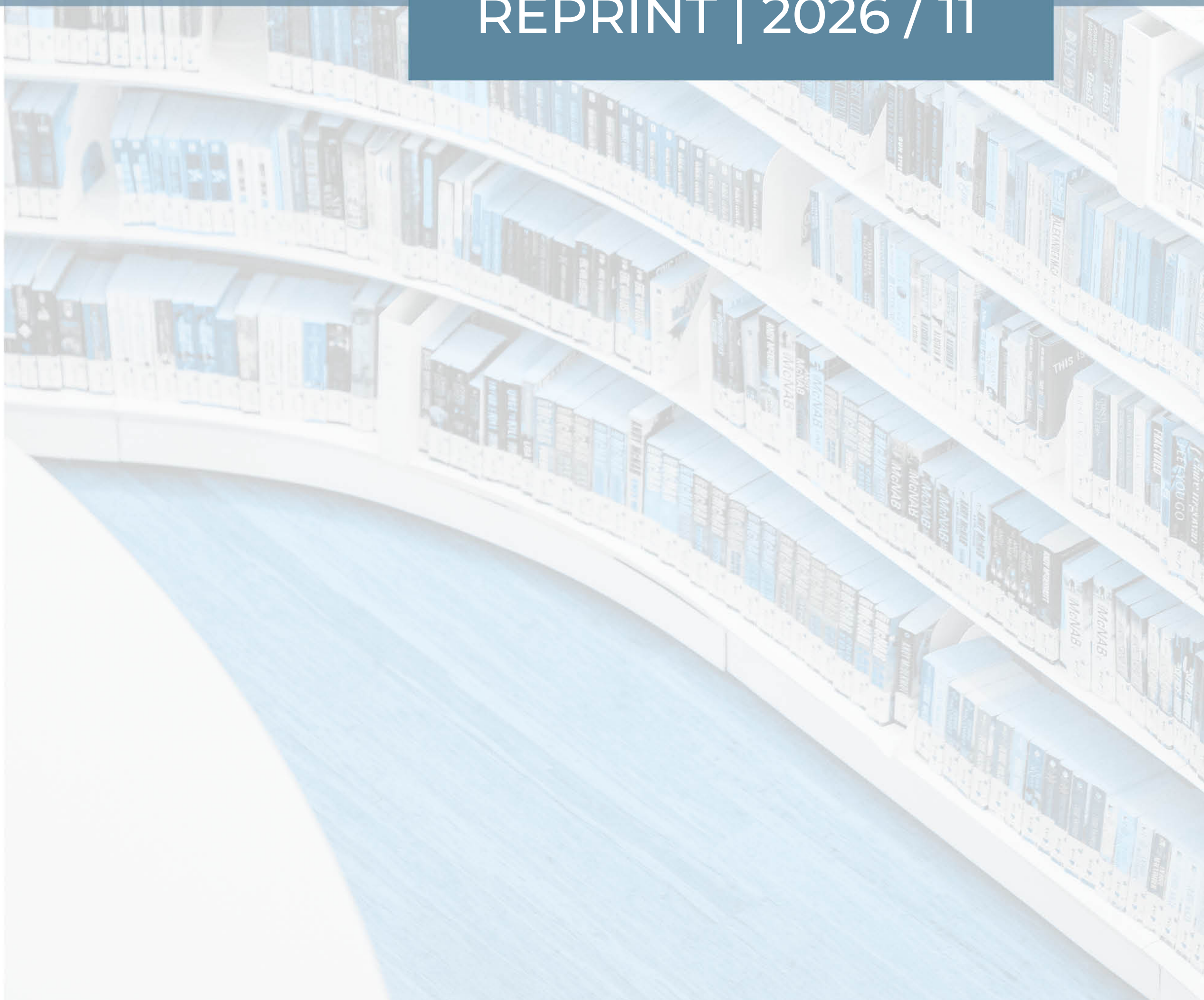
TIME-VARYING DEGREE-CORRECTED STOCHASTIC BLOCK MODELS

Mengxue Li, Rainer von Sachs, Eugen Pircalabelu

REPRINT | 2026 / 11



ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication

Time-varying degree-corrected stochastic block models

Mengxue Li, Rainer von Sachs, Eugen Pircalabelu

Institute of Statistics, Biostatistics and Actuarial Sciences, LIDAM, UCLouvain, Belgium

Abstract

Recent interest has emerged in community detection for dynamic networks which are observed along a trajectory of points in time. In this paper, we present a time-varying (t-v) degree-corrected stochastic block model to fit a dynamic network which allows evolving heterogeneity in the degrees of nodes within a community over time. In order to aggregate network information over time via a sliding window we propose a smoothing-based method that simultaneously allows to recover the: (i) t-v node connection probabilities; (ii) t-v community memberships; and (iii) t-v node degree parameters. As a novelty compared to existing literature we provide asymptotic theory for all of these three goals, i.e. we derive rates of consistency of our smooth estimators for degree parameters and communities using a time-localised profile-likelihood approach as well as strong consistency of community membership estimation at every time point. Extensive simulation studies and applications to two different real data sets complete our work.

Keywords: Dynamic network, Community detection, Time-localised profile-likelihood, Nonparametric curve estimation

1 Introduction

Networks or graphs are used to encode the interactions among a collection of individuals in a complex system. Each node in a network represents an individual, while an edge describes the interactions between each pair of individuals. A fundamental problem in network analysis is community detection, where the nodes can be partitioned into a collection of cohesive groups, also known as 'communities'. Nodes within a community share specific attributes and exhibit high connectivity, while they are sparsely connected across communities. Recovering the community structure has always received significant attention in many fields, such as identifying alliances in social networks (Gulati, 1998), discovering co-expressed genes in gene networks (Lei and Lin, 2023), or collaborating brain regions in brain networks (Liu, Choi, Xie, and Roeder, 2018; Samdin, Ting, and Ombao, 2019).

A classical model for the network displaying a community structure is the stochastic block model (SBM) introduced in Holland, Laskey, and Leinhardt (1983) and developed by many authors (Abbe, 2018; Bickel and Chen, 2009; Zhao, Levina, and Zhu, 2012). SBM posits that each node in the network belongs to a community. An edge between any pair of nodes is independent of other edges and the probability of an edge occurring only depends on the community assignments of related nodes. In other words, all nodes within the same community are considered stochastically equivalent having the same degree, and thus, the model does not allow for the existence of 'hubs' (nodes with high connectivity) which however can be observed in many applications. To address this problem, Karrer and Newman (2011) proposed the degree-corrected block model (DCSBM), which allows node degree heterogeneity by introducing degree parameters. In a sense, SBM can be regarded as a special case of the DCSBM where the degree parameters are all equal. To date, various methods have been extensively studied in both estimation and asymptotic theory for community detection in SBM and DCSBM. Bickel and Chen (2009) established a theoretical framework for the consistency of community detection based on the profile likelihood and modularity algorithms for SBM. Later, Zhao et al. (2012) extended their theoretical results to the DCSBM. Furthermore, the consistency of other methods, based on pseudo-likelihood estimation (Amini, Chen, Bickel, and Levina, 2013; Wang, Zhang, Liu, Zhu, and Guo, 2023), variational expectation-maximization (Bickel, Choi, Chang, and Zhang, 2013; Wang and Bickel, 2017), information-theoretic

(Gao, Ma, Zhang, and Zhou, 2018), and spectral clustering (Jin, 2015; Qin and Rohe, 2013; Rohe, Chatterjee, and Yu, 2011), were also developed under both the SBM and DCSBM frameworks.

All the aforementioned methods focused mostly on the static network observed as a single snapshot. However, the network can be observed at multiple time points, and can evolve over time. Such a collection of networks is referred to as a dynamic network with n nodes and T time points. While both dynamic and multilayer networks comprise multiple layers, the former explicitly incorporates temporal evolution, whereas the latter generally refers to a collection of networks without time dependence. In this paper, we mainly focus on time-ordered sequences of networks. In recent years, numerous studies have sought to extend community detection procedures to dynamic networks. Huang, Weng, and Feng (2023) implemented spectral clustering on the weighted integration of multilayer SBM. Pensky and Zhang (2019) employed kernel-based aggregation spectral clustering for dynamic SBM and studied estimation consistency. Lei and Lin (2023) proposed a debiased spectral clustering method for the sum-of-squares aggregation and established the consistency of common community estimation, followed by the consistency of time-varying estimation using a kernel debiased method in Lin and Lei (2024). Despite the computational advantages of these spectral clustering methods, they fail to incorporate the node heterogeneity, with only a few studies extending them to dynamic DCSBM. Liu et al. (2018) combined the degree-normalized spectral clustering with eigenvector smoothing to detect communities. Bhattacharyya and Chatterjee (2020) and Agterberg, Lubberts, and Arroyo (2025) generalized spectral clustering to multilayer DCSBM. However, the former required the node heterogeneity and communities to remain the same across networks, whereas the latter was only suitable for the networks with a common community structure. On the other hand, spectral clustering-based approaches (see, e.g. Agterberg et al. 2025; Pensky and Zhang 2019) typically impose a strong assumption on community connectivity matrices, specifically requiring that the smallest eigenvalue of each matrix is bounded away from zero across all time points. This condition intuitively makes the detection problem easier since the communities are well-separated, however, it also restricts such methods only to the assortative network case where connections within a community are more likely to appear. A final line of work is based on probabilistic generative models. Lei, Chen, and Lynch (2020) proposed a least-squares estimator and proved the consistency of common community estimation. Matias and Miele (2017) and Riverain, Fossier, and Nadif (2023) applied Markov chain methods to estimate the dynamic community structures which has proven to be intractable for investigating the theoretical properties of the estimators. Li, Yuan, Zhang, and Qu (2024) allowed for abrupt changes occurring in communities. However, they assumed that all connectivity probabilities remain constant before and after the change-point, which is often unrealistic in practice.

Despite many extensions made to dynamic networks and associated estimations, they still have limited applicability to real-world networks since it is common in reality for a network to be dynamic, heterogeneous in node degrees and variable in community structures and connectivities. For example, our application treating the Travian dataset (in Section 6) involves a network that recorded the communication among players spanning 30 days. Figure 1 provides an illustration of the time-varying communication networks. In this figure, we observe that some players (nodes) departed from their original communities to join new ones, and some players communicated more intensively than others during specific periods. Thus, existing methods struggle to be directly applicable to such dynamic networks, which simultaneously incorporate these time-varying features. In this work, we propose a time-varying DCSBM to offer more flexibility in fitting real-world networks. In contrast to existing dynamic network models, our approach focuses on how both node heterogeneity and community connectivity evolve over time. Moreover, as a novelty compared to the existing reviewed literature, we accompany our approach with asymptotic theory on rates of consistency of the time-varying estimators for degree parameters and communities which are based on a time-localised profile-likelihood approach. Finally we show strong consistency of our community membership estimators at every time point.

A key aspect when working with such dynamic networks is that of how to efficiently aggregate the information from single network snapshots to obtain a more accurate community detection. Hence, in this paper we choose kernel smoothing of adjacent networks over time to perform aggregation of the information that one gets from a 'trajectory' in time of a dynamic network. Obviously, and in particular if the series of networks is sparse (in the number of their nodes), this allows us to reduce the sampling error inherent in estimating the population edge probabilities from observed adjacency matrices. In a second step, in order to establish a connection between our first-step edgewise smoothing in time and retrieving the whole structure of the dynamic network we will use our kernel-smooth estimators of edge probabilities as input to a 'time-localised' likelihood approach for recovering the time-varying network degree parameters and communities. To keep the complexity of our asymptotic derivations manageable we content ourselves to present our work while using a basic boxcar kernel smoother which allows optimal

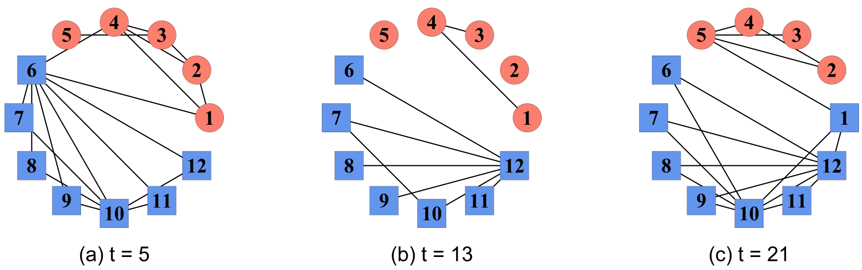


Figure 1: A toy example of a dynamic subnetwork for the Travian dataset with $n = 12$ nodes at three time points $t = 5, 13, 21$. The red circles and blue squares denote $K = 2$ distinct communities given by the ground truth.

rates of convergence essentially for two times differentiable curves (over time). Nevertheless, any other more sophisticated nonparametric curve estimation method - allowing to treat functions of different (higher or lower) regularity in time - could equally well be employed without changing the essential ideas and results of our approach.

Summarizing, the main contributions of our paper are the following. Firstly, we focus on time-varying DCSBM which allows the heterogeneity of node degrees and community structures to vary throughout time. Secondly, we combine nonparametric curve estimation of time-varying edge probabilities with a time-localised profile-likelihood approach to design a smoothing-based method for estimating time-varying degree parameters and community structures. With this, we provide a novel framework for applying nonparametric techniques for estimation of smooth curves to networks observed at discrete time points. Lastly but most importantly, we establish rates of consistency of our smooth estimators of degree parameters and communities under the asymptotic regimes of both semi-sparse and sparse networks.

Our paper is structured as follows. In Section 2, we introduce dynamic networks and profile-likelihood estimation of time-varying degree-corrected stochastic block models. In Section 3, we propose our concrete time-localised profile-likelihood approach for parameter estimation and derive the theoretical rate of convergence of the MSE-optimal bandwidth as well as a data-adaptive bandwidth selection procedure designed to work in practice. Section 4 develops the consistency results for time-varying node degree parameters and community structures. Section 5 studies the performance and robustness of our method through simulations, and Section 6 gives the application to data sets in two different scenarios, one with a constant community structure throughout time and the other with time-varying community structures. Conclusions and discussions are summarized in Section 7. All details of the proofs are provided in the Appendix.

2 Time-varying degree-corrected stochastic block models

2.1 Dynamic network framework

We consider a dynamic network with n nodes where edges between the nodes may appear or disappear as the network evolves over time. Assume that such a dynamic network is recorded as a collection of adjacency matrices $\{\boldsymbol{A}^1, \ldots, \boldsymbol{A}^T\}$ where T is the number of time points over which the network is observed. For simplicity, we focus only on undirected and unweighted networks and allow self-loops, so that each $\mathbf{A}^t$ is an $n \times n$ symmetric binary matrix where $A^t_{ij} = A^t_{ji} = 1$ if there exists an edge between nodes i and j at time point t for any $i, j \in \{1, \ldots, n\}$, otherwise, $A^t_{ij} = A^t_{ji} = 0$ if the nodes i and j are not connected.

Furthermore, we suppose that the n nodes can be grouped into K communities whose structure may vary over time, but the number of communities, denoted as K, remains fixed. Let the set of vectors $\{\mathbf{c}^1, \ldots, \mathbf{c}^T\}$ describe such a time-varying community structure, where $\mathbf{c}^t \in \{1, \ldots, K\}^n$ represents the true assignments of the nodes at time point t, with $c^t_i = k$ indicating that node i belongs to community k.

To illustrate the concepts, we plot a dynamic subnetwork for the Travian dataset (further described in Section 6) in Figure 1 involving $n = 12$ nodes and $K = 2$ different communities (circle group and square group). As can be seen, an edge existed between nodes 6 and 10 at $t = 5$ but disappeared at $t = 13$. Later, after some time evolution, it was formed again at $t = 21$. We also observe that node

1 belonged to the circle group at time points $t = 5$ and $t = 13$ but it appeared in the square group at $t = 21$. In addition, it is notable that some nodes display higher popularity than others even within the same community. We refer to a highly connected node as a hub node, and the variation in connectivity as degree heterogeneity. In the three panels of Figure 1, we observe that nodes 6 and 10 are hub nodes at $t = 5$. However, by $t = 13$, nodes 6 and 10 have a decrease in connectivity while node 12 emerges as a new hub. Subsequently, node 10 becomes active again at $t = 21$ and plays the role of a hub together with node 12. Obviously, the nodes not only exhibit degree heterogeneity in reality but also change with the networks evolving.

2.2 Time-varying degree-corrected stochastic block model

Assume that for each node $i \in \{1, \ldots, n\}$ at time point t a degree parameter $\theta_i^t \in \mathbb{R}_+$ is associated. Its role is to interpret how 'popular' a node in the network is, via a smooth function across time. In this work, we assume that the propensity to form edges in the network, between any two nodes i and j is a function not only of the popularity of nodes i and j, but also of the communities to which these two nodes belong. As such, to account for the influence of the communities on the propensity to form ties between nodes, we consider a $K \times K$ symmetric matrix $\boldsymbol{Z}^t = (Z_{ab}^t)_{K\times K}$, representing the community connectivity parameter, where Z_{ab}^t denotes the probability of an edge between communities a and b when nodes i and j belong to communities a and b, respectively. For simplicity, we suppose that the community connectivity parameters $\boldsymbol{Z}^t$ do not depend on the degree parameters $\boldsymbol{\theta}^t = (\theta_1^t, \ldots, \theta_n^t)^\top$. This assumption is common in the literature of DCSBM.

In the sequel, we propose to model

$$\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(A_{ij}^t\right) = \theta_i^t\theta_j^t Z_{c_i^t c_j^t}^t \;\in [0,1]\;,\quad i,j = 1,\ldots,n,\; t = 1,\ldots,T\;, \tag{1}$$

where for given community labels and node degree parameters, A_{ij}^t's are assumed to be independent Bernoulli random variables, and where $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(\cdot)$ denotes the expectation in a probability model with parameters $\boldsymbol{c}^t$ and $\boldsymbol{\theta}^t$. Let $\mathbf{1}_{\{\cdot\}}$ be the indicator function and further define $n_a(\boldsymbol{c}^t) = \sum_{i=1}^n \mathbf{1}_{\{c_i^t = a\}}$ as the total number of nodes belonging to community a at time point t. For identifiability reasons, throughout this work we assume (i) that θ_i^t is such that $\sum_{i=1}^n \theta_i^t \mathbf{1}_{\{c_i^t=a\}} = n_a(\boldsymbol{c}^t)$ for any $a = 1,\ldots,K$ and (ii) that $Z_{c_i^t c_j^t}^t$ is such that $\theta_i^t\theta_j^t Z_{c_i^t c_j^t}^t \in [0,1]$. Note that as a special case, when $T = 1$, model (1) reduces to a static DCSBM that is well studied for community detection (see, e.g. Gao et al. 2018; Zhao et al. 2012). Moreover, if also $\theta_i^t = 1$ for all nodes in the network, the static DCSBM reduces to the standard static SBM.

Model (1) defines a valid, dynamic degree-corrected stochastic block model. However, in order to evaluate the asymptotic behavior of the model with increasing $n \to \infty$, one needs to address a scaling issue related to the probabilities of edges. If $\boldsymbol{Z}^t$ is fixed and independent of n with positive entries, then the average degree of the nodes will be of order n, making the network overly dense as pointed out by Zhao et al. (2012). Instead, we introduce a baseline parameter ρ_n characterizing the sparsity of the network which depends on n but is constant across time and allow $\boldsymbol{Z}^t$ scaling with n. In view of this remark, we redefine slightly our model as:

$$\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t) = \rho_n\theta_i^t\theta_j^t W_{c_i^t c_j^t}^t\;, \tag{2}$$

where ρ_n stands for the baseline sparsity of the network. In this work, ρ_n is assumed to be constant across time but depending on n via $\rho_n \to 0$ as $n \to \infty$. With W_{ab}^t we denote a fixed entry from a symmetric $K \times K$ matrix $\boldsymbol{W}^t := \rho_n^{-1}\boldsymbol{Z}^t$ describing the connectivity pattern between nodes from any of the K underlying communities. As noted in Bickel and Chen (2009), this parametrization allows the decoupling of the average degree from the inhomogeneity structure.

Remark 1. *Let $\lambda_n = n\rho_n$ be the baseline average node degree of the dynamic network. As commonly defined in the literature, while ρ_n remaining constant in n would refer to a 'dense' network (with unrealistic fast growing connection activity), the regime $\rho_n \to 0$ and $\lambda_n \gg \log n$ is referred to as 'semi-dense'. Next, the case of $\lambda_n \to \infty$ but not faster than $\log n$ is called 'semi-sparse', whereas $\lambda_n = O(1)$ is considered as 'sparse'. We will provide further elaboration on these regimes and their implications for our method in subsequent sections of the paper.*

2.3 Profile-likelihood estimation

Our objective for this work is to estimate the time-varying community structure with heterogeneity of node degrees. There are many related works for community detection, based on graph partitioning

(Newman, 2013), spectral clustering (Lei and Lin, 2023; Pensky and Zhang, 2019) and modularity (Bickel and Chen, 2009; Clauset, Newman, and Moore, 2004; Zhao et al., 2012). As mentioned in Section 1, these methods are relatively less flexible in accommodating node heterogeneity and community structure. One other natural and straightforward approach is the likelihood-based method (Bickel and Chen, 2009; Zhao et al., 2012).

Before discussing this estimation method, we introduce some basic definitions at first. For each time point $t \in \{1, \ldots, T\}$, we consider $d_i^t = \sum_{j=1}^n A_{ij}^t$ as the number of edges from node i to any nodes in the network. Let $\boldsymbol{e}^t = (e_1^t, \ldots, e_n^t)^\top$ be a generic vector defined on $\{1, \ldots, K\}^n$ and e_i^t be the candidate assignment of node i at time point t. For any vector $\boldsymbol{e}^t$, we define

$$O_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^n A_{ij}^t \mathbf{1}_{\{e_i^t = k, e_j^t = l\}}, \qquad O_k(\boldsymbol{e}^t) = \sum_{l=1}^K O_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^n A_{ij}^t \mathbf{1}_{\{e_i^t = k\}}.$$

Here $O_{kl}(\boldsymbol{e}^t)$ denotes the total number of edges connecting nodes from communities k and l when $k \neq l$, implying that $O_{kk}(\boldsymbol{e}^t)$ is twice the number of edges within the community without self-loops in the networks. The quantity $O_k(\boldsymbol{e}^t)$ denotes the total number of marginal connections from nodes in community k to nodes from any community in the network (including itself).

Since it is more complicated to obtain closed form solutions via maximizing the Bernoulli-based likelihood function, following the method mentioned in Karrer and Newman (2011) and Zhao et al. (2012), we also consider using the Poisson-based likelihood to approximate the Bernoulli-based likelihood for simplifying the estimation procedure. Therefore, for any fixed assignment candidates $\boldsymbol{e}^t$ imposing the constraint that $\sum_{i=1}^n \theta_i^t \mathbf{1}_{\{e_i^t = a\}} = n_a(\boldsymbol{e}^t)$ for any $a = 1, \ldots, K$, leads to the following form of the log-likelihood function of the parameters $\boldsymbol{\theta}^t$ and $\boldsymbol{Z}^t$,

$$\log \mathbb{P}(\boldsymbol{A}^1, \ldots, \boldsymbol{A}^T) = 2 \sum_{t=1}^T \sum_{i=1}^n d_i^t \log \theta_i^t + \sum_{t=1}^T \sum_{a,b=1}^K \left(O_{ab}(\boldsymbol{e}^t) \log Z_{ab}^t - n_a(\boldsymbol{e}^t) n_b(\boldsymbol{e}^t) Z_{ab}^t \right),$$

which yields the closed-form MLEs of θ_i^t and $Z_{c_i^t c_j^t}^t$ under the given assignment vector $\boldsymbol{e}^t$ as follows,

$$\hat{\theta}_i^t(\boldsymbol{e}^t) = \frac{n_{e_i^t}(\boldsymbol{e}^t) d_i^t}{O_{e_i^t}(\boldsymbol{e}^t)}, \qquad \hat{Z}_{e_i^t e_j^t}^t(\boldsymbol{e}^t) = \frac{O_{e_i^t e_j^t}(\boldsymbol{e}^t)}{n_{e_i^t}(\boldsymbol{e}^t) n_{e_j^t}(\boldsymbol{e}^t)}. \tag{3}$$

Plugging these MLEs into the log-likelihood function and omitting constants, we can obtain a function of only the community assignments. Therefore, for any generic candidate $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$, the criterion for community detection is defined as

$$Q(O(\boldsymbol{e}^t)) = \sum_{k,l=1}^K \left(O_{kl}(\boldsymbol{e}^t) \log \frac{O_{kl}(\boldsymbol{e}^t)}{O_k(\boldsymbol{e}^t) O_l(\boldsymbol{e}^t)} - O_{kl}(\boldsymbol{e}^t) \right), \tag{4}$$

where the term $O(\boldsymbol{e}^t)$ is used to denote the $K \times K$ matrix with elements $O_{kl}(\boldsymbol{e}^t)$. The term $Q(O(\boldsymbol{e}^t))$ is referred to as the profile-likelihood function in Bickel and Chen (2009) and the estimation of communities can be derived by maximizing this quantity over all possible candidates $\boldsymbol{e}^t$.

Roughly speaking, the estimation based on (4) can be considered as applying the static DCSBM to each individual observation $\boldsymbol{A}^t$ separately. Supposing that K is known and fixed, Zhao et al. (2012) showed that at any fixed time point, optimizing (4) leads to strongly consistent estimators for the communities as long as $\lambda_n / \log n \to \infty$ as $n \to \infty$. In other words, this model would require the networks to satisfy the sparsity condition of $\lambda_n \gg \log n$ at every fixed time point to maintain the strong consistency property when estimating communities. However, this condition is severely constraining in practice. For instance, upon visual inspection of the three original Travian networks in Figure 2 (a)-(c), a notable facet is that while communities persist throughout all time points, the first community is less distinguishable at $t = 5$, and when $t = 13$, at least two communities - the first and the eighth - cannot be recognized. Indeed, this would be tricky in the estimation procedure if only focusing on individual snapshots at $t = 5$ or $t = 13$.

In this case, the undisclosed community structure can however be recovered by aggregation of networks from various time points. To motivate our smoothing-based procedure, discussed further in Section 3, if one were to smooth first these adjacency matrices using the basic kernel function, as defined in (7) below, all communities become more apparent for any individual network as shown in Figure 2 (d)-(f). Moreover, from a statistical sampling perspective, a solitary network is perceived as a noisy sample from

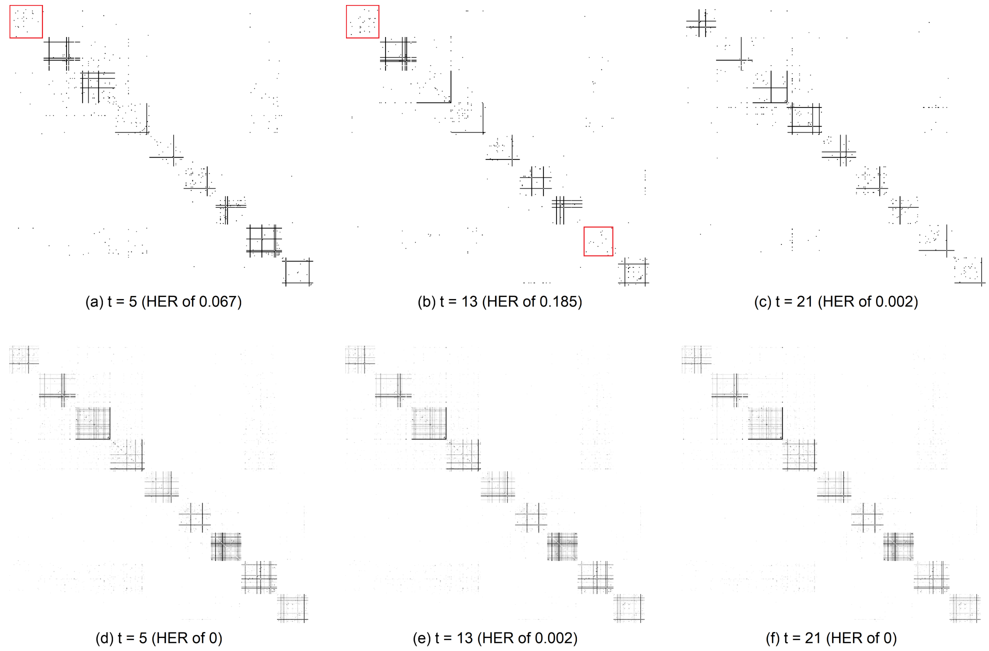


Figure 2: The original adjacency matrices (top row: (a)-(c)) and the smoothed adjacency matrices (bottom row: (d)-(f)) with a bandwidth $h = 0.47$ are presented for three time points $t = 5, 13, 21$ in the Travian dataset among 411 players assigned to 9 communities. The black and white dots represent the presence ($A_{ij}^t = 1$) and absence ($A_{ij}^t = 0$) of an edge, respectively. The non-recognized communities are remarked in red boxes. HER (lower is better) denotes the Hamming error rate which quantifies the proportion of misclassified nodes.

the population, whereas the aggregated network resulting from proper temporal smoothing has obviously a closer resemblance to the population version, thereby parameter estimation based on the smoothed network can alleviate sampling errors inherent in the individual network. The results of the analysis on the Travian dataset corroborate this observation. The HERs (the Hamming error rate, as defined in (20)) based on smoothed adjacency matrices are substantially lower than those using criterion (4) separately for each individual network.

3 A kernel smoothing method to recover time-varying degree parameters and communities

In this section we develop our estimation procedure of smoothing edge probabilities of adjacent networks over time to deal with the community detection problem for the time-varying DCSBM. For this we embed our smoothing approach into the framework of nonparametric curve estimation with a target function observed, in the presence of noise, on a grid of equidistant time points $u^t := t/T, t = 1, \ldots, T$ of the unit interval $u \in [0, 1]$. This is analogous to the fixed design points in a classical nonparametric regression setting. More specifically, suppose that there exist curves $\theta_i(u)$ for any $i \in \{1, \ldots, n\}$ and $W_{ab}(u)$ for any $a, b \in \{1, \ldots, K\}$ with $u \in [0, 1]$, which allow to define our target $\tau_{ij}(u)$ as a curve with

$$\tau_{ij}(u^t) \;=\; \mathbb{E}_{\boldsymbol{c}^t, \boldsymbol{\theta}(u^t)}(A_{ij}^t | u = u^t),$$

where $\boldsymbol{\theta}(u^t) := (\theta_1(u^t), \ldots, \theta_n(u^t))^\top$, such that

$$\tau_{ij}(u^t) \;= \rho_n \theta_i(u^t)\theta_j(u^t)W_{c_i^t c_j^t}(u^t), \tag{5}$$

which relates to model (2). In particular, we note $\theta_i^t = \theta_i(u^t)$ and $W_{ab}^t = W_{ab}(u^t)$. Observe that $\tau_{ij}(u^t)$ denotes the probability of an edge connecting nodes i and j at time point t. Let ε_{ij}^t be defined as the centered version of A_{ij}^t where

$$\varepsilon_{ij}^t = A_{ij}^t - \tau_{ij}(u^t). \tag{6}$$

By construction, the expectation of ε_{ij}^t is zero and its variance satisfies $\sigma_{ij}^2(u^t) = \tau_{ij}(u^t)(1-\tau_{ij}(u^t))$. Moreover, the errors ε_{ij}^t are assumed to be independent across any $i,j \in \{1,\dots,n\}$ and any $t \in \{1,\dots,T\}$.

As we have assumed that $\tau_{ij}(u)$ is not only a curve over time but also an expectation of edge variables, one simple but efficient way to obtain an estimator relies on local smoothing with kernel functions. As such, combining this with the profile-likelihood method from Section 2.3, we propose a two-step time-localised likelihood estimation procedure which first smooths data over time and then estimates the communities and parameters of interest based on the smoothed data.

Step 1: Smoothing. In this step, we estimate the edge probability $\tau_{ij}(u^t)$ by averaging the observations A_{ij}^t over neighbors in time t for any $i,j \in \{1,\dots,n\}$. Let $U(u^t) = [u^t - h, u^t + h]$ be a neighborhood of time point $t \in \{1,\dots,T\}$ for some bandwidth $h \in (0,1]$. Then for any $i,j \in \{1,\dots,n\}$, a basic kernel estimator of $\tau_{ij}(u^t)$ can be written as

$$\hat{\tau}_{ij,h}(u^t) = \frac{\sum_{t'=1}^T A_{ij}^{t'} \mathbf{1}_{\{u^{t'} \in U(u^t)\}}}{\left|\{A_{ij}^{t'} : u^{t'} \in U(u^t)\}\right|} \tag{7}$$

where $|\cdot|$ denotes the size of the set.

Step 2: Community detection. This step involves the estimation of parameters in time-varying DCSBM. For any time point $t \in \{1,\dots,T\}$, we replace the adjacency matrix $\boldsymbol{A}^t$ with the kernel estimator $\hat{\boldsymbol{\tau}}_h(u^t) := (\hat{\tau}_{ij,h}(u^t))_{n\times n}$. Correspondingly, for any vector $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, we define the smoothed counting variables as follows,

$$\hat{O}_{kl,h}(\boldsymbol{e}^t) = \sum_{i,j=1}^n \hat{\tau}_{ij,h}(u^t)\mathbf{1}_{\{e_i^t=k, e_j^t=l\}}, \quad \hat{O}_{k,h}(\boldsymbol{e}^t) = \sum_{l=1}^K \hat{O}_{kl,h}(\boldsymbol{e}^t), \quad \hat{d}_{i,h}^t = \sum_{j=1}^n \hat{\tau}_{ij,h}(u^t). \tag{8}$$

As well, one can obtain the smoothed MLEs of θ_i^t and $Z_{c_i^t c_j^t}^t$ from (3) under a given assignment vector $\boldsymbol{e}^t$ as,

$$\hat{\theta}_{i,h}^t(\boldsymbol{e}^t) = \frac{n_{e_i^t}(\boldsymbol{e}^t)\hat{d}_{i,h}^t}{\hat{O}_{e_i^t,h}(\boldsymbol{e}^t)}, \quad \hat{Z}_{e_i^t e_j^t,h}^t(\boldsymbol{e}^t) = \frac{\hat{O}_{e_i^t e_j^t,h}(\boldsymbol{e}^t)}{n_{e_i^t}(\boldsymbol{e}^t)n_{e_j^t}(\boldsymbol{e}^t)}, \tag{9}$$

where $\hat{O}_{e_i^t e_j^t,h}(\boldsymbol{e}^t) = \sum_{s,m=1}^n \hat{\tau}_{sm,h}(u^t)\mathbf{1}_{\{e_s^t=e_i^t, e_m^t=e_j^t\}}$, and $\hat{O}_{e_i^t,h}(\boldsymbol{e}^t) = \sum_{s,j=1}^n \hat{\tau}_{sj,h}(u^t)\mathbf{1}_{\{e_s^t=e_i^t\}}$. Plugging in these smoothed MLEs, the final estimator for $\tau_{ij}(u^t)$ at time point t is written as

$$\tilde{\tau}_{ij,h}(u^t, \boldsymbol{e}^t) = \hat{\theta}_{i,h}^t(\boldsymbol{e}^t)\hat{\theta}_{j,h}^t(\boldsymbol{e}^t)\hat{Z}_{e_i^t e_j^t,h}^t(\boldsymbol{e}^t) = \frac{\hat{d}_{i,h}^t \hat{d}_{j,h}^t \hat{O}_{e_i^t e_j^t,h}(\boldsymbol{e}^t)}{\hat{O}_{e_i^t,h}(\boldsymbol{e}^t)\hat{O}_{e_j^t,h}(\boldsymbol{e}^t)}, \tag{10}$$

where $\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{e}^t) := (\tilde{\tau}_{ij,h}(u^t, \boldsymbol{e}^t))_{n\times n}$.Thus, the smoothed profile-likelihood criterion (also referred to as the time-localised profile-likelihood criterion) associated only with community assignments is then defined as

$$Q(\hat{O}_h(\boldsymbol{e}^t)) = \sum_{k,l=1}^K \left(\hat{O}_{kl,h}(\boldsymbol{e}^t) \log \frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{\hat{O}_{k,h}(\boldsymbol{e}^t)\hat{O}_{l,h}(\boldsymbol{e}^t)} - \hat{O}_{kl,h}(\boldsymbol{e}^t) \right), \tag{11}$$

with a $K \times K$ matrix $\hat{O}_h(\boldsymbol{e}^t) := (\hat{O}_{kl,h}(\boldsymbol{e}^t))_{K\times K}$. Note that maximizing $Q(\hat{O}_h(\boldsymbol{e}^t))$ over all possible candidate assignments will yield the estimated community assignments $\hat{\boldsymbol{c}}^t$. These estimates are then substituted into (9) and (10) to construct the estimated degree parameters $\hat{\theta}_{i,h}^t(\hat{\boldsymbol{c}}^t)$ and the estimated edge probabilities $\tilde{\tau}_{ij,h}(u^t, \hat{\boldsymbol{c}}^t)$.

Remark 2. *The kernel estimator $\hat{\tau}_{ij,h}(u^t)$ in (7) allows one to aggregate the information from the neighboring time points, but it is edgewise without accounting for any topology of the network. In Step 2, we treat the kernel estimator as a 'weighted' adjacency matrix, combining the likelihood information from the model, to construct the final estimator in (10). A key feature of this approach is that it establishes a bridge between edgewise smoothing and the network structure. First of all, as usual for nonparametric curve estimation, it reduces the variance of the estimators. Moreover, $Q(\hat{O}_h(\boldsymbol{e}^t))$ provides an approximation to the population version $Q(\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O(\boldsymbol{e}^t)))$ by smoothing as well. It has a smaller sampling error than the single-sample $Q(O(\boldsymbol{e}^t))$, enabling the network to gain information from neighboring networks.*

In this context, the selection of the bandwidth h in (7) is of significant importance. Although we smooth the time-varying networks edgewise in our method, we prefer selecting a common bandwidth h for the entire network and for all time points. We will further discuss this later on in Remark 6. On the other hand, the probability matrix $\boldsymbol{\tau}(u^t) := (\tau_{ij}(u^t))_{n\times n}$, serving as the target linked to nonparametric techniques, encapsulates all parameters comprehensively. The goal is to choose a bandwidth that minimizes the estimation error of the smooth estimator $\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$ under the true communities $\boldsymbol{c}^t$ for any $t \in \{1, \ldots, T\}$. For notational simplicity, throughout this work we drop the argument $(\boldsymbol{c}^t)$ and use $\hat{\theta}^t_{i,h}$ and $\tilde{\tau}_{ij,h}(u^t)$ to denote $\hat{\theta}^t_{i,h}(\boldsymbol{c}^t)$ and $\tilde{\tau}_{ij,h}(u^t, \boldsymbol{c}^t)$ when $\boldsymbol{c}^t$ is assumed to be given, unless the argument is explicitly stated otherwise. To facilitate our theoretical development, we state first the following assumptions required to derive an upper bound on the error of $\tilde{\boldsymbol{\tau}}_h(u^t)$.

Assumption 1 (Trajectories of model parameters)**.** *Uniformly over $i \in \{1, \ldots, n\}$, and $a, b \in \{1, \ldots, K\}$, assume that both $\theta_i(u), W_{ab}(u) \in C^2[0, 1]$ over the unit interval $u \in [0, 1]$.*

Assumption 2 (Community structure)**.** *The number of communities K is given and fixed, independent of the number of nodes n and time points T. Assume that there exists a true community vector $\boldsymbol{c}^t \in \{1, \ldots, K\}^n$ such that $n_a(\boldsymbol{c}^t)$ is a positive integer for any community $a \in \{1, \ldots, K\}$ and any time point $t \in \{1, \ldots, T\}$.*

Assumption 3 (Identifiability of parameters)**.** *For any time point $t \in \{1, \ldots, T\}$ and any community $a \in \{1, \ldots, K\}$, the degree parameter vector $\boldsymbol{\theta}^t = (\theta^t_1, \ldots, \theta^t_n)^\top$ satisfies $\sum_{i=1}^n \theta^t_i \mathbf{1}_{\{c^t_i = a\}} = n_a(\boldsymbol{c}^t)$. Moreover, to avoid severely unbalanced degrees, we assume that $\sup_{i\in\{1,\ldots,n\}} \theta^t_i = O(1)$ for any time point t.*

Imposing in Assumption 1 that the temporal trajectories of the degree parameters and community connectivity parameters lie in $C^2[0, 1]$ is merely for convenience of phrasing the usual bias-variance behavior of the (kernel) smoother in nonparametric curve estimation (e.g., assuming the existence of a third derivative in time is imposed only for commodity in phrasing the proofs). The actual amount of smoothness is not essential for the applicability of the second-step of time-localised likelihood estimation of node degrees and communities based on the kernel-smoothed adjacency matrices which estimate $\tau_{ij}(u)$ in the first step. In the same way, one could of course replace the simple boxcar kernel smoother we suggest in (7) by a more refined kernel as to be found in the classical literature. Investigating the selection of the number K of communities, although interesting in itself, it is not the primary focus of this work. Thus, as stated in Assumption 2, K is assumed to be fixed and not varying throughout time. Furthermore, in order for the denominators occurring in the definition of our estimators to be bounded away from zero, there should be at least one node in each community at any given time point. Finally, Assumption 3 states the identifiability conditions for the parameters in model (1). Given the community $\boldsymbol{c}^t$, we observe that model (1) is scale invariant. For example, set constants $C^*_k > 0$ for any $k \in \{1, \ldots, K\}$ and define new parameters $\theta^{t*}_i = \theta^t_i / C^*_{c^t_i}$ and $Z^{t*}_{c^t_i c^t_j} = C^*_{c^t_i} C^*_{c^t_j} Z^t_{c^t_i c^t_j}$, then $(\theta^{t*}_i, Z^{t*}_{c^t_i c^t_j})^\top$ and $(\theta^t_i, Z^t_{c^t_i c^t_j})^\top$ can result in an identical connectivity probability $\tau_{ij}(u^t)$. To address this issue, Assumption 3 is imposed to normalize the degree parameters within each community. Under this constraint, θ^t_i reflects the relative degree propensity of node i within its community.

Throughout this work, we note $a_n \sim b_n$ for any two sequences a_n and b_n if the quantity a_n/b_n tends to a positive constant as $n \to \infty$, and $a_n \asymp b_n$ if both $a_n = O(b_n)$ and $b_n = O(a_n)$ hold simultaneously. Let $n_{\min} = \min_{\substack{a\in\{1,\ldots,K\},\\ t\in\{1,\ldots,T\}}} n_a(\boldsymbol{c}^t)$ be the minimal community size of a dynamic network and $\kappa_n = n_{\min}/n$ describe the level of balance in the community sizes. If the community sizes are balanced for all time points, then $\kappa_n \sim 1$, otherwise $\kappa_n \to 0$ as $n \to \infty$. Further, a smaller κ_n indicates more unbalancedness in the community sizes. Intuitively, this factor impacts directly our result. To evaluate the estimation error we define the scaled mean-squared error (MSE) for $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t)$ as $d\left(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t)\right) := \mathbb{E}\left(\frac{1}{n^2}\|\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t) - \rho_n^{-1}\boldsymbol{\tau}(u^t)\|_F^2\right)$ where $\|\cdot\|_F^2$ denotes the squared Frobenius norm.

To model the behavior of community assignments over time, we introduce a sequence of iid Bernoulli random variables $X^t_i,\ i \in \{1, \ldots, n\},\ t \in \{1, \ldots, T\}$ and a "community stability" parameter $\gamma \in (0, 1]$ defined via

$$\mathbb{E}(X^t_i) = \gamma, \tag{12}$$

which on average controls the community switching rate and implies that $c^t_i = c^{t-1}_i$ if $X^t_i = 1$ and $c^t_i \neq c^{t-1}_i$ otherwise. For our theoretical asymptotic results, we allow this parameter $\gamma := \gamma_T$ to depend on T.

We further treat the community assignments $c_i^1, \ldots, c_i^T$ for each node $i \in \{1, \ldots, n\}$ as fixed but unknown parameters, and the network observations $\boldsymbol{A}^t$ are thus drawn independently for given $\boldsymbol{c}^t$ and $\boldsymbol{\theta}^t$. We next define

$$C_{\sup,1} := \sup_{1\le i,j\le n} \sup_{1\le a,b\le K} \sup_{u\in[0,1]} \left| \frac{\partial^2(\theta_i(u)\theta_j(u)W_{ab}(u))}{\partial u^2} \right|, \quad C_{\sup,2} := \sup_{1\le i,j\le n} \sup_{1\le a,b\le K} \sup_{u\in[0,1]} (\theta_i(u)\theta_j(u)W_{ab}(u)), \tag{13}$$

to be positive constants independent of n and T such that the second derivatives of the function $(\theta_i(u)\theta_j(u)W_{ab}(u))$ and the edge probability curves $\tau_{ij}(u)$ are bounded from above by $C_{\sup,1}$ and $C_{\sup,2}$, respectively, over $i, j \in \{1, \ldots, n\}, a, b \in \{1, \ldots, K\}$ and $u \in [0, 1]$.

Proposition 1 (Pointwise MSE for $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t)$)**.** *Under Assumptions 1-3 and considering γ_T from (12) as a sequence in T, given the true community structure $\boldsymbol{c}^t$ for each time point $t \in \{1, \ldots, T\}$, there exist n_0, T_0 such that for all $n \ge n_0, T \ge T_0$,*

$$d\left(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t)\right) \le C_{\tau,1}h^4 + C_{\tau,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\tau,3}}{n\kappa_n\rho_n Th}\ , \tag{14}$$

with the constant $C_{\tau,1}$ depending on $C_{\sup,1}$, while $C_{\tau,2}$ and $C_{\tau,3}$ depend on $C_{\sup,2}$.

Proof. All details of this proof are given in Appendix A, here we summarize ideas of the essential steps. As presented in (10), the entry $\tilde{\tau}_{ij,h}(u^t, \boldsymbol{e}^t)$ is a function of the smoothed counting variables $\hat{d}_{i,h}^t, \hat{d}_{j,h}^t, \hat{O}_{kl,h}(\boldsymbol{e}^t), \hat{O}_{k,h}(\boldsymbol{e}^t)$ and $\hat{O}_{l,h}(\boldsymbol{e}^t)$ as defined in (8). Lemma A.2 (which itself is based on Lemma A.1 determining the bias and variance of our Step 1-smoother $\hat{\tau}_{ij,h}(u^t)$) provides rates of convergence of these kernel smoothed counting variables to their population quantities $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t), \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{kl}(\boldsymbol{e}^t))$ and $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_k(\boldsymbol{e}^t))$, locally in time point t. We observe that the essential step of the proof is the treatment of a Taylor expansion of $\tilde{\tau}_{ij,h}(u^t)$, given $\boldsymbol{c}^t$, being expressed as a (five-dimensional) function of the smoothed counting variables around its population analog $\tau_{ij}(u^t)$ being expressed as the same function of the corresponding population quantities - see (A.9) in Appendix A. This eventually leads to the desired rates of convergence for the bias, variance and mean-squared error of the scaled $\tilde{\boldsymbol{\tau}}_h(u^t)$. □

Remark 3. *In light of the results of Proposition 1, we next give some insights on the nature of our given assumptions.*

1. *On the scaled mean-squared error: Observe that, since $\tau_{ij}(u^t)$ is proportional to ρ_n, if we considered the non-scaled error of $\tilde{\boldsymbol{\tau}}_h(u^t)$, this quantity would increase as ρ_n grows. This however would contradict our common understanding that smaller estimation errors are typically achieved in denser networks with more information.*

2. *On the bias term: The first term in equation (14) corresponds to the usual smoothing bias caused by the smooth functions $Z_{ab}(u)$ and $\theta_i(u)$ and depends only on the bandwidth h. The second term accounts for the switching bias introduced by community switching in $\boldsymbol{c}^t$ and its magnitude depends on the quantity $Th(1-\gamma_T)$ which is proportional to the expected number of community changes per node within a local window. Intuitively, this quantity is expected to be much smaller than one and to vanish asymptotically, i.e., $\gamma_T \to 1$ such that $Th(1-\gamma_T) \to 0$, where the first requirement, $\gamma_T \to 1$, is imposed to asymptotically maintain the temporal alignability of community assignments. This condition is similar to the existing conditions from the literature. For instance, in Lin and Lei (2024) each node is assumed to switch its community according to an independent Poisson($\tilde{\gamma}$) process. Setting $\tilde{\gamma} = T(1-\gamma_T)$ in their framework has as a direct consequence that their Assumption 5, namely $\tilde{\gamma}/T = o(1)$ is comparable to our first requirement $\gamma_T \to 1$. Moreover, note that the sufficient condition for consistent detection in Lin and Lei (2024) (i.e., for their clustering error bound in Theorem 1 to vanish) involves a decay requirement that matches the latter term in our condition, namely $Th(1-\gamma_T) \to 0$. On the other hand, Pensky and Zhang (2019) posit that at most $s \ll n$ nodes may change communities between two consecutive time points. A direct parallel can be drawn between our work and theirs. The switching rate we consider in our work, namely $1-\gamma_T$, corresponds roughly to the relative frequency s/n from the work of Pensky and Zhang (2019). They further show that the switching bias in the clustering error bound is of order $(n_{\max}/n_{\min})^2 Th(s/n)$, which implies that the requirement for the vanishing clustering error is roughly equivalent to ours when community sizes are balanced; however, for the unbalanced case, their requirement is slightly stronger than ours.*

3. *On node switching: Controlling further the rate with which the term $Th(1-\gamma_T)$ decays, one can also reproduce the same fast bias decay of order h^2 as in usual nonparametric estimation of C^2-curves. In view of Corollary 2 on asymptotically balancing squared bias and variance by the usual AMSE-optimal bandwidth, this rate of bias decay holds, in particular, if the total number of switchings across all nodes, i.e. $nT(1-\gamma_T)$, remains bounded by $Cn^{4/5}(T\rho_n)^{-1/5}$ with some constant C.*

4. *On the variance term: We observe in the upper bound for the MSE of $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t)$ in (14) that the variance term, is inversely proportional to the 'effective local sample size' $n\kappa_n\rho_n Th$, for which we can distinguish two scenarios: If the community sizes are balanced, i.e. if $\kappa_n \sim 1$ since $n_{\min} \asymp n/K$, the variance depends only on the effective local sample size of the network $n\rho_n Th$ needing to diverge. However, a potentially unbalanced situation ($\kappa_n \to 0$) can lead to detectability issues for the probabilities of edges within the smallest community. Hence, it is intuitive that a more stringent assumption is needed, where $n\kappa_n\rho_n Th$ needs to diverge as $\kappa_n, h, \rho_n \to 0$ and $nT \to \infty$. Specifically, the effective local sample size of the smallest community should be large enough to guarantee recovering the connectivity between any pair of communities.*

As previously discussed, one optimal choice for h aims to minimize the mean-sqaured error of $\tilde{\boldsymbol{\tau}}_h(u^t)$ across all time points. Under Proposition 1 we derive the following results for such an optimal bandwidth h_{opt} with respect to two different regimes of community stability γ_T.

Corollary 2 (Bandwidth selection)**.** *Under the conditions of Proposition 1, the optimal bandwidth h_{opt} minimizing the asymptotically leading terms of the global mean-squared error*

$$gMSE := T^{-1}\sum_{t=1}^{T} d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t))$$

has the following rates:

1. *Under low community stability with $T(1-\gamma_T) = O(h^\alpha)$ for some $-1 < \alpha < 1$, then*

$$h_{opt} \sim (n\kappa_n\rho_n T^3(1-\gamma_T)^2)^{-1/3}.$$

2. *Under high community stability with the strengthened condition $T(1-\gamma_T) = O(h)$, then*

$$h_{opt} \sim (n\kappa_n\rho_n T)^{-1/5}.$$

Plugging the rate optimal bandwidth h_{opt} into the upper bounds provided in Proposition 1, leads to a rate of convergence of order $(\frac{1-\gamma_T}{n\kappa_n\rho_n})^{\frac{2}{3}}$ in the low community stability regime and $(n\kappa_n\rho_n T)^{-\frac{4}{5}}$ in the high community stability regime, for both pointwise and global MSE of $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t)$.

Observe that Corollary 2 states the asymptotic behavior of h_{opt} under two distinct regimes, as determined by the leading bias in Proposition 1: the switching bias dominates for low community stability as the nodes are more likely to change their communities, whereas the smoothing bias becomes dominant for high stability regimes. It is intuitive to observe in both cases that the bandwidth h_{opt} will decrease when the number of nodes n or the parameter ρ_n increases, since there is more information in each network as n or ρ_n increase. Recall that the upper bound of the error is inversely proportional to the level of community balance, since there are fewer connections within the minimal community with decreasing κ_n. This means there is more incentive to aggregate information across more networks with increasing h_{opt}. Interestingly, low community stability implies that the bandwidth should shrink at a faster rate to offset the additional bias introduced by community switching.

Remark 4. *In practice, the selection procedure of h_{opt} is challenging because the true value of $\tau_{ij}(u^t)$ is unknown. Inspired by Yang and Jie (2020), we adopt a leave-one-out cross-validation approach to design a data-adaptive bandwidth selection. The procedure is as follows. For any time point $t \in \{1, \ldots, T\}$, let $\mathcal{A}^{-t} = \mathcal{A} \setminus \{\boldsymbol{A}^t\}$ be the training set consisting of other networks surrounding time point t. In Step 1, for each time point $t \in \{1, \ldots, T\}$, we use the training set $\mathcal{A}^{-t}$ as our new observations to obtain the leave-one-out kernel estimator by*

$$\hat{\tau}_{ij,h}(u^t) = \frac{\sum_{t'=1}^{T} A_{ij}^{t'} \mathbf{1}_{\{u^{t'} \in U(u^t) \setminus \{u^t\}\}}}{\left|\{A_{ij}^{t'} : u^{t'} \in U(u^t) \setminus \{u^t\}\}\right|}.$$

Substitute next $\hat{\tau}_{ij,h,-t}(u^t)$ for $\hat{\tau}_{ij,h}(u^t)$ into the time-localised profile-likelihood criterion (11) in Step 2 and compute $\hat{\boldsymbol{c}}_{-t}^t$ by solving the following program over all possible candidates $\boldsymbol{e}^t$,

$$\hat{\boldsymbol{c}}_{-t}^t = \arg\max_{\boldsymbol{e}^t} \sum_{k,l=1}^{K} \left(\hat{O}_{kl,h,-t}(\boldsymbol{e}^t) \log \frac{\hat{O}_{kl,h,-t}(\boldsymbol{e}^t)}{\hat{O}_{k,h,-t}(\boldsymbol{e}^t)\hat{O}_{l,h,-t}(\boldsymbol{e}^t)} - \hat{O}_{kl,h,-t}(\boldsymbol{e}^t) \right)$$

where $\hat{O}_{kl,h,-t}(\boldsymbol{e}^t)$ and $\hat{O}_{k,h,-t}(\boldsymbol{e}^t)$ are obtained by replacing $\hat{\tau}_{ij,h}(u^t)$ with $\hat{\tau}_{ij,h,-t}(u^t)$ in $\hat{O}_{kl,h}(\boldsymbol{e}^t)$, and $\hat{O}_{k,h}(\boldsymbol{e}^t)$, respectively. Following the definition in (10), the leave-one-out estimator $\tilde{\tau}_{ij,h,-t}(u^t, \hat{\boldsymbol{c}}_{-t}^t)$ is constructed based on $\hat{\boldsymbol{c}}_{-t}^t$ as

$$\tilde{\tau}_{ij,h,-t}(u^t, \hat{\boldsymbol{c}}_{-t}^t) = \frac{\hat{d}_{i,h,-t}^t \hat{d}_{j,h,-t}^t \hat{O}_{\hat{c}_{i,-t}^t \hat{c}_{j,-t}^t,h,-t}(\hat{\boldsymbol{c}}_{-t}^t)}{\hat{O}_{\hat{c}_{i,-t}^t,h,-t}(\hat{\boldsymbol{c}}_{-t}^t)\hat{O}_{\hat{c}_{j,-t}^t,h,-t}(\hat{\boldsymbol{c}}_{-t}^t)}, \tag{15}$$

where $\hat{d}_{i,h,-t}^t$ is obtained similarly by replacing $\hat{\tau}_{ij,h}(u^t)$ with $\hat{\tau}_{ij,h,-t}(u^t)$ in $\hat{d}_{i,h}^t$. As we discussed in (6), the observed network $\boldsymbol{A}^t$ can be interpreted as a perturbation of the population $\boldsymbol{\tau}(u^t)$. Intuitively, $\tilde{\boldsymbol{\tau}}_{h,-t}(u^t, \hat{\boldsymbol{c}}_{-t}^t) := (\tilde{\tau}_{ij,h,-t}(u^t, \hat{\boldsymbol{c}}_{-t}^t))_{n\times n}$ can be considered as a prediction for $\boldsymbol{A}^t$. We then derive the cross-validation value for each time point $t \in \{1, \ldots, T\}$ as $CV_t(h) = \|\tilde{\boldsymbol{\tau}}_{h,-t}(u^t, \hat{\boldsymbol{c}}_{-t}^t) - \boldsymbol{A}^t\|_F^2$ via the squared Frobenius norm. Finally, the optimal bandwidth is chosen by minimizing the average $CV_t(h)$ across all time points, that is

$$h_{loocv} = \arg\min_{h\in(0,1]} \frac{1}{T} \sum_{t=1}^{T} CV_t(h).$$

Härdle, Hall, and Marron (1988) suggested that, under general conditions in nonparametric regression, the bandwidth selected through the leave-one-out cross-validation criterion consistently estimates the asymptotically optimal bandwidth that minimizes the global mean squared error. In our setting, the elementwise smoothing of the independently observed adjacency matrices over time satisfies the required regularity conditions in the case of high community stability. Therefore, h_{loocv} converges to h_{opt} in probability as well.

4 Consistency of kernel smooth estimators

Recall that in Step 2, the smooth estimator of θ_i^t for any $i \in \{1, \ldots, n\}$ is given by (9), while the smoothed profile-likelihood criterion $Q(\hat{O}_h(\boldsymbol{e}^t))$ is provided in (11). We now investigate theoretical properties for the smooth estimators in the time-varying DCSBM. As the sum of edges in a network is proportional to $n^2\rho_n$ for all time points under the asymptotic regime (with $\rho_n \to 0$ as $n \to \infty$), we consider the scaled time-localised criterion as $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$. Then the estimated community structure for any $t \in \{1, \ldots, T\}$ is obtained as,

$$\hat{\boldsymbol{c}}^t = \arg\max_{\boldsymbol{e}^t\in\{1,\ldots,K\}^n} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right). \tag{16}$$

We first investigate the consistency in estimating the node degree parameters. With slight abuse of the same notation $d(\cdot,\cdot)$, we define the scaled mean-squared error for $\hat{\boldsymbol{\theta}}_h^t$ as $d(\hat{\boldsymbol{\theta}}_h^t, \boldsymbol{\theta}^t) := \mathbb{E}\left(\frac{1}{n}\|\hat{\boldsymbol{\theta}}_h^t - \boldsymbol{\theta}^t\|_2^2\right)$ where $\|\cdot\|_2^2$ denotes the squared Euclidean norm.

Theorem 3 (Consistency of degree parameters)**.** *Under Assumptions 1-3, consider the true community structure $\boldsymbol{c}^t$ for each time point $t \in \{1, \ldots, T\}$ to be given. Let $\hat{\boldsymbol{\theta}}_h^t = \left(\hat{\theta}_{1,h}^t, \ldots, \hat{\theta}_{n,h}^t\right)^\top$ be the smooth estimators of the degree parameters where $\hat{\theta}_{i,h}^t$ is defined in (9). Provided that $\gamma_T \to 1$, $\rho_n, h \to 0$ as $n, T \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$, the estimator $\hat{\boldsymbol{\theta}}_h^t$ is mean-squared error consistent for $\boldsymbol{\theta}^t$ at any time point $t \in \{1, \ldots, T\}$, and satisfies*

$$d(\hat{\boldsymbol{\theta}}_h^t, \boldsymbol{\theta}^t) \le C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th}\ , \tag{17}$$

with the constant $C_{\theta,1}$ depending on $C_{\sup,1}$, while $C_{\theta,2}$ and $C_{\theta,3}$ depend on $C_{\sup,2}$.

Proof. The main technical component of this proof is similar to the proof of Proposition 1 following the idea that $\hat{\theta}_{i,h}^t$, given $\boldsymbol{c}^t$, can be seen as a (two-dimensional) function of $\frac{\hat{d}_{i,h}^t}{n\rho_n}$ and $\frac{\hat{O}_{c_i^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}$. Again, Taylor expansion of this function around its population limit delivers the aforementioned rates of convergence via those of Lemma A.2. For details, we refer to Appendix B. □

Theorem 3 provides an upper bound for the estimation of node degree parameters at time point t. Recall that θ_i^t measures the popularity of node i, assumed to be independent of the community connectivity parameter in this context. Therefore, for consistency of $\hat{\boldsymbol{\theta}}_h^t$ in Theorem 3, note that no additional assumptions of balancedness in community sizes are necessary (unlike for Proposition 1 where the bound for $\tilde{\boldsymbol{\tau}}_h(u^t)$ relies on κ_n). Additionally, we will later show that these bounds in Proposition 1 and Theorem 3 remain valid when replacing $\boldsymbol{c}^t$ by the estimator $\hat{\boldsymbol{c}}^t$ provided that $\hat{\boldsymbol{c}}^t$ converges to $\boldsymbol{c}^t$ almost surely.

Remark 5. *Most of the existing work on dynamic networks has focused on recovering community structures rather than on node-specific degree propensities, either in terms of estimating the trajectories of $\boldsymbol{\theta}^t$, or providing rigorous convergence rates for the corresponding estimators. Spectral-based procedures can in principle accommodate degree heterogeneity via normalization approaches; however, the heterogeneity is treated as a 'nuisance' component and is hidden in the spectral embedding, which makes it difficult to track the evolution of node-specific propensities over time. Even when plug-in estimators of $\boldsymbol{\theta}^t$ are considered, for example, layerwise maximum profile likelihood estimators as in Agterberg et al. (2025), the resulting error bounds do not benefit from the multilayer or temporal structure and essentially coincide with the static single network rate of order $(n\rho_n)^{-1/2}$ (up to logarithmic factors). In our work, we instead develop a smooth estimator for $\boldsymbol{\theta}^t$ that combines information from neighboring networks with the likelihood structure, thereby achieving a sharper variance rate.*

Remark 6. *In practice, it could be interesting to assign a bandwidth to each individual edge in the dynamic network. However, implementing such an approach would incur significant computational costs. Furthermore, the rate as stated in Theorem 3 would become difficult to derive, as it is not straightforward how to associate the $n\times n$ elements of the bandwidth matrix to the components of the n-dimensional vector $\hat{\boldsymbol{\theta}}_h^t$. Given the above considerations, we prefer to adopt a common bandwidth throughout this study.*

We now investigate the consistency of the estimated community structure $\hat{\boldsymbol{c}}^t$. To simplify the discussion, we introduce some extra notation. Considering the true assignment $\boldsymbol{c}^t$ and any candidate assignment $\boldsymbol{e}^t$, define $\mathcal{N}_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) = \{1 \le i \le n : c_i^t = a, e_i^t = k\}$ to be the index set of nodes whose true community is a and candidate community is k. Further, let $n_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) = |\mathcal{N}_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)|$ be the size of this set and $\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) = \sum_{i=1}^n \theta_i^t \mathbf{1}_{\{c_i^t=a,e_i^t=k\}}$ be the degree-corrected weights with respect to $\mathcal{N}_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)$ which satisfy that $\sum_{k=1}^K \Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) = n_a(\boldsymbol{c}^t)$. We define next $H_{kl}(\boldsymbol{e}^t)$ as the population version of the scaled $O_{kl}(\boldsymbol{e}^t)$ where

$$H_{kl}(\boldsymbol{e}^t) := \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{O_{kl}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \sum_{a,b=1}^K W_{ab}^t \times \frac{\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n} \times \frac{\Lambda_{b(l)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n}, \tag{18}$$

$$H_k(\boldsymbol{e}^t) := \sum_{l=1}^K H_{kl}(\boldsymbol{e}^t) = \sum_{a,b=1}^K \frac{n_b(\boldsymbol{c}^t)W_{ab}^t}{n} \times \frac{\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n}.$$

Let $H(\boldsymbol{e}^t)$ denote the $K\times K$ matrix with entries $H_{kl}(\boldsymbol{e}^t)$. Similarly to Zhao et al. (2012) and Sengupta and Chen (2018), we also define the population version of the criterion as

$$Q\left(H(\boldsymbol{e}^t)\right) := \sum_{k,l=1}^K \left(H_{kl}(\boldsymbol{e}^t)\log\frac{H_{kl}(\boldsymbol{e}^t)}{H_k(\boldsymbol{e}^t)H_l(\boldsymbol{e}^t)} - H_{kl}(\boldsymbol{e}^t)\right).$$

Since we use the time-localised criterion to approximate the population version in this work, our first strategy to prove the consistency of $\hat{\boldsymbol{c}}^t$ is to measure the disparity between these two terms.

Lemma 4 (Uniform convergence of $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$)**.** *Under Assumptions 1-3, if $\gamma_T \to 1$, $\rho_n, h \to 0$ with both n and T diverging such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$, then over $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, the* *scaled time-localised criterion* $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$ *is uniformly close to the population version* $Q\left(H(\boldsymbol{e}^t)\right)$ *for any time point* $t \in \{1,\dots,T\}$*, that is, there exists* $\epsilon_n^t \to 0$ *such that*

$$\mathbb{P}\left(\max_{\boldsymbol{e}^t}\left|Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(H(\boldsymbol{e}^t)\right)\right| < \epsilon_n^t\right) \to 1, \quad \text{as} \quad \gamma_T \to 1,\ \rho_n, h \to 0,\ n\rho_n Th \to \infty, Th(1-\gamma_T) \to 0. \tag{19}$$

Proof. For the proof we refer to Appendix B. □

Next, we quantify the error rate for community detection by using the Hamming distance,

$$\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) := \min_{\pi(\boldsymbol{e}^t)} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{e_i^t \neq c_i^t\}}, \tag{20}$$

where $\pi(\boldsymbol{e}^t)$ denotes the set of all permutations of the assignment vector $\boldsymbol{e}^t$. For convenience, we refer to $\iota(\boldsymbol{e}^t, \boldsymbol{c}^t)$ as the Hamming error rate (HER) at time point t. To show that the estimator $\hat{\boldsymbol{c}}^t$ converges to its true value $\boldsymbol{c}^t$, a natural condition would be the uniqueness of $\boldsymbol{c}^t$, alongside the maximization of $Q\left(H(\boldsymbol{e}^t)\right)$ achieved by $\boldsymbol{c}^t$. Hence, we give the following assumption to guarantee that a unique maximizer of $Q\left(H(\boldsymbol{e}^t)\right)$ exists.

Assumption 4 (Detectability of community structures)**.** *For each time point* $t \in \{1,\dots,T\}$*, we assume that the community connectivity matrix* $\boldsymbol{W}^t$ *does not have two identical columns.*

Lemma 5 (Uniqueness of the maximizer of $Q\left(H(\boldsymbol{e}^t)\right)$)**.** *Under Assumptions 2-4,* $Q\left(H(\boldsymbol{e}^t)\right)$ *is uniquely maximized at the true labels* $\mathbf{c}^t$ *(up to label permutation), that is, for any generic vector* $\mathbf{e}^t \in \{1,\dots,K\}^n$*,*

$$Q\left(H(\boldsymbol{e}^t)\right) \leq Q\left(H(\boldsymbol{c}^t)\right)$$

where equality holds if and only if $\boldsymbol{e}^t = \boldsymbol{c}^t$*. Furthermore, there exist some constants* $0 < M' \leq M$ *such that*

$$M'\theta_{\inf}\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \leq Q\left(H(\boldsymbol{c}^t)\right) - Q\left(H(\boldsymbol{e}^t)\right) \leq M\theta_{\sup}\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \tag{21}$$

with $\theta_{\inf} := \inf_{1\leq i\leq n}\inf_{u\in[0,1]}\theta_i(u)$ *and* $\theta_{\sup} := \sup_{1\leq i\leq n}\sup_{u\in[0,1]}\theta_i(u)$*.*

Proof. For the proof we refer again to Appendix B. □

This Lemma also gives the lower and upper bounds for the population optimality gap for incorrect community assignments. Based on Lemmas 4 and 5, we have the following theorem.

Theorem 6 (Consistency of community detection)**.** *Under Assumptions 1-4, if* $\gamma_T \to 1,\ \rho_n, h \to 0$ *such that* $n\rho_n Th \to \infty$ *and* $Th(1-\gamma_T) \to 0$ *as both* n *and* T *diverge, the community estimators, as defined in (16), are weakly consistent for any time point* $t \in \{1,\dots,T\}$*, that is*

$$\forall\delta > 0, \qquad \mathbb{P}\left(\iota(\hat{\boldsymbol{c}}^t, \boldsymbol{c}^t) < \delta\right) \to 1.$$

Further, if both n *and* T *diverge such that* $(n\rho_n Th)/\log n \to \infty$*, then the community estimators are strongly consistent, that is*

$$\mathbb{P}\left(\iota(\hat{\boldsymbol{c}}^t, \boldsymbol{c}^t) = 0\right) \to 1.$$

Most previous work has shown that for the static (degree-corrected) stochastic block model the weak consistent community estimation is possible if $n\rho_n \to \infty$ and the strong consistency is achievable if $(n\rho_n)/\log n \to \infty$. In the time-varying setting, our time-localized profile-likelihood approach is weakly consistent at each time point as long as $n\rho_n Th \to \infty$ when $n, T \to \infty$. Furthermore, when $n\rho_n Th$ diverges faster than $\log n$, we show that the algorithm is strongly consistent in recovering the communities at each time point. We also observe that pooling information across networks in time is reflected in the rate condition $n\rho_n Th \to \infty$ of our Theorem 6. In other words, Theorem 6 shows that, under smoothness and stability (switching) assumptions, our time-localized likelihood approach can linearly aggregate information within a local window, effectively boosting the signal from $n\rho_n$ to $n\rho_n Th$ which in turn enhances the performance of community recovery. The above results also apply to time-varying SBM as well, by setting all $\theta_i^t = 1$.

Another noteworthy aspect is that we do not impose any restrictions on the baseline average node degree λ_n, allowing our method to be applicable across all regimes of $\rho_n \to 0$ (e.g., sparse, semi-sparse, or semi-dense networks as mentioned in Remark 1). In particular, for semi-sparse or highly sparse networks - such as the regime of $\lambda_n = O(1)$ in Lin and Lei (2024) - our results in Theorem 6 still guarantee a strongly consistent community estimation requiring that T diverges such that $(Th)/\log n \to \infty$ and $\gamma_T = 1 - o\left((\log n)^{-1}\right)$. However, in the semi-dense case when $\lambda_n \gg \log n$, there is no need to pool information across networks since each single network contains sufficient information. This observation is fully supported in the numerical simulation in Section 5 where criterion (11) prevails in sparse and semi-sparse networks with $\lambda_n < \log n$.

Recall from (9) and (10) that $\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t) = \left(\hat{\theta}_{i,h}^t(\boldsymbol{e}^t), \ldots, \hat{\theta}_{n,h}^t(\boldsymbol{e}^t)\right)^\top$ and $\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{e}^t) = (\tilde{\tau}_{ij,h}(u^t, \boldsymbol{e}^t))_{n\times n}$. Based on the consistency results of $\hat{\boldsymbol{c}}^t$ from Theorem 6, the following corollary states that the bounds in Proposition 1 and Theorem 3 continue to hold for the estimators $\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)$ and $\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$ with the community estimate $\hat{\boldsymbol{c}}^t$.

Corollary 7. *Suppose that Assumptions 1 - 4 hold. If $\gamma_T \to 1$, $\rho_n, h \to 0$ such that $(n\rho_n Th)/\log n \to \infty$ and $Th(1-\gamma_T) \to 0$ as both n and T diverge, then the following bounds hold.*

$$
\begin{aligned}
d(\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t), \boldsymbol{\theta}^t) &\le C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th} \ , \\
d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t), \rho_n^{-1}\boldsymbol{\tau}(u^t)) &\le C_{\tau,1}h^4 + C_{\tau,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\tau,3}}{n\kappa_n\rho_n Th}.
\end{aligned}
$$

where the constants $C_{\theta,1}$ and $C_{\tau,1}$ depend on $C_{\sup,1}$ while the constants $C_{\theta,2}, C_{\tau,2}, C_{\theta,3}, C_{\tau,3}$ depend on $C_{\sup,2}$.

5 Numerical simulations

In this section, we explore the performance of our method constructed in Section 3 for recovering the community structures and the degree heterogeneity across time. In Step 2, we use the greedy label-switching algorithm (see, e.g. Glover and Laguna 1997; Zhao et al. 2012) to maximize the smoothed profile-likelihood criterion in (11). Further, we use the data-adaptive procedure based on the leave-one-out cross-validation method to select the optimal bandwidth mentioned in Remark 4.

Time-varying network-generating process. We consider $T = 100$ networks and assign n nodes to $K = 2$ communities. Assume that the community label c_i^t of each node follows a Markov chain with initial probabilities ($\mathbb{P}(c_i^1 = 1) = \mathbb{P}(c_i^1 = 2) = 1/2$) and a state transition matrix $\Gamma = \begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix}$. Roughly speaking, each node will steadily alter the community assignment over time with probability $(1-\gamma)$. Let next $\boldsymbol{Z}^0 = \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.3 \end{pmatrix}$ denote the baseline community connectivity matrix. To introduce temporal variation, we add smooth periodic perturbations to the diagonal and off-diagonal elements via the functions

$$
f_{\text{in}}(x) = 0.1\sin(2\pi x - 0.1\pi), \qquad f_{\text{out}}(x) = -0.1\sin(2\pi x).
$$

Using these functions, the time-varying connectivity matrix at time point t is defined as $Z_{kl}^t = \rho_n(Z_{kl}^0 + f_{\text{in}}(u^t)\mathbf{1}_{\{k=l\}} + f_{\text{out}}(u^t)\mathbf{1}_{\{k\neq l\}})$. For each community, half of the nodes are randomly chosen to have at a given time point t the same node degree parameter θ_i^t given by the function $f_{\boldsymbol{\theta}}(u^t)$ defined below, and the other half of the nodes are chosen accordingly to satisfy the identifiability constraint in Assumption 3 at each time point. Finally, given all parameters $\boldsymbol{c}^t, \boldsymbol{\theta}^t$ and $\boldsymbol{Z}^t$, the edges A_{ij}^t are generated according to model (1). Focusing on the effect of different types of degree heterogeneity, we consider the following scenarios:

- Scenario 1 (SBM): Fix $f_{\boldsymbol{\theta}}(u^t) = 1$ for all $t \in \{1, \ldots, T\}$, as such, the model reduces to a time-varying SBM in which all nodes within the same community are considered equivalent and having the same degrees.
- Scenario 2 (DCSBM with smooth trajectories): Set $f_{\boldsymbol{\theta}}(u^t) = 1.6 + 0.1\sin(2\pi u^t)$ for all $t \in \{1, \ldots, T\}$ where a sinusoidal curve is used to model the trajectories of $\boldsymbol{\theta}^t$; in particular, $f_{\boldsymbol{\theta}}(u^t) \in C^2[0,1]$ and thus Scenario 2 satisfies Assumption 1.

- Scenario 3 (DCSBM with non-smooth trajectories): Set $f_{\boldsymbol{\theta}}(u^t) = 1.6 \times \mathbf{1}_{\{\text{t is odd }\}} + 0.8 \times \mathbf{1}_{\{\text{t is even}\}}$, so that each node alternates between high- and low- degree configurations over time.

For each scenario, we allow varying the network size $n \in \{50, 100\}$, the sparsity parameter $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}$ and the community stability parameter $\gamma \in \{1, 0.95\}$. Beyond these scenarios, we also explored the robustness of our approach to unbalanced community sizes, the number of communities K and to disassortative mixing. Additional numerical results are provided in Appendix C.

Methods. We compare our proposed time-varying DCSBM (TVDCSBM) with several community detection methods for dynamic networks. One natural baseline applies the static DCSBM (SDCSBM) separately to each adjacency matrix $\boldsymbol{A}^t$, relying solely on the single network for estimation. We also include the likelihood-based dynamic SBM (DSBM) of Matias and Miele (2017) which uses Markov chains to recover the temporal evolution of node communities. In addition, several spectral-based algorithms are considered as benchmark, including the kernel debiased sum-of-squares (KD-SoS) method of Lin and Lei (2024) and its degree-corrected variant (KD-DC-SoS) in which the normalized spectral clustering is used to account for degree heterogeneity, and the degree-corrected multiple adjacency spectral embedding approach (DC-MASE) of Agterberg et al. (2025). To evaluate the detection accuracy, we consider the HER, defined in (20), as the measure of the agreement between the estimated and the true communities. This index ranges from 0 to 1, with lower values indicating better agreement. Given the substantive interest of the node-specific degree heterogeneity in applications, we also evaluate the estimation accuracy of $\boldsymbol{\theta}^t$ for competitors that allow for degree heterogeneity, namely, SDCSBM, TVDCSBM, DC-MASE and KD-DC-SoS. Notably, KD-DC-SoS does not explicitly estimate $\boldsymbol{\theta}^t$. Nevertheless, one can still create pseudo-estimates of $\boldsymbol{\theta}^t$ for KD-DC-SoS by plugging its community estimates into formula (3), in the same spirit as formula (5.1) in Agterberg et al. (2025). Although heuristic and possibly suboptimal, this construction enables an empirical assessment of how the community estimates of KD-DC-SoS facilitate the recovery of the degree parameters. For all methods, the estimation error for $\boldsymbol{\theta}^t$ is assessed by the distance $d(\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t), \boldsymbol{\theta}^t)$, where $\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t)$ generally denotes the degree parameter estimates obtained from different methods. In our dynamic framework, the average value is computed over all time points to compare overall performance across methods, and the figures in this section all present the mean results over 50 replications.

Results. The community detection results are shown in Figure 3. As expected, the performance of most methods improves as n or ρ_n increases, reflecting the fact that denser and larger networks provide more information for community recovery. In the weak signal regime with small n or ρ_n, our method achieves substantially lower HER than the competing approaches. This result agrees with the detection threshold established in Theorem 6: for TVDCSBM, a sufficient condition for consistent detection is $n\rho_n Th \gg \log n$, which is less stringent than the requirements imposed for the competing methods. For example, the static SDCSBM needs $n\rho_n \gg \log n$ and KD-SoS of Lin and Lei (2024) needs $n\rho_n\sqrt{Th} \gg \sqrt{\log(Th+n)}$. Thus, TVDCSBM is expected to have a good performance, and aggregation across networks is essential in such regimes. In contrast, in large networks the advantage of TVDCSBM is dissipating with growing ρ_n, since there is sufficient information within each network to accurately recover communities in such scenarios, lessening the necessity of aggregation across networks.

Comparing the three scenarios, it is obvious that DSBM is the most sensitive method to the effect of $\boldsymbol{\theta}^t$. KD-DC-SoS incorporates a regularization step to mitigate node heterogeneity and performs competitively in Scenario 2 which uses the smooth trajectories of $\boldsymbol{\theta}^t$, however, it struggles to accommodate the more complex configurations, such as Scenario 3 which uses non-smooth trajectories. There, the strong violation of the smoothness assumption makes aggregation across networks introduce too much noise which in turn distorts the leading eigenvectors. TVDCSBM also shows a mild degradation in Scenario 3, as expected under this form of misspecification, but its performance is overall superior. In contrast, SDCSBM and DC-MASE show a relatively stable performance across the three scenarios as they do not rely on the smoothness assumption. In addition, DC-MASE posits a time-invariant community structure, which limits its flexibility in applications where communities evolve over time. This limitation is particularly apparent when $\gamma = 0.95$, where as a non-negligible fraction of nodes changes communities between consecutive time points, the performance of DC-MASE no longer improves even when ρ_n and n are large enough.

Table 1 summarizes the estimation accuracy of the degree parameters $\boldsymbol{\theta}^t$ across the three scenarios. The results on estimating the edge probability matrices $\boldsymbol{\tau}(u^t)$ are reported in Table 3 of Appendix C. As a benchmark, we also consider an estimator $\hat{\boldsymbol{\theta}}^t_h(\boldsymbol{c}^t)$ computed with the true communities $\boldsymbol{c}^t$. In both scenarios 1 and 2, the proposed estimator $\hat{\boldsymbol{\theta}}^t_h(\hat{\boldsymbol{c}}^t)$ consistently achieves the lowest estimation errors, nearly matching the oracle performance of $\hat{\boldsymbol{\theta}}^t_h(\boldsymbol{c}^t)$, and this improvement can be attributed to the smoothing step

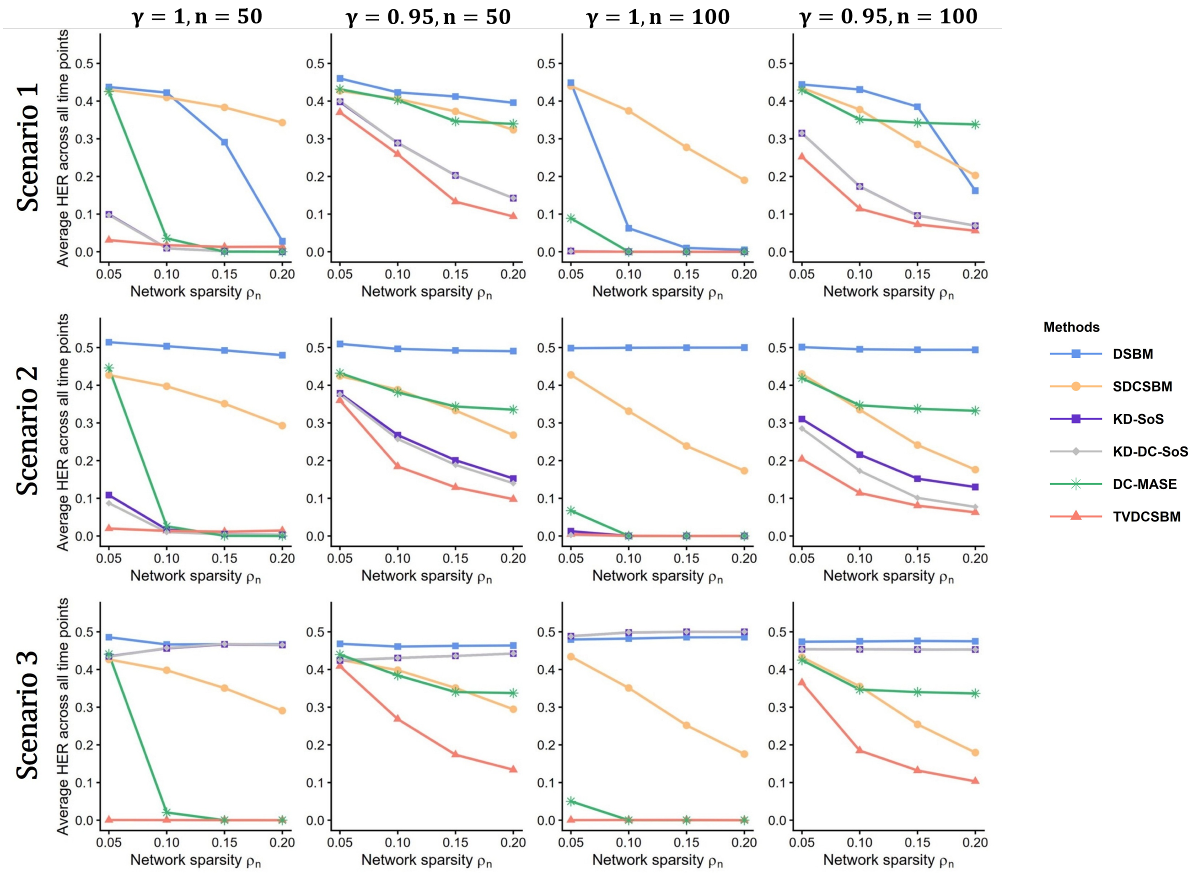


Figure 3: Average Hamming error rates (HER) of different methods with varying $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}, n \in \{50, 100\}$ and $\gamma \in \{1, 0.95\}$. Here $\gamma = 1$ corresponds to networks sharing a common community structure across all time points (while the degree parameters remain time-varying), whereas $\gamma = 0.95$ allows all parameters to vary over time.

which effectively reduces the estimation variance when the true parameters evolve smoothly over time, whereas the spectral-based or static approaches do not take advantage of this smoothness and therefore fail to achieve comparable accuracy in estimating $\boldsymbol{\theta}^t$. This observation provides empirical support for the result of Theorem 3. Furthermore, when the smoothness assumption does not hold, as in Scenario 3, the smoothing step introduces a substantial bias which in turn leads to a larger estimation error for $\boldsymbol{\theta}^t$. Importantly, however, this does not significantly impair the accuracy for community detection. This robustness stems from the explicit modelling of $\boldsymbol{\theta}^t$ in our framework, which effectively decouples the degree variation from the latent community structure and mitigates the impact of model misspecification on clustering performance.

6 Applications

We analyze two types of real-world networks. In the first dataset, the 'Workplace' network shares a common community structure across all time points but with time-varying node degrees. The second application based on the 'Travian' dataset, features evolving community structures and node degrees over time. In all applications, we used our proposed TVDCSBM for community detection, where the bandwidth is selected using the data-adaptive procedure as described in Remark 4, and compared it with the same benchmarks employed in Section 5. The performance of the methods is assessed by the Hamming error rate.

Table 1: Average estimation errors $d(\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t), \boldsymbol{\theta}^t)$ with $n = 100, T = 100$ and $K = 2$. All values, **expressed in units of $\mathbf{10^{-2}}$**, represent means with standard deviations in parentheses across 50 replications. Boldface indicates the smallest value among all competing methods, excluding the unattainable oracle $\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)$ with known true $\boldsymbol{c}^t$.

Method	$\gamma = 1$				$\gamma = 0.95$			
	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$
Scenario 1: time-varying SBM with homogeneous degrees								
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	0.26 (0.07)	0.16 (0.04)	0.12 (0.03)	0.09 (0.02)	0.65 (0.17)	0.62 (0.09)	0.54 (0.06)	0.45 (0.07)
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**0.26** (0.07)	**0.16** (0.04)	**0.12** (0.03)	**0.09** (0.02)	**0.65** (0.18)	**0.62** (0.09)	**0.54** (0.06)	**0.45** (0.07)
SDCSBM	18.45 (0.26)	8.68 (0.11)	5.42 (0.07)	3.79 (0.05)	18.45 (0.26)	8.70 (0.13)	5.45 (0.08)	3.82 (0.05)
DC-MASE	18.44 (0.27)	8.65 (0.12)	5.41 (0.07)	3.78 (0.05)	18.48 (0.27)	8.70 (0.13)	5.47 (0.08)	3.86 (0.06)
KD-DC-SoS	18.42 (0.26)	8.65 (0.12)	5.41 (0.07)	3.78 (0.05)	18.41 (0.27)	8.63 (0.13)	5.41 (0.07)	3.79 (0.05)
Scenario 2: time-varying DCSBM with smooth trajectories								
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	0.51 (0.05)	0.27 (0.02)	0.19 (0.02)	0.14 (0.01)	3.65 (0.08)	2.62 (0.03)	2.12 (0.03)	1.78 (0.02)
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**0.51** (0.05)	**0.27** (0.02)	**0.19** (0.03)	**0.15** (0.01)	**4.02** (0.09)	**2.91** (0.07)	**2.29** (0.07)	**1.94** (0.08)
SDCSBM	18.29 (0.44)	8.31 (0.14)	4.88 (0.10)	3.19 (0.10)	18.36 (0.49)	8.38 (0.17)	4.93 (0.11)	3.19 (0.07)
DC-MASE	17.52 (0.34)	7.65 (0.08)	4.43 (0.07)	2.85 (0.04)	18.31 (1.58)	8.15 (0.25)	4.97 (0.24)	3.35 (0.24)
KD-DC-SoS	17.34 (0.31)	7.65 (0.08)	4.43 (0.07)	2.85 (0.04)	17.74 (0.29)	7.95 (0.12)	4.65 (0.08)	3.01 (0.05)
Scenario 3: time-varying DCSBM with non-smooth trajectories								
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	15.99 (0.05)	15.75 (0.07)	15.60 (0.11)	15.47 (0.07)	15.51 (0.12)	14.40 (0.28)	13.44 (0.03)	13.26 (0.02)
TVDCSBM ($\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**16** (0.04)	15.77 (0.08)	15.6 (0.12)	15.46 (0.1)	**18.16** (0.81)	15.9 (0.69)	14.82 (0.75)	14.18 (0.54)
SDCSBM	18.40 (0.27)	8.50 (0.16)	5.14 (0.08)	3.47 (0.08)	18.39 (0.35)	8.54 (0.13)	**5.18** (0.12)	**3.50** (0.08)
DC-MASE	17.91 (0.24)	**8.12** (0.12)	**4.89** (0.06)	**3.28** (0.05)	18.39 (0.48)	**8.40** (0.18)	5.21 (0.2)	3.57 (0.11)
KD-DC-SoS	43.70 (2.17)	31.35 (0.30)	27.19 (0.11)	25.09 (0.09)	40.46 (0.54)	28.58 (0.27)	24.67 (0.20)	22.73 (0.14)

Table 2: Summary statistics of node degrees across the $T = 10$ time points in the Workplace dataset with $n = 92$ participants partitioned into $K = 5$ communities.

Time	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$	$t = 10$
Mean	4.09	3.30	2.67	4.04	2.24	3.20	3.28	3.48	3.43	2.04
Min	0	0	0	0	0	0	0	0	0	0
Max	15	12	12	14	11	11	12	19	12	10

6.1 Workplace dataset analysis

The first example is a dynamic network of the physical proximity between staff in an office building in France observed for a period of ten days[1]. During this period, 92 participants, distributed across 5 known departments, wore electronic proximity detectors to record the face-to-face interactions at intervals of 20 seconds. We construct one snapshot for each day in which an edge was considered to exist between two participants if there was at least one recorded contact between them on that particular day. In conclusion, we formed a time-varying network comprising 10 adjacency matrices each with 92 nodes. Table 2 reports summary statistics of node degrees. As observed, the maximum degree of the nodes is 19, while the minimum degree is 0 across all time points. In this section, a node with zero degree stands for an 'isolated node' which may be interpreted as a situation in which the corresponding participant is not interacting with the other participants at that moment. However, each selected node must have at least one edge throughout the entire period (i.e. one edge for at least one moment in time) to satisfy the identifiability condition of node communities.

The departments of the 92 nodes are considered as the ground truth for communities with the community assignment of each node remaining constant over time. Another notable observation is that the largest group among these communities comprises 34 nodes, whereas the smallest group has only 4 nodes. Thus, this example also highlights the feature of unbalanced community sizes in the dynamic network. In this application, the optimal bandwidth is determined to be $h = 0.5$. Comparing the partitions found by the three methods with the ground truth, we observe that our approach TVDCSBM performs overall the best with an HER of 0.16, followed by KD-DC-SoS (0.23) and DC-MASE (0.24). The DSBM and static DCSBM methods perform significantly worse, with HERs of 0.36 and 0.41, respectively. As mentioned in Section 2.3, it is a distinct challenge to recover the community structure in sparse networks based on a single network as one is lacking sufficient information. On the contrary, our procedure benefits from smoothing the networks over time, even in cases of unbalanced community sizes.

6.2 Travian game dataset analysis

Travian is a popular browser-based real-time strategy game[2] where players will form alliances (communities) to expand their territory. These alliances are publicly available and can serve as the ground truth for evaluating community detection algorithms. Communication, collaboration, and competition among players may lead to the reorganization or adjustment of the communities. These variations also affect the connections within and between communities.

This dataset collected the message records of 2853 players from Germany spanning 30 days. During this period, there were additions of new players and departures of old players. Considering the stability of the network (where we restrict all participants of our dynamic network to always be present across all time points) and the impact of the community size on community identification, we retained the nodes from those communities that had more than 40 players at all time points to form a subnetwork, upon which further analyses were performed. Consequently, the networks in this study involve a total of 411 nodes from 9 communities corresponding to their alliances. Similarly, we also construct one snapshot for each day in which an edge was considered to exist between two players if there was at least one recorded message between them on that particular day. For simplicity, self-loops are not considered in this case. Finally, a dynamic network is formed by 30 adjacency matrices. Each matrix represents a network corresponding to 411 players at specific time points, with roughly balanced community sizes ranging from 40 to 51 nodes in each community.

[1] http://www.sociopatterns.org/datasets/contacts-in-a-workplace

[2] http://ial.eecs.ucf.edu/travian.php

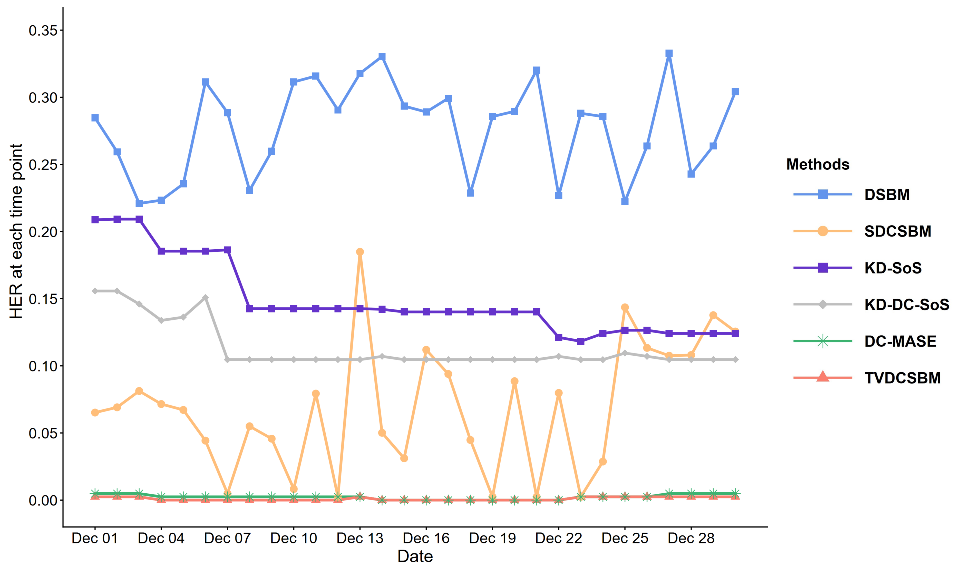


Figure 4: Hamming error rates (HER) at each time point for different methods in the Travian dataset. Lower values indicate more accurate community estimation relative to the ground truth.

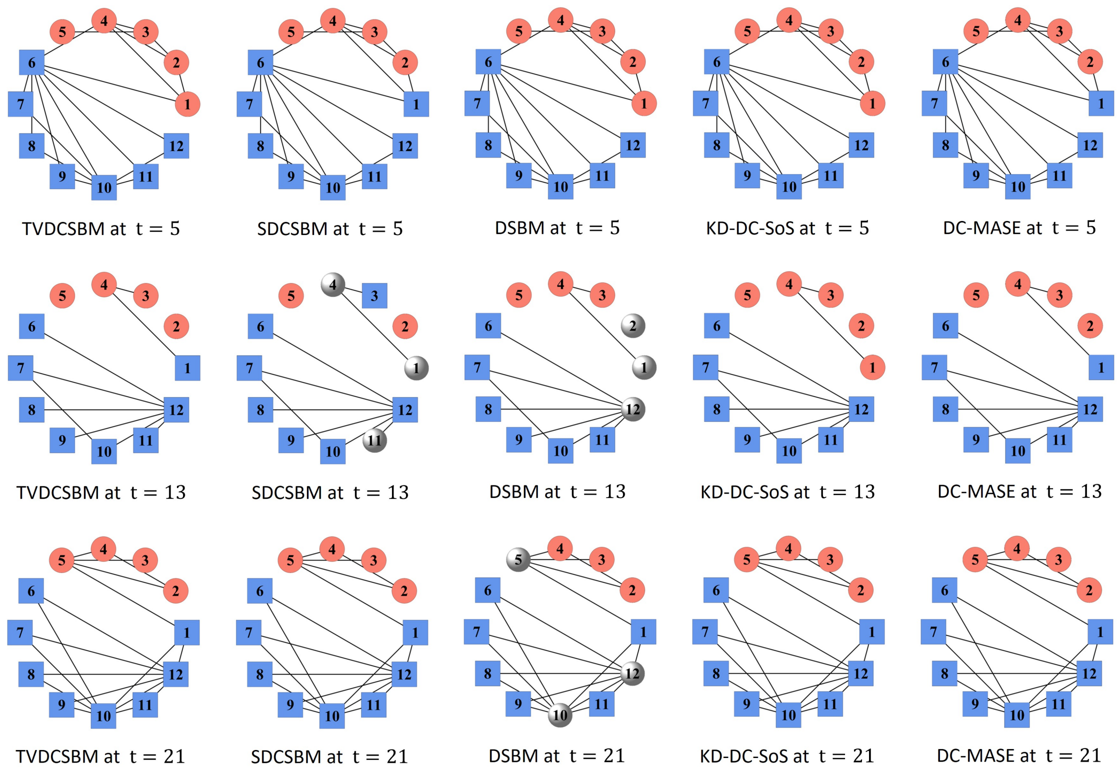


Figure 5: Community estimation results for the Travian subnetwork at three different time points. The red circles and blue squares denote two estimated communities, while the gray spheres indicate the nodes misclassified into communities other than those depicted by the clusters of red circles and blue squares.

As introduced in Figures 1 and 2 of Section 2, the temporal dynamics of community memberships of nodes as well as the connectivity between communities are observed in this dataset. Moreover, the heterogeneity in node degrees not only exists, but also changes as the network evolves. Thus, we use this analysis to validate the applicability of our approach. Figure 4 shows that the TVDCSBM performs better than the other methods. The HERs of our method, as $h = 0.47$, are close to 0 for all time points, and this suggests that even when the node labels change over time, aggregating the information over a neighborhood of a time point is effective in practice. DC-MASE also performs competitively, however, the assumption of a fixed community structure over time prevents it from tracking the nodes that switch communities.

Next, we examine the estimated partitions of the subnetwork presented in Section 2. Indeed, the 12 nodes are divided into two discernible communities, represented by red circles and blue squares as visually evident in Figure 1. Node 1 is a variable node over time and it switches community from the group of red circles to the group of blue squares at time point 14. It is seen from Figure 5 that the subnetwork partitions obtained by the three aggregation methods (TVDCSBM, DC-MASE, and KD-DC-SoS) are close to the ground truth. At time points 5 and 21, TVDCSBM is able to recover the community structure accurately and achieves zero HER. SDCSBM performs well at time points 21 but encounters challenges at time points 5 and 13, particularly in recovering the community of node 1. This result can be explained as the subnetwork displays local sparsity at $t = 13$ and some nodes are isolated without any connections. Consequently, determining the community assignments for these 'isolated nodes' becomes a challenging task if one only focused on the individual snapshots. Since the DSBM method misses the time-evolving heterogeneity among nodes within the same community, the hub nodes are more prone to misclassification. As a result, combining Figures 4 and 5, we observe that DSBM has the worst performance overall.

All in all, our analyses demonstrate that improved performance of community detection for networks with more refined structure, including variation of its features over time, can be achieved by a method that is capable to adapt to these features - such as our proposed TVDCSBM. In particular, in cases where the network lacks sufficient information (or contains many isolated nodes), reliable statistical inference would be rather intractable to achieve if only based on analysing one individual network per time. Therefore, it is of substantial importance to pool information, e.g. via kernel smoothing, across networks over time.

7 Conclusion and discussion

Flexible network models are crucial for accurately modelling real life networks. Our proposal in this paper focuses on the case of dynamic networks and allows time-varying degree heterogeneity in the nodes, as well as time-changing communities of nodes. We propose a simple but efficient two-step algorithm that incorporates kernel-based smoothing techniques into the profile-likelihood estimation, and we provide as well statistical guarantees for the consistency of the method in recovering both node heterogeneity and community structures. For time-varying DCSBM, we demonstrate that weak consistency in community detection can be achieved as $n\rho_n Th \to \infty$, thereby relaxing the constraint on network sparsity imposed by static DCSBM: we recall that $n\rho_n$ is the effective sample size per network, whereas Th is the effective number of networks over time within the smoothing window. Our theoretical results are well-supported by numerical simulations and analyses of real-world networks, and in particular show substantial gains for the case of very sparse time-varying networks.

As extensions of our work, one might possibly allow as well for the number of latent communities, K, to change over time. The smooth change of the number of communities over time also incurs an additional layer of complexity where one meticulously needs to track the appearance/disappearance of a new/old community, or merging/splitting phenomena as is done for e.g. in Greene, Doyle, and Cunningham (2010). Yet another extension could target the incorporation of labelling information and using sequences of networks to predict the health status of patients as in Relión, Kessler, Levina, and Taylor (2019), or the use of node covariate information for improved community structure estimation.

Acknowledgments

The authors would like to thank Hernando Ombao, Christian Hafner, the Associate Editor, and the reviewers for their constructive comments and suggestions, which substantially improved the paper. Mengxue Li acknowledges financial support from the China Scholarship Council.

References

Abbe, E. (2018). Community detection and stochastic block models: recent developments. Journal of Machine Learning Research 18(177), 1–86.

Agterberg, J., Z. Lubberts, and J. Arroyo (2025). Joint spectral clustering in multi-layer degree-corrected stochastic blockmodels. Journal of the American Statistical Association 120(551), 1607–1620.

Amini, A. A., A. Chen, P. J. Bickel, and E. Levina (2013). Pseudo-likelihood methods for community detection in large sparse networks. The Annals of Statistics 41(4), 2097–2122.

Bhattacharyya, S. and S. Chatterjee (2020). Consistent recovery of communities from sparse multi-relational networks: a scalable algorithm with optimal recovery conditions. In Complex Networks XI: Proceedings of the 11th Conference on Complex Networks CompleNet 2020, pp. 92–103. Springer.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman–Girvan and other modularities. Proceedings of the National Academy of Sciences 106(50), 21068–21073.

Bickel, P. J., A. Chen, Y. Zhao, E. Levina, and J. Zhu (2015). Correction to the proof of consistency of community detection. The Annals of Statistics 43(1), 462 – 466.

Bickel, P. J., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. The Annals of Statistics 41(4), 1922–1943.

Clauset, A., M. E. Newman, and C. Moore (2004). Finding community structure in very large networks. Physical Review E 70(6), 066111.

Gao, C., Z. Ma, A. Y. Zhang, and H. H. Zhou (2018). Community detection in degree-corrected block models. The Annals of Statistics 46(5), 2153–2185.

Glover, F. and M. Laguna (1997). Tabu Search. Kluwer Academic Publishers.

Greene, D., D. Doyle, and P. Cunningham (2010). Tracking the evolution of communities in dynamic social networks. In 2010 International Conference on Advances in Social Networks Analysis and Mining, pp. 176–183.

Gulati, R. (1998). Alliances and networks. Strategic Management Journal 19(4), 293–317.

Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. Social networks 5(2), 109–137.

Huang, S., H. Weng, and Y. Feng (2023). Spectral clustering via adaptive layer aggregation for multi-layer networks. Journal of Computational and Graphical Statistics 32(3), 1170–1184.

Härdle, W., P. Hall, and J. S. Marron (1988). How far are automatically chosen regression smoothing parameters from their optimum? Journal of the American Statistical Association 83(401), 86–95.

Jin, J. (2015). Fast community detection by SCORE. The Annals of Statistics 43(1), 57–89.

Karrer, B. and M. E. Newman (2011). Stochastic blockmodels and community structure in networks. Physical Review E 83(1), 016107.

Lei, J., K. Chen, and B. Lynch (2020). Consistent community detection in multi-layer network data. Biometrika 107(1), 61–73.

Lei, J. and K. Z. Lin (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. Journal of the American Statistical Association 118, 2433–2445.

Li, D., Y. Yuan, X. Zhang, and A. Qu (2024). Joint modeling of change-point identification and dependent dynamic community detection. Statistica Sinica 34(3).

Lin, K. Z. and J. Lei (2024). Dynamic clustering for heterophilic stochastic block models with time-varying node memberships. arXiv preprint arXiv:2403.05654.

Liu, F., D. Choi, L. Xie, and K. Roeder (2018). Global spectral clustering in dynamic networks. Proceedings of the National Academy of Sciences 115(5), 927–932.

Matias, C. and V. Miele (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. Journal of the Royal Statistical Society Series B: Statistical

Methodology 79(4), 1119–1141.

Newman, M. E. (2013). Spectral methods for community detection and graph partitioning. Physical Review E 88(10), 042822.

Pensky, M. and T. Zhang (2019). Spectral clustering in the dynamic stochastic block model. Electronic Journal of Statistics 13(1), 678–709.

Qin, T. and K. Rohe (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In Advances in Neural Information Processing Systems, Volume 26 of NIPS'13, Red Hook, NY, USA, pp. 3120–3128. Curran Associates Inc.

Relión, J. D. A., D. Kessler, E. Levina, and S. F. Taylor (2019). Network classification with applications to brain connectomics. The Annals of Applied Statistics 13(3), 1648 – 1677.

Riverain, P., S. Fossier, and M. Nadif (2023). Poisson degree corrected dynamic stochastic block model. Advances in Data Analysis and Classification 17(1), 135–162.

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. The Annals of Statistics 39(4), 1878–1915.

Samdin, S. B., C.-M. Ting, and H. Ombao (2019). Detecting state changes in community structure of functional brain networks using a Markov-switching stochastic block model. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1483–1487.

Sengupta, S. and Y. Chen (2018). A block model for node popularity in networks with community structure. Journal of the Royal Statistical Society Series B: Statistical Methodology 80(2), 365–386.

von Bahr, B. and C.-G. Esseen (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. The Annals of Mathematical Statistics 36(1), 299–303.

Wang, J., J. Zhang, B. Liu, J. Zhu, and J. Guo (2023). Fast network community detection with profile-pseudo likelihood methods. Journal of the American Statistical Association 118(542), 1359–1372.

Wang, Y. X. R. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. The Annals of Statistics 45(2), 500–528.

Yang, J. and P. Jie (2020). Estimating time-varying graphical models. Journal of Computational and Graphical Statistics 29(1), 191–202.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of community detection in networks under degree-corrected stochastic block models. The Annals of Statistics 40(4), 2266 – 2292.

A Proofs of the results from Section 3

Lemma A.1 (Bias and variance of $\hat{\tau}_{ij,h}(u^t)$)**.** *Under Assumption 1, the kernel-based estimator $\hat{\tau}_{ij,h}(u^t)$, as defined in (7), satisfies*

$$\textit{Bias:} \qquad \mathbb{E}(\hat{\tau}_{ij,h}(u^t)) - \tau_{ij}(u^t) \le C_{\sup,1}\rho_n h^2 + C_{\sup,2}\rho_n Th(1-\gamma_T) + O(\rho_n h^3)\ ,$$

$$\textit{Variance:} \qquad Var(\hat{\tau}_{ij,h}(u^t)) \le \frac{C_{\sup,2}\rho_n}{Th} + O\left(\frac{\rho_n^2}{Th}\right)\ ,$$

where the constants $C_{\sup,1}$ and $C_{\sup,2}$ are defined in (13).

This lemma serves as a preparatory step for subsequent proofs.

Proof. Let an even number $2L$ denote the number of observations in the local neighborhood $U(u^t)$. Asymptotically, we have $L = \frac{1}{2}\left|\left\{A_{ij}^{t'}, u^{t'} \in U(u^t)\right\}\right| = Th$, and $\hat{\tau}_{ij,h}(u^t)$ can then be redefined as

$$\hat{\tau}_{ij,h}(u^t) = \frac{1}{2L}\sum_{t'=t-L}^{t+L} A_{ij}^{t'}. \tag{A.1}$$

We further consider a variable $\mathcal{E}_{ij}^t = \sum_{t'=t-L}^{t+L} \mathbf{1}_{\{c_i^{t'} \neq c_i^t \text{ or } c_j^{t'} \neq c_j^t\}}$ to count the number of time points in a local interval from $t-L$ to $t+L$ where either node i or node j (or both) switches communities relative to their communities at time point t. Therefore, the bias of $\hat{\tau}_{ij,h}(u^t)$ can be expressed as,

$$\mathbb{E}(\hat{\tau}_{ij,h}(u^t)) - \tau_{ij}(u^t) = \mathbb{E}_{\mathcal{E}_{ij}^t}\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t) - \tau_{ij}(u^t) = \sum_{m=0}^{2L} \mathbb{P}(\mathcal{E}_{ij}^t = m)\left(\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t = m) - \tau_{ij}(u^t)\right). \tag{A.2}$$

Note from (6) that the error ε_{ij}^t has zero mean for every time point $t = 1, \ldots, T$, and combining with (A.1), we thus have that

$$\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t = m) = \frac{1}{2L}\sum_{t'=t-L}^{t+L}\mathbb{E}\left(\tau_{ij}(u^{t'}) + \varepsilon_{ij}^{t'} \mid \mathcal{E}_{ij}^t = m\right) = \frac{1}{2L}\sum_{t'=t-L}^{t+L}\mathbb{E}\left(\tau_{ij}(u^{t'}) \mid \mathcal{E}_{ij}^t = m\right).$$

Now consider the conditional bias given $\mathcal{E}_{ij}^t = m$ for some $m \in \{1, \ldots, 2L\}$. Suppose that there exists a construction of community assignments $\boldsymbol{c}^{t-L}, \ldots, \boldsymbol{c}^{t+L}$ such that $\mathcal{E}_{ij}^t = m$ holds. The conditional bias can be decomposed as,

$$\begin{aligned}
\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t = m) - \tau_{ij}(u^t) &= \frac{1}{2L}\sum_{t'=t-L}^{t+L}\left(\rho_n\theta_i(u^{t'})\theta_j(u^{t'})W_{c_i^{t'}c_j^{t'}}(u^{t'}) - \rho_n\theta_i(u^t)\theta_j(u^t)W_{c_i^t c_j^t}(u^t)\right) \\
&= \frac{\rho_n}{2L}\sum_{t'=t-L}^{t+L}\left(\theta_i(u^{t'})\theta_j(u^{t'})W_{c_i^{t'}c_j^{t'}}(u^{t'}) - \theta_i(u^t)\theta_j(u^t)W_{c_i^{t'}c_j^{t'}}(u^t)\right) \\
&\quad + \frac{\rho_n}{2L}\sum_{t'=t-L}^{t+L}\left(\theta_i(u^t)\theta_j(u^t)W_{c_i^{t'}c_j^{t'}}(u^t) - \theta_i(u^t)\theta_j(u^t)W_{c_i^t c_j^t}(u^t)\right) = B_1 + B_2\ ,
\end{aligned}$$

where the first term B_1 corresponds to the smoothing part and the second term B_2 measures the bias induced by community switching. Under the condition $\mathcal{E}_{ij}^t = m$ and without loss of generality, if we let $c_i^{t-L} \neq c_i^t, \ldots, c_i^{t-L+m-1} \neq c_i^t, c_i^{t-L+m} = c_i^t, \ldots, c_i^{t+L} = c_i^t$, while $c_j^{t-L} = \ldots = c_j^{t+L} = c_j^t$,

$$\begin{aligned}
B_2 &= \frac{\rho_n}{2L}\sum_{t'=t-L}^{t+L}\theta_i(u^t)\theta_j(u^t)\left(W_{c_i^{t'}c_j^{t'}}(u^t) - W_{c_i^t c_j^t}(u^t)\right)\mathbf{1}_{\{c_i^{t'} \neq c_i^t, \text{ or } c_j^{t'} \neq c_j^t\}} \\
&= \frac{\rho_n}{2L}\sum_{t'=t-L}^{t-L+m-1}\theta_i(u^t)\theta_j(u^t)\left(W_{c_i^{t'}c_j^{t'}}(u^t) - W_{c_i^t c_j^t}(u^t)\right) \le \frac{1}{2L}C_{\sup,2}m\rho_n,
\end{aligned}$$

with the constant $C_{\sup,2} := \sup_{1\le i,j\le n}\sup_{1\le a,b\le K}\sup_{u\in[0,1]}(\theta_i(u)\theta_j(u)W_{ab}(u))$ due to $\theta_i(u), W_{ab}(u) > 0$ for all $i \in \{1,\dots,n\}$, $a,b \in \{1,\dots,K\}$ and $u \in [0,1]$. Concerning the smoothing bias B_1, let $(\theta \cdot W)_{ij}(u) := \theta_i(u)\theta_j(u)W_{c_i^{t'}c_j^{t'}}(u)$ denote the product of three functions $\theta_i(u), \theta_j(u)$ and $W_{c_i^{t'}c_j^{t'}}(u)$. Under Assumption 1 where $\theta_i(u), W_{ab}(u) \in C^2[0,1]$ holds uniformly over $i,j \in \{1,\dots,n\}, a,b \in \{1,\dots,K\}$, we have $(\theta \cdot W)_{ij}(u) \in C^2[0,1]$ as well uniformly over $i,j \in \{1,\dots,n\}$. Applying a Taylor expansion to the function $(\theta \cdot W)_{ij}(u^{t'})$ around point u^t gives that

$$(\theta \cdot W)_{ij}(u^{t'}) = (\theta \cdot W)_{ij}(u^t) + (\theta \cdot W)_{ij}^{'}(u^t)\left(\frac{t'-t}{T}\right) + \frac{(\theta \cdot W)_{ij}^{''}(u^t)}{2}\left(\frac{t'-t}{T}\right)^2 + O\left(\left(\frac{|t'-t|}{T}\right)^3\right),$$

where $(\theta \cdot W)_{ij}^{'}(u^t)$ and $(\theta \cdot W)_{ij}^{''}(u^t)$ represent the first and second derivatives evaluated in u^t. Thus, the term B_1 becomes

$$\begin{aligned}
B_1 &= \frac{\rho_n}{2L}\sum_{t'=t-L}^{t+L}\left[(\theta \cdot W)_{ij}^{'}(u^t)\left(\frac{t'-t}{T}\right) + \frac{(\theta \cdot W)_{ij}^{''}(u^t)}{2}\left(\frac{t'-t}{T}\right)^2 + O\left(\left(\frac{|t'-t|}{T}\right)^3\right)\right] \\
&= \frac{\rho_n(\theta \cdot W)_{ij}^{''}(u^t)}{12} \times \left(\frac{L}{T}\right)^2\left(2 + \frac{3}{L} + \frac{1}{L^2}\right) + O\left(\rho_n\left(\frac{L}{T}\right)^3\right) \le C_{\sup,1}\rho_n h^2 + O\left(\rho_n h^3\right),
\end{aligned}$$

for a constant $C_{\sup,1} := \sup_{1\le i,j\le n}\sup_{1\le a,b\le K}\sup_{u\in[0,1]}\left|\frac{\partial^2(\theta_i(u)\theta_j(u)W_{ab}(u))}{\partial u^2}\right|$. Since the switching indicators X_i^t are iid Bernoulli random variables, each time point $t = 1,\dots,T$ has a probability $1-\gamma_T$ of being a switching point for any node $i = 1,\dots,n$. Therefore, for any integer $L' \le L$ such that $u^{t+L'} \in U(u^t)$, the probability that node i belongs to community c_i^t at time point $t+L'$ is at least $(\gamma_T)^{L'}$, implying that node i has not changed its community from time point t to $t+L'$. Moreover, given the independence of community switching among nodes, the probability of the event $\{\mathcal{E}_{ij}^t \ge 1\}$ to occur is at most $1-(\gamma_T)^{2L}$. Based on this observation, putting B_1 and B_2 together and plugging them into (A.2) immediately yields the bias of $\hat{\tau}_{ij}(u^t)$ as follows:

$$\begin{aligned}
\mathbb{E}(\hat{\tau}_{ij,h}(u^t)) - \tau_{ij}(u^t) &= \mathbb{P}(\mathcal{E}_{ij}^t = 0)\left(\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t = 0) - \tau_{ij}(u^t)\right) \\
&\quad + \sum_{m=1}^{2L}\mathbb{P}(\mathcal{E}_{ij}^t = m)\left(\mathbb{E}(\hat{\tau}_{ij,h}(u^t) \mid \mathcal{E}_{ij}^t = m) - \tau_{ij}(u^t)\right) \\
&\le \mathbb{P}(\mathcal{E}_{ij}^t = 0)\left(C_{\sup,1}\rho_n h^2 + O(\rho_n h^3)\right) \\
&\quad + \sum_{m=1}^{2L}\mathbb{P}(\mathcal{E}_{ij}^t = m)\left(C_{\sup,1}\rho_n h^2 + O(\rho_n h^3) + \frac{1}{2L}C_{\sup,2}m\rho_n\right) \\
&\le C_{\sup,1}\rho_n h^2 + O(\rho_n h^3) + \frac{C_{\sup,2}\rho_n}{2L} \times 2L \times \mathbb{P}(\mathcal{E}_{ij}^t \ge 1) \\
&\le C_{\sup,1}\rho_n h^2 + O(\rho_n h^3) + C_{\sup,2}\rho_n(1-(\gamma_T)^{2L}) \\
&\le C_{\sup,1}\rho_n h^2 + C_{\sup,2}\rho_n Th(1-\gamma_T) + O(\rho_n h^3),
\end{aligned}$$

where the last inequality holds due to the fact that $1-x^n \le (1-x)(1+x+\dots,+x^{n-1}) \le n(1-x)$. We next consider the variance of $\hat{\tau}_{ij,h}(u^t)$. Note that the independence of the error terms ε_{ij}^t across nodes $i,j \in \{1,\dots,n\}$ and time points $t \in \{1,\dots,T\}$ is assumed in our model setup in (6). Specifically, the edges A_{ij}^t's are assumed to be formed independently given community labels and node degree parameters for any $i,j \in \{1,\dots,n\}$ and any $t \in \{1,\dots,T\}$. So (A.1) implies that

$$\mathrm{Var}(\hat{\tau}_{ij,h}(u^t)) = \frac{1}{4L^2}\sum_{t'=t-L}^{t+L}\mathrm{Var}\left(\varepsilon_{ij}^{t'}\right) = \frac{1}{4L^2}\sum_{t'=t-L}^{t+L}\sigma_{ij}^2(u^{t'}).$$

Using the fact that $\sigma_{ij}^2(u^t) = \tau_{ij}(u^t)(1-\tau_{ij}(u^t))$, this variance can be rewritten as

$$\begin{aligned}
\mathrm{Var}(\hat{\tau}_{ij,h}(u^t)) &= \frac{1}{4L^2}\sum_{t'=t-L}^{t+L}\tau_{ij}(u^{t'}) - \frac{1}{4L^2}\sum_{t'=t-L}^{t+L}(\tau_{ij}(u^{t'}))^2 \\
&= \frac{\rho_n \sum_{t'=t-L}^{t+L}\theta_i(u^{t'})\theta_j(u^{t'})W_{c_i^{t'}c_j^{t'}}(u^{t'})}{4T^2h^2} + O\left(\frac{\rho_n^2}{Th}\right) \\
&\le \frac{C_{\sup,2}\rho_n}{Th} + O\left(\frac{\rho_n^2}{Th}\right).
\end{aligned}$$

□

Before we proceed to the proof of Proposition 1 from the main manuscript, we first show that the smoothed counting variables converge to their corresponding population version. Recalling the definitions of d_i^t and $O_{kl}(\boldsymbol{e}^t)$, we observe that the former is proportional to the baseline average node degree, $n\rho_n$, and the latter is proportional to the baseline average number of edges in a network namely, $n^2\rho_n$. Furthermore, as defined in (18), the quantities $H_{kl}(\boldsymbol{e}^t)$ and $H_k(\boldsymbol{e}^t)$ denote the population counterparts of $O_{kl}(\boldsymbol{e}^t)/(n^2\rho_n)$ and $O_k(\boldsymbol{e}^t)/(n^2\rho_n)$, respectively:

$$H_{kl}(\boldsymbol{e}^t) = \sum_{a,b=1}^{K} W_{ab}^t \times \frac{\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n} \times \frac{\Lambda_{b(l)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n}, \qquad H_k(\boldsymbol{e}^t) = \sum_{a,b=1}^{K} \frac{n_b(\boldsymbol{c}^t)W_{ab}^t}{n} \times \frac{\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n},$$

where $\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) = \sum_{i=1}^n \theta_i^t \mathbf{1}_{\{c_i^t=a, e_i^t=k\}}$. Let $n_l(\boldsymbol{e}^t) = \sum_{i=1}^n \mathbf{1}_{\{e_i^t=l\}}$ and $\tilde{\kappa}_l(\boldsymbol{e}^t) = n_l(\boldsymbol{e}^t)/n$. We consider the expectations and variances of the following scaled quantities.

Lemma A.2 (Convergence of the smoothed counting variables)**.** *Under Assumptions 1-3, and for any fixed assignment $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, the smoothed counting variables, defined in (8), satisfy that*

$$\mathbb{E}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) - \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right) \le C_{\sup,1}h^2 + C_{\sup,2}Th(1-\gamma_T) + O(h^3), \tag{A.3}$$

$$Var\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) \le \frac{C_{\sup,2}}{n\rho_n Th} + O\left(\frac{1}{nTh}\right). \tag{A.4}$$

$$\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) - H_{kl}(\boldsymbol{e}^t) \le C_{\sup,1}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)Th(1-\gamma_T) + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3\right), \tag{A.5}$$

$$Var\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) \le \frac{C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2Th}\right). \tag{A.6}$$

$$\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) - H_k(\boldsymbol{e}^t) \le C_{\sup,1}\tilde{\kappa}_k(\boldsymbol{e}^t)h^2 + C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)Th(1-\gamma_T) + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)h^3\right), \tag{A.7}$$

$$Var\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) \le \frac{C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2Th}\right). \tag{A.8}$$

Furthermore, if there exist $\gamma_T \to 1$, $\rho_n, h \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$, then for any $i \in \{1,\dots,n\}, k,l \in \{1,\dots,K\}$ and any vector $\boldsymbol{e}^t \in \{1,\dots,K\}^n$,

$$\frac{\hat{d}_{i,h}^t}{n\rho_n} \overset{qm}{\to} \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right), \quad \frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} \overset{qm}{\to} H_{kl}(\boldsymbol{e}^t), \quad \frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n} \overset{qm}{\to} H_k(\boldsymbol{e}^t)$$

where the notation '$\overset{qm}{\to}$' denotes the convergence in quadratic mean.

Proof. Note that our smoothing-based method discussed in Section 3 is edgewise across time, as such the $\hat{\tau}_{ij,h}(u^t)$'s are independent variables over all i and j since A_{ij}^t's are independent given $\boldsymbol{\theta}^t$ and $\boldsymbol{c}^t$. Relying on the results of Lemma A.1, it holds that

$$\mathbb{E}\left(\frac{\hat{d}^t_{i,h}}{n\rho_n}\right) = \frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}(\hat{\tau}_{ij,h}(u^t)) \le \frac{1}{n\rho_n}\sum_{j=1}^n \left[\tau_{ij}(u^t) + C_{\sup,1}\rho_n h^2 + C_{\sup,2}\rho_n Th(1-\gamma_T) + O(\rho_n h^3)\right]$$
$$= \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d^t_i}{n\rho_n}\right) + C_{\sup,1}h^2 + C_{\sup,2}Th(1-\gamma_T) + O(h^3),$$

and

$$\mathrm{Var}\left(\frac{\hat{d}^t_{i,h}}{n\rho_n}\right) = \frac{1}{n^2\rho_n^2}\sum_{j=1}^n \mathrm{Var}(\hat{\tau}_{ij,h}(u^t)) \le \frac{1}{n^2\rho_n^2}\sum_{j=1}^n \left[\frac{C_{\sup,2}\rho_n}{Th} + O\left(\frac{\rho_n^2}{Th}\right)\right] = \frac{C_{\sup,2}}{n\rho_n Th} + O\left(\frac{1}{nTh}\right).$$

Similarly, the expectation and variance of $\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}$ can be written as follows,

$$\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^2\rho_n}\sum_{i,j=1}^n \mathbb{E}(\hat{\tau}_{ij,h}(u^t))\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}$$
$$\le \frac{1}{n^2\rho_n}\sum_{i,j=1}^n \left[\tau_{ij}(u^t) + C_{\sup,1}\rho_n h^2 + C_{\sup,2}\rho_n Th(1-\gamma_T) + O(\rho_n h^3)\right]\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}$$
$$= \frac{1}{n^2}\sum_{a,b=1}^K \frac{Z^t_{ab}}{\rho_n}\sum_{i=1}^n \theta^t_i \mathbf{1}_{\{e^t_i=k,c^t_i=a\}}\sum_{j=1}^n \theta^t_j \mathbf{1}_{\{e^t_j=l,c^t_j=b\}} + \frac{\sum_{i,j=1}^n C_{\sup,1}\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}}{n^2}h^2$$
$$+ \frac{\sum_{i,j=1}^n C_{\sup,2}\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}}{n^2}Th(1-\gamma_T) + \frac{\sum_{i,j=1}^n \mathbf{1}_{\{e^t_i=k,e^t_j=l\}}O(h^3)}{n^2}$$
$$= H_{kl}(\boldsymbol{e}^t) + C_{\sup,1}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)Th(1-\gamma_T) + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3\right),$$

and

$$\mathrm{Var}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^n \mathrm{Var}(\hat{\tau}_{ij,h}(u^t))\mathbf{1}_{\{e^t_i=k,e^t_j=l\}} = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^n \left[\frac{C_{\sup,2}\rho_n}{Th} + O\left(\frac{\rho_n^2}{Th}\right)\right]\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}$$
$$= \frac{\sum_{i,j=1}^n C_{\sup,2}\mathbf{1}_{\{e^t_i=k,e^t_j=l\}}}{n^2}\times\frac{1}{n^2\rho_n Th} + O\left(\frac{n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)}{n^4 Th}\right)$$
$$= \frac{C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2 Th}\right).$$

Again, we consider the quantity $\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}$ satisfying that

$$\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^2\rho_n}\sum_{i,j=1}^n \mathbb{E}(\hat{\tau}_{ij,h}(u^t))\mathbf{1}_{\{e^t_i=k\}}$$
$$\le \frac{1}{n^2\rho_n}\sum_{i,j=1}^n \left[\tau_{ij}(u^t) + C_{\sup,1}\rho_n h^2 + C_{\sup,2}\rho_n Th(1-\gamma_T) + O(\rho_n h^3)\right]\mathbf{1}_{\{e^t_i=k\}}$$
$$= \frac{1}{n^2}\sum_{a,b=1}^K \frac{Z^t_{ab}}{\rho_n}\sum_{i=1}^n \theta^t_i \mathbf{1}_{\{e^t_i=k,c^t_i=a\}}\sum_{j=1}^n \theta^t_j \mathbf{1}_{\{c^t_j=b\}} + \frac{\sum_{i,j=1}^n C_{\sup,1}\mathbf{1}_{\{e^t_i=k\}}}{n^2}h^2$$
$$+ \frac{\sum_{i,j=1}^n C_{\sup,2}\mathbf{1}_{\{e^t_i=k\}}}{n^2}Th(1-\gamma_T) + \frac{\sum_{i,j=1}^n \mathbf{1}_{\{e^t_i=k\}}O(h^3)}{n^2}$$
$$= H_k(\boldsymbol{e}^t) + C_{\sup,1}\tilde{\kappa}_k(\boldsymbol{e}^t)h^2 + C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)Th(1-\gamma_T) + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)h^3\right),$$

and

$$\mathrm{Var}\left(\frac{\hat{O}_{k,h}(\boldsymbol{c}^t)}{n^2\rho_n}\right) = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^n \mathrm{Var}(\hat{\tau}_{ij,h}(u^t))\mathbf{1}_{\{e^t_i=k\}} = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^n \left[\frac{C_{\sup,2}\rho_n}{Th} + O\left(\frac{\rho_n^2}{Th}\right)\right]\mathbf{1}_{\{e^t_i=k\}}$$
$$= \frac{\sum_{i,j=1}^n C_{\sup,2}\mathbf{1}_{\{e^t_i=k\}}}{n^2}\times\frac{1}{n^2\rho_n Th} + O\left(\frac{n_k(\boldsymbol{e}^t)}{n^3 Th}\right) = \frac{C_{\sup,2}\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2 Th}\right).$$

Hence, the results of (A.5) through (A.8) are obtained. By the mean-variance decomposition, the MSE of each smoothed counting variable tends to 0 since

$$
\begin{aligned}
\mathbb{E}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n} - \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right)\right)^2 &= O\left(h^4 + T^2h^2(1-\gamma_T)^2 + \frac{1}{n\rho_n Th}\right) \to 0,\\
\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} - H_{kl}(\boldsymbol{e}^t)\right)^2 &= O\left(\tilde{\kappa}_k^2(\boldsymbol{e}^t)\tilde{\kappa}_l^2(\boldsymbol{e}^t)h^4 + \tilde{\kappa}_k^2(\boldsymbol{e}^t)\tilde{\kappa}_l^2(\boldsymbol{e}^t)T^2h^2(1-\gamma_T)^2 + \frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_n Th}\right) \to 0,\\
\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n} - H_{k}(\boldsymbol{e}^t)\right)^2 &= O\left(\tilde{\kappa}_k^2(\boldsymbol{e}^t)h^4 + \tilde{\kappa}_k^2(\boldsymbol{e}^t)T^2h^2(1-\gamma_T)^2 + \frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_n Th}\right) \to 0,
\end{aligned}
$$

as $\gamma_T \to 1$, $\rho_n, h \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$. Thus, each smoothed counting variable converges in quadratic mean to its corresponding population version. □

Proof of Proposition 1. To verify Proposition 1, we first inspect the individual edge probability $\tilde{\tau}_{ij,h}(u^t)$ for any $i, j \in \{1, \ldots, n\}$. Define

$$
\hat{\boldsymbol{\nu}} = \left(\frac{\hat{d}_{i,h}^t}{n\rho_n}, \frac{\hat{d}_{j,h}^t}{n\rho_n}, \frac{\hat{O}_{c_i^t c_j^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}, \frac{\hat{O}_{c_i^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}, \frac{\hat{O}_{c_j^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}\right)^\top,
$$

where $\hat{\nu}_m$ corresponds to the m-th entry in $\hat{\boldsymbol{\nu}}$ for any $m \in \{1, \ldots, 5\}$. We notice that $\tilde{\tau}_{ij,h}(u^t)$, defined in (10), can be treated as a function of $\hat{\boldsymbol{\nu}}$,

$$
\tilde{\tau}_{ij,h}(u^t) = f(\hat{\boldsymbol{\nu}}) = \frac{\rho_n \hat{\nu}_1 \hat{\nu}_2 \hat{\nu}_3}{\hat{\nu}_4 \hat{\nu}_5}.
$$

Correspondingly, let

$$
\boldsymbol{\nu} = \left(\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right), \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_j^t}{n\rho_n}\right), H_{c_i^t c_j^t}(\boldsymbol{c}^t), H_{c_i^t}(\boldsymbol{c}^t), H_{c_j^t}(\boldsymbol{c}^t)\right)^\top,
$$

and ν_m be the m-th entry in $\boldsymbol{\nu}$ for any $m \in \{1, \ldots, 5\}$.

Remark 7. *Let $\kappa_{c_i^t} = n_{c_i^t}(\boldsymbol{c}^t)/n$ be the ratio of the size of the community to which node i belongs at time point t relative to the total number of nodes. One can show that as n diverges, the scaled terms $(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3, \kappa_{c_i^t}^{-1}\nu_4$ and $\kappa_{c_j^t}^{-1}\nu_5$ are bounded below and above by positive constants. Recalling from (13) the definition of $C_{\sup,2}$, it is similarly natural to allow for a lower bound on $\theta_i(u)$ and $W_{ab}(u)$ for any $i \in \{1, \ldots, n\}, a, b \in \{1, \ldots, K\}$ and $u \in [0,1]$. Specifically, define*

$$
C_{\inf} := \inf_{1\le i,j\le n} \inf_{1\le a,b\le K} \inf_{u\in[0,1]} \left(\theta_i(u)\theta_j(u)W_{ab}(u)\right),
$$

to be a positive constant independent of n and T such that $\rho_n C_{\inf} \le \tau_{ij}(u^t) = \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t) \le \rho_n C_{\sup,2}$ over $i, j \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, T\}$. It is easy to see that

$$
C_{\inf} \le \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right) = \frac{\sum_{j=1}^n \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t)}{n\rho_n} \le C_{\sup,2},
$$

$$
\kappa_{c_i^t}\kappa_{c_j^t} C_{\inf} \le H_{c_i^t c_j^t}(\boldsymbol{c}^t) = \frac{\sum_{i,j=1}^n \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t)\mathbf{1}_{\{e_i^t=k, e_j^t=l\}}}{n^2\rho_n} \le \kappa_{c_i^t}\kappa_{c_j^t} C_{\sup,2},
$$

$$
\kappa_{c_i^t} C_{\inf} \le H_{c_i^t}(\boldsymbol{c}^t) = \frac{\sum_{i,j=1}^n \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t)\mathbf{1}_{\{e_i^t=k\}}}{n^2\rho_n} \le \kappa_{c_i^t} C_{\sup,2},
$$

which implies that ν_1, ν_2 and the scaled terms $(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3, \kappa_{c_i^t}^{-1}\nu_4$ and $\kappa_{c_j^t}^{-1}\nu_5$ are bounded below and above by positive constants.

To simplify the notations, define the index set $\mathcal{N}_a^t = \{1 \le i \le n : c_i^t = a\}$ of nodes whose true community is a at time point t. With this definition, it is clear that the identifiability of θ_i^t in Assumption 3 is equivalent to $\sum_{s\in\mathcal{N}^t_{c_i^t}} \theta_s^t = n_{c_i^t}(\boldsymbol{c}^t)$. By a Taylor expansion around the point $\boldsymbol{\nu}$, we have that

$$\tilde{\tau}_{ij,h}(u^t) = f(\boldsymbol{\nu}) + \nabla f(\boldsymbol{\xi})^\top (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})$$

where $\nabla f(\boldsymbol{\xi})^\top (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})$ is the Taylor remainder term. The term $\nabla f(\boldsymbol{\xi})$ denotes the vector of first derivatives of the function $f(\cdot)$ evaluated in $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)^\top$ where $\xi_m = \hat{\nu}_m + \alpha(\hat{\nu}_m - \nu_m)$ for some $\alpha \in (0,1)$ and any $m \in \{1, \ldots, 5\}$. Concerning the first term $f(\boldsymbol{\nu})$, it is not difficult to show that $f(\boldsymbol{\nu})$ is equivalent to $\tau_{ij}(u^t)$. Given the community labels $\boldsymbol{c}^t$ and the identifiability of $\boldsymbol{\theta}^t$ via Assumption 3, we have that

$$\begin{aligned}
f(\boldsymbol{\nu}) &= \frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_j^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t c_j^t}(\boldsymbol{c}^t))}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_j^t}(\boldsymbol{c}^t))} \\
&= \frac{\left(\sum_{j=1}^n \theta_i^t\theta_j^t Z^t_{c_i^t c_j^t}\right)\left(\sum_{i=1}^n \theta_i^t\theta_j^t Z^t_{c_i^t c_j^t}\right)\left(\sum_{s\in\mathcal{N}^t_{c_i^t}, m\in\mathcal{N}^t_{c_j^t}} \theta_s^t\theta_m^t Z^t_{c_i^t c_j^t}\right)}{\left(\sum_{s\in\mathcal{N}^t_{c_i^t}}\sum_{j=1}^n \theta_s^t\theta_j^t Z^t_{c_i^t c_j^t}\right)\left(\sum_{m\in\mathcal{N}^t_{c_j^t}}\sum_{i=1}^n \theta_m^t\theta_i^t Z^t_{c_i^t c_j^t}\right)} \\
&= \frac{\theta_i^t\left(\sum_{j=1}^n \theta_j^t Z^t_{c_i^t c_j^t}\right) \times \theta_j^t\left(\sum_{i=1}^n \theta_i^t Z^t_{c_i^t c_j^t}\right) \times \left(\sum_{s\in\mathcal{N}^t_{c_i^t}} \theta_s^t\right)\left(\sum_{m\in\mathcal{N}^t_{c_j^t}} \theta_m^t\right) Z^t_{c_i^t c_j^t}}{\left(\sum_{s\in\mathcal{N}^t_{c_i^t}} \theta_s^t\right)\left(\sum_{j=1}^n \theta_j^t Z^t_{c_i^t c_j^t}\right)\left(\sum_{m\in\mathcal{N}^t_{c_j^t}} \theta_m^t\right)\left(\sum_{i=1}^n \theta_i^t Z^t_{c_i^t c_j^t}\right)} \\
&= \theta_i^t\theta_j^t Z^t_{c_i^t c_j^t} = \tau_{ij}(u^t).
\end{aligned} \tag{A.9}$$

Next we consider the Taylor remainder term $\nabla f(\boldsymbol{\xi})^\top (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu})$ and break it up as $(R_{f1} + R_{f2} + R_{f3} - R_{f4} - R_{f5})$ where

$$R_{f1} = \frac{\rho_n\xi_2\xi_3}{\xi_4\xi_5}(\hat{\nu}_1 - \nu_1), \quad R_{f2} = \frac{\rho_n\xi_1\xi_3}{\xi_4\xi_5}(\hat{\nu}_2 - \nu_2), \quad R_{f3}\frac{\rho_n\xi_1\xi_2}{\xi_4\xi_5}(\hat{\nu}_3 - \nu_3),$$
$$R_{f4} = \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4^2\xi_5}(\hat{\nu}_4 - \nu_4), \quad R_{f5} = \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4\xi_5^2}(\hat{\nu}_5 - \nu_5).$$

Combining it with (A.9), the Taylor expansion for $\tilde{\tau}_{ij,h}(u^t)$ becomes

$$\tilde{\tau}_{ij,h}(u^t) = \tau_{ij}(u^t) + R_{f1} + R_{f2} + R_{f3} - R_{f4} - R_{f5}. \tag{A.10}$$

To show the asymptotic behavior of this remainder term, we need to show the convergence of each part R_{fm}. Since R_{f2} has the same behavior as R_{f1} and as well R_{f5} has the same behavior as R_{f4}, it suffices to focus on R_{f1}, R_{f3} and R_{f4} separately in the following steps.

Step 1: R_{f1} (same as for R_{f2}). Lemma A.2 suggests that for any $m \in \{1, \ldots, 5\}$, $\hat{\nu}_m$ converges to its population version ν_m in quadratic mean which implies that $\xi_m \in (\hat{\nu}_m, \nu_m)$ also converges to ν_m as $\gamma_T \to 1,\ h, \rho_n \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$. From Remark 7, we know that $\kappa_{c_i^t}^{-1}\nu_4$ and $\kappa_{c_j^t}^{-1}\nu_5$ are positive constants. Applying Slutsky's Theorem, we have that

$$(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\xi_3 \xrightarrow{P} (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3, \quad \frac{1}{\kappa_{c_i^t}^{-1}\xi_4} \xrightarrow{P} \frac{1}{\kappa_{c_i^t}^{-1}\nu_4}, \quad \text{and} \quad \frac{1}{\kappa_{c_j^t}^{-1}\xi_5} \xrightarrow{P} \frac{1}{\kappa_{c_j^t}^{-1}\nu_5} \tag{A.11}$$

as $\gamma_T \to 1,\ h, \rho_n \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$, which implies that

$$\begin{aligned}
R_{f1} = \frac{\rho_n\xi_2\xi_3}{\xi_4\xi_5}(\hat{\nu}_1 - \nu_1) &= \frac{\xi_2 \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\xi_3}{\left(\kappa_{c_i^t}^{-1}\xi_4\right)\left(\kappa_{c_j^t}^{-1}\xi_5\right)} \times \rho_n(\hat{\nu}_1 - \nu_1) \\
&= \left[\frac{\nu_2 \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3}{\left(\kappa_{c_i^t}^{-1}\nu_4\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)} + o_p(1)\right] \times \rho_n(\hat{\nu}_1 - \nu_1) \\
&= M_{\nabla f_1}\rho_n(\hat{\nu}_1 - \nu_1) + o_p(\rho_n(\hat{\nu}_1 - \nu_1)),
\end{aligned} \tag{A.12}$$

where $M_{\nabla f_1} = \frac{\nu_2 \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3}{\left(\kappa_{c_i^t}^{-1}\nu_4\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)} = \frac{\nu_2\nu_3}{\nu_4\nu_5}$ is a constant. We next take the expectation and variance on both sides of (A.12). By (A.3),

$$\begin{aligned}\mathbb{E}(R_{f1}) &= M_{\nabla f_1}\rho_n\mathbb{E}\left(\hat{\nu}_1 - \nu_1\right) + o(\rho_n\mathbb{E}(\hat{\nu}_1 - \nu_1))\\ &\leq M_{\nabla f_1}\rho_n C_{\sup,1}h^2 + M_{\nabla f_1}\rho_n C_{\sup,2}Th(1-\gamma_T) + O(\rho_n h^3) + o(\rho_n h^2 + \rho_n Th(1-\gamma_T))\\ &= M_{\nabla f_1}C_{\sup,1}\rho_n h^2 + M_{\nabla f_1}\rho_n C_{\sup,2}Th(1-\gamma_T) + o(\rho_n h^2 + \rho_n Th(1-\gamma_T)),\end{aligned} \tag{A.13}$$

and also substituting the variance bound from (A.4), we have that

$$\begin{aligned}\mathrm{Var}(R_{f1}) &= M_{\nabla f_1}^2\rho_n^2\mathrm{Var}\left(\hat{\nu}_1 - \nu_1\right) + o(\rho_n^2\mathrm{Var}(\hat{\nu}_1 - \nu_1))\\ &\leq M_{\nabla f_1}^2\rho_n^2 \times \frac{C_{\sup,2}}{n\rho_n Th} + O\left(\frac{\rho_n^2}{nTh}\right) + o\left(\frac{\rho_n}{nTh}\right) = \frac{M_{\nabla f_1}^2 C_{\sup,2}\rho_n}{nTh} + o\left(\frac{\rho_n}{nTh}\right),\end{aligned} \tag{A.14}$$

where the final $o\left(\frac{\rho_n}{nTh}\right)$ holds because the $O\left(\frac{\rho_n^2}{nTh}\right)$ term is also of order $o\left(\frac{\rho_n}{nTh}\right)$.

Step 2: R_{f3}. We use next the same technique to handle R_{f3}. By expression (A.11) and Slutsky's Theorem, it is easy to show that

$$\begin{aligned}R_{f3} &= \frac{\rho_n\xi_1\xi_2}{\xi_4\xi_5}\left(\hat{\nu}_3 - \nu_3\right) = \frac{\xi_1\xi_2}{\left(\kappa_{c_i^t}^{-1}\xi_4\right)\left(\kappa_{c_j^t}^{-1}\xi_5\right)} \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n(\hat{\nu}_3 - \nu_3)\\ &= \left[\frac{\nu_1\nu_2}{\left(\kappa_{c_i^t}^{-1}\nu_4\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)} + o_p(1)\right] \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n(\hat{\nu}_3 - \nu_3)\\ &= M_{\nabla f_3}(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n(\hat{\nu}_3 - \nu_3) + o_p\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n(\hat{\nu}_3 - \nu_3)\right)\end{aligned} \tag{A.15}$$

with a constant $M_{\nabla f_3} = \frac{\nu_1\nu_2}{\left(\kappa_{c_i^t}^{-1}\nu_4\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)}$. Following (A.15), (A.5) and (A.6), we have that

$$\begin{aligned}\mathbb{E}(R_{f3}) &= M_{\nabla f_3}(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n\mathbb{E}\left(\hat{\nu}_3 - \nu_3\right) + o\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n\mathbb{E}(\hat{\nu}_3 - \nu_3)\right)\\ &\leq M_{\nabla f_3}(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n \times C_{\sup,1}\kappa_{c_i^t}\kappa_{c_j^t}h^2 + M_{\nabla f_3}(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n \times C_{\sup,2}\kappa_{c_i^t}\kappa_{c_j^t}Th(1-\gamma_T)\\ &\quad + O\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n \times \kappa_{c_i^t}\kappa_{c_j^t}h^3\right) + o\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\rho_n \times \kappa_{c_i^t}\kappa_{c_j^t}\left(h^2 + Th(1-\gamma_T)\right)\right)\\ &= M_{\nabla f_3}C_{\sup,1}\rho_n h^2 + M_{\nabla f_3}C_{\sup,2}\rho_n Th(1-\gamma_T) + o(\rho_n h^2 + \rho_n Th(1-\gamma_T)),\end{aligned} \tag{A.16}$$

$$\begin{aligned}\mathrm{Var}(R_{f3}) &= M_{\nabla f_3}^2(\kappa_{c_i^t}\kappa_{c_j^t})^{-2}\rho_n^2\mathrm{Var}\left(\hat{\nu}_3 - \nu_3\right) + o\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-2}\rho_n^2\mathrm{Var}(\hat{\nu}_3 - \nu_3)\right)\\ &\leq M_{\nabla f_3}^2(\kappa_{c_i^t}\kappa_{c_j^t})^{-2}\rho_n^2 \times \frac{C_{\sup,2}\kappa_{c_i^t}\kappa_{c_j^t}}{n^2\rho_n Th} + O\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-2}\rho_n^2 \times \frac{\kappa_{c_i^t}\kappa_{c_j^t}}{n^2 Th}\right) + o\left((\kappa_{c_i^t}\kappa_{c_j^t})^{-2}\rho_n^2 \times \frac{\kappa_{c_i^t}\kappa_{c_j^t}}{n^2\rho_n Th}\right)\\ &= \frac{M_{\nabla f_3}^2 C_{\sup,2}\rho_n}{n^2\kappa_{c_i^t}\kappa_{c_j^t}Th} + o\left(\frac{\rho_n}{n^2\kappa_{c_i^t}\kappa_{c_j^t}Th}\right).\end{aligned} \tag{A.17}$$

Step 3: R_{f4} (same as for R_{f5}). Considering R_{f4}, we apply Slutsky's Theorem once more, and we have that

$$\begin{aligned}R_{f4} &= \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4^2\xi_5}\left(\hat{\nu}_4 - \nu_4\right) = \frac{\xi_1\xi_2 \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\xi_3}{\left(\kappa_{c_i^t}^{-2}\xi_4^2\right)\left(\kappa_{c_j^t}^{-1}\xi_5\right)} \times \kappa_{c_i^t}^{-1}\rho_n(\hat{\nu}_4 - \nu_4)\\ &= \left[\frac{\nu_1\nu_2 \times (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3}{\left(\kappa_{c_i^t}^{-2}\nu_4^2\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)} + o_p(1)\right] \times \kappa_{c_i^t}^{-1}\rho_n(\hat{\nu}_4 - \nu_4)\\ &= M_{\nabla f_4}\kappa_{c_i^t}^{-1}\rho_n(\hat{\nu}_4 - \nu_4) + o_p\left(\kappa_{c_i^t}^{-1}\rho_n(\hat{\nu}_4 - \nu_4)\right)\end{aligned} \tag{A.18}$$

with a constant $M_{\nabla f_4} = \frac{\nu_1\nu_2\times(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3}{\left(\kappa_{c_i^t}^{-2}\nu_4^2\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)}$. By(A.18), (A.7) and (A.8), we calculate that

$$\begin{aligned}\mathbb{E}(R_{f4}) &= M_{\nabla f_4}\kappa_{c_i^t}^{-1}\rho_n\mathbb{E}\left(\hat{\nu}_4-\nu_4\right)+o\left(\kappa_{c_i^t}^{-1}\rho_n\mathbb{E}(\hat{\nu}_4-\nu_4)\right)\\ &\le M_{\nabla f_4}\kappa_{c_i^t}^{-1}\rho_n\times C_{\sup,1}\kappa_{c_i^t}h^2+M_{\nabla f_4}\kappa_{c_i^t}^{-1}\rho_n\times C_{\sup,2}\kappa_{c_i^t}Th(1-\gamma_T)\\ &\qquad +O\left(\kappa_{c_i^t}^{-1}\rho_n\times\kappa_{c_i^t}h^3\right)+o\left(\kappa_{c_i^t}^{-1}\rho_n\times\kappa_{c_i^t}\left(h^2+Th(1-\gamma_T)\right)\right)\\ &= M_{\nabla f_4}C_{\sup,1}\rho_nh^2+C_{\sup,2}\rho_nTh(1-\gamma_T)+o(\rho_nh^2+\rho_nTh(1-\gamma_T)),\end{aligned} \tag{A.19}$$

and that

$$\begin{aligned}\mathrm{Var}(R_{f4}) &= M_{\nabla f_4}^2\kappa_{c_i^t}^{-2}\rho_n^2\mathrm{Var}\left(\hat{\nu}_4-\nu_4\right)+o\left(\kappa_{c_i^t}^{-2}\rho_n^2\mathrm{Var}(\hat{\nu}_4-\nu_4)\right)\\ &\le M_{\nabla f_4}^2\kappa_{c_i^t}^{-2}\rho_n^2\times\frac{C_{\sup,2}\kappa_{c_i^t}}{n^2\rho_nTh}+O\left(\kappa_{c_i^t}^{-2}\rho_n^2\times\frac{\kappa_{c_i^t}}{n^2Th}\right)+o\left(\kappa_{c_i^t}^{-2}\rho_n^2\times\frac{\kappa_{c_i^t}}{n^2\rho_nTh}\right)\\ &= \frac{M_{\nabla f_4}^2C_{\sup,2}\rho_n}{n^2\kappa_{c_i^t}Th}+o\left(\frac{\rho_n}{n^2\kappa_{c_i^t}Th}\right).\end{aligned} \tag{A.20}$$

Note that, analogously to R_{f1} and R_{f4}, we define the constants $M_{\nabla f_2}$ and $M_{\nabla f_5}$ corresponding to R_{f2} and R_{f5} as follows,

$$M_{\nabla f_2}=\frac{\nu_1\nu_3}{\nu_4\nu_5}\quad\text{and}\quad M_{\nabla f_5}=\frac{\nu_1\nu_2\times\left(\kappa_{c_i^t}\kappa_{c_j^t}\right)^{-1}\nu_3}{\kappa_{c_i^t}^{-1}\nu_4\kappa_{c_j^t}^{-2}\nu_5^2}.$$

Therefore, taking the expectation on both sides of equation (A.10), and then combining (A.13), (A.16) and (A.19), the bias of $\tilde{\tau}_{ij,h}(u^t)$ is determined as

$$\begin{aligned}\mathbb{E}(\tilde{\tau}_{ij,h}(u^t))-\tau_{ij}(u^t) &= \mathbb{E}(R_{f1})+\mathbb{E}(R_{f2})+\mathbb{E}(R_{f3})-\mathbb{E}(R_{f4})-\mathbb{E}(R_{f5})\\ &\le (|M_{\nabla f_1}C_{\sup,1}|+|M_{\nabla f_2}C_{\sup,1}|+|M_{\nabla f_3}C_{\sup,1}|+|M_{\nabla f_4}C_{\sup,1}|\\ &\quad +|M_{\nabla f_5}C_{\sup,1}|)\,\rho_nh^2+(|M_{\nabla f_1}C_{\sup,2}|+|M_{\nabla f_2}C_{\sup,2}|+|M_{\nabla f_3}C_{\sup,2}|\\ &\quad +|M_{\nabla f_4}C_{\sup,2}|+|M_{\nabla f_5}C_{\sup,2}|)\,\rho_nTh(1-\gamma_T)+o(\rho_nh^2+\rho_nTh(1-\gamma_T))\\ &= C_{\tau,1}\rho_nh^2+C_{\tau,2}\rho_nTh(1-\gamma_T)+o(\rho_nh^2+\rho_nTh(1-\gamma_T)),\end{aligned} \tag{A.21}$$

where the constant $C_{\tau,1}$ depends on $C_{\sup,1}$, while the constant $C_{\tau,2}$ depends on $C_{\sup,2}$.

We next investigate the variance of $\tilde{\tau}_{ij,h}(u^t)$. Using equation (A.10), for any $i\ne j\in\{1,\dots,n\}$, the variance can be expressed as,

$$\begin{aligned}\mathrm{Var}(\tilde{\tau}_{ij,h}(u^t)) &= \mathrm{Var}\left(R_{f1}+R_{f2}+R_{f3}-R_{f4}-R_{f5}\right)\le\mathbb{E}\left(\sum_{m=1}^{5}|R_{fm}-\mathbb{E}(R_{fm})|\right)^2\\ &\le 5\sum_{m=1}^{5}\mathbb{E}\left(R_{fm}-\mathbb{E}(R_{fm})\right)^2=5\sum_{m=1}^{5}\mathrm{Var}\left(R_{fm}\right).\end{aligned} \tag{A.22}$$

Note that the C_r inequality states that $\left(\sum_{i=1}^n|x_i|\right)^r\le C_r\sum_{i=1}^n|x_i|^r$ with $C_r=n^{r-1}$ for any $r>1$ (see, e.g. von Bahr and Esseen 1965). Thus, we apply the C_r inequality by setting $r=2$ and $n=5$ to conclude the inequality in (A.22). Next combining (A.14), (A.17) and (A.20), inequality (A.22) implies that

$$\begin{aligned}&\mathrm{Var}(\tilde{\tau}_{ij,h}(t/T))\\ &\le\frac{5M_{\nabla f_1}^2C_{\sup,2}\rho_n}{nTh}+\frac{5M_{\nabla f_2}^2C_{\sup,2}\rho_n}{nTh}+\frac{5M_{\nabla f_3}^2C_{\sup,2}\rho_n}{n^2\kappa_{c_i^t}\kappa_{c_j^t}Th}+\frac{5M_{\nabla f_4}^2C_{\sup,2}\rho_n}{n^2\kappa_{c_i^t}Th}+\frac{5M_{\nabla f_5}^2C_{\sup,2}\rho_n}{n^2\kappa_{c_j^t}Th}\\ &\quad+o\left(\frac{\rho_n}{nTh}\right)+o\left(\frac{\rho_n}{n^2\kappa_{c_i^t}\kappa_{c_j^t}Th}\right)+o\left(\frac{\rho_n}{n^2\kappa_{c_i^t}Th}\right)+o\left(\frac{\rho_n}{n^2\kappa_{c_j^t}Th}\right)\\ &\le\frac{5(M_{\nabla f_1}^2C_{\sup,2}+M_{\nabla f_2}^2C_{\sup,2})\rho_n}{nTh}+\frac{5M_{\nabla f_3}^2C_{\sup,2}\rho_n}{n^2\kappa_n^2Th}\\ &\quad+\frac{5(M_{\nabla f_4}^2C_{\sup,2}+M_{\nabla f_5}^2C_{\sup,2})\rho_n}{n^2\kappa_nTh}+o\left(\frac{\rho_n}{nTh}\right)+o\left(\frac{\rho_n}{n^2\kappa_n^2Th}\right),\end{aligned}$$

where we define $\kappa_n = \min_{i,t} n_{c_i^t}(\boldsymbol{c}^t)/n$ to measure the level of balance in community sizes. The last inequality holds as $\kappa_n \le \kappa_{c_i^t}, \forall i \in \{1,\dots,n\}$. Additionally, since $\kappa_n \le 1$ and $n\kappa_n \ge 1$, we have that $\frac{\rho_n}{nTh} \le \frac{\rho_n}{n\kappa_n Th}$, $\frac{\rho_n}{n^2\kappa_n^2 Th} \le \frac{\rho_n}{n\kappa_n Th}$ and $\frac{\rho_n}{n^2\kappa_n Th} \le \frac{\rho_n}{n\kappa_n Th}$. Therefore, all leading terms can be bounded by $\frac{C\rho_n}{n\kappa_n Th}$ for some constant C. Then the variance of $\tilde{\tau}_{ij,h}(u^t)$ satisfies

$$\operatorname{Var}(\tilde{\tau}_{ij,h}(u^t)) \le \frac{C_{\tau,3}\rho_n}{n\kappa_n Th} + o\left(\frac{\rho_n}{n\kappa_n Th}\right), \tag{A.23}$$

where the constant $C_{\tau,3}$ depends on $C_{\sup,2}$. To investigate the essential issue of the impact of network sparsity on parameter estimation, we consider the scaled quantity $(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t))$ since $\tau_{ij}(u^t)$ is proportional to ρ_n. Recalling the definition of $d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t))$, we have that

$$\begin{aligned} d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t)) =& \frac{1}{n^2\rho_n^2}\sum_{i,j=1}^n \left\{\mathbb{E}\left[\tilde{\tau}_{ij,h}(u^t) - \mathbb{E}(\tilde{\tau}_{ij,h}(u^t))\right]^2 + \left[\mathbb{E}\left(\tilde{\tau}_{ij,h}(u^t)\right) - \tau_{ij}(u^t)\right]^2\right\} \\ =& \Delta_{1,h}^t + \Delta_{2,h}^t \ , \end{aligned}$$

where $\Delta_{1,h}^t = \frac{1}{n^2\rho_n^2}\sum_{i,j=1}^n \mathbb{E}\left[\tilde{\tau}_{ij,h}(u^t) - \mathbb{E}(\tilde{\tau}_{ij,h}(u^t))\right]^2, \Delta_{2,h}^t = \frac{1}{n^2\rho_n^2}\sum_{i,j=1}^n \left[\mathbb{E}\left(\tilde{\tau}_{ij,h}(u^t)\right) - \tau_{ij}(u^t)\right]^2$. Note that $\Delta_{1,h}^t$ and $\Delta_{2,h}^t$ represent the variance and the squared bias, respectively. Combining (A.21) and (A.23), we obtain that the leading term of $d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t))$ implies

$$d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(u^t), \rho_n^{-1}\boldsymbol{\tau}(u^t)) \le C_{\tau,1}h^4 + C_{\tau,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\tau,3}}{n\kappa_n\rho_n Th} \ .$$

□

B Proofs of the results from Section 4

Proof of Theorem 3. This proof is similar to the proof of Proposition 1. As $\hat{\theta}_{i,h}^t$, defined in (9) can be seen as a function of $\hat{\nu}_1 = \frac{\hat{d}_{i,h}^t}{n\rho_n}$ and $\hat{\nu}_4 = \frac{\hat{O}_{c_i^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}$, we have that

$$\hat{\theta}_{i,h}^t = g(\hat{\nu}_1, \hat{\nu}_4) = \frac{n_{c_i^t}(\boldsymbol{c}^t)\hat{\nu}_1}{n\hat{\nu}_4} = \frac{\hat{\nu}_1}{\kappa_{c_i^t}^{-1}\hat{\nu}_4}.$$

Note that, conditional on $\boldsymbol{c}^t$ and under the assumption $\sum_{s\in\mathcal{N}_{c_i^t}^t}\theta_s^t = n_{c_i^t}(\boldsymbol{c}^t)$, we have that

$$g(\nu_1,\nu_4) = \frac{n_{c_i^t}(\boldsymbol{c}^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(O_{c_i^t}(\boldsymbol{c}^t)\right)} = \frac{n_{c_i^t}(\boldsymbol{c}^t)\sum_{j=1}^n\theta_i^t\theta_j^t Z_{c_i^tc_j^t}^t}{\sum_{s\in\mathcal{N}_{c_i^t}^t}\sum_{j=1}^n\theta_s^t\theta_j^t Z_{c_i^tc_j^t}^t} = \frac{n_{c_i^t}(\boldsymbol{c}^t)\times\theta_i^t\left(\sum_{j=1}^n\theta_j^t Z_{c_i^tc_j^t}^t\right)}{\left(\sum_{s\in\mathcal{N}_{c_i^t}^t}\theta_s^t\right)\left(\sum_{j=1}^n\theta_j^t Z_{c_i^tc_j^t}^t\right)} = \theta_i^t \ . \tag{B.1}$$

We use again a Taylor expansion of $\hat{\theta}_{i,h}^t$ around the point $(\nu_1,\nu_4)^\top$ implying that

$$\begin{aligned} \hat{\theta}_{i,h}^t =& g(\nu_1,\nu_4) + \nabla g(\xi_1',\xi_4')^\top\begin{pmatrix}\hat{\nu}_1-\nu_1\\ \hat{\nu}_4-\nu_4\end{pmatrix} \\ =& \theta_i^t + \frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}(\hat{\nu}_1-\nu_1) - \frac{\xi_1'}{\kappa_{c_i^t}^{-1}(\xi_4')^2}(\hat{\nu}_4-\nu_4) \\ =& \theta_i^t + R_{g1} - R_{g4} \end{aligned} \tag{B.2}$$

where the term $\nabla g(\xi_1',\xi_4')^\top\binom{\hat{\nu}_1-\nu_1}{\hat{\nu}_4-\nu_4}$ denotes the Taylor remainder term for $\hat{\theta}_{i,h}^t$, and $\nabla g(\xi_1',\xi_4')$ is the vector of first derivatives of $g(\cdot)$ evaluated in $(\xi_1',\xi_4')^\top$ with $\xi_m' = \hat{\nu}_m + \alpha'(\hat{\nu}_m - \nu_m)$ for some $\alpha' \in (0,1)$ and any $m \in \{1,4\}$. We then rewrite the remainder term as $(R_{g1} - R_{g4})$ where

$$R_{g1} = \frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}(\hat{\nu}_1-\nu_1), \quad R_{g4} = \frac{\xi_1'}{\kappa_{c_i^t}^{-1}(\xi_4')^2}(\hat{\nu}_4-\nu_4).$$

In analogy to the Taylor expansion for $f(\hat{\boldsymbol{\nu}})$, we analyse next each of the terms R_{g1} and R_{g4}.

Given that $\xi'_m \in (\hat{\nu}_m, \nu_m), \forall m \in \{1,4\}$, we conclude from Lemma A.2 that $(\xi'_1, \xi'_4)^\top$ converges to $(\nu_1, \nu_4)^\top$ in quadratic mean provided that $\gamma_T \to 1,\ \rho_n, h \to 0$ as $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$. Considering that $\kappa_{c_i^t}^{-1}\nu_4$ is a non-zero constant, as stated in Remark 7, we have that

$$\frac{1}{\kappa_{c_i^t}^{-1}\xi'_4} \xrightarrow{P} \frac{1}{\kappa_{c_i^t}^{-1}\nu_4}, \quad \frac{1}{\kappa_{c_i^t}^{-2}(\xi'_4)^2} \xrightarrow{P} \frac{1}{\kappa_{c_i^t}^{-2}\nu_4^2} \quad \text{as } \gamma_T \to 1,\ \rho_n, h \to 0,\ n\rho_n Th \to \infty \text{ and } Th(1-\gamma_T) \to 0. \tag{B.3}$$

For R_{g1}, letting $M_{\nabla g_1} = \frac{1}{\kappa_{c_i^t}^{-1}\nu_4}$, we have that

$$R_{g1} = \frac{1}{\kappa_{c_i^t}^{-1}\xi'_4}(\hat{\nu}_1 - \nu_1) = [M_{\nabla g_1} + o_p(1)]\,(\hat{\nu}_1 - \nu_1) = M_{\nabla g_1}(\hat{\nu}_1 - \nu_1) + o_p(\hat{\nu}_1 - \nu_1).$$

Taking the expectation on both sides and plugging in the expectation bound from (A.3), we obtain that

$$\mathbb{E}(R_{g1}) = M_{\nabla g_1}\mathbb{E}(\hat{\nu}_1 - \nu_1) + o\left(\mathbb{E}(\hat{\nu}_1 - \nu_1)\right) \le M_{\nabla g_1}C_{\sup,1}h^2 + M_{\nabla g_1}C_{\sup,2}Th(1-\gamma_T) + o\left(h^2 + Th(1-\gamma_T)\right), \tag{B.4}$$

furthermore, by (A.4),

$$\begin{aligned}
\mathrm{Var}(R_{g1}) &= M_{\nabla g_1}^2 \mathrm{Var}(\hat{\nu}_1 - \nu_1) + o\left(\mathrm{Var}(\hat{\nu}_1 - \nu_1)\right) \\
&\le \frac{M_{\nabla g_1}^2 C_{\sup,2}}{n\rho_n Th} + O\left(\frac{1}{nTh}\right) + o\left(\frac{1}{n\rho_n Th}\right) \\
&= \frac{M_{\nabla g_1}^2 C_{\sup,2}}{n\rho_n Th} + o\left(\frac{1}{n\rho_n Th}\right).
\end{aligned} \tag{B.5}$$

As for R_{g4}, applying Slutsky's Theorem once again, expression (B.3) yields that

$$\begin{aligned}
R_{g4} &= \frac{\xi'_1}{\kappa_{c_i^t}^{-1}(\xi'_4)^2}(\hat{\nu}_4 - \nu_4) = \frac{\xi'_1}{\kappa_{c_i^t}^{-2}(\xi'_4)^2} \times \kappa_{c_i^t}^{-1}(\hat{\nu}_4 - \nu_4) \\
&= \left[\frac{\nu_1}{\kappa_{c_i^t}^{-2}\nu_4^2} + o_p(1)\right] \times \kappa_{c_i^t}^{-1}(\hat{\nu}_4 - \nu_4) \\
&= M_{\nabla g_4}\kappa_{c_i^t}^{-1}(\hat{\nu}_4 - \nu_4) + o_p\left(\kappa_{c_i^t}^{-1}(\hat{\nu}_4 - \nu_4)\right),
\end{aligned}$$

where $M_{\nabla g_4} = \frac{\nu_1}{\kappa_{c_i^t}^{-2}\nu_4^2}$. Then, by (A.7) and (A.8), the expectation and variance can be obtained as follows,

$$\begin{aligned}
\mathbb{E}(R_{g4}) &= M_{\nabla g_4}\kappa_{c_i^t}^{-1}\mathbb{E}(\hat{\nu}_4 - \nu_4) + o\left(\kappa_{c_i^t}^{-1}\mathbb{E}(\hat{\nu}_4 - \nu_4)\right) \\
&\le M_{\nabla g_4}C_{\sup,1}h^2 + M_{\nabla g_4}C_{\sup,2}Th(1-\gamma_T) + o\left(h^2 + Th(1-\gamma_T)\right),
\end{aligned} \tag{B.6}$$

and

$$\begin{aligned}
\mathrm{Var}(R_{g4}) &= M_{\nabla g_4}^2\kappa_{c_i^t}^{-2}\mathrm{Var}(\hat{\nu}_4 - \nu_4) + o\left(\kappa_{c_i^t}^{-2}\mathrm{Var}(\hat{\nu}_4 - \nu_4)\right) \\
&\le M_{\nabla g_4}^2\kappa_{c_i^t}^{-2} \times \frac{C_{\sup,2}\kappa_{c_i^t}}{n^2\rho_n Th} + O\left(\kappa_{c_i^t}^{-2} \times \frac{\kappa_{c_i^t}}{n^2 Th}\right) + o\left(\kappa_{c_i^t}^{-2} \times \frac{\kappa_{c_i^t}}{n^2\rho_n Th}\right) \\
&= \frac{M_{\nabla g_4}^2 C_{\sup,2}}{n^2\kappa_{c_i^t}\rho_n Th} + o\left(\frac{1}{n^2\kappa_{c_i^t}\rho_n Th}\right).
\end{aligned} \tag{B.7}$$

Taking expectation on both sides in (B.2) and combining (B.4) and (B.6), we have that

$$\begin{aligned}
\mathbb{E}(\hat{\theta}_{i,h}^t) - \theta_i^t &= \mathbb{E}(R_{g1}) - \mathbb{E}(R_{g4}) \\
&\le \left[|M_{\nabla g_1}C_{\sup,1}| + |M_{\nabla g_4}C_{\sup,1}|\right]h^2 \\
&\quad + \left[|M_{\nabla g_1}C_{\sup,2}| + |M_{\nabla g_4}C_{\sup,2}|\right]Th(1-\gamma_T) + o\left(h^2 + Th(1-\gamma_T)\right) \\
&= C_{\theta,1}h^2 + C_{\theta,2}Th(1-\gamma_T) + o(h^2 + Th(1-\gamma_T)),
\end{aligned}$$

where the constant $C_{\theta,1}$ depends on $C_{\sup,1}$ while the constant $C_{\theta,2}$ depends on $C_{\sup,2}$.

Using again the C_r inequality with $r=2$ and $n=2$, we have $\mathrm{Var}(x_1+x_2) \le 2\big(\mathrm{Var}(x_1)+\mathrm{Var}(x_2)\big)$, and from (B.2), we conclude that

$$\mathrm{Var}(\hat{\theta}^t_{i,h}) = \mathrm{Var}\left(R_{g1}-R_{g4}\right) \le 2\left[\mathrm{Var}(R_{g1})+\mathrm{Var}(R_{g4})\right],$$

and then substitute (B.5) and (B.7) to get that

$$\begin{aligned}\mathrm{Var}(\hat{\theta}^t_{i,h}) &\le \frac{2M^2_{\nabla g_1}C_{\sup,2}}{n\rho_n Th} + \frac{2M^2_{\nabla g_4}C_{\sup,2}}{n^2\kappa_{c_i^t}\rho_n Th} + o\left(\frac{1}{n\rho_n Th}\right) + o\left(\frac{1}{n^2\kappa_{c_i^t}\rho_n Th}\right)\\ &= \frac{C_{\theta,3}}{n\rho_n Th} + o\left(\frac{1}{n\rho_n Th}\right),\end{aligned}$$

where the constant $C_{\theta,3}$ depends on $C_{\sup,2}$. Note that the last equality holds due to $n \le n^2\kappa_{c_i^t}$ for any $\kappa_{c_i^t}$. We now turn back to the distance $d(\hat{\boldsymbol{\theta}}^t_h, \boldsymbol{\theta}^t)$, which by the bias and variance of $\hat{\theta}^t_{i,h}$, satisfies

$$d(\hat{\boldsymbol{\theta}}^t_h, \boldsymbol{\theta}^t) = \frac{1}{n}\sum_{i=1}^n \left[\mathbb{E}\left(\hat{\theta}^t_{i,h} - \mathbb{E}(\hat{\theta}^t_{i,h})\right)^2 + \left(\mathbb{E}(\hat{\theta}^t_{i,h}) - \theta^t_i\right)^2\right] \le C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th},$$

where the constant $C_{\theta,1}$ depends on $C_{\sup,1}$, while $C_{\theta,2}, C_{\theta,3}$ depend on $C_{\sup,2}$. □

Proof of Lemma 4. For any generic matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, let $\|\boldsymbol{A}\|_\infty = \max_{i,j=1,\dots,n}|A_{ij}|$ and $\|\boldsymbol{A}\|_1 = \sum_{i,j=1}^n |A_{ij}|$. Since $Q(\cdot)$ is a Lipschitz function in all arguments, by this Lipschitz continuity and the triangle inequality, we have that

$$\begin{aligned}\left|Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(H(\boldsymbol{e}^t)\right)\right| &\le L\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty\\ &\le L\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty + L\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty \qquad \text{(B.8)}\end{aligned}$$

where L is the constant occurring in the Lipschitz continuity of $Q(\cdot)$. To show (19), we need to establish uniform convergence of both ℓ_∞ norms over $\boldsymbol{e}^t \in \{1,\dots,K\}^n$.

Bernstein's inequality states that if Y_i's are independent, $|Y_i| \le B, \mathbb{E}(Y_i)=0$ and $S_n = \sum_{i=1}^n Y_i$, then it holds that $\mathbb{P}(|S_n| \ge w) \le 2\exp\left\{-\frac{w^2/2}{\mathrm{Var}(S_n)+Bw/3}\right\}$. Relying on Bernstein's inequality, we deal with the first term. Combining the definitions of $\hat{\tau}_{ij,h}(u^t)$ and $\hat{O}_{kl,h}(\boldsymbol{e}^t)$ and letting

$$S_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^n\sum_{t'=1}^T A^{t'}_{ij}\mathbf{1}_{\{e^t_i=k, e^t_j=l\}}\mathbf{1}_{\{u^{t'}\in U(u^t)\}},$$

we have $\hat{O}_{kl,h}(\boldsymbol{e}^t)/(n^2\rho_n) = S_{kl}(\boldsymbol{e}^t)/(2n^2\rho_n Th)$. Notice that A^t_{ij}'s are independent for all time points and $|A^t_{ij}| \le 1$. For any fixed $\varepsilon > 0$, letting $B=2$, $w = 2n^2\rho_n Th\varepsilon$ and $\alpha = \frac{3\varepsilon^2}{3C_{\sup,2}+2\varepsilon}$, we then have that

$$\begin{aligned}\mathbb{P}\left(\left|\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n}\right| \ge \varepsilon\right) &= \mathbb{P}\left(\left|\frac{S_{kl}(\boldsymbol{e}^t)}{2n^2\rho_n Th} - \frac{\mathbb{E}(S_{kl}(\boldsymbol{e}^t))}{2n^2\rho_n Th}\right| \ge \varepsilon\right)\\ &= \mathbb{P}\left(\left|S_{kl}(\boldsymbol{e}^t) - \mathbb{E}(S_{kl}(\boldsymbol{e}^t))\right| \ge 2n^2\rho_n Th\varepsilon\right)\\ &\le 2\exp\left\{-\frac{2\varepsilon^2 n^4\rho_n^2T^2h^2}{\mathrm{Var}(S_{kl}(\boldsymbol{e}^t)) + 4n^2\rho_n Th\varepsilon/3}\right\}\\ &\le 2\exp\left\{-\frac{\varepsilon^2 n^4\rho_n^2T^2h^2}{C_{\sup,2}n^2\rho_n Th + 2n^2\rho_n Th\varepsilon/3}\right\} \le 2\exp\left\{-\alpha n^2\rho_n Th\right\}\end{aligned} \qquad \text{(B.9)}$$

given that

$$\mathrm{Var}(S_{kl}(\boldsymbol{e}^t)) = \sum_{i,j=1}^n\sum_{t'=1}^T \mathrm{Var}(A^{t'}_{ij})\mathbf{1}_{\{e^t_i=k, e^t_j=l\}}\mathbf{1}_{\{u^{t'}\in U(u^t)\}} \le 2C_{\sup,2}n^2\rho_n Th.$$

Since there exist K^n possible assignments of $\boldsymbol{e}^t$ and K^2 elements in the matrix $\hat{O}_h(\boldsymbol{e}^t)$, using the union bound, we conclude that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{\boldsymbol{e}^t}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty \geq \varepsilon\right) \leq 2K^{n+2}\exp\left\{-\alpha n^2\rho_n Th\right\} = 2K^2\left[K\exp\left\{-\alpha n\rho_n Th\right\}\right]^n \to 0, \tag{B.10}$$

where the last convergence follows when $n\rho_n Th \to \infty$. This means that the first term of (B.8) converges to zero in probability if $n\rho_n Th \to \infty$. To inspect the second term of (B.8), by (A.5) we have that

$$\begin{aligned}\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty &\leq \left\|\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) - H_{kl}(\boldsymbol{e}^t)\right\|_1 \\ &\leq \sum_{k,l=1}^K \tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)\left(C_{\sup,1}h^2 + C_{\sup,2}Th(1-\gamma_T) + O\left(h^3\right)\right) = O(h^2 + Th(1-\gamma_T)) \to 0,\end{aligned} \tag{B.11}$$

as $\gamma_T \to 1$ and $h \to 0$ such that $Th(1-\gamma_T) \to 0$. Combining (B.10) and (B.11) implies that the left-hand side of (B.8) converges to zero in probability if $\gamma_T \to 1$, $\rho_n, h \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$. □

Proof of Lemma 5. Recalling the definitions of $H_{kl}(\boldsymbol{e}^t)$ and $H_k(\boldsymbol{e}^t)$, we note that when $\boldsymbol{e}^t = \boldsymbol{c}^t$,

$$H_{kl}(\boldsymbol{c}^t) = \frac{n_k(\boldsymbol{c}^t)n_l(\boldsymbol{c}^t)W_{kl}^t}{n^2}, \qquad H_k(\boldsymbol{c}^t) = \frac{n_k(\boldsymbol{c}^t)\sum_{l=1}^K n_l(\boldsymbol{c}^t)W_{kl}^t}{n^2}.$$

Let $G_{kl}^t = \frac{H_{kl}(\boldsymbol{e}^t)}{H_k(\boldsymbol{e}^t)H_l(\boldsymbol{e}^t)}$ and $B_{ab}^t = \frac{n^2 W_{ab}^t}{(\sum_{s=1}^K n_s(\boldsymbol{c}^t)W_{as}^t)(\sum_{s'=1}^K n_{s'}(\boldsymbol{c}^t)W_{bs'}^t)}$. Applying that $\arg\max_G(C_1' \log G - C_2' G) = C_1'/C_2'$, we can obtain that for any generic vector $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$,

$$\begin{aligned} Q\left(H(\boldsymbol{e}^t)\right) &= \sum_{k,l=1}^K \Bigg[\frac{\sum_{a,b=1}^K \Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)W_{ab}^t}{n^2} \times \log G_{kl}^t \\ &\qquad - \frac{\sum_{a,s=1}^K n_s(\boldsymbol{c}^t)W_{as}^t\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n^2} \times \frac{\sum_{b,s'=1}^K n_{s'}(\boldsymbol{c}^t)W_{bs'}^t\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n^2} \times G_{kl}^t\Bigg] \\ &= \frac{1}{n^2}\sum_{k,l=1}^K\sum_{a,b=1}^K \Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\left[W_{ab}^t\log G_{kl}^t - \frac{W_{ab}^t}{B_{ab}^t}\times G_{kl}^t\right] \\ &\leq \frac{1}{n^2}\sum_{k,l=1}^K\sum_{a,b=1}^K \Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\left[W_{ab}^t\log B_{ab}^t - W_{ab}^t\right] \\ &= \sum_{a,b=1}^K \Bigg\{\frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2}\log\Bigg[\frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2}\times\left(\frac{n_a(\boldsymbol{c}^t)\sum_{s=1}^K n_s(\boldsymbol{c}^t)W_{as}^t}{n^2}\right)^{-1} \\ &\qquad \times\left(\frac{n_b(\boldsymbol{c}^t)\sum_{s'=1}^K n_{s'}(\boldsymbol{c}^t)W_{bs'}^t}{n^2}\right)^{-1}\Bigg] - \frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2}\Bigg\} \\ &= \sum_{a,b=1}^K\left[H_{ab}(\boldsymbol{c}^t)\log\frac{H_{ab}(\boldsymbol{c}^t)}{H_a(\boldsymbol{c}^t)H_b(\boldsymbol{c}^t)} - H_{ab}(\boldsymbol{c}^t)\right] = Q\left(H(\boldsymbol{c}^t)\right), \end{aligned} \tag{B.12}$$

where one goes from line 2 to line 3 by plugging in the maximizer of the function and where the equality holds if and only if $G_{kl}^t = B_{ab}^t$ whenever $\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t) > 0$ due to the uniqueness of the maximizer of this concave function $C_1'\log G - C_2' G$.

Next, we show that $Q\left(H(\boldsymbol{e}^t)\right) = Q\left(H(\boldsymbol{c}^t)\right)$ if and only if $\boldsymbol{e}^t = \boldsymbol{c}^t$. The 'if' part is immediate after plugging $\boldsymbol{e}^t = \boldsymbol{c}^t$ in the definition of $Q\left(H(\boldsymbol{e}^t)\right)$. Now we need to prove that

$$Q\left(H(\boldsymbol{e}^t)\right) = Q\left(H(\boldsymbol{c}^t)\right) \text{ implies the event } \{\boldsymbol{e}^t = \boldsymbol{c}^t\}.$$

Suppose that $Q(H(\boldsymbol{e}^t)) = Q(H(\boldsymbol{c}^t))$. Notice that when $\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) > 0$, the index set $\mathcal{N}^t_{a(k)}$ must be a non-empty set, indicating $n_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) > 0$. Thus, the equality in (B.12) gives that $G^t_{kl} = B^t_{ab}$ whenever $n_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)n_{b(l)}(\boldsymbol{c}^t, \boldsymbol{e}^t) > 0$ for $k, l, a, b \in \{1, \ldots, K\}$. Since $\boldsymbol{B}^t := (B^t_{ab})_{K\times K}$ and $\boldsymbol{G}^t := (G^t_{kl})_{K\times K}$ are constructed with $\boldsymbol{W}^t$ and $H(\boldsymbol{e}^t)$, both matrices $\boldsymbol{B}^t, \boldsymbol{G}^t$ are symmetric. Under Assumption 4, we can further conclude that $\boldsymbol{B}^t$ does not have two identical columns. Finally, combining the above conditions with Assumption 2, and subsequently applying Lemma A.2 from Zhao et al. (2012), the matrix $\boldsymbol{n}(\boldsymbol{c}^t, \boldsymbol{e}^t) := \left(n_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\right)_{K\times K}$ or its permutation is a diagonal matrix with all positive elements, which implies that all nodes are accurately assigned to their true communities (i.e. $\boldsymbol{c}^t = \boldsymbol{e}^t$). Hence, the sufficiency part is proved.

To finish this proof, we evaluate the bound on $Q(H(\boldsymbol{c}^t)) - Q(H(\boldsymbol{e}^t))$. Let

$$\left\|\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)\right\|_1 = \sum_{a,k=1}^{K} |\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)|.$$

Note by Assumptions (b) and (c) of Zhao et al. (2012) that $Q(\cdot)$ has continuous and bounded first derivatives in a neighborhood of $H(\boldsymbol{c}^t)$ implying that there exist some positive constants $M' \le M$ such that

$$\frac{1}{2}M'\left\|\frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t)}{n} - \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n}\right\|_1 \le Q\left(H(\boldsymbol{c}^t)\right) - Q\left(H(\boldsymbol{e}^t)\right) \le \frac{1}{2}M\left\|\frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t)}{n} - \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n}\right\|_1 \tag{B.13}$$

since, from (18), $H(\boldsymbol{e}^t)$ is a function depending on $\frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n}$. Next by the definition of $\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)$, it is easy to see that $\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t)$ is a diagonal matrix satisfying $\sum_{a=1}^{K}\Lambda_{a(a)}(\boldsymbol{c}^t, \boldsymbol{c}^t) = n$. Thus,

$$\begin{aligned}\left\|\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t) - \Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)\right\|_1 &= \sum_{a=1}^{K}\left|\Lambda_{a(a)}(\boldsymbol{c}^t, \boldsymbol{c}^t) - \Lambda_{a(a)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\right| + \sum_{a=1}^{K}\sum_{k\ne a}\left|\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{c}^t) - \Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\right| \\ &= 2\sum_{a=1}^{K}\sum_{k\ne a}\left|\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\right| = 2\sum_{i=1}^{n}\theta^t_i \mathbf{1}_{\{e^t_i \ne c^t_i\}}\end{aligned}$$

from which we can conclude that

$$2\theta_{\inf}\sum_{i=1}^{n}\mathbf{1}_{\{e^t_i\ne c^t_i\}} \le \left\|\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t) - \Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)\right\|_1 \le 2\theta_{\sup}\sum_{i=1}^{n}\mathbf{1}_{\{e^t_i\ne c^t_i\}} \tag{B.14}$$

with $\theta_{\inf} := \inf_{1\le i\le n}\inf_{u\in[0,1]}\theta_i(u)$ and $\theta_{\sup} := \sup_{1\le i\le n}\sup_{u\in[0,1]}\theta_i(u)$ and after plugging it into (B.13) inequality (21) follows. □

Lemma B.3. *For any $m_{\text{mis}} = 1, \ldots, n$ and any constant $C > 0$, if $\rho_n, h \to 0$ and $nT \to \infty$ such that $(n\rho_n Th)/\log n \to \infty$, then the following convergence holds,*

$$\mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n_\iota(\boldsymbol{e}^t, \boldsymbol{c}^t)\le m_{\text{mis}}\}}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}\left(\hat{O}_h(\boldsymbol{e}^t)\right)}{n^2\rho_n} + \frac{\mathbb{E}\left(\hat{O}_h(\boldsymbol{c}^t)\right)}{n^2\rho_n}\right\|_\infty > \frac{Cm_{\text{mis}}}{n}\right) \to 0. \tag{B.15}$$

Proof. The technique of proving this result is similar to Lemma A.1 of Zhao et al. (2012). Letting $S_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^{n}\sum_{t'=1}^{T}A^{t'}_{ij}\mathbf{1}_{\{e^t_i=k, e^t_j=l\}}\mathbf{1}_{\{u^{t'}\in U(u^t)\}}$, we have $\hat{O}_{kl,h}(\boldsymbol{e}^t)/(n^2\rho_n) = S_{kl}(\boldsymbol{e}^t)/(2n^2\rho_n Th)$. Note that when $e^t_1 \ne c^t_1, \ldots, e^t_{m_{\text{mis}}} \ne c^t_{m_{\text{mis}}}, e^t_{m_{\text{mis}}+1} = c^t_{m_{\text{mis}}+1}, \ldots, e^t_n = c^t_n$,

$$\begin{aligned}S_{kl}(\boldsymbol{e}^t) - S_{kl}(\boldsymbol{c}^t) = \sum_{t'=1}^{T}\mathbf{1}_{\{u^{t'}\in U(u^t)\}}\Bigg[&\sum_{i=1}^{m_{\text{mis}}}A^{t'}_{ii}\left(\mathbf{1}_{\{e^t_i=k\}} - \mathbf{1}_{\{c^t_i=k\}}\right) + 2\sum_{i<j}^{m_{\text{mis}}}A^{t'}_{ij}\left(\mathbf{1}_{\{e^t_i=k,e^t_j=l\}} - \mathbf{1}_{\{c^t_i=k,c^t_j=l\}}\right) \\ &+ 2\sum_{i=1}^{m_{\text{mis}}}\sum_{j=m_{\text{mis}}+1}^{n}A^{t'}_{ij}\left(\mathbf{1}_{\{e^t_i=k,e^t_j=l\}} - \mathbf{1}_{\{c^t_i=k,c^t_j=l\}}\right)\Bigg],\end{aligned}$$

$\text{Var}\left(S_{kl}(\boldsymbol{e}^t) - S_{kl}(\boldsymbol{c}^t)\right) \le 8C_{\sup,2}m_{\text{mis}}n\rho_n Th.$

Apply Bernstein's inequality again. Setting $B = 2$ and $w = 2n^2\rho_n Th\varepsilon$, for $\varepsilon \le 6C_{\mathrm{sup},2}m_{\mathrm{mis}}/n$,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\hat{O}_{kl,h}(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n} + \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{c}^t))}{n^2\rho_n}\right| > \varepsilon\right)\\
&= \mathbb{P}\left(\left|S_{kl}(\boldsymbol{e}^t) - S_{kl}(\boldsymbol{c}^t) - \mathbb{E}(S_{kl}(\boldsymbol{e}^t)) + \mathbb{E}(S_{kl}(\boldsymbol{c}^t))\right| > 2n^2\rho_n Th\varepsilon\right)\\
&\le 2\exp\left\{-\frac{2\varepsilon^2 n^4\rho_n^2 T^2h^2}{\mathrm{Var}\left(S_{kl}(\boldsymbol{e}^t) - S_{kl}(\boldsymbol{c}^t)\right) + 4n^2\rho_n Th\varepsilon/3}\right\}\\
&\le 2\exp\left\{-\frac{\varepsilon^2 n^4\rho_n^2 T^2h^2}{4C_{\mathrm{sup},2}m_{\mathrm{mis}}n\rho_n Th + 2n^2\rho_n Th\varepsilon/3}\right\}\\
&\le 2\exp\left\{-\frac{n^3\rho_n Th\varepsilon^2}{8C_{\mathrm{sup},2}m_{\mathrm{mis}}}\right\},
\end{aligned}
$$

on the other hand, for $\varepsilon > 6C_{\mathrm{sup},2}m_{\mathrm{mis}}/n$,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\hat{O}_{kl,h}(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n} + \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{c}^t))}{n^2\rho_n}\right| > \varepsilon\right)\\
&\le 2\exp\left\{-\frac{\varepsilon^2 n^4\rho_n^2 T^2h^2}{4C_{\mathrm{sup},2}m_{\mathrm{mis}}n\rho_n Th + 2n^2\rho_n Th\varepsilon/3}\right\}\\
&\le 2\exp\left\{-\frac{3n^2\rho_n Th\varepsilon}{4}\right\}.
\end{aligned}
$$

As $n\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \le m_{\mathrm{mis}}$, there exist $\binom{n}{m_{\mathrm{mis}}}K^{m_{\mathrm{mis}}}$ possible assignments of $\boldsymbol{e}^t$ and K^2 elements in the matrix $\hat{O}_h(\boldsymbol{e}^t)$. Define P_1 as the probability on the left-hand side of (B.15), and let $\varepsilon = (Cm_{\mathrm{mis}})/n$ for some constant C. Using the union bound, if $C \le 6C_{\mathrm{sup},2}$,

$$
\begin{aligned}
P_1 \le 2\binom{n}{m_{\mathrm{mis}}}K^{m_{\mathrm{mis}}+2}\exp\left\{-\frac{n^3\rho_n Th\varepsilon^2}{8C_{\mathrm{sup},2}m_{\mathrm{mis}}}\right\} &\le 2n^{m_{\mathrm{mis}}}K^{m_{\mathrm{mis}}+2}\exp\left\{-\frac{n^3\rho_n Th\varepsilon^2}{8C_{\mathrm{sup},2}m_{\mathrm{mis}}}\right\}\\
&= 2K^2\left[K\exp\left\{\log n - \frac{C^2 n\rho_n Th}{8C_{\mathrm{sup},2}}\right\}\right]^{m_{\mathrm{mis}}}\\
&\to 0
\end{aligned}
$$

as $(n\rho_n Th)/\log n \to \infty$. Otherwise, if $C > 6C_{\mathrm{sup},2}$,

$$
\begin{aligned}
P_1 \le 2\binom{n}{m_{\mathrm{mis}}}K^{m_{\mathrm{mis}}+2}\exp\left\{-\frac{3n^2\rho_n Th\varepsilon}{4}\right\} &\le 2n^{m_{\mathrm{mis}}}K^{m_{\mathrm{mis}}+2}\exp\left\{-\frac{3n^2\rho_n Th\varepsilon}{4}\right\}\\
&= 2K^2\left[K\exp\left\{\log n - \frac{3}{4}n\rho_n Th\right\}\right]^{m_{\mathrm{mis}}}\\
&\to 0
\end{aligned}
$$

as $(n\rho_n Th)/\log n \to \infty$. Consequently, the convergence (B.15) holds for any constant C if $(n\rho_n Th)/\log n \to \infty$. $\square$

Proof of Theorem 6. **Step 1:** we first show the weak consistency. Let $\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{e_i^t \neq c_i^t\}}$. As defined in (16), $\hat{\boldsymbol{c}}^t$ is the maximizer of $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$. Thus, in order to show weak consistency, we only need to show that there exists $\delta_n^t \to 0$ such that

$$
\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\ge\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) > Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right) \to 0. \tag{B.16}
$$

By the continuity and the uniqueness of $Q(H(\boldsymbol{e}^t))$ in Lemma 5, we can always find a specific ϵ_n^t such that

$Q(H(\boldsymbol{c}^t)) > \max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q(H(\boldsymbol{e}^t)) + 2\epsilon_n^t$. Then

$$
\begin{aligned}
&\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) > Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right)\\
&\leq \mathbb{P}\left(\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) + Q(H(\boldsymbol{c}^t)) > Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) + \max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q(H(\boldsymbol{e}^t)) + 2\epsilon_n^t\right)\\
&= \mathbb{P}\left(\left[Q(H(\boldsymbol{c}^t)) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right] + \left[\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - \max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q(H(\boldsymbol{e}^t))\right] > 2\epsilon_n^t\right)\\
&\leq \mathbb{P}\left(\left\{Q\left(H(\boldsymbol{c}^t)\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) > \epsilon_n^t\right\} \cup \left\{\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - \max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\geq\delta_n^t\}} Q(H(\boldsymbol{e}^t)) > \epsilon_n^t\right\}\right)\\
&\leq \mathbb{P}\left(\left|Q\left(H(\boldsymbol{c}^t)\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right| > \epsilon_n^t\right) + \mathbb{P}\left(\max_{\boldsymbol{e}^t}\left|Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q(H(\boldsymbol{e}^t))\right| > \epsilon_n^t\right) \to 0
\end{aligned}
$$

as $\gamma_T \to 1, \rho_n, h \to 0$ such that $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$, and where the last convergence holds by (19) in Lemma 4.

Step 2: To show strong consistency, we need to prove that

$$
\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)>0\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) > Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right) \to 0.
$$

Since in (B.16) we analyze the behavior of the $Q(\cdot)$ function on the set $\{\boldsymbol{e}^t : \iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \geq \delta_n^t\}$, we now focus on its complement set and it is thus sufficient to show that

$$
\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:1\leq n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) > 0\right) \to 0. \tag{B.17}
$$

Since $Q(\cdot)$ is a continuous function, we can establish from (A.5) that $Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)$ converges to $Q(H(\boldsymbol{e}^t))$ for any $\boldsymbol{e}^t \in \{1,\ldots,K\}^n$ as $\gamma_T \to 1, h \to 0$ and $n\rho_n Th \to \infty$ and $Th(1-\gamma_T) \to 0$. Also note that by Lemma 5, $Q(H(\boldsymbol{e}^t)) \leq Q(H(\boldsymbol{c}^t))$, thus there exists an η_n^t such that

$$
Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) \leq Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)
$$

also holds for any $\boldsymbol{e}^t$ when $\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty < \eta_n^t$ with T large enough. Using a first order Taylor expansion of $Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)$ around $\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}$ implies that

$$
\begin{aligned}
&Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\\
&= -\mathrm{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)\right)^\top \mathrm{vec}\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) + o\left(\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_\infty\right)\\
&\geq M_1\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_1 + o\left(\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_\infty\right),
\end{aligned} \tag{B.18}
$$

where $\mathrm{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)\right)$ is the vector of first derivatives of the function $Q(\cdot)$ evaluated in $\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}$ and M_1 is a positive constant since $Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) \geq 0$ as we just showed above. We now consider bounding the ℓ_1 norm $\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_1$ and rewrite it as $\|R_1 - R_2\|_1$ where

$$
R_1 = \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t) - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n} + H(\boldsymbol{c}^t), \qquad R_2 = H(\boldsymbol{c}^t) - H(\boldsymbol{e}^t).
$$

From the expectation in (A.5) the vanishing bias between $\frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n}$ and $H_{kl}(\boldsymbol{e}^t)$, for any $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, is given by

$$C_{\text{sup},1}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + C_{\text{sup},2}\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)Th(1-\gamma_T) + O(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3).$$

Thus, without loss of generality if we let $n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t) = m_{\text{mis}}$ and $e_1^t \neq c_1^t,\dots,e_{m_{\text{mis}}}^t \neq c_{m_{\text{mis}}}^t, e_{m_{\text{mis}}+1}^t = c_{m_{\text{mis}}+1}^t,\dots,e_n^t = c_n^t$,

$$\begin{aligned}
&\frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n} - H_{kl}(\boldsymbol{e}^t) - \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{c}^t))}{n^2\rho_n} + H_{kl}(\boldsymbol{c}^t)\\
&\quad\leq \frac{\rho_n(C_{\text{sup},1}h^2 + C_{\text{sup},2}Th(1-\gamma_T))}{n^2\rho_n}\sum_{i=1}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k\}} - \mathbf{1}_{\{c_i^t=k\}}\right) + \frac{O(h^3)}{n^2}\sum_{i=1}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k\}} - \mathbf{1}_{\{c_i^t=k\}}\right)\\
&\qquad + \frac{2\rho_n(C_{\text{sup},1}h^2 + C_{\text{sup},2}Th(1-\gamma_T))}{n^2\rho_n}\sum_{i<j}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\quad + \frac{O(h^3)}{n^2}\sum_{i<j}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\qquad + \frac{2\rho_n(C_{\text{sup},1}h^2 + C_{\text{sup},2}Th(1-\gamma_T))}{n^2\rho_n}\sum_{i=1}^{m_{\text{mis}}}\sum_{j=m_{\text{mis}}+1}^{n}\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\qquad + \frac{O(h^3)}{n^2}\sum_{i=1}^{m_{\text{mis}}}\sum_{j=m_{\text{mis}}+1}^{n}\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\quad\leq \frac{4m_{\text{mis}}(C_{\text{sup},1}h^2 + C_{\text{sup},2}Th(1-\gamma_T))}{n}\\
&\quad= o\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right),
\end{aligned}$$

since $\gamma_T \to 1,\ h \to 0$ as $n, T \to \infty$ satisfying $Th(1-\gamma_T) \to 0$. As such, $\|R_1\|_1 \leq K^2\|R_1\|_\infty = o\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right)$ given that K is fixed. Considering R_2, as, from (18), $H(\boldsymbol{e}^t)$ is a function depending on $\frac{\Lambda(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n}$, $\|R_2\|_1 \sim \iota(\boldsymbol{e}^t,\boldsymbol{c}^t)$ follows by substituting the bounds in (B.14). By the (reverse) triangle inequality we have that

$$\left|\|R_2\|_1 - \|R_1\|_1\right| \leq \|R_1 - R_2\|_1 \leq \|R_1\|_1 + \|R_2\|_1 \ ,$$

which implies that $\|R_1 - R_2\|_1 \sim \iota(\boldsymbol{e}^t,\boldsymbol{c}^t)$. Applying the same argument under the ℓ_∞ norm yields $\|R_1 - R_2\|_\infty \sim \iota(\boldsymbol{e}^t,\boldsymbol{c}^t)$. Therefore, plugging these quantities into (B.18) we conclude that there exists a positive constant M'' such that

$$Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) \geq M''\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\ , \tag{B.19}$$

when $\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty < \eta_n^t$ with T large enough for any $\boldsymbol{e}^t$. The remainder of our proof is similar to the proof of statement (1.1) in Bickel, Chen, Zhao, Levina, and Zhu (2015). For any generic vector $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, we use a Taylor expansion for $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$ around $\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}$ and then obtain that

$$\begin{aligned}
Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) &= Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) + \text{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\right)^\top \text{vec}\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\\
&\quad + O_p\left(\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty^2\right)
\end{aligned} \tag{B.20}$$

where $\text{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\right)$ represents the vector of first derivatives of the function $Q(\cdot)$ evaluated in $\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}$. Since $Q(\cdot)$ has continuous second derivatives in a neighborhood of $\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}$ as stated in Assumption (b) of Zhao et al. (2012), we have that

$$\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) = \nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) + O\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right). \tag{B.21}$$

Using the Taylor expansion from (B.20) for $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$ and $Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)$, then based on (B.21), we have that

$$
\begin{aligned}
& Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) \\
&= Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) \\
&= \operatorname{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)\right)^\top \operatorname{vec}\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} + \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) \\
&\quad + O\left(\iota(\boldsymbol{e}^t, \boldsymbol{c}^t)\right) \operatorname{vec}\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) + O_p\left(\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty^2\right) \\
&\quad + O_p\left(\left\|\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_\infty^2\right) + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) \\
&\le C_1 \left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} + \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_\infty + C_2\iota(\boldsymbol{e}^t, \boldsymbol{c}^t)\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty \\
&\quad + C_3\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty^2 + C_4\left\|\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_\infty^2 + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) \\
&= C_1I_1 + C_2I_2 + C_3I_3 + C_4I_4 + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right).
\end{aligned}
$$

Using this bound, we have that

$$
\begin{aligned}
& \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: 1\le n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}} \sum_{m=1}^4 C_m I_m + Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) > 0\right) \\
&\le \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: 1\le n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}} \sum_{m=1}^4 C_m I_m - M''\iota(\boldsymbol{e}^t,\boldsymbol{c}^t) > 0\right) \\
&\le \sum_{m=1}^3 \sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} I_m > \frac{M'' m_{\text{mis}}}{4C_m n}\right) + \sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(I_4 > \frac{M'' m_{\text{mis}}}{4C_4 n}\right),
\end{aligned}
$$

where the first inequality holds when plugging in the bound from (B.19). To show (B.17), it is enough to show that $\sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} I_m > \frac{M'' m_{\text{mis}}}{4C_m n}\right)$ for any $m = 1, 2, 3$, and $\sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(I_4 > \frac{M'' m_{\text{mis}}}{4C_4 n}\right)$ converge to 0.

For I_1: Let $C = \frac{M''}{4C_1}$ and $(n\rho_n Th)/\log n \to \infty$, then Lemma B.3 yields that

$$
\begin{aligned}
\sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} I_1 > \frac{M'' m_{\text{mis}}}{4C_1 n}\right) &\le \sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}} I_1 > \frac{M'' m_{\text{mis}}}{4C_1 n}\right) \\
&\le n \max_{m_{\text{mis}}=1,\ldots,n} \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}} I_1 > \frac{M'' m_{\text{mis}}}{4C_1 n}\right) \to 0
\end{aligned}
$$

due to the exponential decay implied by the use of Bernstein's inequality.

For I_2: Let $\varepsilon = \frac{M''}{4C_2}$ and apply the bound in (B.10), hence, as $(n\rho_n Th)/\log n \to \infty$,

$$
\sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} I_2 > \frac{M''}{4nC_2}\right) \le \sum_{m_{\text{mis}}=1}^n \mathbb{P}\left(\max_{\boldsymbol{e}^t}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty > \frac{M''}{4C_2}\right) \to 0.
$$

For I_3: For any $1 \le m_{\text{mis}} \le n$, set $\varepsilon = \sqrt{\frac{M'' m_{\text{mis}}}{4C_3 n}}$ and $\alpha = \frac{3\varepsilon^2}{3C_{\text{sup},2}+2\varepsilon}$. Applying the union bound, we conclude from inequality (B.9) that

$$\begin{aligned}
&\mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}-\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty > \sqrt{\frac{M''m_{\text{mis}}}{4C_3 n}}\right)\\
&\le 2\binom{n}{m_{\text{mis}}}K^{m_{\text{mis}}+2}\exp\left\{-\alpha n^2\rho_n Th\right\}\\
&\le 2n^{m_{\text{mis}}}K^{m_{\text{mis}}+2}\exp\left\{-\frac{3\varepsilon^2 n^2\rho_n Th}{3C_{\text{sup},2}+\sqrt{\frac{M''}{C_3}}}\right\}\\
&= 2K^2\left[K\exp\left\{\log n-\frac{3M''n\rho_n Th}{12C_{\text{sup},2}C_3+4\sqrt{M''C_3}}\right\}\right]^{m_{\text{mis}}}.
\end{aligned}$$

Thus, when $(n\rho_n Th)/\log n\to\infty$, we have that

$$\begin{aligned}
&\sum_{m_{\text{mis}}=1}^{n}\mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} I_3 > \frac{M''m_{\text{mis}}}{4C_3 n}\right)\\
&\le \sum_{m_{\text{mis}}=1}^{n}\mathbb{P}\left(\max_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}-\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty > \sqrt{\frac{M''m_{\text{mis}}}{4C_3 n}}\right)\to 0.
\end{aligned}$$

For I_4: For any $1\le m_{\text{mis}}\le n$, let $\varepsilon=\sqrt{\frac{M''m_{\text{mis}}}{4C_4 n}}$. Since (B.9) implies that

$$\mathbb{P}\left(I_4>\frac{M''m_{\text{mis}}}{4C_4 n}\right)\le 2K^2\exp\left\{-\frac{3\varepsilon^2 n^2\rho_n Th}{3C_{\text{sup},2}+\sqrt{\frac{M''}{C_4}}}\right\}=2K^2\exp\left\{-\frac{3M''m_{\text{mis}}n\rho_n Th}{12C_{\text{sup},2}C_4+4\sqrt{M''C_4}}\right\},$$

the convergence $\sum_{m_{\text{mis}}=1}^{n}\mathbb{P}\left(I_4>\frac{M''m_{\text{mis}}}{4C_4 n}\right)\to 0$ follows due to the exponential decay as $(n\rho_n Th)/\log n\to\infty$. □

Proof of Corollary 7. To proceed, we define the following fractions for an arbitrary vector $\boldsymbol{e}^t\in\{1,\dots,K\}^n$,

$$\vartheta_i^t(\boldsymbol{e}^t):=\frac{n_{e_i^t}(\boldsymbol{e}^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))},\qquad \psi_{ij}^t(\boldsymbol{e}^t):=\frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_j^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t e_j^t}(\boldsymbol{e}^t))}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_j^t}(\boldsymbol{e}^t))}.$$

Let $\boldsymbol{\vartheta}^t(\boldsymbol{e}^t):=(\vartheta_1^t(\boldsymbol{e}^t),\dots,\vartheta_n^t(\boldsymbol{e}^t))^\top$ and $\boldsymbol{\psi}^t(\boldsymbol{e}^t):=(\psi_{ij}^t(\boldsymbol{e}^t))_{n\times n}$. Observe from (B.1) and (A.9) that $\vartheta_i^t(\boldsymbol{c}^t)=\theta_i^t$ and $\psi_{ij}^t(\boldsymbol{c}^t)=\tau_{ij}(u^t)$ for any $i,j\in\{1,\dots,n\}$ when $\boldsymbol{e}^t=\boldsymbol{c}^t$, we thus have $\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)=\boldsymbol{\theta}^t$ and $\boldsymbol{\psi}^t(\boldsymbol{c}^t)=\boldsymbol{\tau}(u^t)$. Here we prove only the result for $\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)$, since the argument for $\tilde{\boldsymbol{\tau}}_h(u^t,\hat{\boldsymbol{c}}^t)$ follows in the same way. Notice that $d(\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t),\boldsymbol{\theta}^t)=\mathbb{E}\left(\frac{1}{n}\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\theta}^t\|_2^2\right)$. By the law of iterated expectations, we obtain that

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n}\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\theta}^t\|_2^2\right) &= \frac{1}{n}\mathbb{E}\left[\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\mid\hat{\boldsymbol{c}}^t\right)\right]\\
&= \frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\mid\hat{\boldsymbol{c}}^t\in\pi(\boldsymbol{c}^t)\right)\mathbb{P}(\hat{\boldsymbol{c}}^t\in\pi(\boldsymbol{c}^t))\\
&\qquad +\frac{1}{n}\sum_{\boldsymbol{e}^t\notin\pi(\boldsymbol{c}^t)}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\mid\hat{\boldsymbol{c}}^t=\boldsymbol{e}^t\right)\mathbb{P}(\hat{\boldsymbol{c}}^t=\boldsymbol{e}^t)\\
&= \frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\mathbb{P}(\iota(\hat{\boldsymbol{c}}^t,\boldsymbol{c}^t)=0)\\
&\qquad +\frac{1}{n}\sum_{m_{\text{mis}}=1}^{n}\sum_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\hat{\boldsymbol{c}}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\mid\hat{\boldsymbol{c}}^t=\boldsymbol{e}^t\right)\mathbb{P}(\hat{\boldsymbol{c}}^t=\boldsymbol{e}^t)\\
&= \frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\mathbb{P}(\iota(\hat{\boldsymbol{c}}^t,\boldsymbol{c}^t)=0)\\
&\qquad +\frac{1}{n}\sum_{m_{\text{mis}}=1}^{n}\sum_{\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\mathbb{P}(\hat{\boldsymbol{c}}^t=\boldsymbol{e}^t)\\
&= E_1+E_2,
\end{aligned}$$

where $\pi(\boldsymbol{c}^t)$ represents the set of all permutations of the true assignment vector $\boldsymbol{c}^t$.

For E_1: Theorem 1 states that, given $\boldsymbol{c}^t$, $\mathbb{E}\left(\frac{1}{n}\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{c}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\leq C_{\theta,1}h^4+C_{\theta,2}T^2h^2(1-\gamma_T)^2+\frac{C_{\theta,3}}{n\rho_n Th}$, impliying that the first term E_1 is bounded by

$$C_{\theta,1}h^4+C_{\theta,2}T^2h^2(1-\gamma_T)^2+\frac{C_{\theta,3}}{n\rho_n Th}, \tag{B.22}$$

owing to $\mathbb{P}(\iota(\hat{\boldsymbol{c}}^t,\boldsymbol{c}^t)=0)\leq 1$.

For E_2: Now the remaining task is to quantify the second term E_2. To this end, consider that

$$\frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\leq\frac{2}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{e}^t)\|_2^2\right)+\frac{2}{n}\|\boldsymbol{\vartheta}^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2. \tag{B.23}$$

Here $\boldsymbol{e}^t$ represents an arbitrary vector within the set $\{\boldsymbol{e}^t: n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}$ and thus, the second term is deterministic. This decomposition allows to separate the contribution of the deviation induced by the estimation error under the same community assignments, and the deviation induced by community misclassification. Following the same strategy as in the proof of Theorem 1, since Lemma A.2 holds for any vector $\boldsymbol{e}^t$, it is easy to show that

$$\frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{e}^t)\|_2^2\right)\leq C_{\theta,1}h^4+C_{\theta,2}T^2h^2(1-\gamma_T)^2+\frac{C_{\theta,3}}{n\rho_n Th}. \tag{B.24}$$

To bound the second term in (B.23), observe that for each node i the absolute deviation due to misclassification satisfies that

$$\begin{aligned}
|\vartheta_i^t(\boldsymbol{e}^t)-\vartheta_i^t(\boldsymbol{c}^t)| &= \left|\frac{n_{e_i^t}(\boldsymbol{e}^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))}-\frac{n_{c_i^t}(\boldsymbol{c}^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))}\right|\\
&= \frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))}\left[\left|n_{e_i^t}(\boldsymbol{e}^t)-n_{c_i^t}(\boldsymbol{c}^t)\right|+n_{c_i^t}(\boldsymbol{c}^t)\left|1-\frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))}\right|\right]\\
&= \frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))}\left[\left|n_{e_i^t}(\boldsymbol{e}^t)-n_{c_i^t}(\boldsymbol{c}^t)\right|+n_{c_i^t}(\boldsymbol{c}^t)\frac{\left|\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))-\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))\right|}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))}\right].
\end{aligned}$$

Consider that $n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)$ denotes the total number of misclassified nodes at time point t. The misclassification of these nodes can potentially affect at most $C_{\text{sup},2}n(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))\rho_n$ edges. Intuitively, it holds that

$$\left|n_{e_i^t}(\boldsymbol{e}^t)-n_{c_i^t}(\boldsymbol{c}^t)\right|\leq n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t),\qquad \left|\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))-\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))\right|\leq C_{\text{sup},2}n(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))\rho_n.$$

Thus,

$$|\vartheta_i^t(\boldsymbol{e}^t)-\vartheta_i^t(\boldsymbol{c}^t)|\leq\frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))}\left[n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)+\frac{C_{\text{sup},2}n_{c_i^t}(\boldsymbol{c}^t)n(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))\rho_n}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))}\right].$$

Since $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)=\sum_{j=1}^n\tau_{ij}(u^t)\leq C_{\text{sup},2}n\rho_n$ and $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{e_i^t}(\boldsymbol{e}^t))=\sum_{s,j=1}^n\tau_{sj}(u^t)\mathbf{1}_{\{e_s^t=e_i^t\}}\geq C_{\inf}n_{e_i^t}(\boldsymbol{e}^t)n\rho_n$, we further obtain that

$$|\vartheta_i^t(\boldsymbol{e}^t)-\vartheta_i^t(\boldsymbol{c}^t)|\leq\frac{C_{\text{sup},2}n\rho_n}{C_{\inf}n_{e_i^t}(\boldsymbol{e}^t)n\rho_n}\left[n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)+\frac{C_{\text{sup},2}n_{c_i^t}(\boldsymbol{c}^t)n(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))\rho_n}{C_{\inf}nn_{c_i^t}(\boldsymbol{c}^t)\rho_n}\right]=\frac{C'n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)}{n_{e_i^t}(\boldsymbol{e}^t)}$$

for some constant C'. Aggregating across all nodes yields that

$$\|\boldsymbol{\vartheta}^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2=\sum_{i=1}^n|\vartheta_i^t(\boldsymbol{e}^t)-\vartheta_i^t(\boldsymbol{c}^t)|^2\leq\sum_{i=1}^n\left(\frac{C'n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)}{n_{e_i^t}(\boldsymbol{e}^t)}\right)^2\leq C'^2n(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))^2. \tag{B.25}$$

Substituting (B.24) and (B.25) into (B.23) gives that

$$\frac{1}{n}\mathbb{E}\left(\|\hat{\boldsymbol{\theta}}_h^t(\boldsymbol{e}^t)-\boldsymbol{\vartheta}^t(\boldsymbol{c}^t)\|_2^2\right)\leq 2\left(C_{\theta,1}h^4+C_{\theta,2}T^2h^2(1-\gamma_T)^2+\frac{C_{\theta,3}}{n\rho_n Th}+C'^2(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))^2\right).$$

which implies that the term E_2 fulfills that

$$
\begin{aligned}
E_2 &\leq \sum_{m_{\text{mis}}=1}^{n} \sum_{\{\boldsymbol{e}^t : n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} 2\left(C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th} + C'^2(n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t))^2\right)\mathbb{P}(\hat{\boldsymbol{c}}^t = \boldsymbol{e}^t)\\
&= \sum_{m_{\text{mis}}=1}^{n} \sum_{\{\boldsymbol{e}^t : n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} 2\left(C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th} + C'^2 m_{\text{mis}}^2\right)\mathbb{P}(\hat{\boldsymbol{c}}^t = \boldsymbol{e}^t)\\
&\leq 2\left(C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th} + C'^2 n^2\right)\sum_{m_{\text{mis}}=1}^{n} \sum_{\{\boldsymbol{e}^t : n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}} \mathbb{P}(\hat{\boldsymbol{c}}^t = \boldsymbol{e}^t)\\
&= 2\left(C_{\theta,1}h^4 + C_{\theta,2}T^2h^2(1-\gamma_T)^2 + \frac{C_{\theta,3}}{n\rho_n Th} + C'^2 n^2\right)\left(1 - \mathbb{P}(\iota(\hat{\boldsymbol{c}}^t,\boldsymbol{c}^t)=0)\right).
\end{aligned}
$$

Theorem 2 ensures that exact recovery occurs with high probability

$$\mathbb{P}(\iota(\hat{\boldsymbol{c}}^t,\boldsymbol{c}^t)=0) \geq 1 - C_1' \exp\{C_2' \log n - C_3' n\rho_n Th\}$$

with some constants C_1', C_2', C_3'. Hence,

$$E_2 \leq 4C'^2 C_1' \exp\{(C_2'+2)\log n - C_3' n\rho_n Th\}.$$

which decays at an exponential rate, faster than that of E_1. □

C Additional simulation results

C.1 Estimation errors for edge probability matrices

Let $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ generally denote the edge probability estimates obtained from different methods. In this section, we assess the estimation accuracy of $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ using the scaled mean-squared error $d\left(\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t), \boldsymbol{\tau}(u^t, \boldsymbol{c}^t)\right)$ as defined in Section 3. For comparison, we also include the oracle estimator $\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$ computed with the true community assignments $\boldsymbol{c}^t$ as a benchmark. Table 3 reports the results for Scenarios 1-3 from Section 5 of the main text.

C.2 Unbalanced community sizes

Previous studies in the main text all involve balanced community sizes. In this section, we investigate the performance of our method for unbalanced community sizes. We follow the same data generating process as in Section 5 with $T = 100$ and $n = 100$, and keep the same smooth trajectories for degree parameters $\boldsymbol{\theta}^t$ and connectivity matrices $\boldsymbol{Z}^t$. The initial community assignments are drawn independently with $\mathbb{P}(c_i^1 = 1) = 0.8$ and $\mathbb{P}(c_i^1 = 2) = 0.2$. We then consider $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}$, and $\gamma \in \{1, 0.95\}$. When $\gamma = 1$, the community structure is time-invariant with 78 nodes in the first community and 22 nodes in the second community. When $\gamma = 0.95$, community assignments will evolve over time according to the same Markov switching process of Section 5. It is worth noting that since the community assignment process of each node is initialized away from its stationary distribution, the community proportions are nonstationary at early time points and gradually converge toward their stationary regime (at around $t = 40$). This setup therefore represents a setting with unbalanced communities and a nonstationary-to-stationary transition. The results are summarized in Table 4. Relative to Figure 3 and Table 1 in the main text, all methods are slightly sensitive to unbalanced community sizes, however, KD-(DC)-SoS and DC-MASE are more affected than TVDCSBM. This observation is aligned with the fact that the theoretical analyses of KD-(DC)-SoS and DC-MASE assume the community sizes to be roughly balanced (see Assumption 3 of Lin and Lei (2024) and Assumption 1 of Agterberg et al. (2025)). These discrepancies are gradually diminishing as ρ_n grows.

C.3 Number of communities

To evaluate the performance of our method under a more complex community structure, we extend the baseline simulation from Section 5 to the case with $K = 4$ communities. We first construct a four-block state transition matrix Γ and the baseline connectivity matrix $\boldsymbol{Z}^0$ as follows,

$$\Gamma = \begin{pmatrix} 0.95 & 0.02 & 0.02 & 0.01 \\ 0.02 & 0.95 & 0.01 & 0.02 \\ 0.02 & 0.01 & 0.95 & 0.02 \\ 0.01 & 0.02 & 0.02 & 0.95 \end{pmatrix}, \qquad \boldsymbol{Z}^0 = \begin{pmatrix} 1.3 & 0.7 & 0.75 & 0.8 \\ 0.7 & 1.3 & 0.7 & 0.75 \\ 0.75 & 0.7 & 1.3 & 0.8 \\ 0.8 & 0.75 & 0.8 & 1.3 \end{pmatrix},$$

and then retain the same way as described in Section 5 to simulate a sequence of $T = 100$ DCSBMs on $n = 100$ nodes where the degree parameters $\boldsymbol{\theta}^t$ follow a sinusoidal trajectory $f_{\boldsymbol{\theta}}(u^t) = 1.6 + 0.1\sin(2\pi u^t)$ for all $t \in \{1, \ldots, T\}$. The results are then shown in Table 5. We see that all methods have larger Hamming error rates in the four-community setting compared to the two-community case as shown in Figure 3 of Section 5. This deterioration is expected, as each community contains fewer nodes (about 25 when $n = 100$ and $K = 4$), which results in less information available for community recovery. Nevertheless, TVDCSBM is still able to achieve smaller Hamming error rates for community detection and lower estimation errors for the parameters than the competing methods.

C.4 Disassortative networks

As noted in the introduction, a positivity assumption imposed on $\boldsymbol{Z}^t$ in spectral clustering methods restricts the applicability to assortative networks. In contrast, our proposed method does not require any such positivity constraint and can therefore accommodate both assortative and disassortative structures. Previous simulations in this work mainly focus on assortative networks. To generalise now this, we design the following simulations to evaluate the performance of our method for disassortative networks. Specifically, we generate a sequence of $T = 100$ DCSBMs with $n = 100$ nodes. The degree parameters $\boldsymbol{\theta}^t$ follow the sinusoidal trajectory $f_{\boldsymbol{\theta}}(u^t) = 1.6 + 0.1\sin(2\pi u^t), \forall t \in \{1, \ldots, T\}$. The connectivity matrices $\boldsymbol{Z}^t$ are generated as in Section 5; the only difference is the baseline matrix $\boldsymbol{Z}^0$ which is set according to the two scenarios below.

- Scenario 1 ($K = 2$): Swapping of diagonal and off-diagonal entries of $\boldsymbol{Z}^0$ in the data generation process of Section 5.
- Scenario 2 ($K = 4$): Swapping the within-community and between-community values for the first and second communities of $\boldsymbol{Z}^0$ in Section C.3. That is,

$$\boldsymbol{Z}^0 = \begin{pmatrix} 0.7 & 1.3 & 0.75 & 0.8 \\ 1.3 & 0.7 & 0.7 & 0.75 \\ 0.75 & 0.7 & 1.3 & 0.8 \\ 0.8 & 0.75 & 0.8 & 1.3 \end{pmatrix}.$$

By construction, the networks generated under Scenario 1 with two communities stand for the case of complete disassortative mixing, in which connections within the same communities are less frequent than the connections between communities. Scenario 2 generalises this setting to $K = 4$ communities and incorporates heterogeneity in mixing patterns: the first and second communities are disassortative, while the thrid and fourth communities retain assortative structures.

The results are shown in Table 6. Interestingly, we observe that our proposed method still achieves the best performance under both disassortative scenarios. In addition, TVDCSBM with estimated labels also yields the lowest estimation errors for both $\boldsymbol{\theta}^t$ and $\boldsymbol{\tau}(u^t)$ among all competing methods. This observation is expected, as the estimation procedure in TVDCSBM is based on likelihood maximization, which does not impose any prior assumption on the positivity of community connectivity matrices $\boldsymbol{Z}^t$ (or $\boldsymbol{W}^t$) beyond the identifiability conditions in Assumption 4. Moreover, neither the theoretical guarantees in Theorem 2 nor the Tabu search algorithm used in practice relies on the assortativity assumption of networks. A similar property also holds for SDCSBM and KD-(DC)-SoS; however, these competing methods become less competitive in the sparse regime.

Table 3: Average estimation errors for $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ with $n = 100, T = 100$ and $K = 2$. All values, **expressed in units of $\mathbf{10^{-2}}$**, represent means with standard deviations in parentheses across 50 replications. Boldface indicates the smallest value among all competing methods, excluding the unattainable oracle $\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$ with known true $\boldsymbol{c}^t$.

Method	$\gamma = 1$				$\gamma = 0.95$			
	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$
Scenario 1: time-varying SBM with homogeneous degrees								
TVDCSBM	0.90	0.53	0.38	0.29	4.76	3.10	2.36	1.93
($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	(0.08)	(0.04)	(0.03)	(0.02)	(0.18)	(0.11)	(0.08)	(0.07)
TVDCSBM	**0.91**	**0.54**	**0.39**	**0.30**	**9.75**	**6.56**	**4.99**	**4.12**
($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	(0.08)	(0.06)	(0.05)	(0.03)	(0.28)	(0.22)	(0.18)	(0.26)
SDCSBM	97.57	45.55	26.88	17.15	97.62	45.77	27.26	17.34
	(1.03)	(0.61)	(0.53)	(0.41)	(1.07)	(0.50)	(0.56)	(0.31)
DC-MASE	48.18	20.35	12.52	8.69	50.74	27.20	19.96	16.47
	(1.48)	(0.32)	(0.18)	(0.13)	(0.66)	(0.33)	(0.18)	(0.11)
KD-DC-SoS	45.34	20.35	12.52	8.69	50.61	24.88	15.92	11.42
	(0.80)	(0.32)	(0.18)	(0.13)	(0.73)	(0.37)	(0.24)	(0.18)
Scenario 2: time-varying DCSBM with smooth trajectories								
TVDCSBM	1.84	1.02	0.67	0.48	13.15	9.18	7.42	6.06
($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	(0.13)	(0.07)	(0.04)	(0.03)	(0.31)	(0.26)	(0.35)	(0.07)
TVDCSBM	**1.95**	**1.02**	**0.74**	**0.52**	**22.65**	**15.29**	**11.18**	**8.95**
($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	(0.23)	(0.07)	(0.05)	(0.05)	(0.71)	(0.47)	(0.62)	(0.76)
SDCSBM	131.88	55.66	27.68	15.68	131.93	56.16	28.08	15.94
	(1.66)	(1.40)	(0.80)	(0.52)	(1.46)	(1.22)	(0.85)	(0.48)
DC-MASE	57.54	24.37	13.90	8.82	69.51	38.58	29.12	24.53
	(0.82)	(0.28)	(0.19)	(0.14)	(0.79)	(0.38)	(0.24)	(0.19)
KD-DC-SoS	57.35	24.36	13.90	8.82	66.67	30.55	17.85	11.24
	(0.89)	(0.28)	(0.19)	(0.14)	(1.19)	(0.45)	(0.26)	(0.26)
Scenario 3: time-varying DCSBM with non-smooth trajectories								
TVDCSBM	40.91	40.23	39.82	39.47	42.14	38.78	35.90	35.49
($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	(0.22)	(0.21)	(0.29)	(0.19)	(0.63)	(0.99)	(0.10)	(0.07)
TVDCSBM	**40.93**	40.28	39.84	39.47	**49.48**	43.45	39.94	37.88
($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	(0.21)	(0.22)	(0.31)	(0.26)	(0.50)	(1.01)	(1.47)	(0.78)
SDCSBM	116.63	52.16	28.06	16.49	116.47	52.58	28.33	**16.75**
	(1.18)	(1.02)	(0.79)	(0.50)	(1.09)	(0.90)	(0.78)	(0.47)
DC-MASE	53.75	**22.66**	**13.41**	**8.89**	60.63	**33.15**	**24.64**	20.50
	(1.03)	(0.38)	(0.16)	(0.12)	(0.70)	(0.32)	(0.22)	(0.14)
KD-DC-SoS	60.64	34.36	26.02	21.91	60.35	33.95	25.57	21.46
	(1.04)	(0.34)	(0.15)	(0.11)	(0.68)	(0.29)	(0.20)	(0.12)

Table 4: Average estimation errors under the unbalanced community size setting with varying $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}, \gamma \in \{1, 0.95\}$ and fixed $n = 100, T = 100$ and $K = 2$. All values, **expressed in units of 10^{-2}**, represent means with standard deviations in parentheses across 50 replications. Boldface indicates the smallest value among all competing methods, excluding the unattainable oracle estimators with known true $\boldsymbol{c}^t$.

Method	$\gamma = 1$				$\gamma = 0.95$			
	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$
Average Hamming error rates for $\hat{\boldsymbol{c}}^t$ across all time points								
TVDCSBM	**1.06**	**0.40**	**0.19**	0.24	**20.82**	**11.77**	**8.33**	**6.25**
DSBM	49.89	49.73	49.84	49.88	50.12	49.46	49.39	49.40
SDCSBM	43.98	41.71	36.56	28.74	43.03	34.87	25.82	18.36
DC-MASE	38.78	9.44	0.26	**0**	41.41	35.55	34.29	33.65
KD-SoS	14.39	3.31	1.98	2.84	31.18	19.04	16.25	15.18
KD-DC-SoS	19.30	5.91	1.91	0.95	30.23	16.53	12.98	11.55
Average estimation errors for $\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM ($\hat{\boldsymbol{\theta}}^t_h(\boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	0.49 (0.05)	0.25 (0.01)	0.20 (0.03)	0.14 (0.01)	0.53 (0.08)	2.59 (0.04)	2.12 (0.02)	1.89 (0.02)
TVDCSBM ($\hat{\boldsymbol{\theta}}^t_h(\hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**0.53** (0.08)	**0.27** (0.07)	**0.20** (0.04)	**0.16** (0.05)	**3.89** (0.1)	**2.79** (0.06)	**2.25** (0.05)	**1.88** (0.05)
SDCSBM	19.01 (0.48)	10.09 (0.28)	6.86 (0.27)	4.80 (0.25)	18.54 (0.44)	8.60 (0.18)	5.11 (0.11)	3.32 (0.11)
DC-MASE	19.52 (2.21)	7.98 (0.94)	4.06 (0.07)	2.58 (0.05)	18.47 (1.03)	8.30 (0.2)	5.13 (0.19)	3.60 (0.21)
KD-DC-SoS	19.61 (1.46)	8.24 (0.74)	4.28 (0.34)	2.68 (0.14)	18.03 (0.38)	8.17 (0.15)	4.93 (0.1)	3.31 (0.07)
Average estimation errors for $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM ($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	2.06 (0.15)	1.13 (0.06)	0.79 (0.05)	0.51 (0.02)	13.13 (0.35)	9.34 (0.30)	7.73 (0.10)	7.05 (0.07)
TVDCSBM ($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**2.26** (0.21)	**1.18** (0.19)	**0.83** (0.07)	**0.55** (0.09)	**21.83** (0.68)	**14.50** (0.48)	**10.98** (0.45)	**8.44** (0.52)
SDCSBM	135.90 (1.40)	61.96 (0.68)	35.26 (0.83)	19.96 (0.65)	132.26 (1.49)	57.59 (1.03)	29.40 (0.68)	16.45 (0.49)
DC-MASE	64.40 (0.97)	27.23 (1.96)	14.19 (0.23)	8.80 (0.17)	68.76 (0.85)	37.49 (0.38)	28.17 (0.27)	23.69 (0.24)
KD-DC-SoS	61.57 (1.70)	25.82 (0.39)	14.40 (0.42)	8.91 (0.21)	67.06 (1.24)	33.03 (0.44)	22.37 (0.30)	17.12 (0.20)

Table 5: Average estimation errors for four-community networks with varying $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}, \gamma \in \{1, 0.95\}$ and fixed $n = 100$ and $T = 100$. All values, **expressed in units of $\mathbf{10^{-2}}$**, represent means with standard deviations in parentheses across 50 replications. Boldface indicates the smallest value among all competing methods, excluding the unattainable oracle estimators with known true $\boldsymbol{c}^t$.

Method	$\gamma = 1$				$\gamma = 0.95$			
	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$
Average Hamming error rates for $\hat{\boldsymbol{c}}^t$ across all time points								
TVDCSBM	**3.12**	**1.50**	**0.71**	**0.24**	**53.84**	**37.55**	**28.32**	**22.25**
DSBM	67.07	65.54	58.63	52.80	65.33	66.12	66.43	66.08
SDCSBM	64.81	63.66	61.48	58.14	64.55	63.42	61.64	58.72
DC-MASE	55.54	20.76	4.54	0.34	64.55	62.30	55.69	51.59
KD-SoS	42.45	14.47	8.27	4.47	62.13	52.08	42.04	34.63
KD-DC-SoS	39.00	12.58	7.16	4.33	62.22	51.84	41.24	33.80
Average estimation errors for $\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM	0.52	0.30	0.20	0.15	4.23	3.21	2.66	2.28
($\hat{\boldsymbol{\theta}}^t_h(\boldsymbol{c}^t)$, known c^t)	(0.05)	(0.02)	(0.01)	(0.01)	(0.08)	(0.05)	(0.05)	(0.09)
TVDCSBM	**0.63**	**0.34**	**0.22**	**0.16**	**6.18**	**4.73**	**3.88**	**3.36**
($\hat{\boldsymbol{\theta}}^t_h(\hat{\boldsymbol{c}}^t)$, estimated $\hat{c}^t$)	(0.08)	(0.04)	(0.02)	(0.01)	(0.32)	(0.16)	(0.12)	(0.11)
SDCSBM	22.05	11.84	8.84	7.63	22.09	11.99	9.02	7.94
	(0.55)	(0.38)	(0.48)	(0.45)	(0.65)	(0.47)	(0.48)	(0.46)
DC-MASE	25.53	10.47	5.53	3.46	23.74	10.51	7.06	4.89
	(4.06)	(1.39)	(0.40)	(0.07)	(3.23)	(1.10)	(0.94)	(0.58)
KD-DC-SoS	21.56	9.74	5.90	4.04	21.28	10.47	7.00	5.34
	(1.94)	(0.91)	(0.87)	(0.86)	(1.32)	(0.74)	(0.41)	(0.31)
Average estimation errors for $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM	1.51	0.84	0.58	0.45	11.50	8.44	6.90	5.83
($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	(0.11)	(0.07)	(0.03)	(0.02)	(0.27)	(0.21)	(0.16)	(0.30)
TVDCSBM	**1.94**	**0.98**	**0.62**	**0.46**	**22.98**	**16.27**	**12.46**	**10.02**
($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	(0.17)	(0.07)	(0.03)	(0.02)	(0.37)	(0.38)	(0.32)	(0.26)
SDCSBM	178.44	81.02	49.97	34.04	179.62	80.88	50.22	34.39
	(2.31)	(0.69)	(0.50)	(0.38)	(2.17)	(0.88)	(0.61)	(0.37)
DC-MASE	75.39	32.29	16.21	9.47	74.81	38.38	28.05	22.48
	(22.31)	(9.08)	(4.21)	(0.42)	(16.97)	(5.71)	(6.56)	(5.68)
KD-DC-SoS	59.46	24.47	14.31	9.27	66.09	32.95	20.37	13.67
	(1.76)	(0.89)	(0.78)	(0.44)	(2.18)	(0.75)	(0.45)	(0.28)

Table 6: Average estimation errors for disassortative networks with varying $\rho_n \in \{0.05, 0.1, 0.15, 0.2\}, K = \{2, 4\}$ and fixed $n = 100, T = 100$ and $\gamma = 0.95$. All values, **expressed in units of $\mathbf{10^{-2}}$**, represent means with standard deviations in parentheses across 50 replications. Boldface indicates the smallest value among all competing methods, excluding the unattainable oracle estimators with known true $\boldsymbol{c}^t$.

Method	Scenario 1 ($K = 2$)				Scenario 1 ($K = 4$)			
	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$	$\rho_n = 0.05$	$\rho_n = 0.10$	$\rho_n = 0.15$	$\rho_n = 0.20$
Average Hamming error rates for $\hat{\boldsymbol{c}}^t$ across all time points								
TVDCSBM	**19.89**	**11.34**	**8.20**	**6.46**	**54.74**	**35.49**	**24.14**	**17.88**
DSBM	50.12	49.54	49.42	49.42	65.39	66.21	66.69	66.93
SDCSBM	42.28	33.76	22.79	15.02	64.67	63.50	61.85	58.93
DC-MASE	41.60	35.54	34.60	33.61	64.50	62.10	55.20	50.27
KD-SoS	27.83	15.72	9.93	7.86	62.02	53.37	43.80	36.80
KD-DC-SoS	27.27	14.85	9.18	7.19	62.32	53.13	43.18	35.84
Average estimation errors for $\hat{\boldsymbol{\theta}}^t(\hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM ($\hat{\boldsymbol{\theta}}^t_h(\boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	3.63 (0.09)	2.60 (0.04)	2.12 (0.03)	1.78 (0.02)	4.42 (0.10)	3.20 (0.04)	2.67 (0.04)	2.29 (0.10)
TVDCSBM ($\hat{\boldsymbol{\theta}}^t_h(\hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**4.05** (0.13)	**2.83** (0.08)	**2.30** (0.05)	**1.95** (0.08)	**6.28** (0.26)	**4.72** (0.16)	**3.78** (0.12)	**3.21** (0.09)
SDCSBM	18.56 (0.44)	8.60 (0.22)	5.19 (0.17)	3.39 (0.15)	22.17 (0.59)	12.10 (0.41)	9.29 (0.39)	8.13 (0.52)
DC-MASE	18.55 (1.49)	8.25 (0.28)	4.99 (0.26)	3.29 (0.20)	23.98 (2.46)	10.53 (1.03)	6.94 (1.05)	4.85 (0.61)
KD-DC-SoS	18.01 (0.47)	8.08 (0.14)	4.75 (0.08)	3.10 (0.06)	21.32 (1.33)	10.61 (0.77)	7.91 (0.49)	7.18 (0.56)
Average estimation errors for $\tilde{\boldsymbol{\tau}}(u^t, \hat{\boldsymbol{c}}^t)$ across all time points								
TVDCSBM ($\tilde{\boldsymbol{\tau}}_h(u^t, \boldsymbol{c}^t)$, known $\boldsymbol{c}^t$)	12.93 (0.33)	9.05 (0.29)	7.36 (0.31)	6.02 (0.08)	11.44 (0.31)	8.33 (0.19)	6.86 (0.12)	5.81 (0.33)
TVDCSBM ($\tilde{\boldsymbol{\tau}}_h(u^t, \hat{\boldsymbol{c}}^t)$, estimated $\hat{\boldsymbol{c}}^t$)	**21.73** (0.64)	**14.29** (0.54)	**10.98** (0.55)	**8.89** (0.64)	**23.23** (0.35)	**16.09** (0.39)	**11.98** (0.28)	**9.45** (0.26)
SDCSBM	129.95 (1.47)	58.01 (0.94)	29.54 (0.85)	15.58 (0.51)	178.80 (1.83)	80.70 (0.78)	50.01 (0.55)	34.53 (0.38)
DC-MASE	68.42 (0.80)	37.37 (0.41)	27.78 (0.29)	23.22 (0.23)	76.33 (15.16)	36.60 (4.17)	28.32 (8.05)	23.16 (5.41)
KD-DC-SoS	66.69 (0.91)	30.63 (0.49)	17.79 (0.27)	11.24 (0.38)	66.10 (2.08)	33.53 (0.68)	21.40 (0.52)	14.74 (0.28)