. IEEE Transactions on Cybernetics, 2015, 45(11): 2585–2598
12. Shi X H, Lu H T, He Y C, He S. Community detection in social network with pairwisely constrained symmetric non-negative matrix factorization. In: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. 2015, 541–546
13. Liu X, Wang W J, He D X, Jiao P F, Jin D, Cannistraci C V. Semi-supervised community detection based on non-negative matrix factorization with node popularity. Information Sciences, 2017, 381: 304–321
14. Liu W Y, Yue K, Liu H, Zhang P. Associative categorization of frequent patterns based on the probabilistic graphical model. Frontiers of Computer Science, 2014, 8(2): 265–278
15. Combe D, Largeron C, Egyed-Zsigmond E, Géry M. Combining relations and text in scientific network clustering. In: Proceedings of International Conference on Advances in Social Networks Analysis and Mining. 2012, 1248–1253
16. Dang T, Viennet E. Community detection based on structural and attribute similarities. In: Proceedings of International Conference on Digital Society. 2012, 7–12
17. Neville J, Adler M, Jensen D. Clustering relational data using attribute and link information. In: Proceedings of International Joint Conference on Text Mining and Link Analysis Workshop. 2003
18. Muslim N. A combination approach to community detection in social networks by utilizing structural and attribute data. Social Networking, 2016, 5(1): 11–15
19. Elhadi H, Agam G. Structure and attributes community detection: comparative analysis of composite, ensemble and selection methods. In: Proceedings of Workshop on Social Network Mining and Analysis. 2013, 1–10
20. Strehl A, Ghosh J. Cluster ensembles — a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 2003, 3(3): 583–617
21. Xu Z Q, Ke Y P, Wang Y, Cheng H. A model-based approach to attributed graph clustering. In: Proceedings of ACM Sigmod International Conference on Management of Data. 2012, 505–516
22. Xu Z Q, Ke Y P, Wang Y, Cheng H, Cheng J. GBAGC: a general bayesian framework for attributed graph clustering. ACM Transactions on Knowledge Discovery from Data, 2014, 9(1): 1–43
23. Yu L, Wu B, Wang B. Topic model-based link community detection with adjustable range of overlapping. In: Proceedings of International Conference on Advances in Social Networks Analysis and Mining. 2013, 1437–1438
24. Liu L, Peng T. Clustering-based topical Web crawling using CFu-tree guided by link-context. Frontiers of Computer Science, 2014, 8(4): 581–595
25. Zhu S H, Yu K, Chi Y, Gong Y H. Combining content and link for classification using matrix factorization. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2007, 487–494
26. Yang T B, Jin R, Chi Y, Zhu S J. Combining link and content for community detection. In: Proceedings of Encyclopedia of Social Network Analysis and Mining. 2017, 1–10
27. Liu D, Liu X, Wang W J, Bai H Y. Semi-supervised community detection based on discrete potential theory. Physica A: Statistical Mechanics and Its Applications, 2014, 416: 173–182
28. Ma X K, Gao L, Yong X R, Fu L D. Semi-supervised clustering algorithm for community structure detection in complex networks. Physica A: Statistical Mechanics and Its Applications, 2010, 389(1): 187–197
29. Deng X L, Wen Y, Chen Y H. Highly efficient epidemic spreading model based LPA threshold community detection method. Neurocomputing, 2016, 210: 3–12
30. Wang X, Cui P, Wang J, Pei J. Community preserving network embedding. In: Proceedings of AAAI Conference on Artificial Intelligence. 2017
31. Wang W J, Liu X, Jiao P F, Chen X, Jin D. A unified weakly supervised framework for community detection and semantic matching. In: Proceedings of Pacific-asia Conference on Knowledge Discovery and Data Mining. 2018, 218–230
32. Brunet J P, Tamayo P, Golub T R, Mesirov J P. Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the National Academy of Sciences, 2004, 101(12): 4164–4169
33. Cavallari S, Zheng W S, Cai H Y, Chang C C. Learning community embedding with community detection and node embedding on graphs. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017, 377–386
34. Eaton E, Mansbach R. A spin-glass model for semi-supervised community detection. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence. 2012, 900–906
35. Jin H, Yu W, Li S J. Graph regularized nonnegative matrix tri-factorization for overlapping community detection. Physica A: Statistical Mechanics and Its Applications, 2019, 515: 376–387
36. Pei Y, Chakraborty N, Sycara K P. Nonnegative matrix tri-factorization with graph regularization for community detection in social networks. In: Proceedings of International Conference on Artificial Intelligence. 2015
37. Zhu S H, Yu K, Chi Y, Gong Y H. Combining content and link for classification using matrix factorization. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2007, 487–494
38. Wang X, Jin D, Cao X C, Yang L. Semantic community identification in large attribute networks. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. 2016, 265–271
39. Wu Q Y, Wang Z Y, Li C S, Ye Y M. Protein functional properties prediction in sparsely-label PPI networks through regularized non-negative matrix factorization. BMC Systems Biology, 2015, 9(S1): S9
40. Wang R S, Zhang S H, Wang Y, Zhang X S, Chen L N. Clustering complex networks and biological networks by nonnegative matrix factorization with various similarity measures. Neurocomputing, 2008, 72(1–3): 134–141
41. Zhang Y, Du N, Ge L, Jia K B. A collective NMF method for detecting protein functional module from multiple data sources. In: Proceedings of ACM Conference on Bioinformatics. 2012, 655–660
42. Chin P, Rao A, Vu V. Stochastic block model and community detection in the sparse graphs: a spectral algorithm with optimal rate of recovery. In: Proceedings of Conference on Learning Theory. 2015, 391–423
43. Cao J X, Jin D, Yang L, Dang J W. Incorporating network structure with node contents for community detection on large networks using deep learning. Neurocomputing, 2018, 297: 71–81
44. Wang D D, Li T, Zhu S G, Ding C H Q. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2008, 307–314
45. Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014, 701–710
46. Tang J, Qu M, Wang M Z, Zhang M. Line: large-scale information net-

work embedding. In: Proceedings of the 24th International Conference on World Wide Web. 2015, 1067–1077

47. Grover A, Leskovec J. node2vec: scalable feature learning for networks. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 855–864
48. Zhang Z Y, Sun K D, Wang S Q. Enhanced community structure detection in complex networks with partial background information. Scientific Reports, 2013, 3(1): 3241



Di Jin received his BS, MS, and PhD degrees in computer science from Jilin University, China in 2005, 2008, and 2012, respectively. He was a post-doctoral research fellow at the School of Design, Engineering, and Computer, Bournemouth University, U.K., from 2013 to 2014. He is currently an associate professor with the College of Intelligence and Computing, Tianjin University, China. He has published more than 50 papers in international journals and conferences in the areas of community detection, social network analysis, and machine learning.

Jing He received the BS degrees from Guangxi University, China in 2016 and MS degree in computer science and technology from Tianjin University, China in 2020. Her research interests are mainly related to community detection on social networks.

Bianfang Chai received the BE and ME degrees in computer science from Hebei University, China in 2002 and 2006 respectively, and the PhD degree from Beijing Jiaotong University, China in 2015. She is an associate professor with the School of Information Engineering, Hebei GEO University, China. Her current research interests include community detection, semi-supervised clustering, probabilistic graphical model, and complex network analysis.

Dongxiao He received her BS, MS, and PhD degrees in computer science from Jilin University, China in 2007, 2010, and 2014, respectively. She was a post-doctoral research fellow in Department of Computer Science, Dresden University of Technology, Germany from 2014 to 2015. She is an associate professor with the College of Intelligence and Computing, Tianjin University, China. She has published over 40 international journal and conference papers. Her current research interests include data mining and analysis of complex networks.