

Article

Studying Asymmetric Structure in Directed Networks by Overlapping and Non-Overlapping Models

Huan Qing 

School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China;
qinghuan@cumt.edu.cn

Abstract: We consider the problem of modeling and estimating communities in directed networks. Models to this problem in the previous literature always assume that the sending clusters and the receiving clusters have non-overlapping property or overlapping property simultaneously. However, previous models cannot model the directed network in which nodes in sending clusters have overlapping property, while nodes in receiving clusters have non-overlapping property, especially for the case when the number of sending clusters is no larger than that of the receiving clusters. This kind of directed network exists in the real world for its randomness, and by the fact that we have little prior knowledge of the community structure for some real-world directed networks. To study the asymmetric structure for such directed networks, we propose a flexible and identifiable Overlapping and Non-overlapping model (ONM). We also provide one model as an extension of ONM to model the directed network, with a variation in node degree. Two spectral clustering algorithms are designed to fit the models. We establish a theoretical guarantee on the estimation consistency for the algorithms under the proposed models. A small scale computer-generated directed networks are designed and conducted to support our theoretical results. Four real-world directed networks are used to illustrate the algorithms, and the results reveal the existence of highly mixed nodes and the asymmetric structure for these networks.



Citation: Qing, H. Studying Asymmetric Structure in Directed Networks by Overlapping and Non-Overlapping Models. *Entropy* **2022**, *24*, 1216. <https://doi.org/10.3390/e24091216>

Academic Editor: Adam Lipowski

Received: 4 July 2022

Accepted: 27 August 2022

Published: 30 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: community detection; directed network; network analysis; spectral clustering

1. Introduction

Community detection is a powerful tool in studying social networks with a latent structure of community [1–4]. The goal of community detection is to estimate a node's community information from the network. In the study of social networks, various models have been proposed for community detection to model different networks with different community structures [5]. Due to the extremely intensive studies on community detection, we only focus on identifiable models that are closely relevant to our study in this paper.

The Stochastic Blockmodel (SBM) [6] is a classical and widely used model for an undirected network. SBM assumes that the probability of an edge between two nodes only depends on the clusters they belong to, and this assumption is not realistic because nodes have various degrees in real-world networks. To model real-world un-directed networks in which nodes degrees vary, the Degree-Corrected Stochastic Blockmodel (DCSBM) [7] extends SBM by introducing degree heterogeneities. Under SBM and DCSBM, all nodes are pure, such that each node only belongs to one community. However, in real cases, some nodes may belong to multiple communities, and such nodes have overlapping (also known as mixed membership) property. To model undirected networks in which nodes have an overlapping property, Ref. [8] designs the Mixed Membership Stochastic Blockmodel (MMSB). Ref. [9] introduces the Degree-Corrected Mixed Membership model (DCMM), which extends MMSB by considering degree heterogeneities. Ref. [10] designs the Overlapping Continuous Community Assignment model (OCCAM), which equals

DCMM actually. Spectral methods with consistent estimations under the above models are provided in [9,11–17].

For directed networks in which all nodes have a non-overlapping property, Ref. [18] proposes a model called Stochastic co-Blockmodel (ScBM) and its extension, the Degree-Corrected Stochastic co-Blockmodel (DCScBM), by considering the degree heterogeneity, where ScBM (DCScBM) is an extension of SBM (DCSBM) from an un-directed network to a directed network. ScBM and DCScBM can model non-overlapping directed networks in which row nodes belong to K_r sending clusters (we also use community to denote cluster occasionally) and column nodes belong to K_c receiving clusters, where row nodes can differ from column nodes, and K_r can differ from K_c . Ref. [19] studies the consistency of some adjacency-based spectral algorithms under ScBM. Ref. [20] studies the consistency of the spectral method D-SCORE under DCScBM when $K_r = K_c$. Ref. [21] designs the Directed Mixed Membership Stochastic Blockmodel (DiMMSB) as an extension of ScBM and MMSB to model directed networks in which all nodes have overlapping property. Meanwhile, DiMMSB can also be seen as an extension of the two-way blockmodels with a Bernoulli distribution of [22]. All of the above models are identifiable under certain conditions. The identifiability of ScBM and DCScBM holds even for the case when $K_r \neq K_c$. DiMMSB is identifiable only when $K_r = K_c$. Sure, SBM, DCSBM, MMSB, DCMM, and OCCAM are identifiable when $K_r = K_c$, since they model undirected networks. For all the above models, row nodes and column nodes have symmetric structural information such that they always have non-overlapping property or overlapping property simultaneously. As shown by the identifiability of DiMMSB, to model a directed network in which all nodes have overlapping property, the identifiability of the model requires $K_r = K_c$. Naturally, there is a bridge model from ScBM to DiMMSB such that the bride model can model a directed network in which the row nodes and column nodes have asymmetric structural information such that they have different overlapping property. In this paper, we introduce this model and name it the Overlapping and Non-overlapping model.

Our contributions in this paper are as follows. We propose an identifiable model for directed networks, the Overlapping and Non-overlapping model (ONM for short). ONM allows that nodes in a directed network can have different overlapping properties. Without a loss of generality, in a directed network, we let the row nodes have overlapping property while the column nodes do not. The proposed model is identifiable when $K_r \leq K_c$. Recall that the identifiability of ScBM modeling non-overlapping directed networks holds even for the case $K_r \neq K_c$, and that DiMMSB modeling overlapping directed networks is identifiable only when $K_r = K_c$, this is the reason for why we call ONM modeling directed networks, in which row nodes have different overlapping properties to column nodes, as a bridge model from ScBM to DiMMSB. We also propose an identifiable model, Overlapping and Degree-Corrected Non-overlapping model (ODCNM), as an extension of ONM, by considering the degree heterogeneity. We construct two spectral algorithms to fit ONM and ODCNM. We show that our methods enjoy consistent estimations under mild conditions. Especially, our theoretical results under ODCNM match those under ONM when ODCNM reduces to ONM. The numerical results of simulated directed networks generated under ONM and ODCNM support our theoretical findings, and the results on four real-world directed networks demonstrate the advantages of our algorithms in studying the asymmetric structure between the sending and receiving clusters.

Notations. We take the following general notations in this paper. For any positive integer m , let $[m] := \{1, 2, \dots, m\}$, and let I_m denote the $m \times m$ identity matrix. For a vector x and fixed $q > 0$, $\|x\|_q$ denotes its l_q -norm. For a matrix M , M' denotes the transpose of the matrix M , $\|M\|$ denotes the spectral norm, $\|M\|_F$ denotes the Frobenius norm, and $\|M\|_{2 \rightarrow \infty}$ denotes the maximum l_2 -norm of all the rows of M . Let $\sigma_i(M)$ be the i -th largest singular value of matrix M , and let $\lambda_i(M)$ denote the i -th largest eigenvalue of the matrix M ordered by the magnitude. $M(i, :)$ and $M(:, j)$ denote the i -th row and the j -th column of matrix M , respectively. $M(S_r, :)$ and $M(:, S_c)$ denote the rows and columns in the index sets S_r and S_c of matrix M , respectively. For any matrix M , we simply use $Y = \max(0, M)$

to represent $Y_{ij} = \max(0, M_{ij})$ for any i, j . For any matrix $M \in \mathbb{R}^{m \times m}$, let $\text{diag}(M)$ be the $m \times m$ diagonal matrix whose i -th diagonal entry is $M(i, i)$, and let $\text{rank}(M)$ be M 's rank. $\mathbf{1}$ is a column vector with all entries being the value 1. e_i is a column vector whose i -th entry is 1, while other entries are zero. In this paper, C is a positive constant which may occasionally be different.

2. The Overlapping and Non-Overlapping Model

Consider a directed network $\mathcal{N} = (V_r, V_c, E)$, where $V_r = \{1, 2, \dots, n_r\}$ is the set of row nodes, $V_c = \{1, 2, \dots, n_c\}$ is the set of column nodes, and E is the set of edges from the row nodes to the column nodes. Note that since the row nodes can be different from the column nodes, we may have $V_r \cap V_c = \emptyset$ (i.e., there are no common nodes between V_r and V_c), and V_r may not be equal to V_c (i.e., the row nodes are different from the column nodes), which is a more general case than $V_r = V_c$ (i.e., all row nodes are same as column nodes), where $\emptyset$ denotes the null set, and such a directed network $\mathcal{N}$ is also known as a bipartite graph (or bipartite network) in [18,19]. In this paper, we use the subscript r and c to distinguish the terms for the row nodes and column nodes, where works in [18,19,23–26] also consider the general bipartite setting, such that the row nodes may differ from the column nodes. Let $A \in \{0, 1\}^{n_r \times n_c}$ be the bi-adjacency matrix of directed network $\mathcal{N}$, such that $A(i_r, i_c) = 1$ if there is a directional edge from row node i_r to column node i_c , and $A(i_r, i_c) = 0$ otherwise. For convenience, we call the community that the row nodes belong to as the row community (or sending cluster occasionally), and the community that the column nodes belong to as the column community (or receiving cluster occasionally).

We propose a new blockmodel which we call the Overlapping and Non-overlapping model (ONM for short). ONM can model directed networks whose row nodes belong to K_r overlapping row communities, while the column nodes belong to K_c non-overlapping column communities. For row nodes, let $\Pi_r \in \mathbb{R}^{n_r \times K_r}$ be the membership matrix, such that

$$\Pi_r(i_r, \cdot) \geq 0, \|\Pi_r(i_r, \cdot)\|_1 = 1 \quad \text{for } i_r \in [n_r]. \quad (1)$$

Call row node i_r *pure* if $\Pi_r(i_r, \cdot)$ degenerates (i.e., one entry is 1, all others $K_r - 1$ entries are 0), and *mixed* otherwise. From such a definition, row node i_r has mixed membership and may belong to more than one row communities for $i_r \in [n_r]$.

For column nodes, let ℓ be the $n_c \times 1$ vector whose i_c -th entry $\ell(i_c) = k$ if column node i_c belongs to the k -th column community, and $\ell(i_c)$ takes value from $\{1, 2, \dots, K_c\}$ for $i_c \in [n_c]$. Let $\Pi_c \in \mathbb{R}^{n_c \times K_c}$ be the membership matrix of column nodes, such that for $i_c \in [n_c], k \in [K_c]$,

$$\Pi_c(i_c, k) = 1 \text{ when } \ell(i_c) = k, \text{ and } 0 \text{ otherwise, and } \|\Pi_c(i_c, \cdot)\|_1 = 1. \quad (2)$$

From such a definition, column node i_c belongs to *exactly* one of the K_c column communities for $i_c \in [n_c]$. Sure, all of the column nodes are pure nodes.

In this paper, we assume that

$$K_r \leq K_c. \quad (3)$$

Equation (3) is required for the identifiability of ONM. Let $P \in \mathbb{R}^{K_r \times K_c}$ be the probability matrix, such that

$$0 \leq P(k, l) \leq \rho \leq 1 \quad \text{for } k \in [K_r], l \in [K_c], \quad (4)$$

where ρ controls the network sparsity and is called the sparsity parameter in this paper. For convenience, set $P = \rho \tilde{P}$, where $\tilde{P}(k, l) \in [0, 1]$ for $k \in [K_r], l \in [K_c]$, and

$\max_{k \in [K_r], l \in [K_c]} \tilde{P}(k, l) = 1$ for model identifiability. For all pairs of (i_r, i_c) with $i_r \in [n_r], i_c \in [n_c]$, our model assumes that $A(i_r, i_c)$ are independent Bernoulli random variables satisfying

$$\Omega := \Pi_r P \Pi_c', \quad A(i_r, i_c) \sim \text{Bernoulli}(\Omega(i_r, i_c)), \quad (5)$$

where $\Omega = \mathbb{E}[A]$, and we call it the population adjacency matrix in this paper.

Definition 1. Call model (1)–(5) the Overlapping and Non-overlapping model (ONM), and denote it with $\text{ONM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$.

Remark 1. Under $\text{ONM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, for $i_r \in [n_r], j_c \in [n_c]$, since $\mathbb{P}(A(i_r, j_c) = 1) = \Omega(i_r, j_c) = \rho \Pi(i_r, :) \tilde{P} \Pi_c'(j_c, :)$, we see that increasing ρ increases the probability to generate an edge from row node i_r to column node j_c , i.e., the sparsity of the network is governed by ρ .

The following conditions are sufficient for the identifiability of ONM:

- (I1) $\text{rank}(P) = K_r$, $\text{rank}(\Pi_r) = K_r$, and $\text{rank}(\Pi_c) = K_c$.
- (I2) There is at least one pure row node for each of the K_r row communities.

For $k \in [K_r]$, let $\mathcal{I}_r^{(k)} = \{i \in [n_r] : \Pi_r(i, k) = 1\}$. By condition (I2), $\mathcal{I}_r^{(k)}$ is non-empty for all $k \in [K_r]$. For $k \in [K_r]$, select one row node from $\mathcal{I}_r^{(k)}$ to construct the index set $\mathcal{I}_r$; i.e., $\mathcal{I}_r$ is the indices of row nodes corresponding to K_r pure row nodes, one from each row community. Without loss of generality, let $\Pi_r(\mathcal{I}_r, :) = I_{K_r}$ (Lemma 2.1 [17] also has a similar setting to design their spectral algorithm under MMSB). $\mathcal{I}_c$ is defined similarly for the column nodes, such that $\Pi_c(\mathcal{I}_c, :) = I_{K_c}$. The next proposition guarantees the identifiability of ONM.

Proposition 1. If conditions (I1) and (I2) hold, ONM is identifiable: For eligible (P, Π_r, Π_c) and $(\tilde{P}, \tilde{\Pi}_r, \tilde{\Pi}_c)$, if $\Pi_r P \Pi_c' = \tilde{\Pi}_r \tilde{P} \tilde{\Pi}_c'$, then $P = \tilde{P}$, $\Pi_r = \tilde{\Pi}_r$, and $\Pi_c = \tilde{\Pi}_c$.

All proofs of propositions, lemmas, and theorems are provided in Appendix B and Appendix C of this paper. Compared to some previous models, ONM models different directed networks.

- When the row nodes are the same as the column nodes, $K_r = K_c$, and all nodes are pure, ONM degenerates to SBM. However, ONM can model directed networks where row nodes enjoy mixed memberships, while SBM only models un-directed networks.
- When all row nodes are pure, our ONM reduces to ScBM with K_r row clusters and K_c column clusters [18]. However, ONM allows for row nodes to have overlapping memberships, while ScBM does not. Meanwhile, for model identifiability, ScBM does not require $\text{rank}(P) = K_r$ that ONM requires, and this can be seen as the cost of ONM when modeling the overlapping row nodes.
- Though DiMMSB [21] can model directed networks whose row and column nodes have overlapping memberships, DiMMSB requires $K_r = K_c$ for model identifiability. For comparison, our ONM allows $K_r \leq K_c$ at the cost of losing the overlapping property of the column nodes.

2.1. A Spectral Algorithm for Fitting ONM

The primary goal of the proposed algorithm is to estimate the row membership matrix Π_r and the column membership matrix Π_c from the observed adjacency matrix A with a given K_r and K_c . We now discuss our intuition for the design of our algorithm to fit ONM.

Under conditions (I1) and (I2), by basic algebra, we have $\text{rank}(\Omega) = K_r$. Let $\Omega = U_r \Lambda U_c'$ be the compact singular value decomposition of Ω , where $U_r \in \mathbb{R}^{n_r \times K_r}, \Lambda \in \mathbb{R}^{K_r \times K_r}, U_c \in \mathbb{R}^{n_c \times K_r}, U_r' U_r = I_{K_r}, U_c' U_c = I_{K_r}$, and I_{K_r} is a $K_r \times K_r$ identity matrix. Let $n_{c,k} = |\{i_c : \ell(i_c) = k\}|$ be the size of the k -th column community for $k \in [K_c]$. Let $n_{c,\max} = \max_{k \in [K_c]} n_{c,k}$ and $n_{c,\min} = \min_{k \in [K_c]} n_{c,k}$. Meanwhile, without causing confusion,

let n_{c,K_r} be the K_r -th largest size among all column communities. The following lemma guarantees that U_r enjoys an ideal simplex structure and U_c has K_c distinct rows.

Lemma 1. Under $ONM_{n_r,n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, there exists a unique $K_r \times K_r$ matrix B_r and a unique $K_c \times K_r$ matrix B_c , such that

- $U_r = \Pi_r B_r$, where $B_r = U_r(\mathcal{I}_r, :)$. Meanwhile, $U_r(i_r, :) = U_r(\bar{i}_r, :)$ when $\Pi_r(i_r, :) = \Pi_r(\bar{i}_r, :)$ for $i_r, \bar{i}_r \in [n_r]$.
- $U_c = \Pi_c B_c$. Meanwhile, $U_c(i_c, :) = U_c(\bar{i}_c, :)$ when $\ell(i_c) = \ell(\bar{i}_c)$ for $i_c, \bar{i}_c \in [n_c]$, i.e., U_c has K_c distinct rows. Furthermore, when $K_r = K_c = K$, we have $\|B_c(k, :) - B_c(l, :)\|_F = \sqrt{\frac{1}{n_{c,k}} + \frac{1}{n_{c,l}}}$ for all $1 \leq k < l \leq K$.

Lemma 1 says that the rows of U_c form a K_r -simplex in $\mathbb{R}^{K_r}$, which we call the Ideal Simplex (IS), with the K_r rows of B_r being the vertices. This IS is also found in [9,17,21]. Meanwhile, Lemma 1 says that U_c has K_c distinct rows, and if two column nodes i_c and $\bar{i}_c$ are from the same column community, then $U_c(i_c, :) = U_c(\bar{i}_c, :)$.

Under ONM, to recover Π_c from U_c , since U_c has K_c distinct rows, applying the k-means algorithm on all rows of U_c returns true column communities by Lemma 1. Since U_c has K_c distinct rows, we can set $\delta_c = \min_{k \neq l} \|B_c(k, :) - B_c(l, :)\|_F$ to measure the minimum center separation of B_c . By Lemma 1, $\delta_c \geq \sqrt{\frac{2}{n_{c,\max}}}$ when $K_r = K_c = K$ under $ONM_{n_r,n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$. However, when $K_r < K_c$, it is a challenge to obtain a positive lower bound of δ_c ; see the proof of Lemma 1 for details.

Under ONM, to recover Π_r from U_r , since B_r is full rank, if U_r and B_r are known in advance ideally, we can exactly recover Π_r by setting $\Pi_r = U_r B_r' (B_r B_r')^{-1}$ via Lemma 1. Set $Y_r = U_r B_r' (B_r B_r')^{-1}$. Since $Y_r \equiv \Pi_r$ and $\|\Pi_r(i_r, :)\|_1 = 1$ for $i_r \in [n_r]$, we have

$$\Pi_r(i_r, :) = \frac{Y_r(i_r, :)}{\|Y_r(i_r, :)\|_1}, i_r \in [n_r].$$

With a given U_r , since it enjoys IS structure $U_r = \Pi_r B_r \equiv \Pi_r U_r(\mathcal{I}_r, :)$, as long as we can obtain the row corner matrix $U_r(\mathcal{I}_r, :)$ (i.e., B_r), we can recover Π_r exactly. As mentioned in [9,17,21], for such an ideal simplex, the successive projection (SP) algorithm [27] (for details of SP, see Algorithm A1) can be applied to U_r with K_r row communities to find $U_r(\mathcal{I}_r, :)$.

Based on the above analysis, we are now ready to give the following algorithm which we call Ideal ONA. Input Ω , K_r , and K_c with $K_r \leq K_c$. Outputs: Π_r and ℓ .

- Let $\Omega = U_r \Lambda U_c'$ be the compact SVD of Ω , such that $U_r \in \mathbb{R}^{n_r \times K_r}$, $U_c \in \mathbb{R}^{n_c \times K_r}$, $\Lambda \in \mathbb{R}^{K_r \times K_r}$, $U_r' U_r = I_{K_r}$, and $U_c' U_c = I_{K_r}$.
- For the row nodes,
 - Run the SP algorithm on all rows of U_r , assuming there are K_r row communities to obtain $U_r(\mathcal{I}_r, :)$. Set $B_r = U_r(\mathcal{I}_r, :)$.
 - Set $Y_r = U_r B_r' (B_r B_r')^{-1}$. Recover Π_r by setting $\Pi_r(i_r, :) = \frac{Y_r(i_r, :)}{\|Y_r(i_r, :)\|_1}$ for $i_r \in [n_r]$.

For the column nodes,

- Run k-means on U_c assuming that there are K_c column communities, i.e., find the solution to the following optimization problem

$$M^* = \operatorname{argmin}_{M \in M_{n_c, K_r, K_c}} \|M - U_c\|_F^2,$$

where M_{n_c, K_r, K_c} denotes the set of $n_c \times K_r$ matrices with only K_c different rows.

- Use M^* to obtain the labels vector ℓ of the column nodes. Note that since M^* has K_c distinct rows, two different column nodes, $i_c, \bar{i}_c \in [n_c]$, are in the same column community if $M^*(i_c, :) = M^*(\bar{i}_c, :)$.

Following a similar proof of Theorem 1 of [21], the Ideal ONA exactly recovers row nodes memberships and column nodes labels, and this also verifies the identifiability of ONM in turn. For convenience, call the two steps for column nodes “run k-means on U_c assuming there are K_c column communities to obtain ℓ ”.

We now extend the ideal case to the real case. Set $\tilde{A} = \hat{U}_r \hat{\Lambda} \hat{U}_c'$ be the top- K_r -dimensional SVD of A , such that $\hat{U}_r \in \mathbb{R}^{n_r \times K_r}$, $\hat{U}_c \in \mathbb{R}^{n_c \times K_r}$, $\hat{\Lambda} \in \mathbb{R}^{K_r \times K_r}$, $\hat{U}_r' \hat{U}_r = I_{K_r}$, $\hat{U}_c' \hat{U}_c = I_{K_r}$, and $\hat{\Lambda}$ contains the top K_r singular values of A . For the real case, we use $\hat{B}_r, \hat{B}_c, \hat{Y}_r, \hat{\Pi}_r, \hat{\Pi}_c$ given in Algorithm 1 to estimate $B_r, B_c, Y_r, \Pi_r, \Pi_c$, respectively. Algorithm 1, called the Overlapping and Non-overlapping algorithm (ONA for short), is a natural extension of the Ideal ONA to the real case. In ONA, we set the negative entries of $\hat{Y}_r$ as 0 by setting $\hat{Y}_r = \max(0, \hat{Y}_r)$, for the reason that the weights for any row node should be non-negative while there may exist some negative entries of $\hat{U}_r \hat{B}_r' (\hat{B}_r \hat{B}_r')^{-1}$. Note that in a directed network, if the column nodes have an overlapping property while row nodes do not, to perform community detection for such a directed network, the transpose of the adjacency matrix should be set as input when applying our algorithm.

Algorithm 1 Overlapping and Non-overlapping Algorithm (ONA)

Require: The adjacency matrix $A \in \mathbb{R}^{n_r \times n_c}$ of a directed network, the number of row communities K_r , and the number of column communities K_c with $K_r \leq K_c$.

Ensure: The estimated $n_r \times K_r$ membership matrix $\hat{\Pi}_r$ for row nodes, and the estimated $n_c \times 1$ labels vector $\hat{\ell}$ for column nodes.

- 1: Compute $\hat{U}_r \in \mathbb{R}^{n_r \times K_r}$ and $\hat{U}_c \in \mathbb{R}^{n_c \times K_r}$ from the top- K_r -dimensional SVD of A .
- 2: For row nodes:
 - Apply SP algorithm (i.e., Algorithm A1) on the rows of $\hat{U}_r$ assuming there are K_r row clusters to obtain the near-corners matrix $\hat{U}_r(\hat{\mathcal{I}}_r, :) \in \mathbb{R}^{K_r \times K_r}$, where $\hat{\mathcal{I}}_r$ is the index set returned by SP algorithm. Set $\hat{B}_r = \hat{U}_r(\hat{\mathcal{I}}_r, :)$.
 - Compute the $n_r \times K_r$ matrix $\hat{Y}_r$ such that $\hat{Y}_r = \hat{U}_r \hat{B}_r' (\hat{B}_r \hat{B}_r')^{-1}$. Set $\hat{Y}_r = \max(0, \hat{Y}_r)$ and estimate $\Pi_r(i_r, :)$ by $\hat{\Pi}_r(i_r, :) = \frac{\hat{Y}_r(i_r, :)}{\|\hat{Y}_r(i_r, :)\|_1}, i_r \in [n_r]$.

For column nodes: run k-means on $\hat{U}_c$ assuming there are K_c column communities to obtain $\hat{\ell}$.

2.2. Main Results for ONA

In this section, we show the consistency of our algorithm for fitting the ONM as the number of row nodes n_r and the number of column nodes n_c increases. Throughout this paper, K_r and K_c are two known integers. First, we assume that:

Assumption 1. $\rho \max(n_r, n_c) \geq \log(n_r + n_c)$.

Assumption 1 controls the sparsity of the directed network considered for theoretical study. When building an estimation consistency of the spectral clustering methods in community detection, the sparsity assumption is common; see [13,14,17,18,20,21]. Especially, when ONM reduces to SBM, the sparsity requirement in Assumption 1 is consistent with that of Theorem 3.1 in [13], which guarantees the theoretical optimality on the sparsity condition of this paper. To measure the performance of ONA for row nodes memberships, since row nodes have mixed memberships, naturally, we use the l_1 norm difference between Π_r and $\hat{\Pi}_r$. Since the column nodes are all pure nodes, we consider the performance criterion defined in [15] to measure the estimation error of ONA on the column nodes. We introduce this measurement of estimation error below.

Let $\mathcal{T}_c = \{\mathcal{T}_{c,1}, \mathcal{T}_{c,2}, \dots, \mathcal{T}_{c,K_c}\}$ be the true partition of column nodes $\{1, 2, \dots, n_c\}$ obtained from ℓ , such that $\mathcal{T}_{c,k} = \{i_c \in [n_c] : \ell(i_c) = k\}$ for $k \in [K_c]$. Let $\hat{\mathcal{T}}_c = \{\hat{\mathcal{T}}_{c,1}, \hat{\mathcal{T}}_{c,2}, \dots, \hat{\mathcal{T}}_{c,K_c}\}$ be the estimated partition of column nodes $\{1, 2, \dots, n_c\}$ obtained

from $\hat{\ell}$ of ONA, such that $\hat{\mathcal{T}}_{c,k} = \{i_c \in [n_c] : \hat{\ell}(i_c) = k\}$ for $k \in [K_c]$. The criterion is defined as

$$\hat{f}_c = \min_{\pi \in S_{K_c}} \max_{k \in [K_c]} \frac{|\mathcal{T}_{c,k} \cap \hat{\mathcal{T}}_{c,\pi(k)}^c| + |\mathcal{T}_{c,k}^c \cap \hat{\mathcal{T}}_{c,\pi(k)}|}{n_{c,k}},$$

where S_{K_c} is the set of all permutations of $\{1, 2, \dots, K_c\}$, and the superscript c denotes the complementary set. As mentioned in [15], $\hat{f}_c$ measures the maximum proportion of column nodes in the symmetric difference of $\mathcal{T}_{c,k}$ and $\hat{\mathcal{T}}_{c,\pi(k)}$.

The next theorem gives the theoretical bounds on the estimations of memberships for both the row and column nodes, which is the main theoretical result for ONA.

Theorem 1. Under $ONM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, when Assumption 1 holds, suppose that $\sigma_{K_r}(\Omega) \geq C\sqrt{\rho(n_r + n_c)\log(n_r + n_c)}$, with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$ for any $\alpha > 0$,

- For row nodes, there exists a permutation matrix $\mathcal{P}_r$ such that

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega \kappa(\Pi_r' \Pi_r) K_r \sqrt{\lambda_1(\Pi_r' \Pi_r)}),$$

where $\omega = \|\hat{U}_r \hat{U}_r' - U_r U_r'\|_{2 \rightarrow \infty}$ is the row-wise singular eigenvector error.

- For column nodes, $\hat{f}_c = O(\frac{K_r K_c \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 \sigma_{K_r}^2(\Pi_r) n_{c, K_r} n_{c, \min}})$. Especially, when $K_r = K_c = K$,

$$\hat{f}_c = O(\frac{K^2 \max(n_r, n_c) n_{c, \max} \log(n_r + n_c)}{\sigma_K^2(\tilde{P}) \rho \sigma_K^2(\Pi_r) n_{c, \min}^2}).$$

Adding conditions similar to Corollary 3.1 in [17], we have the following corollary.

Corollary 1. Under $ONM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, suppose conditions in Theorem 1 hold, and further, suppose that $\lambda_{K_r}(\Pi_r' \Pi_r) = O(\frac{n_r}{K_r})$, $n_{c, \min} = O(\frac{n_c}{K_c})$, with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$,

- For row nodes, when $K_r = K_c = K$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\frac{K^2 (\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\sigma_K(\tilde{P}) \sqrt{\rho n_c}}).$$

- For column nodes, $\hat{f}_c = O(\frac{K_r^2 K_c^3 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 n_r n_c^2})$. When $K_r = K_c = K$,

$$\hat{f}_c = O(\frac{K^4 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_K^2(\tilde{P}) \rho n_r n_c}).$$

Especially, when $n_r = O(n)$, $n_c = O(n)$, $K_r = O(1)$, and $K_c = O(1)$,

- For row nodes, when $K_r = K_c = K$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\frac{\sqrt{\log(n)}}{\sigma_K(\tilde{P}) \sqrt{\rho n}}).$$

- For column nodes, $\hat{f}_c = O(\frac{\log(n)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 n^2})$. When $K_r = K_c = K$,

$$\hat{f}_c = O(\frac{\log(n)}{\sigma_K^2(\tilde{P}) \rho n}).$$

When $n_r = O(n)$, $n_c = O(n)$, $K_r = K_c = K = O(1)$ in Corollary 1, the bounds for the row and column nodes are $O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{n}})$ and $O(\frac{1}{\sigma_K^2(\tilde{P})} \frac{\log(n)}{\rho n})$, respectively, and we see that ONA yields a stable and consistent community detection for both the row and column nodes, since the error rates go to zero as $n \rightarrow \infty$ when $\tilde{P}$ is fixed. Especially, for the row nodes with mixed memberships, when the DCMM proposed in [9] reduces to MMSB and $K = O(1)$, the error bound of the Mixed-SCORE in Theorem 2.2 of [9] is also $O(\frac{1}{\sigma_K(\tilde{P})} \sqrt{\frac{\log(n)}{n}})$, which guarantees the theoretical optimality of our analysis for the row nodes. For the column nodes, when every column community enjoys similar sizes and $K = O(1)$, our bound $O(\frac{1}{\sigma_K^2(\tilde{P})} \frac{\log(n)}{\rho n})$ matches Corollary 3.2 in [13] up to a logarithmic factor, which guarantees the theoretical optimality of our analysis for column nodes. Furthermore, the optimality of our requirement on network sparsity and the theoretical upper bounds of ONA's error rates is also supported by using the separation condition and sharp threshold criterion developed in [28].

3. The Overlapping and Degree-Corrected Non-Overlapping Model

In this section, we propose an extension of ONM by considering the degree heterogeneity, and we build theoretical guarantees for algorithm fitting our model.

Let θ_c be an $n_c \times 1$ vector whose i_c -th entry is the degree heterogeneity of column node i_c , for $i_c \in [n_c]$. Let Θ_c be an $n_c \times n_c$ diagonal matrix whose i_c -th diagonal element is $\theta_c(i_c)$. For $i_r \in [n_r]$, $i_c \in [n_c]$, the extended model for generating A is:

$$\Omega := \Pi_r \Pi_c' \Theta_c, \quad A(i_r, i_c) \sim \text{Bernoulli}(\Omega(i_r, i_c)). \quad (6)$$

Definition 2. Call model (1)–(4), (6) the Overlapping and Degree-Corrected Non-overlapping model (ODCNM), and denote it by $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$.

Note that, under ODCNM, the maximum element of P can be larger than 1, since $\max_{i_c \in [n_c]} \theta_c(i_c)$ also controls the sparsity of directed network $\mathcal{N}$. The following proposition guarantees that ODCNM is identifiable in terms of P, Π_r , and Π_c , and such identifiability is similar to that of DCSBM.

Proposition 2. If conditions (I1) and (I2) hold, ODCNM is identifiable for the membership matrices: For eligible $(P, \Pi_r, \Pi_c, \Theta_c)$ and $(\check{P}, \check{\Pi}_r, \check{\Pi}_c, \check{\Theta}_c)$, if $\Pi_r \Pi_c' \Theta_c = \check{\Pi}_r \check{\Pi}_c' \check{\Theta}_c$, then $\Pi_r = \check{\Pi}_r$ and $\Pi_c = \check{\Pi}_c$.

Remark 2. By setting $\theta_c(i_c) = \rho$ for $i_c \in [n_c]$, ODCNM reduces to ONM, and this is the reason for why ODCNM can be seen as an extension of ONM. Meanwhile, though DCScBM [18] can model directed networks with degree heterogeneities for both row and column nodes, DCScBM does not allow the overlapping property for row nodes. For comparison, our ODCNM allows row nodes to have an overlapping property at the cost of losing the degree heterogeneities and requiring $K_r \leq K_c$ for model identifiability.

3.1. A Spectral Algorithm for Fitting ODCNM

We now discuss our intuition for the design of our algorithm to fit ODCNM. Without causing confusion, we also use $U_r, U_c, B_r, B_c, \delta_c, Y_r$ under ODCNM. Let $U_{c,*} \in \mathbb{R}^{n_c \times K_r}$ be the row-normalized version of U_c , such that $U_{c,*}(i_c, :) = \frac{U_c(i_c, :)}{\|U_c(i_c, :)\|_F}$ for $i_c \in [n_c]$. Then, clustering the rows of $U_{c,*}$ using the k-means algorithm can return perfect clustering for column nodes, and this is guaranteed by the following lemma.

Lemma 2. Under $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$, there exists a unique $K_r \times K_r$ matrix B_r and a unique $K_c \times K_r$ matrix B_c , such that

- $U_r = \Pi_r B_r$, where $B_r = U_r(\mathcal{I}_r, :)$. Meanwhile, $U_r(i_r, :) = U_r(\bar{i}_r, :)$ when $\Pi_r(i_r, :) = \Pi_r(\bar{i}_r, :)$ for $i_r, \bar{i}_r \in [n_r]$.

- $U_{c,*} = \Pi_c B_c$. Meanwhile, $U_{c,*}(i_c, :) = U_{c,*}(\bar{i}_c, :)$ when $\ell(i_c) = \ell(\bar{i}_c)$ for $i_c, \bar{i}_c \in [n_c]$. Furthermore, when $K_r = K_c = K$, we have $\|B_c(k, :) - B_c(l, :)\|_F = \sqrt{2}$ for all $1 \leq k < l \leq K$.

Recall that we set $\delta_c = \min_{k \neq l} \|B_c(k, :) - B_c(l, :)\|_F$ by Lemma 2; $\delta_c = \sqrt{2}$ when $K_r = K_c = K$ under $ODCNM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$. However, when $K_r < K_c$, it is a challenge to obtain a positive lower bound of δ_c ; see the proof of Lemma 2 for details.

Under ODCNM, to recover Π_c from U_c , since $U_{c,*}$ has K_c distinct rows, applying the k-means algorithm on all rows of $U_{c,*}$ returns true column communities by Lemma 2. To recover Π_r from U_r , the same idea as that of under ONM can be followed.

Based on the above analysis, we are now ready to present the following algorithm, which we call Ideal ODCNA. Input Ω, K_r, K_c with $K_r \leq K_c$. Output: Π_r and ℓ .

- Let $\Omega = U_r \Lambda U_c'$ be the compact SVD of Ω , such that $U_r \in \mathbb{R}^{n_r \times K_r}, U_c \in \mathbb{R}^{n_c \times K_r}, \Lambda \in \mathbb{R}^{K_r \times K_r}, U_r' U_r = I_{K_r}, U_c' U_c = I_{K_r}$. Let $U_{c,*}$ be the row-normalization of U_c .
- For row nodes, they are the same as that of Ideal ONA.
For column nodes: run k-means on $U_{c,*}$ assuming there are K_c column communities to obtain ℓ .

We now extend the ideal case to the real case. Let $\hat{U}_{c,*} \in \mathbb{R}^{n_c \times K_r}$ be the row-normalized version of $\hat{U}_c$, such that $\hat{U}_{c,*}(i_c, :) = \frac{\hat{U}_c(i_c, :)}{\|\hat{U}_c(i_c, :)\|_F}$ for $i_c \in [n_c]$. The Overlapping and Degree-Corrected Non-overlapping Algorithm (ODCNA for short) is a natural extension of the Ideal ODCNA to the real case, where all steps of ODCNA are the same as ONA except for those for column nodes. ODCNA applies k-means on $\hat{U}_{c,*}$ to obtain $\hat{\ell}$.

3.2. Main Results for ODCNA

Set $\theta_{c,\max} = \max_{i_c \in [n_c]} \theta_c(i_c)$, $\theta_{c,\min} = \min_{i_c \in [n_c]} \theta_c(i_c)$, and $P_{\max} = \max_{k \in [K_r], l \in [n_c]} P(k, l)$. Assume that

Assumption 2. $P_{\max} \max(\theta_{c,\max} n_r, \|\theta_c\|_1) \geq \log(n_r + n_c)$.

The next theorem is the main theoretical result for ODCNA, where we also use the same measurements as ONA to measure the performances of ODCNA.

Theorem 2. Under $ODCNM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$, when Assumption 2 holds, suppose $\sigma_{K_r}(\Omega) \geq C \sqrt{\theta_{c,\max}(n_r + n_c) \log(n_r + n_c)}$, with a probability at least $1 - o((n_r + n_c)^{-\alpha})$,

- For the row nodes,

$$\max_{i_r \in [n_r]} \|e'_{i_r} (\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega \kappa(\Pi_r' \Pi_r) K_r \sqrt{\lambda_1(\Pi_r' \Pi_r)}).$$

- For the column nodes,

$$\hat{f}_c = O\left(\frac{\theta_{c,\max}^2 K_r K_c \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_{K_r}^2(P) \theta_{c,\min}^4 \delta_c^2 m_{V_c}^2 \sigma_{K_r}^2(\Pi_r) n_{c,K_r} n_{c,\min}}\right),$$

where m_{V_c} is a parameter defined in the proof of this theorem, and it is 1 when $K_r = K_c$. Especially, when $K_r = K_c = K$,

$$\hat{f}_c = O\left(\frac{\theta_{c,\max}^2 K^2 \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_K^2(P) \theta_{c,\min}^4 \sigma_K^2(\Pi_r) n_{c,\min}^2}\right).$$

Adding some conditions on model parameters, we have the following corollary.

Corollary 2. Under $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$, suppose that conditions in Theorem 2 hold, and further, suppose that $\lambda_{K_r}(\Pi_r' \Pi_r) = O(\frac{n_r}{K_r})$, $n_{c, \min} = O(\frac{n_c}{K_c})$, with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$,

- For row nodes, when $K_r = K_c = K$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{K^2 \sqrt{\theta_{c, \max}} \left(\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)}\right)}{\theta_{c, \min} \sigma_K(P) \sqrt{n_c}}\right).$$

- For column nodes, $\hat{f}_c = O\left(\frac{\theta_{c, \max}^2 K_r^2 K_c^2 \max(\theta_{c, \max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}{\sigma_{K_r}^2(P) \theta_{c, \min}^4 \delta_c^2 m_{V_c}^2 n_r n_c}\right)$. When $K_r = K_c = K$,

$$\hat{f}_c = O\left(\frac{\theta_{c, \max}^2 K^4 \max(\theta_{c, \max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}{\sigma_K^2(P) \theta_{c, \min}^4 n_r n_c}\right).$$

Especially, when $n_r = O(n)$, $n_c = O(n)$, $K_r = O(1)$ and $K_c = O(1)$,

- For row nodes, when $K_r = K_c$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{\sqrt{\theta_{c, \max} \log(n)}}{\theta_{c, \min} \sigma_K(P) \sqrt{n}}\right).$$

- For column nodes, $\hat{f}_c = O\left(\frac{\theta_{c, \max}^2 \max(\theta_{c, \max} n_r, \|\theta_c\|_1) \log(n)}{\sigma_{K_r}^2(P) \theta_{c, \min}^4 \delta_c^2 m_{V_c}^2 n^2}\right)$. When $K_r = K_c = K$,

$$\hat{f}_c = O\left(\frac{\theta_{c, \max}^2 \max(\theta_{c, \max} n_r, \|\theta_c\|_1) \log(n)}{\sigma_K^2(P) \theta_{c, \min}^4 n^2}\right).$$

If we further set $\theta_{c, \max} = O(\rho)$ and $\theta_{c, \min} = O(\rho)$, we have the below corollary.

Corollary 3. Under $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$, suppose that the conditions in Theorem 2 hold, and further, suppose that $\lambda_{K_r}(\Pi_r' \Pi_r) = O(\frac{n_r}{K_r})$, $n_{c, \min} = O(\frac{n_c}{K_c})$ and $\theta_{c, \max} = O(\rho)$, $\theta_{c, \min} = O(\rho)$, with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$,

- For row nodes, when $K_r = K_c = K$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{K^2 \left(\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)}\right)}{\sigma_K(P) \sqrt{\rho n_c}}\right).$$

- For column nodes, $\hat{f}_c = O\left(\frac{K_r^2 K_c^2 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(P) \rho \delta_c^2 m_{V_c}^2 n_r n_c}\right)$. When $K_r = K_c = K$,

$$\hat{f}_c = O\left(\frac{K^4 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_K^2(P) \rho n_r n_c}\right).$$

Especially, when $n_r = O(n)$, $n_c = O(n)$, $K_r = O(1)$ and $K_c = O(1)$,

- For row nodes, when $K_r = K_c$,

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{\sqrt{\log(n)}}{\sigma_K(P) \sqrt{\rho n}}\right).$$

- For column nodes, $\hat{f}_c = O\left(\frac{\log(n)}{\sigma_{K_r}^2(P) \rho \delta_c^2 m_{V_c}^2 n}\right)$. When $K_r = K_c = K$,

$$\hat{f}_c = O\left(\frac{\log(n)}{\sigma_K^2(P) \rho n}\right).$$

By setting $\Theta_c = \rho I$, ODCNM degenerates to ONM. By comparing Corollaries 1 and 3, we see that theoretical results under ODCNM are consistent with those under ONM when ODCNM degenerates to ONM for the case where $K_r = K_c = K$.

4. Simulations

In this section, we present some simulations to investigate the performances of the two proposed algorithms. We measure their performances using the Mixed-Hamming error rate (MHamm for short) for row nodes, and the Hamming error rate (Hamm for short) for the column nodes defined below

$$\text{MHamm} = \frac{\min_{\pi \in S_{K_r}} \|\hat{\Pi}_r \pi - \Pi_r\|_1}{n_r}, \text{Hamm} = \frac{\min_{\pi \in S_{K_c}} \|\hat{\Pi}_c \pi - \Pi_c\|_0}{n_c},$$

where S_{K_r} is the set of all permutations of $\{1, 2, \dots, K_r\}$, S_{K_c} is the set of all permutations of $\{1, 2, \dots, K_c\}$; $\hat{\Pi}_c \in \mathbb{R}^{n_c \times K_c}$ is defined as $\hat{\Pi}_c(i_c, k) = 1$ if $\hat{\ell}(i_c) = k$, and 0 otherwise for $i_c \in [n_c], k \in [K_c]$.

For all simulations in this section, the parameters $(n_r, n_c, K_r, K_c, P, \rho, \Pi_r, \Pi_c, \Theta_c)$ are set as below. Unless specified, set $n_r = 400, n_c = 300, K_r = 3, K_c = 4$. For the column nodes, generate Π_c by setting each column node belonging to one of the column communities with equal probability. Let each row community have 100 pure nodes, and let all the mixed row nodes have memberships $(0.6, 0.3, 0.1)$. $P = \rho \tilde{P}$ is set independently under ONM and ODCNM. Under ONM, ρ is 0.5 in Experiment 1, and we study the influence of ρ in Experiment 2. Under ODCNM, for $z_c \geq 1$, we generate the degree parameters for the column nodes as below: let $\theta_c \in \mathbb{R}^{n_c \times 1}$, such that $1/\theta_c(i_c) \stackrel{iid}{\sim} U(1, z_c)$ for $i_c \in [n_c]$, where $U(1, z_c)$ denotes the uniform distribution on $[1, z_c]$. We study the influences of z_c and ρ under ODCNM in Experiments 3 and 4, respectively. For all settings, we report the averaged MHamm and the averaged Hamm over 50 repetitions.

Experiment 1: Changing n_c under ONM. Let n_c range over $\{50, 100, 150, \dots, 300\}$. For this experiment, P is set as

$$P = \rho \begin{bmatrix} 1 & 0.3 & 0.2 & 0.3 \\ 0.2 & 0.9 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.8 & 0.3 \end{bmatrix}.$$

Let $\rho = 0.5$ for this experiment designed under ONM. The numerical results are shown in panels (a) and (b) of Figure 1. The results show that as n_c increases, ONA and ODCNA perform better. For the row nodes, since both ONA and ODCNA apply the SP algorithm on $\hat{U}$ to estimate Π_r , the estimated row membership matrices of ONA and ODCNA are same, and hence, MHamm for ONA is always equal to that of ODCNA.

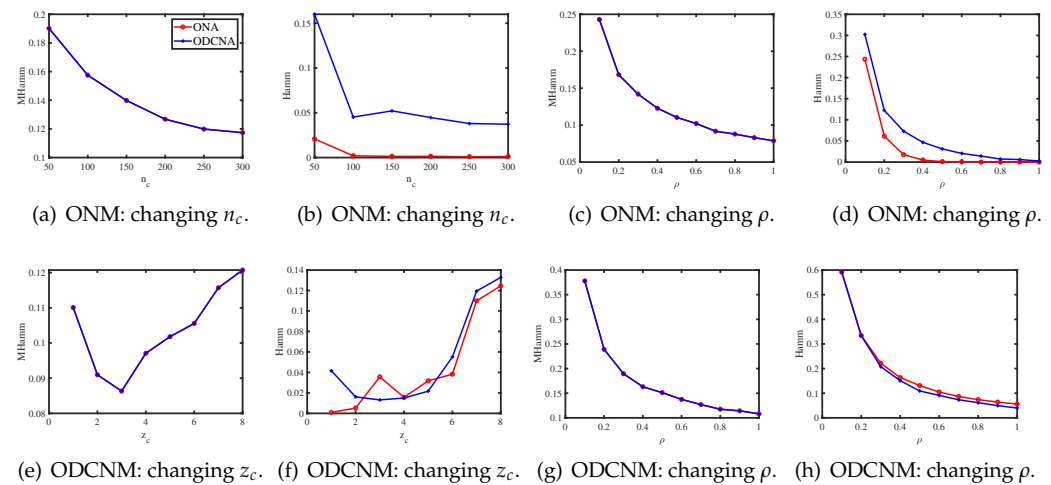


Figure 1. Estimation errors of ONA and ODCNA.

Experiment 2: Changing ρ under ONM. P is set the same as in Experiment 1, and we let the range of ρ be $\{0.1, 0.2, \dots, 1\}$ to study the influence of ρ on the performances of ONA and ODCNA under ONM. The results are displayed in panels (c) and (d) of Figure 1. From the results, we can see that both methods perform better as ρ increases, since a larger ρ gives more edges generated in a directed network.

Experiment 3: Change z_c under ODCNM. P is set to be the same as Experiment 1, and $\rho = 0.5$. Let z_c range in $\{1, 2, \dots, 8\}$. Increasing z_c decreases the edges generated under ODCNM. Panels (e) and (f) in Figure 1 display the simulation results of this experiment. The results show that generally, increasing the variability of the node degrees makes it harder to detect the node memberships for both ONA and ODCNA. Though ODCNA is designed under ODCNM, it holds similar performances as ONA for directed networks in which column nodes have various degrees in this experiment, and this is consistent with our theoretical findings in Corollaries 1 and 2.

Experiment 4: Change ρ under ODCNM. Setting $z_c = 3$, P is set to be the same as in Experiment 1, and let ρ range in $\{0.1, 0.2, \dots, 1\}$ under ODCNM. Panels (g) and (h) in Figure 1 display the simulation results of this experiment. The performances of the two proposed methods are similar as those of Experiment 2.

Remark 3. For visibility, we plot A generated under ONM. Let $n_r = 24, n_c = 20, K_r = 2, K_c = 2$, and

$$P = \begin{bmatrix} 1 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}.$$

For the row nodes, let $\Pi_r(i_r, 1) = 1$ for $1 \leq i_r \leq 8$, $\Pi_r(i_r, 2) = 1$ for $9 \leq i_r \leq 16$, and $\Pi_r(r_r, :) = [0.7 \ 0.3]$ for $17 \leq i_r \leq 24$. For the column nodes, let $\ell(i_c) = 1$ for $1 \leq i_c \leq 10$, and $\ell(i_c) = 2$ for $11 \leq i_c \leq 20$. For the above setting, we generate two random adjacency matrices in Figure 2, where we also report the error rates of ONA and ODCNA. Note that, since the adjacency matrices are shown in Figure 2, and as Π_r, ℓ, K_r , and K_c are known here, readers can apply ONA and ODCNA to A in Figure 2 to check the effectiveness of ONA and ODCNA.

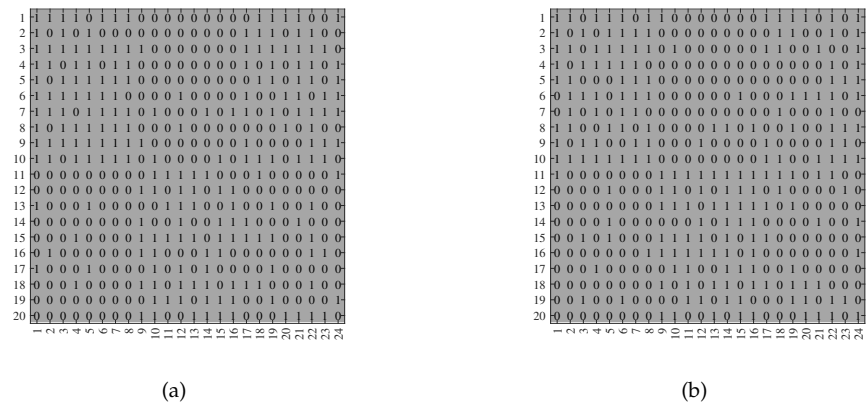


Figure 2. For adjacency matrix in panel (a), MHamm and Hamm for ONA are 0.0544 and 0, respectively. For adjacency matrix in panel (b), MHamm and Hamm for ONA are 0.1004 and 0, respectively. ODCNA enjoys same error rates as ONA. x-axis: row nodes; y-axis: column nodes.

Remark 4. For visuality, we also plot a directed network as well as its adjacency matrix generated under ONM. Let $n_r = 30, n_c = 30, K_r = 2, K_c = 3$, and

$$P = 0.9 \begin{bmatrix} 1 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0.1 \end{bmatrix}.$$

For row nodes, let $\Pi_r(i_r, 1) = 1$ for $1 \leq i_r \leq 10$, $\Pi_r(i_r, 2) = 1$ for $11 \leq i_r \leq 20$, and $\Pi_r(i_r, :) = [0.7 \ 0.3]$ for $21 \leq i_r \leq 30$. For column nodes, let $\ell(i_c) = 1$ for $1 \leq i_c \leq 10$, $\ell(i_c) = 2$ for $11 \leq i_c \leq 20$, and $\ell(i_c) = 3$ for $21 \leq i_c \leq 30$. For the above setting, we generate one adjacency matrix in panel (a) of Figure 3, where we also report the error rates of ONA and ODCNA. Furthermore, panels (b) and (c) of Figure 3 show the sending pattern and receiving pattern sides of this simulated directed network, respectively.

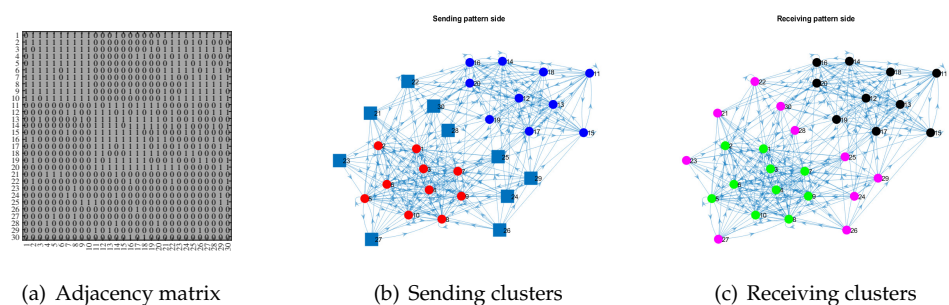


Figure 3. Illustration of a simulated directed network generated under ONM. Panels (a–c) show the adjacency matrix, the sending clusters, and the receiving clusters of this simulated directed network, respectively. For this directed network, MHamm and Hamm for ONA (and ODCNA) are 0.0615 (0.0615) and 0 (0.2333), respectively. In panels (b,c), the dots in the same color are pure nodes in the same sending (receiving) clusters, and the square indicates the mixed nodes with weight 0.7 belonging to red sending clusters, and weight 0.3 belonging to blue sending clusters, where the sending and receiving clusters are obtained by Π_r and ℓ provided in Remark 4.

5. Real Data Analysis

For real-world directed networks, since nodes always have various degrees, we only apply ODCNA to deal with real-world datasets in this section. For the real-world directed networks analyzed in this section, the row nodes are always same as the column nodes, so we do not use subscript r and c to distinguish the row and column nodes here, and we let $n_r = n_c = n$. Meanwhile, the number of row communities is equal to that of the column communities; i.e., $K_r = K_c = K$ for real data, where we always set $K_r = K_c = K$, as analyzed

in [18], since it is a challenge to determine the number of row (column) communities for real-world directed networks without prior knowledge. When the row nodes are the same as the column nodes, $A(i, j) = 1$ means that a directed edge is sent from node i to node j . Thus, for any node, it has two patterns, the sending pattern and the receiving pattern. For the sending (receiving) pattern, we use the sending (receiving) cluster to denote the prior row (column) community, where we use the sending and receiving patterns to distinguish the behaviors of any node having the two patterns, as was performed in [18].

For $\hat{\Pi}_r$ obtained from ODCNA, we call node i a highly mixed node if $0.8 \geq \max_{1 \leq k \leq K} \hat{\Pi}_r(i, k)$, where 0.8 is a threshold. Here, 0.8 is a moderate value to define highly mixed nodes, and we can also choose 0.9, 0.95, or some other values in $(0, 1)$. However, we choose 0.8 as the threshold, because setting the threshold to be larger (or lesser) than 0.8 may be too restrictive (loose) to define highly mixed nodes. The definition of highly mixed node is important, since it tells us whether a node only belongs to one community or belongs to multiple communities. Let τ be the proportion of highly mixed row nodes among all nodes, to measure the mixability of a directed network, i.e., $\tau = \frac{|\{i \in [n] : i \text{ is a highly mixed node}\}|}{n}$. Meanwhile, we let $\hat{\ell}_r$ be an $n \times 1$ vector, such that $\hat{\ell}_r(i) = \arg\max_{1 \leq k \leq K} \hat{\Pi}_r(i, k)$, where we use $\hat{\ell}_r(i)$ to denote the home base sending pattern cluster of node i . Set

$$\text{Hamm}_{rc} = \frac{\min_{\pi \in S_K} \|\hat{\Pi}_c \pi - \hat{\Pi}_r\|_0}{n},$$

where S_K is the set of all permutations of $\{1, 2, \dots, K\}$; $\hat{\Pi}_r \in \mathbb{R}^{n \times K}$ is defined as $\hat{\Pi}_r(i, k) = 1$ if $\hat{\ell}_r(i) = k$, and 0 otherwise for $i \in [n], k \in [K]$. Hamm_{rc} is defined to measure the difference between the sending and receiving pattern clusters. After defining τ and Hamm_{rc} , we see that a larger τ indicates a directed network in which a large proportion of nodes are highly mixed nodes with a sending pattern, and a larger Hamm_{rc} indicates that the sending pattern differs a lot with the receiving pattern. For $i \in [n]$, let $d_{\text{sending}}(i) = \sum_{j=1}^n A(i, j)$ denote the total number of edges sent by node i , and let $d_{\text{receiving}}(i) = \sum_{j=1}^n A(j, i)$ denote the total number of edges that are received by node i . Call $d_{\text{sending}}(i)$ and $d_{\text{receiving}}(i)$ the sending degree and receiving degree of node i , respectively. For real-world directed networks, we find that there are many nodes whose sending degree or receiving degree are zero, and so we need the following pre-processing steps before analyzing the real data:

- Set $S_{\text{sending},0} = \{i \in [n] : d_{\text{sending}}(i) = 0\}$ and $S_{\text{receiving},0} = \{i \in [n] : d_{\text{receiving}}(i) = 0\}$.
- Set $S_{\text{degree},0} = S_{\text{sending},0} \cup S_{\text{receiving},0}$.
- Update A by removing the nodes in $S_{\text{degree},0}$.
- Repeat (a)–(c) until $S_{\text{degree},0}$ is a null set.
- After obtaining A through the above four steps, obtain the largest connected component of A .

We now describe the real-world directed networks analyzed in this paper:

Metabolic: This is a directed network representing the metabolic reactions of E.coli bacteria. In this data, node means metabolite, and a directed edge from node i to node j means that there is a reaction where node i is an input and node j is a product [29]. These data can be downloaded from <http://networksciencebook.com/translations/en/resources/data.html>. The original data has 1039 nodes; after preprocessing, $A \in \{0, 1\}^{893 \times 893}$. To estimate K , we plot the leading 20 singular values of A , and panel (a) of Figure 4 shows the result that suggests that $K = 2$ for these data, where [18] also applies the idea of an eigengap to estimate K for real-world directed networks with an unknown number of communities.

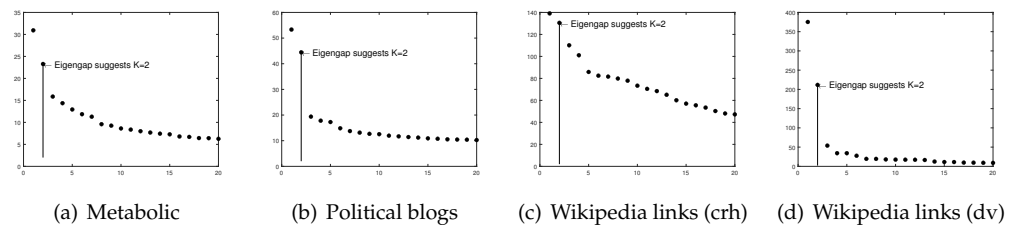


Figure 4. Leading 20 singular values of adjacency matrices for real-world directed networks used in this paper.

Political blogs: this is a directed network of hyperlinks between weblogs on US politics [30], and it can be downloaded from <http://www-personal.umich.edu/~mejn/netdata/>. Political blogs send and receive hyperlinks to and from blogs for the same political persuasion [18], so node means blog and edge means hyperlink in these data. The original network has 1490 nodes. After removing nodes with zero degrees via pre-processing steps, there are 814 nodes left; i.e., $A \in \{0, 1\}^{813 \times 813}$ for these data. Since there are two parties, “liberal” and “conservative”, K is 2 for both the sending and receiving pattern communities for these data. [18] applies their DI-SIM algorithm to the Political blogs network, assuming that all nodes have non-overlapping property. In this paper, we apply our ODCNA algorithm on these data to study its asymmetric structure on the overlapping property.

Wikipedia links (crh): This directed network represents the wikilinks of the Wikipedia website in the Crimean Turkish language (crh). Node means article, and the directed edge between two articles is the wikilink [31]. These data can be downloaded from http://konect.cc/networks/wikipedia_link_crh/. The original data have 8286 nodes. After pro-processing, $A \in \{0, 1\}^{3555 \times 3555}$. Panel (c) of Figure 4 suggests $K = 2$ for this data.

Wikipedia links (dv): These data represent the wikilinks of the Wikipedia website in the Divehi language (dv), where node means article and the directed edge is a wikilink [31]. These data can be downloaded from http://konect.cc/networks/wikipedia_link_dv/. The original data has 4266 nodes. After removing nodes with zero degrees, $A \in \{0, 1\}^{2394 \times 2394}$. Panel (d) of Figure 4 suggests $K = 2$ for these data.

The proportion of highly mixed nodes and Hamm_{rc} for these directed networks are reported in Table 1 when assuming that nodes in sending (receiving) clusters having an overlapping (non-overlapping) property. For the Metabolic network, the results show that the sending pattern differs a lot with the receiving pattern, since $\text{Hamm}_{rc} = 0.2497$ is quite large, and there are $893 \times 0.1209 \approx 108$ highly mixed nodes in the sending pattern. For the Political blogs network, there is a slight asymmetric structure between the sending pattern and the receiving pattern, since $\text{Hamm}_{rc} = 0.0443$ is small. Meanwhile, for the sending pattern of Political blogs, there are $813 \times 0.0246 \approx 20$ highly mixed nodes. Thus, we may conclude that though there is a slight asymmetric structure in sending and receiving patterns for Political blogs, there are 20 highly mixed nodes in the sending pattern. For the Wikipedia links (crh), they have a slight asymmetric structure between sending and receiving patterns, and there are $3555 \times 0.0444 \approx 158$ highly mixed nodes in the sending pattern. For the Wikipedia links (dv) network, it has a large number of highly mixed nodes for its large τ , and a heavy asymmetric structure in sending and receiving patterns for its large Hamm_{rc} . Generally, Table 1 suggests that if there are a large number of highly mixed nodes in the sending pattern, there is a heavy asymmetric structure between the sending and receiving patterns, and vice versa.

Table 1. The proportion of highly mixed nodes and the asymmetric structure measured by Hamm_{rc} for real-world directed networks considered in this paper when ODCNA's input adjacency matrix is A ; i.e., the case when assuming that nodes in a sending (receiving) pattern have overlapping (non-overlapping) property.

Data	τ	Hamm_{rc}
Metabolic	0.1209	0.2497
Political blogs	0.0246	0.0443
Wikipedia links (crh)	0.0444	0.0307
Wikipedia links (dv)	0.4089	0.1466

For visualization, we plot the sending clusters and receiving clusters detected by ODCNA for these directed networks when assuming that nodes in the sending (receiving) clusters have an overlapping (non-overlapping) property, i.e., when the input adjacency matrix of the ODCNA approach is A . The results are shown in Figures 5–8, where we also show the highly mixed nodes in sending clusters detected by ODCNA. We see that there exists a clear asymmetric structure between the sending and receiving patterns for Metabolic and Wikipedia links (dv), as shown in Figures 5 and 8, while there is a slight asymmetric structure between the sending and receiving patterns for Political blogs and Wikipedia links (crh), as shown in Figures 6 and 7. Furthermore, most nodes are in the same sending (receiving) cluster for Metabolic and Wikipedia links (crh), while the two sending (receiving) clusters for Political blogs and Wikipedia links (crh) have close sizes. The results also show that most highly mixed nodes have many edges, while some highly mixed nodes have only a few edges, where such a phenomenon can be explained easily, since nodes with many edges tend to have an overlapping property, while it is difficult to detect a community for nodes with only a few edges, and ODCNA tends to treat such nodes as highly mixed nodes.

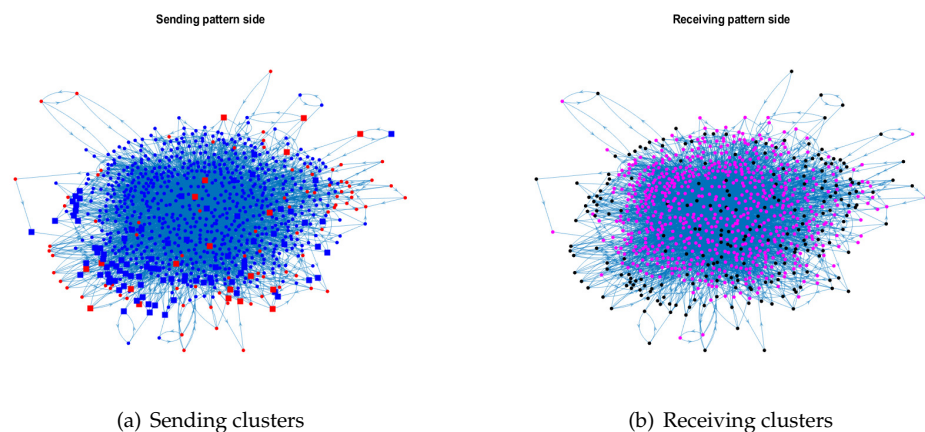


Figure 5. Sending and receiving clusters detected by ODCNA for Metabolic network when assuming that nodes in a sending (receiving) pattern have an overlapping (non-overlapping) property. Colors indicate clusters detected using ODCNA, and squares indicate highly mixed nodes, where sending clusters are obtained using $\hat{\ell}_r$, the home base sending pattern community, and receiving clusters are obtained by $\hat{\ell}$ from ODCNA.

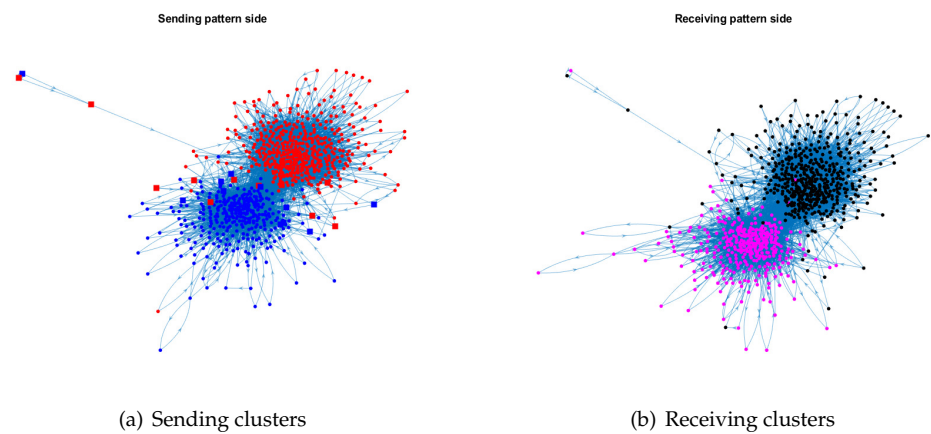


Figure 6. Sending and receiving clusters detected by ODCNA for Political blogs network. Colors indicate clusters and square indicates highly mixed nodes.

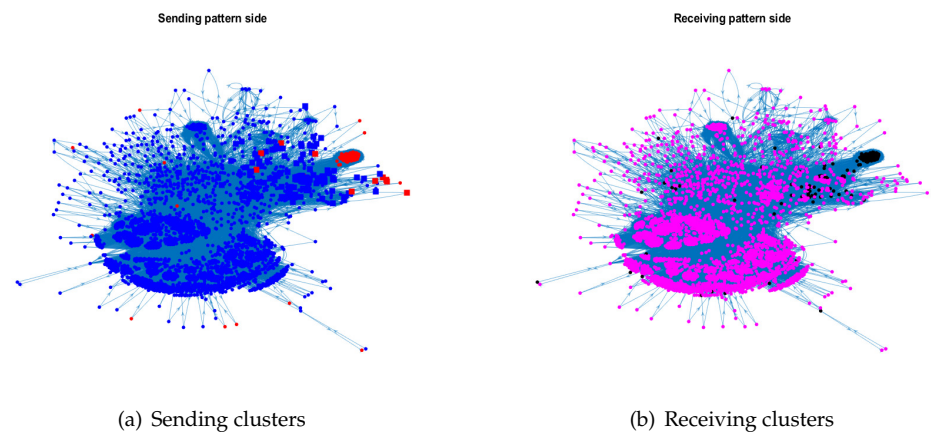


Figure 7. Sending and receiving clusters detected by ODCNA for Wikipedia links (crh) network. Colors indicate clusters and square indicates highly mixed nodes.

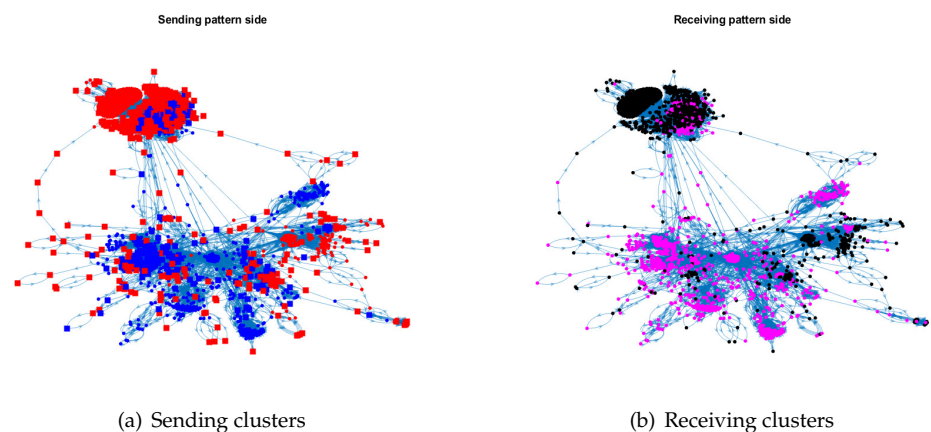


Figure 8. Sending and receiving clusters detected by ODCNA for Wikipedia links (dv) network. Colors indicate clusters and square indicates highly mixed nodes.

Furthermore, for real-world directed networks, since we have no prior knowledge on whether nodes in the sending pattern side or the receiving pattern side or both sides have overlapping property, simply inputting A with K sending (receiving) pattern communities

in our ODCNA algorithm is not enough. To solve this problem, we also apply ODCNA on A' , and the numerical results are provided in Table 2, where the results show that there also exist highly mixed nodes in the receiving pattern for these directed networks, and there also exists a heavy asymmetric structure between the sending and receiving clusters for the Metabolic and Wikipedia links (dv), while there also exists a slight asymmetric structure between the sending and receiving clusters for the Political blogs and Wikipedia links (crh).

Table 2. The proportion of highly mixed nodes and the asymmetric structure measured by Hamm_{rc} for real-world directed networks considered in this paper when ODCNA's input adjacency matrix is A' , i.e., the case when assuming that nodes in sending (receiving) pattern have non-overlapping (overlapping) property.

Data	τ	Hamm_{rc}
Metabolic	0.0594	0.2945
Political blogs	0.1365	0.0443
Wikipedia links (crh)	0.1308	0.0543
Wikipedia links (dv)	0.3492	0.2059

6. Discussion

In this paper, we introduced Overlapping and Non-overlapping models and their extension, by considering the degree heterogeneity. The models can model a directed network with K_r row communities and K_c column communities, in which the row node can belong to multiple sending clusters, while the column node only belongs to one of the receiving clusters. The proposed models are identifiable under the case when $K_r \leq K_c$, and some other popular constraints on the connectivity matrix and membership matrices. For comparison, modeling a directed network in which the row nodes have overlapping property while column nodes do not, with $K_r > K_c$, is unidentifiable. Meanwhile, since previous works have found that modeling directed networks in which both row and column nodes have an overlapping property with $K_r \neq K_c$ is unidentifiable, our identifiable ONM and ODCNM supply a gap in modeling overlapping directed networks when $K_r \neq K_c$. These models provide exploratory tools for studying community structure in directed networks with one side overlapping while another side is non-overlapping. Two spectral algorithms are designed to fit ONM and ODCNM. We also showed an estimation consistency under mild conditions for our methods. Especially, when ODCNM reduces to ONM, our theoretical results under ODCNM are consistent with those under ONM. The numerical results for the simulated directed networks generated under ONM and ODCNM support our theoretical results, and the results for real-world directed networks reveal the existence of highly mixed nodes and an asymmetric structure between the sending and receiving clusters.

The models and algorithms introduced in this paper are useful tools for studying the asymmetric structure for directed networks, and we wish that they can be widely applied in network science. However, perhaps the main limitation of the models is that K_r and K_c in the directed network are assumed as givens, and such a limitation also holds for the spectral clustering algorithms developed under the ScBM and DCScBM studied in [18–20]. In most community problems, the number of row communities and the number of column communities are unknown; therefore, a complete calculation and theoretical study requires not only the algorithms and their theoretically consistent estimations described in this paper, but also a method for estimating K_r and K_c . A possible solution to this problem may be a combination of algorithms developed in this paper and the modularity for the directed networks developed in [32]. Meanwhile, our idea can be extended in many ways. In this paper, we only consider modeling an un-weighted directed network, and it is possible to extend our work to a weighted directed network. Our algorithms are designed based on the adjacency matrix, and it is possible to design spectral algorithms to fit ONM and ODCNM by applying the regularized Laplace matrix used in [11,12]. When detecting large-scale directed networks, the random projection-based and the random sampling-based spectral

clustering ideas in [33] may be applied to accelerate our algorithms. We leave the studies of these problems to our future work.

Funding: This research was funded by the scientific research start-up fund of China University of Mining and Technology, NO. 102520253, and the high-level personal project of Jiangsu Province, NO.JSSCBS20211218.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available within the article.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SBM	Stochastic Blockmodel
DCSBM	Degree-Corrected Stochastic Blockmodel
MMSB	Mixed Membership Stochastic Blockmodel
DCMM	Degree-Corrected Mixed Membership model
OCCAM	Overlapping Continuous Community Assignment model
ScBM	Stochastic co-Blockmodel
DCScBM	Degree-Corrected Stochastic co-Blockmodel
DiMMSB	Directed Mixed Membership Stochastic Blockmodel
ONM	Overlapping and Non-overlapping model
ODCNM	Overlapping and Degree-Corrected Non-overlapping model
SP	successive projection algorithm
ONA	Overlapping and Non-overlapping algorithm
ODCNA	Overlapping and Degree-Corrected Non-overlapping Algorithm

Appendix A. Successive Projection Algorithm

Algorithm A1 is the Successive Projection algorithm.

Algorithm A1 Successive Projection (SP) [27]

Require: Near-separable matrix $Y_{sp} = S_{sp}M_{sp} + Z_{sp} \in \mathbb{R}_+^{m \times n}$, where S_{sp}, M_{sp} should satisfy Assumption 1 [27], the number r of columns to be extracted.

Ensure: Set of indices $\mathcal{K}$ such that $Y(\mathcal{K}, :) \approx S$ (up to permutation)

- 1: Compute $\hat{U}_r \in \mathbb{R}^{n_r \times K_r}$ and $\hat{U}_c \in \mathbb{R}^{n_c \times K_r}$ from the top- K_r -dimensional SVD of A .
- 2: Let $R = Y_{sp}, \mathcal{K} = \{\}, k = 1$.
- 3: **While** $R \neq 0$ and $k \leq r$ **do**
- 4: $k_* = \arg\max_k \|R(k, :)\|_F$.
- 5: $u_k = R(k_*, :)$.
- 6: $R \leftarrow (I - \frac{u_k u_k'}{\|u_k\|_F^2})R$.
- 7: $\mathcal{K} = \mathcal{K} \cup \{k_*\}$.
- 8: $k = k + 1$.
- 9: **end while**

Appendix B. Proofs under ONM

Appendix B.1. Proof of Proposition 1

Proof. By Lemma 1, let $U_r \Lambda U_c'$ be the compact SVD of Ω , such that $\Omega = U_r \Lambda U_c'$; since $\Omega = \Pi_r P \Pi_c' = \check{\Pi}_r \check{P} \check{\Pi}_c'$, we have $\Omega(\mathcal{I}_r, \mathcal{I}_c) = P = \check{P}$, which gives $P = \check{P}$. By Lemma 1, since $U_r = \Pi_r U_r(\mathcal{I}_r, :) = \check{\Pi}_r U_r(\mathcal{I}_r, :)$, we have $\Pi_r = \check{\Pi}_r$ where we have used the fact that the inverse of $U_r(\mathcal{I}_r, :)$ exists. Since $\Omega = \Pi_r P \Pi_c' = \check{\Pi}_r \check{P} \check{\Pi}_c' = \Pi_r P \check{\Pi}_c'$, we have

$\Pi_r P \Pi'_c = \Pi_r P \check{\Pi}'_c$. By Lemma 7 of [21], we have $P \Pi'_c = P \check{\Pi}'_c$, i.e., $\Pi_c X = \check{\Pi}_c X$, where we set $X = P' \in \mathbb{R}^{K_c \times K_r}$. Let $\check{\ell}$ be the $n_c \times 1$ vector of column nodes labels obtained from $\check{\Pi}_c$. For $i_c \in [n_c], k \in [K_r]$, from $\Pi_c X = \check{\Pi}_c X$, we have $(\Pi_c X)(i_c, k) = \Pi_c(i_c, :) X(:, k) = X(\ell(i_c), k) = X(\check{\ell}(i_c), k)$, which means that we must have $\ell(i_c) = \check{\ell}(i_c)$ for all $i_c \in [n_c]$, i.e., $\ell = \check{\ell}$ and $\Pi_c = \check{\Pi}_c$. Note that for the special case $K_r = K_c = K$, $\Pi_c = \check{\Pi}_c$ can be obtained easily: since $P \Pi'_c = P \check{\Pi}'_c$ and $P \in \mathbb{R}^{K \times K}$ is assumed to be full rank, we have $\Pi_c = \check{\Pi}_c$. Thus, the proposition holds. $\square$

Appendix B.2. Proof of Lemma 1

Proof. For U_r , since $\Omega = U_r \Lambda U'_c$ and $U'_c U_c = I_{K_r}$, we have $U_r = \Omega U_c \Lambda^{-1}$. Recall that $\Omega = \Pi_r P \Pi'_c$; we have $U_r = \Pi_r P \Pi'_c U_c \Lambda^{-1} = \Pi_r B_r$, where we set $B_r = P \Pi'_c U_c \Lambda^{-1}$. Since $U_r(\mathcal{I}_r, :) = \Pi_r(\mathcal{I}_r, :) B_r = B_r$, we have $B_r = U_r(\mathcal{I}_r, :)$. For $i_r \in [n_r]$, $U_r(i_r, :) = e'_{i_r} \Pi_r B_r = \Pi_r(i_r, :) B_r$, so we have $U_r(i_r, :) = U_r(\tilde{i}_r, :)$ when $\Pi_r(i_r, :) = \Pi_r(\tilde{i}_r, :)$.

For U_c , following a similar analysis as for U_r , we have $U_c = \Pi_c B_c$, where $B_c = P' \Pi'_r U_r \Lambda^{-1}$. Note that $B_c \in \mathbb{R}^{K_c \times K_r}$. Sure, $U_c(i_c, :) = U_c(\tilde{i}_c, :)$ when $\ell(i_c) = \ell(\tilde{i}_c)$ for $i_c, \tilde{i}_c \in [n_c]$.

Now, we focus on the case where $K_r = K_c = K$. For this case, since $B_c \in \mathbb{R}^{K_c \times K_r}$, B_c is full rank when $K_r = K_c$. Since $I_{K_r} = I_K = U'_c U_c = B'_c \Pi'_c \Pi_c B_c$, we have $\Pi'_c \Pi_c = (B_c B'_c)^{-1}$. Since $\Pi'_c \Pi_c = \text{diag}(n_{c,1}, n_{c,2}, \dots, n_{c,K})$, we have $B_c B'_c = \text{diag}(\frac{1}{n_{c,1}}, \frac{1}{n_{c,2}}, \dots, \frac{1}{n_{c,K}})$. When $K_r = K_c = K$, we have $B_c(k, :) B'_c(l, :) = 0$ for any $k \neq l$ and $k, l \in [K]$. Then, we have $B_c B'_c = \text{diag}(\|B_c(1, :)\|_F^2, \|B_c(2, :)\|_F^2, \dots, \|B_c(K, :)\|_F^2) = \text{diag}(\frac{1}{n_{c,1}}, \frac{1}{n_{c,2}}, \dots, \frac{1}{n_{c,K}})$, and the lemma follows.

Note that when $K_r < K_c$, since B_c is not full rank now, we cannot obtain $\Pi'_c \Pi_c = (B_c B'_c)^{-1}$ from $I_{K_r} = B'_c \Pi'_c \Pi_c B_c$. Therefore, when $K_r < K_c$, the equality $\|B_c(k, :) - B_c(l, :)\|_F = \sqrt{\frac{1}{n_{c,k}} + \frac{1}{n_{c,l}}}$ does not hold for any $k \neq l$. Additionally, we can only know that U_c has K_c distinct rows when $K_r < K_c$, but have no knowledge about the minimum distance between any two distinct rows of U_c . $\square$

Appendix B.3. Proof of Theorem 1

Proof. First, by Lemma 4 of [21], we have the below lemma.

Lemma A1. (Row-wise singular eigenvector error) Under $ONM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, when Assumption 1 holds, suppose $\sigma_{K_r}(\Omega) \geq C \sqrt{\rho(n_r + n_c) \log(n_r + n_c)}$, with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$,

$$\|\hat{U}_r \hat{U}'_r - U_r U'_r\|_{2 \rightarrow \infty} = O\left(\frac{\sqrt{K_r}(\kappa(\Omega)) \sqrt{\frac{\max(n_r, n_c) \mu}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)}}{\sqrt{\rho} \sigma_{K_r}(\tilde{P}) \sigma_{K_r}(\Pi_r) \sqrt{n_{c, K_r}}}\right),$$

where μ is the incoherence parameter defined as $\mu = \max(\frac{n_r \|U_r\|_{2 \rightarrow \infty}^2}{K_r}, \frac{n_c \|U_c\|_{2 \rightarrow \infty}^2}{K_r})$.

For the row nodes, when conditions in Lemma A1 hold, by Theorem 2 of [21], with a probability of at least $1 - o((n_r + n_c)^{-\alpha})$ for any $\alpha > 0$, there exists a permutation matrix $\mathcal{P}_r$ such that, for $i_r \in [n_r]$, we have

$$\|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega \kappa(\Pi'_r \Pi_r) K_r \sqrt{\lambda_1(\Pi'_r \Pi_r)}).$$

Next, we focus on the column nodes. By the Proof of Lemma 3 in [19], there exists an orthogonal matrix $\hat{O}$ such that

$$\|\hat{U}_c \hat{O} - U_c\|_F \leq \frac{2\sqrt{2K_r} \|A - \Omega\|}{\sqrt{\lambda_{K_r}(\Omega' \Omega)}}. \quad (\text{A1})$$

Under $ONM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$, by Lemma 10 of [21], we have

$$\sqrt{\lambda_{K_r}(\Omega' \Omega)} \geq \rho \sigma_{K_r}(\tilde{P}) \sigma_{K_r}(\Pi_r) \sigma_{K_r}(\Pi_c). \quad (A2)$$

Since all column nodes are pure, $\sigma_{K_r}(\Pi_c) = \sqrt{n_{c, K_r}}$. By Lemma 3 of [21], when Assumption 1 holds with a probability at least $1 - o((n_r + n_c)^{-\alpha})$, we have

$$\|A - \Omega\| = O(\sqrt{\rho \max(n_r, n_c) \log(n_r + n_c)}). \quad (A3)$$

Substituting the two bounds in Equations (A2) and (A3) into Equation (A1), we have

$$\|\hat{U}_c \hat{O} - U_c\|_F \leq C \frac{\sqrt{K_r \max(n_r, n_c) \log(n_r + n_c)}}{\sigma_{K_r}(\tilde{P}) \sqrt{\rho} \sigma_{K_r}(\Pi_r) \sqrt{n_{c, K_r}}}. \quad (A4)$$

Let $\varsigma > 0$ be a small quantity; by Lemma 2 in [15], if

$$\frac{\sqrt{K_c}}{\varsigma} \|U_c - \hat{U}_c \hat{O}\|_F \left(\frac{1}{\sqrt{n_{c, k}}} + \frac{1}{\sqrt{n_{c, l}}} \right) \leq \|B_c(k, :) - B_c(l, :)\|_F, \text{ for each } 1 \leq k \neq l \leq K_c, \quad (A5)$$

then the clustering error $\hat{f}_c = O(\varsigma^2)$. Recall that we set $\delta_c = \min_{k \neq l} \|B_c(k, :) - B_c(l, :)\|_F$ to measure the minimum center separation of B_c . Setting $\varsigma = \frac{2}{\delta_c} \sqrt{\frac{K_c}{n_{c, \min}}} \|U_c - \hat{U}_c \hat{O}\|_F$ makes Equation (A5) hold for all $1 \leq k \neq l \leq K_c$. Then, we have $\hat{f}_c = O(\varsigma^2) = O(\frac{K_c \|U_c - \hat{U}_c \hat{O}\|_F^2}{\delta_c^2 n_{c, \min}})$. By Equation (A4), we have

$$\hat{f}_c = O\left(\frac{K_r K_c \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 \sigma_{K_r}^2(\Pi_r) n_{c, K_r} n_{c, \min}}\right).$$

Especially, when $K_r = K_c = K$, $\delta_c \geq \sqrt{\frac{2}{n_{c, \max}}}$ under $ONM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c)$ by Lemma 1. When $K_r = K_c = K$, we have

$$\hat{f}_c = O\left(\frac{K^2 \max(n_r, n_c) n_{c, \max} \log(n_r + n_c)}{\sigma_K^2(\tilde{P}) \rho \sigma_K^2(\Pi_r) n_{c, \min}^2}\right).$$

□

Appendix B.4. Proof of Corollary 1

Proof. For the row nodes, under the conditions of Corollary 1, we have

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega K_r \sqrt{\frac{n_r}{K_r}}) = O(\omega \sqrt{K n_r}).$$

Under the conditions of Corollary 1, $\kappa(\Omega) = O(1)$ and $\mu \leq C$ for some $C > 0$ by the proof of Corollary 1 [21]. Then, by Lemma A1, we have

$$\begin{aligned} \omega &= O\left(\frac{\sqrt{K}(\kappa(\Omega) \sqrt{\frac{\max(n_r, n_c) \mu}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\sqrt{\rho} \sigma_K(\tilde{P}) \sigma_K(\Pi_r) \sqrt{n_{c, K_r}}}\right) = O\left(\frac{\sqrt{K}(\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\sqrt{\rho} \sigma_K(\tilde{P}) \sigma_K(\Pi_r) \sqrt{n_{c, \min}}}\right) \\ &= O\left(\frac{K^{1.5}(\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\sigma_K(\tilde{P}) \sqrt{\rho n_r n_c}}\right), \end{aligned}$$

which gives that

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{K^2\left(\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)}\right)}{\sigma_K(\tilde{P})\sqrt{\rho n_c}}\right).$$

Note that, when $K_r < K_c$, we cannot draw a conclusion that $\mu \leq C$. This is because, when $K_r < K_c$, the inverse of $B_c B'_c$ does not exist, since $B_c \in \mathbb{R}^{K_c \times K_r}$. Therefore, Lemma 8 of [21] does not hold, and we cannot obtain the upper bound of $\|U_c\|_{2 \rightarrow \infty}$, causing the impossibility of obtaining the upper bound of μ , and this is the reason for why we only consider the case for when $K_r = K_c$, for the row nodes here.

For the column nodes, under the conditions of Corollary 1, we have

$$\begin{aligned} \hat{f}_c &= O\left(\frac{K_r K_c \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 \sigma_{K_r}^2(\Pi_r) n_{c, K_r} n_{c, \min}}\right) = O\left(\frac{K_r K_c \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 (n_r / K_r) (n_c / K_c) (n_c / K_c)}\right) \\ &= O\left(\frac{K_r^2 K_c^3 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_{K_r}^2(\tilde{P}) \rho \delta_c^2 n_r n_c^2}\right). \end{aligned}$$

For the special case $K_r = K_c = K$, since $\frac{n_{c, \max}}{n_{c, \min}} = O(1)$ when $n_{c, \min} = O(\frac{n_c}{K})$, we have

$$\hat{f}_c = O\left(\frac{K^4 \max(n_r, n_c) \log(n_r + n_c)}{\sigma_K^2(\tilde{P}) \rho n_r n_c}\right).$$

When $n_r = O(n)$, $n_c = O(n)$, $K_r = O(1)$, and $K_c = O(1)$, the corollary follows immediately by basic algebra. $\square$

Appendix C. Proofs under ODCNM

Appendix C.1. Proof of Proposition 2

Proof. Since $\Omega = \Pi_r P \Pi'_c \Theta_c = \check{\Pi}_r \check{P} \check{\Pi}'_c \check{\Theta}_c = U_r \Lambda U'_c$, we have $U_r = \Pi_r U_r(\mathcal{I}_r, :) = \check{\Pi}_r U_r(\mathcal{I}_r, :)$ by Lemma 2, which gives that $\Pi_r = \check{\Pi}_r$. Since $U_{c,*} = \Pi_c B_c = \Pi_c U_{c,*}(\mathcal{I}_c, :)$ by Lemma 2, we have $\Pi_c = \check{\Pi}_c$. $\square$

Appendix C.2. Proof of Lemma 2

Proof.

- For U_r : since $\Omega = U_r \Lambda U'_c$ and $U'_c U_c = I_{K_r}$, we have $U_r = \Omega U_c \Lambda^{-1}$. Recall that $\Omega = \Pi_r P \Pi'_c \Theta_c$ under ODCNM; we have $U_r = \Pi_r P \Pi'_c \Theta_c U_c \Lambda^{-1} = \Pi_r B_r$, where $B_r = P \Pi'_c \Theta_c U_c \Lambda^{-1}$. Sure, $U_r(i_r, :) = U_r(\tilde{i}_r, :)$ holds when $\Pi_r(i_r, :) = \Pi_r(\tilde{i}_r, :)$ for $i_r, \tilde{i}_r \in [n_r]$.
- For U_c : let D_c be a $K_c \times K_c$ diagonal matrix, such that $D_c(k, k) = \frac{\|\Theta_c \Pi_c(:, k)\|_F}{\|\theta_c\|_F}$ for $k \in [K_c]$. Let Γ_c be an $n_c \times K_c$ matrix, such that $\Gamma_c(:, k) = \frac{\Theta_c \Pi_c(:, k)}{\|\Theta_c \Pi_c(:, k)\|_F}$ for $k \in [K_c]$. For such D_c and Γ_c , we have $\Gamma'_c \Gamma_c = I_{K_c}$ and $\Omega = \Pi_r P \|\theta_c\|_F D_c \Gamma'_c$, i.e., $\Theta_c \Pi_c = \|\theta_c\|_F \Gamma_c D_c$. Since $\Omega = U_r \Lambda U'_c$ and $U'_r U_r = I_{K_r}$, we have $U_c = \Theta_c \Pi_c P' \Pi'_r U_r \Lambda^{-1}$. Since $\Theta_c \Pi_c = \|\theta_c\|_F \Gamma_c D_c$, we have $U_c = \Gamma_c \|\theta_c\|_F D_c P' \Pi'_r U_r \Lambda^{-1} = \Gamma_c V_c$, where we set $V_c = \|\theta_c\|_F D_c P' \Pi'_r U_r \Lambda^{-1} \in \mathbb{R}^{K_c \times K_r}$. Note that since $U'_c U_c = I_{K_r} = V'_c \Gamma'_c \Gamma_c V_c = V'_c V_c$, we have $V'_c V_c = I_{K_r}$. Now, for $i_c \in [n_c]$, $k \in [K_r]$, we have

$$\begin{aligned} U_c(i_c, k) &= e'_{i_c} U_c e_k = e'_{i_c} \Gamma_c V_c e_k = \Gamma_c(i_c, :) V_c e_k \\ &= \theta_c(i_c) \left[\frac{\Pi_c(i_c, 1)}{\|\Theta_c \Pi_c(:, 1)\|_F} \frac{\Pi_c(i_c, 2)}{\|\Theta_c \Pi_c(:, 2)\|_F} \cdots \frac{\Pi_c(i_c, K_c)}{\|\Theta_c \Pi_c(:, K_c)\|_F} \right] V_c e_k \\ &= \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} V_c(\ell(i_c), k), \end{aligned}$$

which gives that

$$\begin{aligned} U_c(i_c, :) &= \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} [V_c(\ell(i_c), 1) \ V_c(\ell(i_c), 2) \ \dots \ V_c(\ell(i_c), K_r)] \\ &= \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} V_c(\ell(i_c), :). \end{aligned}$$

Then, we have

$$U_{c,*}(i_c, :) = \frac{V_c(\ell(i_c), :)}{\|V_c(\ell(i_c), :)\|_F}. \quad (\text{A6})$$

Sure, we have $U_{c,*}(i_c, :) = U_{c,*}(\bar{i}_c, :)$ when $\ell(i_c) = \ell(\bar{i}_c)$ for $i_c, \bar{i}_c \in [n_c]$. Let $B_c \in \mathbb{R}^{K_c \times K_r}$, such that $B_c(l, :) = \frac{V_c(l, :)}{\|V_c(l, :)\|_F}$ for $l \in [K_c]$. Equation (A6) gives $U_{c,*} = \Pi_c B_c$, which guarantees the existence of B_c .

Now, we consider the case for when $K_r = K_c = K$. Since $V_c \in \mathbb{R}^{K_c \times K_r}$ and $U_c = \Gamma_c V_c \in \mathbb{R}^{n_c \times K_r}$, we have $V_c \in \mathbb{R}^{K \times K}$ and $\text{rank}(V_c) = K$. Since $V_c' V_c = I_{K_r}$, we have $V_c' V_c = I_K$ when $K_r = K_c = K$. Then, we have

$$V_c' V_c = I_K \Rightarrow V_c' V_c V_c' = V_c' \Rightarrow V_c' (V_c V_c' - I_K) = 0 \xrightarrow{\text{rank}(V_c)=K} V_c V_c' = I_K. \quad (\text{A7})$$

Since $V_c V_c' = V_c' V_c = I_K$, we have $U_{c,*}(i_c, :) = V_c(\ell(i_c), :)$ by Equation (A6), and $\|U_{c,*}(i_c, :) - U_{c,*}(\bar{i}_c, :)\|_F = \|V_c(\ell(i_c), :) - V_c(\ell(\bar{i}_c), :)\|_F = \sqrt{2}$ when $\ell(i_c) \neq \ell(\bar{i}_c)$ for $i_c, \bar{i}_c \in [n_c]$, i.e., $\|B_c(k, :) - B_c(l, :)\|_F = \sqrt{2}$ for $k \neq l \in [K]$.

Note that, when $K_r < K_c$, since $\text{rank}(V_c) = K_r$ and $V_c \in \mathbb{R}^{K_c \times K_r}$, the inverse of V_c does not exist, which causes that the last equality in Equation (A7) does not hold and $\|B_c(k, :) - B_c(l, :)\| \neq \sqrt{2}$ for all $k \neq l \in [K_c]$.

□

Appendix C.3. Proof of Theorem 2

Proof. First, by the proof of Lemma 4.3 of [25], we have the below lemma.

Lemma A2. (Row-wise singular eigenvector error) Under $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \text{and } \Theta_c)$, when Assumption 2 holds, suppose that $\sigma_{K_r}(\Omega) \geq C\sqrt{\theta_{c, \max}(n_r + n_c)\log(n_r + n_c)}$, with a probability at least $1 - o((n_r + n_c)^{-\alpha})$,

$$\|\hat{U}_r \hat{U}_r' - U_r U_r'\|_{2 \rightarrow \infty} = O\left(\frac{\sqrt{\theta_{c, \max} K_r}(\kappa(\Omega) \sqrt{\frac{\max(n_r, n_c)\mu}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\theta_{c, \min} \sigma_{K_r}(P) \sigma_{K_r}(\Pi_r) \sqrt{n_{c, K_r}}}\right).$$

For the row nodes, when the conditions in Lemma A2 hold, by Theorem 2 of [21], we have

$$\max_{i_r \in [n_r]} \|e_{i_r}'(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega \kappa(\Pi_r' \Pi_r) K_r \sqrt{\lambda_1(\Pi_r' \Pi_r)}).$$

Next, we focus on the column nodes. By the proof of Lemma 3 in [19], there is an orthogonal matrix $\hat{O}$, such that

$$\|\hat{U}_c \hat{O} - U_c\|_F \leq \frac{2\sqrt{2K_r} \|A - \Omega\|}{\sqrt{\lambda_{K_r}(\Omega' \Omega)}}. \quad (\text{A8})$$

Under $\text{ODCNM}_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \text{and } \Theta_c)$, by Lemma 4 of [25], we have

$$\sqrt{\lambda_{K_r}(\Omega' \Omega)} \geq \theta_{c, \min} \sigma_{K_r}(P) \sigma_{K_r}(\Pi_r) \sqrt{n_{c, K_r}}. \quad (\text{A9})$$

By Lemma 4.2 of [25], when Assumption 2 holds, with a probability at least $1 - o((n_r + n_c)^{-\alpha})$, we have

$$\|A - \Omega\| = O(\sqrt{\max(\theta_{c,\max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}). \quad (\text{A10})$$

Substituting the two bounds in Equations (A9) and (A10) into Equation (A8), we have

$$\|\hat{U}_c \hat{O} - U_c\|_F \leq C \frac{\sqrt{K_r \max(\theta_{c,\max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}}{\sigma_{K_r}(P) \theta_{c,\min} \sigma_{K_r}(\Pi_r) \sqrt{n_{c,K_r}}}. \quad (\text{A11})$$

For $i_c \in [n_c]$, by basic algebra, we have

$$\|\hat{U}_{c,*}(i_c, :) \hat{O} - U_{c,*}(i_c, :)\|_F \leq \frac{2\|\hat{U}_c(i_c, :) \hat{O} - U_c(i_c, :)\|_F}{\|U_c(i_c, :)\|_F}.$$

Setting $m_c = \min_{1 \leq i_c \leq n_c} \|U_c(i_c, :)\|_F$, we have

$$\|\hat{U}_{c,*} \hat{O} - U_{c,*}\|_F = \sqrt{\sum_{i_c=1}^{n_c} \|\hat{U}_{c,*}(i_c, :) \hat{O} - U_{c,*}(i_c, :)\|_F^2} \leq \frac{2\|\hat{U}_c \hat{O} - U_c\|_F}{m_c}.$$

Next, we provide the lower bounds of m_c . By the proof of Lemma 2, we have

$$\begin{aligned} \|U_c(i_c, :)\|_F &= \left\| \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} V_c(\ell(i_c), :)\right\|_F = \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} \|V_c(\ell(i_c), :)\|_F \\ &\geq \frac{\theta_c(i_c)}{\|\Theta_c \Pi_c(:, \ell(i_c))\|_F} m_{V_c} \geq \frac{\theta_{c,\min}}{\theta_{c,\max} \sqrt{n_{c,\max}}} m_{V_c}, \end{aligned}$$

where we set $m_{V_c} = \min_{k \in [K_c]} \|V_c(k, :)\|_F$. Note that when $K_r = K_c = K$, by the Proof of Lemma 2, we know that $V_c V_c' = I_K$, which gives that $\|V_c(k, :)\|_F = 1$ for $k \in [K]$; i.e., $m_{V_c} = 1$ when $K_r = K_c = K$. However, when $K_r < K_c$, it is challenge to obtain a positive lower bound of m_{V_c} . Hence, we have $\frac{1}{m_c} \leq \frac{\theta_{c,\max} \sqrt{n_{c,\max}}}{\theta_{c,\min} m_{V_c}}$. Then, by Equation (A11), we have

$$\|\hat{U}_{c,*} \hat{O} - U_{c,*}\|_F = O\left(\frac{\theta_{c,\max} \sqrt{K_r \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}}{\sigma_{K_r}(P) \theta_{c,\min}^2 m_{V_c} \sigma_{K_r}(\Pi_r) \sqrt{n_{c,K_r}}}\right).$$

Let $\varsigma > 0$ be a small quantity; by Lemma 2 in [15], if

$$\frac{\sqrt{K_c}}{\varsigma} \|U_{c,*} - \hat{U}_{c,*} \hat{O}\|_F \left(\frac{1}{\sqrt{n_{c,k}}} + \frac{1}{\sqrt{n_{c,l}}} \right) \leq \|B_c(k, :) - B_c(l, :)\|_F, \text{ for each } 1 \leq k \neq l \leq K_c, \quad (\text{A12})$$

then the clustering error $\hat{f}_c = O(\varsigma^2)$. Setting $\varsigma = \frac{2}{\delta_c} \sqrt{\frac{K_c}{n_{c,\min}}} \|U_{c,*} - \hat{U}_{c,*} \hat{O}\|_F$ makes Equation (A12) hold for all $1 \leq k \neq l \leq K_c$. Then, we have $\hat{f}_c = O(\varsigma^2) = O\left(\frac{K_c \|U_{c,*} - \hat{U}_{c,*} \hat{O}\|_F^2}{\delta_c^2 n_{c,\min}}\right)$. By Equation (A11), we have

$$\hat{f}_c = O\left(\frac{\theta_{c,\max}^2 K_r K_c \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_{K_r}^2(P) \theta_{c,\min}^4 \delta_c^2 m_{V_c}^2 \sigma_{K_r}^2(\Pi_r) n_{c,K_r} n_{c,\min}}\right).$$

Especially, when $K_r = K_c = K$, $\delta_c = \sqrt{2}$ under $ODCNM_{n_r, n_c}(K_r, K_c, P, \Pi_r, \Pi_c, \Theta_c)$ by Lemma 2, and $m_{V_c} = 1$. When $K_r = K_c = K$, we have

$$\hat{f}_c = O\left(\frac{\theta_{c,\max}^2 K^2 \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_K^2(P) \theta_{c,\min}^4 \sigma_K^2(\Pi_r) n_{c,\min}^2}\right).$$

□

Appendix C.4. Proof of Corollary 2

Proof. For the row nodes, under the conditions of Corollary 2, we have

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O(\omega K_r \sqrt{\frac{n_r}{K_r}}) = O(\omega \sqrt{K n_r}).$$

Under the conditions of Corollary 2, $\kappa(\Omega) = O(1)$ and $\mu \leq C \frac{\theta_{c,\max}^2}{\theta_{c,\min}^2} \leq C$ for some $C > 0$ by Lemma 2 of [25]. Then, by Lemma A2, we have

$$\begin{aligned} \omega &= O\left(\frac{\sqrt{\theta_{c,\max} K_r}(\kappa(\Omega) \sqrt{\frac{\max(n_r, n_c) \mu}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\theta_{c,\min} \sigma_{K_r}(P) \sigma_{K_r}(\Pi_r) \sqrt{n_{c,K_r}}}\right) \\ &= O\left(\frac{\sqrt{\theta_{c,\max} K}(\kappa(\Omega) \sqrt{\frac{\max(n_r, n_c) \mu}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\theta_{c,\min} \sigma_K(P) \sigma_K(\Pi_r) \sqrt{n_{c,\min}}}\right) \\ &= O\left(\frac{K^{1.5} \sqrt{\theta_{c,\max}} (\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\theta_{c,\min} \sigma_K(P) \sqrt{n_r n_c}}\right), \end{aligned}$$

which gives that

$$\max_{i_r \in [n_r]} \|e'_{i_r}(\hat{\Pi}_r - \Pi_r \mathcal{P}_r)\|_1 = O\left(\frac{K^2 \sqrt{\theta_{c,\max}} (\sqrt{\frac{C \max(n_r, n_c)}{\min(n_r, n_c)}} + \sqrt{\log(n_r + n_c)})}{\theta_{c,\min} \sigma_K(P) \sqrt{n_c}}\right).$$

The reason for why we do not consider the case when $K_r < K_c$ for row nodes is similar as that of Corollary 1, and we omit it here.

For column nodes, under conditions of Corollary 2, we have

$$\begin{aligned} \hat{f}_c &= O\left(\frac{\theta_{c,\max}^2 K_r K_c \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_{K_r}^2(P) \theta_{c,\min}^4 \delta_c^2 m_{V_c}^2 \sigma_{K_r}^2(\Pi_r) n_{c,K_r} n_{c,\min}}\right) \\ &= O\left(\frac{\theta_{c,\max}^2 K_r^2 K_c^2 \max(\theta_{c,\max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}{\sigma_{K_r}^2(P) \theta_{c,\min}^4 \delta_c^2 m_{V_c}^2 n_r n_c}\right). \end{aligned}$$

For the case $K_r = K_c = K$, we have

$$\begin{aligned} \hat{f}_c &= O\left(\frac{\theta_{c,\max}^2 K^2 \max(\theta_{c,\max} n_r, \|\theta_c\|_1) n_{c,\max} \log(n_r + n_c)}{\sigma_K^2(P) \theta_{c,\min}^4 \sigma_K^2(\Pi_r) n_{c,\min}^2}\right) \\ &= O\left(\frac{\theta_{c,\max}^2 K^4 \max(\theta_{c,\max} n_r, \|\theta_c\|_1) \log(n_r + n_c)}{\sigma_K^2(P) \theta_{c,\min}^4 n_r n_c}\right). \end{aligned}$$

When $n_r = O(n)$, $n_c = O(n)$, $K_r = O(1)$ and $K_c = O(1)$, the corollary follows immediately by basic algebra. □

References

1. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826.
2. Newman, M.E. The structure and function of complex networks. *Siam Rev.* **2003**, *45*, 167–256.
3. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174.
4. Fortunato, S.; Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **2016**, *659*, 1–44.
5. Goldenberg, A.; Zheng, A.X.; Fienberg, S.E.; Airolidi, E.M. A survey of statistical network models. *Found. Trends Mach. Learn.* **2010**, *2*, 129–233.
6. Holland, P.W.; Laskey, K.B.; Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **1983**, *5*, 109–137.

7. Karrer, B.; Newman, M.E.J. Stochastic blockmodels and community structure in networks. *Phys. Rev.* **2011**, *83*, 16107.
8. Airoldi, E.M.; Blei, D.M.; Fienberg, S.E.; Xing, E.P. Mixed Membership Stochastic Blockmodels. *J. Mach. Learn. Res.* **2008**, *9*, 1981–2014.
9. Jin, J.; Ke, Z.T.; Luo, S. Estimating network memberships by simplex vertex hunting. *arXiv* **2017**. arXiv:1708.07852.
10. Zhang, Y.; Levina, E.; Zhu, J. Detecting Overlapping Communities in Networks Using Spectral Methods. *Siam J. Math. Data Sci.* **2020**, *2*, 265–283.
11. Rohe, K.; Chatterjee, S.; Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Stat.* **2011**, *39*, 1878–1915.
12. Qin, T.; Rohe, K. Regularized spectral clustering under the degree-corrected stochastic blockmodel. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3120–3128.
13. Lei, J.; Rinaldo, A. Consistency of spectral clustering in stochastic block models. *Ann. Stat.* **2015**, *43*, 215–237.
14. Jin, J. Fast community detection by SCORE. *Ann. Stat.* **2015**, *43*, 57–89.
15. Joseph, A.; Yu, B. Impact of regularization on spectral clustering. *Ann. Stat.* **2016**, *44*, 1765–1791.
16. Mao, X.; Sarkar, P.; Chakrabarti, D. Overlapping Clustering Models, and One (class) SVM to Bind Them All. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December, 2018; Volume 31, pp. 2126–2136.
17. Mao, X.; Sarkar, P.; Chakrabarti, D. Estimating mixed memberships with sharp eigenvector deviations. *J. Am. Stat. Assoc.* **2020**, *116*, 1–13.
18. Rohe, K.; Qin, T.; Yu, B. Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12679–12684.
19. Zhou, Z.; Amini, A.A. Analysis of spectral clustering algorithms for community detection: The general bipartite setting. *J. Mach. Learn. Res.* **2019**, *20*, 1–47.
20. Wang, Z.; Liang, Y.; Ji, P. Spectral algorithms for community detection in directed networks. *J. Mach. Learn. Res.* **2020**, *21*, 1–45.
21. Qing, H.; Wang, J. Directed mixed membership stochastic blockmodel. *arXiv* **2021**. arXiv:2101.02307v2.
22. Airoldi, E.M.; Wang, X.; Lin, X. Multi-way blockmodels for analyzing coordinated high-dimensional responses. *Ann. Appl. Stat.* **2013**, *7*, 2431–2457.
23. Razaee, Z.S.; Amini, A.A.; Li, J.J. Matched bipartite block model with covariates. *J. Mach. Learn. Res.* **2019**, *20*, 1174–1217.
24. Zhou, Z.; Amini, A.A. Optimal bipartite network clustering. *J. Mach. Learn. Res.* **2020**, *21*, 1–68.
25. Qing, H. Directed degree corrected mixed membership model and estimating community memberships in directed networks. *arXiv* **2021**. arXiv:2109.07826.
26. Ndaoud, M.; Sigalla, S.; Tsybakov, A.B. Improved clustering algorithms for the bipartite stochastic block model. *IEEE Trans. Inf. Theory* **2021**, *68*, 1960–1975.
27. Gillis, N.; Vavasis, S.A. Semidefinite programming based preconditioning for more robust near-separable nonnegative matrix factorization. *Siam J. Optim.* **2015**, *25*, 677–698.
28. Qing, H. A useful criterion on studying consistent estimation in community detection. *Entropy* **2022**, *24*, 1098.
29. Schellenberger, J.; Park, J.O.; Conrad, T.M.; Palsson, B.Ø. BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinform.* **2010**, *11*, 1–10.
30. Adamic, L.A.; Glance, N. The political blogosphere and the 2004 US election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, Chicago, IL, USA, 21–25 August 2005; pp. 36–43.
31. Kunegis, J. Konect: The koblenz network collection. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 1343–1350.
32. Leicht, E.A.; Newman, M.E. Community structure in directed networks. *Phys. Rev. Lett.* **2008**, *100*, 118703.
33. Zhang, H.; Guo, X.; Chang, X. Randomized spectral clustering in large-scale stochastic block models. *J. Comput. Graph. Stat.* **2022**, 1–20. <https://doi.org/10.1080/10618600.2022.2034636>.