2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

Joint Role and Community Detection in Networks via $L_{2,1}$ Norm Regularized Nonnegative Matrix Tri-Factorization

Yulong Pei, George Fletcher, Mykola Pechenizkiy
Department of Mathematics and Computer Science
Eindhoven University of Technology, Eindhoven, the Netherlands
Email: {y.pei.1,g.h.l.fletcher,m.pechenizkiy}@tue.nl

***Abstract*—Role discovery and community detection in networks are two essential tasks in network analytics where the role denotes the global structural patterns of nodes in networks and the community represents the local connections of nodes in networks. Previous studies viewed these two tasks orthogonally and solved them independently while the relation between them has been totally neglected. However, it is intuitive that roles and communities in a network are correlated and complementary to each other. In this paper, we propose a novel model for simultaneous roles and communities detection (REACT) in networks. REACT uses non-negative matrix tri-factorization (NMTF) to detect roles and communities and utilizes $L_{2,1}$ norm as the regularization to capture the diversity relation between roles and communities. The proposed model has several advantages comparing with other existing methods: (1) it incorporates the diversity relation between roles and communities to detect them simultaneously using a unified model, and (2) it provides extra information about the interaction patterns between roles and between communities using NMTF. To analyze the performance of REACT, we conduct experiments on several real-world SNs from different domains. By comparing with state-of-the-art community detection and role discovery methods, the obtained results demonstrate REACT performs best for both role and community detection tasks. Moreover, our model provides a better interpretation for the interaction patterns between communities and between roles.**

I. Introduction

Role discovery and community detection in networks are two important tasks in network analytics and recent years have witnessed numerous approaches to solve both problems. Role discovery mines the structural properties of networks from a global perspective and clusters nodes into groups with different structural patterns [1]. For example, bridge nodes in a social network play an important role in information diffusion. By contrast, community detection analyzes networks from the local perspective and aims to group together nodes that are closely connected to each other [2]. For instance, an online social group consists of users having the same interests and interacting with each other frequently. A comparison of roles and communities is shown in Fig. 1. Discovering roles and communities in networks can shed light on numerous graph mining tasks such as link prediction [3], social recommendation [4], graph transferring [5] and anomaly detection [6].

ASONAM '19, August 27-30, 2019, Vancouver, Canada

ACM ISBN 978-1-4503-6868-1/19/08/$15.00
https://doi.org/10.1145/3341161.3342886

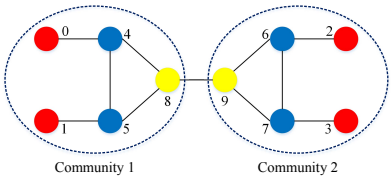


Fig. 1. An example of ten nodes belonging to (1) three groups (different colors indicate different groups) based on global structural information, i.e., *roles* and (2) two groups (groups are shown by the dashed ellipses) based on local structural information, i.e., *communities*. These exists a diversity relation between roles and communities. For example, in both community 1 and 2, nodes have all three roles.

Previous studies viewed the task of role discovery and community detection orthogonally and solved them independently. The relation between them has been totally neglected in most previous studies: (1) most role discovery methods only exploit global structural features but ignore the community structures. For instance, RolX [5] extracts global features for role detection. (2) Most community detection methods only focus on the local structural patterns of networks without considering roles. For example, NMF-based community detection method [7] decomposes the adjacency matrix which only captures the local connections between nodes. However, it is intuitive that roles and communities in a network are correlated and complementary to each other. An example is shown in Fig. 1. It can be observed that roles and communities analyze networks from the global and local perspectives, respectively. There is a diversity relation between these two concepts, i.e., the role assignment inside each community are diverse. This information has been neglected in most previous studies.

There are limited number of studies exploring the problem of joint detection of roles and communities, e.g., RC-Joint [8] and MMCR [9]. But there are some limitations in these studies. In RC-Joint, although the diversity relation between roles and communities has been integrated into the framework explicitly, it is incapable of learning the community interaction and role interaction patters. MMCR [9] extends MMSB by integrating the community-aware role assignment in a Bayesian frame-

work. Thus, the community-aware role interactions can be learned same to MMSB but the interaction patterns between communities have not been studied.

In order to overcome these limitations in previous studies, we propose a novel model for simultaneous role and community detection (REACT). REACT combines role discovery and community detection using NMTF components respectively. The role discovery component factorize the pairwise RoleSim similarity matrix [10] which models the global structural information and the community detection component decomposes the adjacency matrix which captures the local structural information of network. To model the diversity relation between roles and communities, we use the $L_{2,1}$ norm as a regularization term. REACT is advantageous over previous methods because it (1) detects roles and communities simultaneously in a unified model via $L_{2,1}$ norm to explicitly model the diversity relation between roles and communities, and (2) provides interaction patterns for roles and communities respectively via the extra latent factor in the output of NMTF.

The contributions of our work are summarized as follows:

- We propose a novel model to discover roles and communities simultaneously (REACT) in networks. REACT combines role discovery and community detection using two NMTF components and integrates the diversity relation between roles and communities using the $L_{2,1}$ norm.
- We derive efficient updating rules to learn the parameters of REACT, and propose a selection model to automatically determine the number of roles and communities.
- The conducted experimental study on several real-world networks from different domains demonstrate the effectiveness of REACT for both role and community detection. It also provides extra information in interaction patterns for communities and roles.

The rest of the paper is organized as follows. Section II provides an overview of the related work. Notations and problem formulation are given in Section III. Section IV explains the proposed REACT model. In Section V we then discuss our experimental study. Finally, in Section VI we draw conclusions and outline directions for future work.

II. Related Work

A. Role Discovery

The problem of role discovery was first studied in sociology to analyze and explain the functions of individuals in society [11], [12]. In the data mining community, role discovery is the process of partitioning the nodes into different classes based on their structural patterns. Methods for role discovery can be categorized into graph-based methods and feature-based methods [1]. Graph-based methods extract roles directly from the graph. Representative methods include different blockmodels such as stochastic blockmodels [13] and mixed-membership stochastic blockmodels (MMSB) [14]. However, the efficiency issue exists in these methods and they are not appropriate in large-scale graphs. For feature-based methods, RolX [5] has been proposed to extract structural roles from graphs. RolX extracts node features in a recursive method. Then the feature matrix V is factorized using the non-negative matrix factorization (NMF). RolX is an unsupervised method for role discovery which does not incorporate external guidance, e.g., the expertise and expectation from domain experts. Therefore, a guided learning framework for role discovery, GLRD [15]. Three types of supervision, i.e., sparsity, diversity and alternativeness in the roles, have been integrated into the RolX framework.

Some studies extended the structural role discovery task to social networks to identify the social roles [16], [17]. SRS [16] incorporates theories from sociology including five types of social patterns as features for role learning. DUMR [17] combines regularization from graph view and attribute view respectively and the proposed dual uncertainty regularization and this combination optimization function can capture both the local profile and the global network characteristics.

All of the methods introduced above only work on static networks. In practice, dynamics in real networks is ubiquitous. Based on RolX, DBMM [6] and DyNMF [18] have been proposed for modeling dynamic behaviors in evolving graphs. DBMM discovers roles of nodes in every time step and a transition matrix is learned based on the roles at each step. DyNMF explicitly models temporal information by introducing a role transition matrix. LAP [19] was proposed to extract latent features and then extended to learn the dynamic patterns and make predictions on object roles. Evolutionary clustering model [20] exploited user influence and user blockage for user similarity calculation and an ensemble clustering method is used for node clustering.

B. Community Detection

Communities are groups of vertices which probably share common properties and/or play similar roles within the graph [2]. 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 [21], k-core decomposition [22]. There are some methods aiming at explore the graph structures. Modularity maximization aims to find communities which can optimize a predefined measures [23]. Spectral graph clustering makes use of the eigenvectors of Laplacian matrices to group nodes [2].

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 [7] where it aims to factorize the adjacency matrix of the given network. Bayesian version of NMF for community detection has been proposed in [24] 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 specific bounds has been proposed in [25] to detect overlapping communities. NMTF with graph regularization has been used in [26] to discovery communities in attributed social networks. We refer the reader to [2] for more details in community detection research.

Two exceptions which joint detect roles and communities are RC-Joint [8] and MMCR [9]. RC-Joint also integrates

TABLE I. SUMMARY OF THE NOTATIONS.

Notation	Description
n	Number of nodes.
r	Number of roles.
c	Number of communities.
$A_{n\times n}$	Adjacency matrix of the given network.
$S_{n\times n}$	RoleSim similarity matrix of the given network.
$C_{n\times c}$	Community membership matrix.
$R_{n\times r}$	Role membership matrix.
$B_{c\times c}$	Community interaction matrix.
$M_{r\times r}$	Role interaction matrix.
λ	Trade-off parameter for diversity relation.

the diversity relation between roles and communities and updates the assignment of roles and community iteratively. MMCR extends MMSB by integrating the community-aware role assignment in a Bayesian framework. However, both studies could not provide the interaction patterns both between roles and between communities.

III. NOTATIONS AND BACKGROUNDS

We first summarize some notations in Table I and then introduce the backgrounds of the techniques we will use in this paper.

A. Non-negative Matrix Tri-factorization (NMTF)

Non-negative matrix factorization (NMF) [27] is a popular model in multivariate analysis and linear algebra where a matrix is factorized into two matrices, with the property that all three matrices have no negative elements. There are several advantages in NMF including ease of implementing inference and ease of interpreting results. Non-negative Matrix Tri-Factorization (NMTF) extends NMF by factorizing a matrix into three matrices so that it provides an extra latent matrix (the middle matrix) to denote the interaction patterns between clusters. Hence MNTF has been used in community detection recently [7], [26] to detect communities and learn community interactions. The objective function for NMTF is defined:

$$\min_{H,S} \|X - HSH^T\|_F^2, \quad s.t.\ H \geq 0, S \geq 0, H^T H = I \quad (1)$$

where $\|\cdot\|_F^2$ is the Frobenius norm. The non-negativity constraint in Eq. (1) makes the sparsity of the representation of the original data which is easier to interpret and more semantically meaningful compared with other factorization methods, e.g., SVD and PCA [28]. The orthogonal constraint can improve the interpretability of the clustering results. Using multiplicative update rules [29], the solution for Eq. (1) is shown as follows:

$$H_{jk} \leftarrow H_{jk} \frac{(W^T HS)_{jk}}{(HH^T W^T HS)_{jk}}, \quad (2)$$
$$S_{ik} \leftarrow S_{ik} \frac{(H^T WH)_{ik}}{(H^T HSH^T H)_{ik}}.$$

B. $L_{2,1}$ Norm

$L_{2,1}$ norm of a matrix is proposed as the rotational invariant of $L1$ norm [30]. Because of its robustness to noise and outliers, It has been widely used in many machine learning tasks, e.g., clustering [31], classification [32], and feature selection [33]. Given a matrix $M \in \mathbb{R}^{N\times K}$, $L_{2,1}$ norm is formally defined as:

$$\|M\|_{2,1} = \sum_{i=1}^{N} \left(\sum_{k=1}^{K} |M_{ik}|^2 \right)^{1/2} = \sum_{i=1}^{n} \|m_i\|_2, \quad (3)$$

where $\|\cdot\|_2$ is the L2 norm. $L_{2,1}$ norm controls the capacity of M and also ensures M to be sparse in rows so it is robust to noise and outliers. However, the non-smoothness of this norm makes it difficult for optimization.

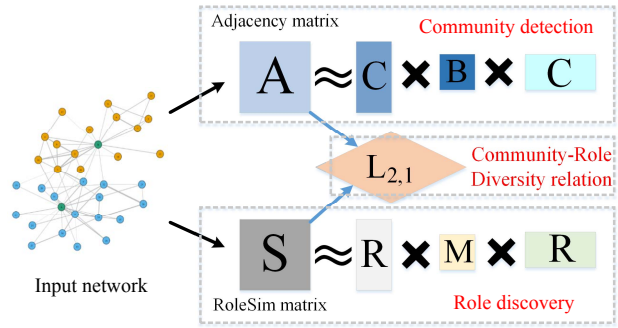


Fig. 2. Our proposed REACT model.

IV. OUR PROPOSED MODEL

A. REACT Model

In most previous studies on role discovery and community detection, one major problem is that they solved these problems separately and neglected the relation between roles and communities. To deal with this problem, we propose a novel model for joint role and community detection (REACT). REACT consists of three components: role discovery, community detection, and diversity relation component. The first and second components utilize NMTF to cluster nodes respectively and third component employs the $L_{2,1}$ norm to capture the diversity relation between roles and communities. A graphical representation of REACT is shown in Fig. 2.

1) Role discovery component: Role discovery clusters nodes according to the global structural information of the given network. Previous methods extract predefined features to capture the global structural information, For example, RolX [5] uses a recursive way to extract features, SRS [16] extracts features based on social theories. However, this type of methods has the some limitations: (1) they require prior knowledge to define the extracted features; and (2) they are incapable of learning role interaction patterns. To overcome these limitations, we select NMTF to factorize the pairwise RoleSim similarity matrix. RoleSim can measure the global similarity between two nodes which is based on the structural information instead of prior knowledge on features and has proved to be effective in role discovery [34]. NMTF factorizes a matrix into three matrices where the middle latent matrix/factor denote the interactions between clusters.

Formally, RoleSim similarity $S(u,v)$ between two nodes u and v is defined as:

$$S(u,v) = (1-\beta) \max_{M(u,v)} \frac{\sum_{(x,y)\in M(u,v)} S(x,y)}{|N(u)| + |N(v)| - |M(u,v)|} + \beta \quad (4)$$

where $N(u)$ and $N(v)$ are neighbors of node u and v, respectively. $M(u,v)$ is a matching between $N(u)$ and $N(v)$, i.e., $M(u,v) \subseteq N(u) \times N(v)$ is a bijection between $N(u)$ and $N(v)$. The parameter β is a decay factor where $0 < \beta < 1$. The intuition of RoleSim is that two nodes are structurally similar if their corresponding neighbors are also structurally similar. This intuition is consistent with the notion of automorphic and regular equivalence [11].

Then the objective function to decompose the RoleSim matrix for role discovery is defined as:

$$\min_{R,M} \|S - RMR^T\|_F^2, \quad s.t.\ R \geq 0,\ M \geq 0,\ R^T R = I. \tag{5}$$

2) Community detection component: Community detection aims to cluster nodes from the local perspective. Therefore, following [7], we directly factorize the adjacency matrix because the adjacency matrix captures the first-order local connections of nodes in the network. In order to obtain the community interaction patterns, we select NMTF to factorize the adjacency matrix. Formally, the objective function to factorize the adjacency matrix for community detection is defined as:

$$\min_{C,B} \|A - CBC^T\|_F^2, \quad s.t.\ C \geq 0,\ B \geq 0,\ C^T C = I. \tag{6}$$

The nonnegativity of NMTF ensures the explainability of node-community matrix and community interaction matrix. The latent factor B denotes the community interaction patterns.

3) Community-role relation component: As we mentioned in Section I, roles and communities are correlated to play two complementary roles in network analysis. In order to jointly detect roles and communities in networks, we need to model the relation between roles and communities. In this work, we aim to explicitly model the diversity relation between them (shown in Fig. 1). In specific, the distribution of role assignment inside each community should be as diverse as possible so we aim to minimize the product of role assignment membership matrix and community membership matrix which is similar to [8]. To achieve this goal, different types of norms can be employed. Considering (1) the noise and outliers in networks and (2) the possible inaccurate assignment in role and community detection, we select $L_{2,1}$ norm as the regularization because it is well-known to be robust to noise [31], [35]. Formally, we define the regularization as:

$$Reg = \|C^T R\|_{2,1}. \tag{7}$$

After defining two NMTF components for role discovery and community detection for networks, we propose to utilize both components jointly, and also integrate the diversity relation between roles and communities as the regularization. So we can discover roles and communities simultaneously:

$$L = \min_{C,B,R,M} \underbrace{\|S - RMR^T\|_F^2}_{role\ discovery} + \underbrace{\|A - CBC^T\|_F^2}_{community\ detection} + \underbrace{\lambda \|C^T R\|_{2,1}}_{diversity\ relation\ regularization} \quad s.t.\ C \geq 0, B \geq 0, R \geq 0, M \geq 0, C^T C = I, R^T R = I. \tag{8}$$

Optimization. The objective function in Eq. (8) is not convex for all parameters R, M C, and B simultaneously. We use the multiplicative update rules to solve this optimization problem due to its good compromise between speed and ease of implementation [28]. The optimization is done by iterating the four following steps until the convergence (or the number of iteration exceeds a given threshold): (1) fix matrices M, C and B to update R, (2) fix matrices R, C and B to update M, (3) fix matrices R, M and B to update C and (4) fix matrices R, M and C to update B (Algorithm 1). Note that we follow [33] to calculate the derivative of the objective function with $L_{2,1}$ norm.

Using the Karush-Kuhn-Tucker (KKT) complementary condition, the update rules of the objective function (8) are:

$$R_{jk} \leftarrow R_{jk} \frac{(S^T RM)_{jk}}{(RR^T SRM + \lambda CDC^T R)_{jk}} \tag{9}$$

$$M_{jk} \leftarrow M_{jk} \frac{(R^T SR)_{jk}}{(R^T RMR^T R)_{jk}} \tag{10}$$

$$C_{jk} \leftarrow C_{jk} \frac{(A^T CB)_{jk}}{(CC^T ACB + \lambda RR^T CD)_{jk}} \tag{11}$$

$$B_{jk} \leftarrow B_{jk} \frac{(C^T AC)_{jk}}{(C^T CBC^T C)_{jk}} \tag{12}$$

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

$$D_{jj} = \frac{1}{\|(C^T R)_j\|_2} \tag{13}$$

Algorithm 1 Optimization Algorithm

Input: Adjacency matrix A, number of roles r, number of communities c, trade-off parameter λ
Output: Role membership matrix R, community membership matrix C, role interaction matrix M, and community interaction matrix B

1: Calculate RoleSim similarity S according to Eq. (4)
2: Initialize R, M, C and B
3: **while** not converge **do**
4: Update R according to Eq. (9)
5: Update M according to Eq. (10)
6: Update C according to Eq. (11)
7: Update B according to Eq. (12)
8: **end while**

Complexity analysis of REACT. 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 4 - 7) is $O(n^2 r + nr^2 + r^3 + n^2 c + nc^2 + c^3)$. Since c and r can be viewed as the input constant and we also have $c \ll n$ and $r \ll n$, the complexity can be reduced to $O(n^2 + nr^2 + r^3 + nc^2 + c^3)$. By considering the number of iteration i the number of snapshots t the complexity is $O(i(n^2 + nr^2 + r^3 + nc^2 + c^3))$. Besides, the complexity of RoleSim (Line 1) is $O(kn^2 d)$, where n is the number of nodes, k is the number of iterations and d is the average of $y \log y$ over all node-pair bipartite graph in G [10].

B. Model Selection

In practice, the number of roles and communities may be not available beforehand. Therefore, how to determine the

TABLE II. SUMMARY OF DATA SETS USED IN THE EXPERIMENTS. NA MEANS THAT THE INFORMATION IS NOT AVAILABLE.

Type	Dataset	n	e	c	r
with community labels	Citeseer	3312	4732	6	NA
	Email	1005	25579	42	NA
with role labels	Brazil-air	131	1038	NA	4
	Europe-air	399	5995	NA	4
	USA-air	1190	13599	NA	4

suitable numbers is challenging. To tackle this problem, in this paper we follow [5] to use a model selection method to determine the number of roles and communities in the discovery process because this method has better generalization in different networks.

We adopt the model selection method proposed in [5]. It uses the Minimum Description Length (MDL) [36] to decide on the number of roles. Without loss of generality, we consider a general NMTF problem, i.e., $\|X - HSH^T\|_F^2$. The selected model, i.e., the appropriate number of clusters r, is the one that minimizes the description length $\mathcal{L}$, where $\mathcal{L}$ is sum of the model description cost $\mathcal{E}$ and the coding cost $\mathcal{M}$, i.e., $\mathcal{L} = \mathcal{M} + \mathcal{E}$. $\mathcal{M}$ is defined as $br(n+f)$, where b is the bits used for each element. $\mathcal{E}$ is defined as the KL divergence based error:

$$\mathcal{E} = \sum_{i,j} \left(X_{i,j} \log \frac{V_{i,j}}{(HSH^T)_{i,j}} - X_{i,j} + (HSH^T)_{i,j} \right) \quad (14)$$

In our case, we have two NMTF components so we select the suitable number of roles and communities separately.

V. EXPERIMENTAL STUDIES

A. Experimental Settings

To validate the effectiveness of our proposed REACT model for joint role and community detection, we conduct experiments on several real-world networks from different domains. We first evaluate REACT and state-of-the-art algorithms on the role discovery task (Section V-E, and compare REACT with previous community detection methods (Section V-F). We then investigate the influence of the trade-off parameter λ on the algorithms performance (Section V-G). We also empirically show that our method can provide extra community interaction and role interaction information using visualization (Section V-H).

B. Datasets

We conduct experiments on two types of network datasets: networks with ground-truth role labels[1] and community labels[2]. Community detection data consists of two networks and role discovery data consists of three networks. A brief summary of these datasets is shown in Table II.

C. Baselines

As our model aims to jointly solve the problem of community detection and role discovery, we compare REACT with three types of state-of-the-art methods: role discovery methods, community detection methods, and joint community and role detection methods. For role discovery methods, we use the following baselines:

- NMTF [7]: We use the same NMTF model in [7] but to factorize the RoleSim matrix instead of the adjacency matrix. Note that this is the role discovery component in our proposed model.
- RolX [5]: It is also a NMF based method and it decomposes a feature matrix which is extracted from the network based on some predefined operations.
- MMSB [14]: It is a Bayesian model which treats the role of each node as a latent variable and then uses Bayesian statistics to infer these latent variables.

For community detection methods, we use

- NMTF [7]: It is the basic NMTF method for community detection which factorizes the adjacency matrix into three matrices. Note that this is the community detection component in our proposed model.
- BigClam [38]: BigClam models the affiliation strength of each node to each community and assigns each node-community pair a nonnegative latent factor as the degree of community membership. Note that it is designed for overlapping community detection.
- BNMF [24]: It is the Bayesian version of NMF model to detect communities.

For joint methods:

- RC-Joint [8]: It aims to simultaneously identify community and structural role assignments in a network. In a nonparametric fashion, RC-Joint updates and improves community and role assignment iteratively.
- MMCR [9]: It detects the latent community and role of each node at the same time. By extending the role interaction probability to community-aware role interaction probabilities, MMCR can explicitly models the relation between communities and roles.

D. Evaluation Metrics

To evaluate the experimental results, we use purity and normalized mutual information (NMI) as the evaluation metrics. These metrics are widely used in the evaluation of clustering methods with ground-truth labels.

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:

$$Purity = \frac{1}{N} \sum_{i=1}^{k} max_j |c_i \cap t_j|, \quad (15)$$

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.

[1]The networks are from [37].

[2]These datasets are from http://snap.stanford.edu/data/index.html and https://linqs.soe.ucsc.edu/data.

TABLE III. ROLE DISCOVERY RESULTS.

	Brazil		Europe		USA	
	NMI	Purity	NMI	Purity	NMI	Purity
NMTF [7]	0.5133	0.7328	0.2813	0.5037	0.2533	0.5066
RolX [5]	0.5032	0.6867	0.2642	0.4862	0.2767	0.4967
MMSB [14]	0.3796	0.4012	0.2851	0.5088	0.2356	0.4578
RC-Joint [8]	0.3625	0.3864	0.2565	0.4862	0.2056	0.4186
MMCR [9]	0.3856	0.4213	0.2732	0.4987	0.2461	0.4780
REACT (actual)	0.5302	0.7557	0.2976	0.5388	0.3180	0.5891
REACT (inferred)	**0.5302**	**0.7557**	**0.2976**	**0.5388**	**0.3180**	**0.5891**

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:

$$NMI(\mathcal{C},\mathcal{D}) = \frac{2 * \mathcal{I}(\mathcal{C},\mathcal{D})}{\mathcal{H}(\mathcal{C}) + \mathcal{H}(\mathcal{D})}, \tag{16}$$

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.

E. Role Discovery

The results for role discovery is shown in Table III. For the baselines, we use the actual numbers of roles in each network as the input parameter. For our method, we report the results using different numbers of roles: (1) the actual numbers of roles as the input (REACT (actual)) and (2) using the method introduced in Section IV-B to infer the role numbers (REACT (inferred)). Note that due to the network characteristics, in this experiment the inferred number of roles is the same to the actual role number, i.e., 4 roles in all three networks.

From these results, it can be observed that:

- Our proposed method, REACT, outperforms other state-of-the-art methods in role discovery task which demonstrates the effectiveness of our method. REACT performs better than NMTF which is the role discovery component in our model. It indicates that by taking the diversity relation between roles and communities into consideration, we can achieve better performance.
- These two joint role and community detection methods do not perform well in role discovery task. This may result from that these methods perform better in optimizing different objective: (1) RC-Joint aims to maximize the likelihood of the network, and (2) MMCR aims to minimize the perplexity of the model.
- An interesting observation is that the inferred number of roles is exactly the same to the actual number of role. In fact, smaller number of roles can achieve better performance (we will present the empirically evidence below). This is because roles are defined based on the global structural patterns and the possible patterns are limited, i.e., they will not change with the increase/decrease of nodes/edges in a network.

Besides, we explore the effect on the performance with different numbers of roles and the results are shown in Fig. 3. As we mentioned before, smaller number of roles can achieve better performance. In these networks, the optimal number of roles is 4 which is the actual number of roles.

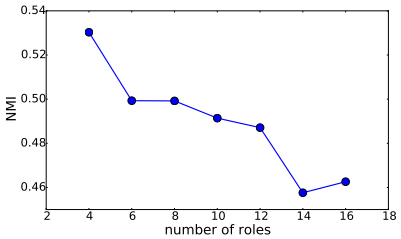


Fig. 3. Effect of number of roles on Brazil-airport network.

TABLE IV. COMMUNITY DETECTION RESULTS.

	Email		Citeseer	
	NMI	Purity	NMI	Purity
NMTF [7]	0.6059	0.5652	0.0790	0.2648
BNMF [24]	0.5862	0.5085	0.0812	0.2385
BigClam [38]	0.5705	0.5883	0.0735	0.1543
RC-Joint [8]	0.3021	0.3644	0.0956	0.2178
MMCR [39]	0.5675	0.5330	0.1194	0.2967
REACT (actual)	0.6194	0.5920	0.1158	0.3077
REACT (inferred)	**0.6560**	**0.6547**	**0.1557**	**0.3373**

F. Community Detection

The results for community detection is shown in Table IV. Same to the role discovery task, for the baselines, we use the actual numbers of communities in each network as the input parameter. For our method, we report the results using the actual numbers of communities and using the method introduced in Section IV-B to infer the community numbers. Note that the inferred number of communities for Email and Citesser network is 80 and 60, respectively.

From these results, it can be observed that:

- REACT performs better than other state-of-the-art methods in community detection task which demonstrates the effectiveness of our method. Similar, REACT performs better than NMTF which is the community detection component in our model. It indicates that by taking the diversity relation between roles and communities into consideration, we can achieve better performance.
- Same to the role discovery experiment, these two joint role and community detection methods do not perform well. The reason is similar to that in Section V-E.
- Different from role discovery, larger community number is preferred in REACT for community detection task. This conclusion may stem from the definition of community: with the increase of network scale, the number of dense connected subgraphs may increase correspondingly. More empirical evidence will be introduced below.

G. Influence of the Trade-off Parameter

Furthermore, we investigate the influence of the trade-off parameter λ on the REACTs performance of role discovery (Fig. 5) and community detection (Fig. 6). We change the value of the trade-off parameter λ from 0.1 to 1.0 and compare the corresponding NMI values. From these results, a conclusion

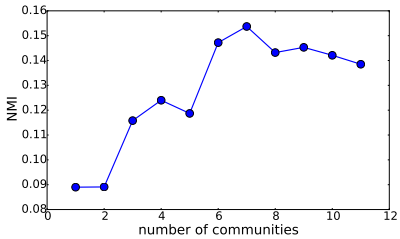
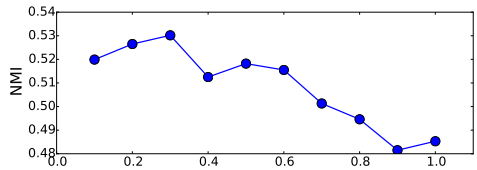
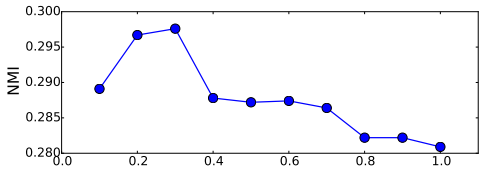


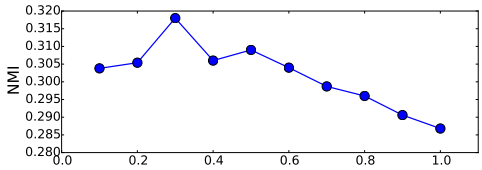
Fig. 4. Effect of number of communities on Citeseer network.



(a) Brazil-air data set



(b) Europe-air data set



(c) USA-air data set

Fig. 5. Effect of different trade-off parameters on the role discovery task.

for optimal λ select can be drawn: The value of λ should not be too small or too large. In practice, $\lambda = 0.3$ performs best on both tasks in most networks. The only exception is community detection task on Citeseer network, and $\lambda = 0.4$ achieve the best performance.

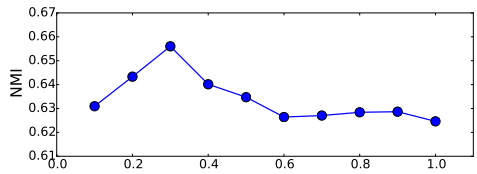
H. Community and Role Interaction Patterns

As we mentioned in the introduction, one advantage of our proposed method is that it can provide the information of community interaction patterns and role interaction patters while other baselines fail to. In this experiment, we use the Email and USA-airport networks for case study to visualize these interaction patterns. The visualization results are shown in Fig. 7 and 8. From these results, we can observe that the community interaction matrix is approximately diagonal where more interactions happen inside each community. By contrast, the role interaction matrix reflects more complicated and global patterns that is consistent to the definitions of roles.

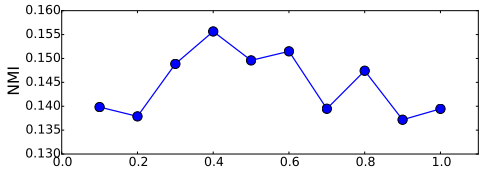
VI. Concluding Remarks

In this work we proposed a novel joint role and community detection approach named REACT. REACT consists of three components: role discovery, community detection and community-role relation. The first two components are based on nonnegative matrix tri-factorization (NMTF) and the last component is a regularization term to capture the diversity relation between roles and communities which is based on $L_{2,1}$ norm. We also extended MDL to determine the number of roles and communities automatically. We evaluated REACT in both role discovery and community detection compared to state-of-the-art methods. The results indicate the effectiveness of our proposed method in both tasks. We also investigated the effect of the trade-off parameter for community-role relation on both tasks. Besides, we empirically showed the interaction patters for roles and communities REACT can provide.

In future work we will exploit more relations between roles and communities in networks. We will also apply our method in different types of networks, e.g., dynamic and attributed. In the optimization, we will explore more advanced optimization methods for NMTF with regularization.

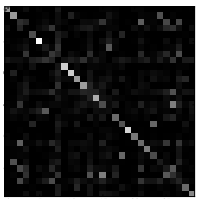


(a) NMI vs. λ on Email network.

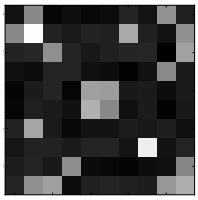


(b) NMI vs. λ on Citeseer network.

Fig. 6. Effect of different trade-off parameters on the community detection task.

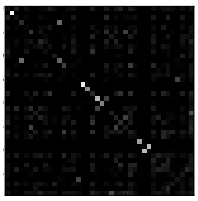


(a) Community Interaction

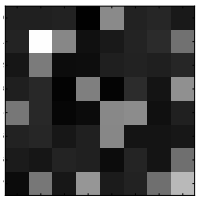


(b) Role Interaction

Fig. 7. Community and role interaction matrices in Email network.



(a) Community Interaction



(b) Role Interaction

Fig. 8. Community and role interaction matrices in USA-airport network.

REFERENCES

[1] R. Rossi, N. K. Ahmed *et al.*, "Role discovery in networks," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 27, no. 4, pp. 1112–1131, 2015.

[2] S. Fortunato, "Community detection in graphs," *Physics reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[3] A. Clauset, C. Moore, and M. E. Newman, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, no. 7191, p. 98, 2008.

[4] X. Wang, W. Lu, M. Ester, C. Wang, and C. Chen, "Social recommendation with strong and weak ties," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 5–14.

[5] K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li, "Rolx: structural role extraction & mining in large graphs," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1231–1239.

[6] R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson, "Modeling dynamic behavior in large evolving graphs," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 667–676.

[7] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, "Community discovery using nonnegative matrix factorization," *Data Mining and Knowledge Discovery*, vol. 22, no. 3, pp. 493–521, 2011.

[8] Y. Ruan and S. Parthasarathy, "Simultaneous detection of communities and roles from large networks," in *Proceedings of the second edition of the ACM conference on Online social networks*. ACM, 2014, pp. 203–214.

[9] T. Chen, L.-A. Tang, Y. Sun, Z. Chen, H. Chen, and G. Jiang, "Integrating community and role detection in information networks," in *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 2016, pp. 72–80.

[10] R. Jin, V. E. Lee, and H. Hong, "Axiomatic ranking of network role similarity," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 922–930.

[11] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.

[12] A. Ferligoj, P. Doreian, and V. Batagelj, "Positions and roles," *Sage handbook of social network analysis*, pp. 434–446, 2011.

[13] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social networks*, vol. 5, no. 2, pp. 109–137, 1983.

[14] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *Advances in Neural Information Processing Systems*, 2009, pp. 33–40.

[15] S. Gilpin, T. Eliassi-Rad, and I. Davidson, "Guided learning for role discovery (glrd): framework, algorithms, and applications," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 113–121.

[16] Y. Zhao, G. Wang, P. S. Yu, S. Liu, and S. Zhang, "Inferring social roles and statuses in social networks," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 695–703.

[17] Y. Cheng, A. Agrawal, A. Choudhary, H. Liu, and T. Zhang, "Social role identification via dual uncertainty minimization regularization," in *Data Mining (ICDM), 2014 IEEE International Conference on*. IEEE, 2014, pp. 767–772.

[18] Y. Pei, J. Zhang, G. Fletcher, and M. Pechenizkiy, "Dynmf: role analytics in dynamic social networks," in *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. AAAI Press, 2018, pp. 3818–3824.

[19] K. Li, S. Guo, N. Du, J. Gao, and A. Zhang, "Learning, analyzing and predicting object roles on dynamic networks," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 2013, pp. 428–437.

[20] S. Choobdar, P. Ribeiro, S. Parthasarathy, and F. Silva, "Dynamic inference of social roles in information cascades," *Data Mining and Knowledge Discovery*, pp. 1–26, 2014.

[21] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp. 359–392, 1998.

[22] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, "K-core organization of complex networks," *Physical review letters*, vol. 96, no. 4, p. 040601, 2006.

[23] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[24] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon, "Overlapping community detection using bayesian non-negative matrix factorization," *Physical Review E*, vol. 83, no. 6, p. 066114, 2011.

[25] Y. Zhang and D.-Y. Yeung, "Overlapping community detection via bounded nonnegative matrix tri-factorization," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 606–614.

[26] Y. Pei, N. Chakraborty, and K. Sycara, "Nonnegative matrix tri-factorization with graph regularization for community detection in social networks," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[27] H. Kim and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 2, pp. 713–730, 2008.

[28] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[29] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006, pp. 126–135.

[30] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 281–288.

[31] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using l21-norm," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 673–682.

[32] X. Cai, F. Nie, H. Huang, and C. Ding, "Multi-class l2, 1-norm support vector machine," in *2011 IEEE 11th International Conference on Data Mining*. IEEE, 2011, pp. 91–100.

[33] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint 2, 1-norms minimization," in *Advances in neural information processing systems*, 2010, pp. 1813–1821.

[34] Y. Pei, X. Du, J. Zhang, G. Fletcher, and M. Pechenizkiy, "struc2gauss: Structure preserving network embedding via gaussian embedding," *arXiv preprint arXiv:1805.10043*, 2018.

[35] W. Shao, L. He, and S. Y. Philip, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with $l_{2,1}$ regularization," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2015, pp. 318–334.

[36] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[37] L. F. Ribeiro, P. H. Saverese, and D. R. Figueiredo, "struc2vec: Learning node representations from structural identity," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 385–394.

[38] J. Yang and J. 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*. ACM, 2013, pp. 587–596.

[39] C. Chen, D. Gao, W. Li, and Y. Hou, "Inferring topic-dependent influence roles of twitter users," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 2014, pp. 1203–1206.