TIME-VARYING DEGREE-CORRECTED STOCHASTIC BLOCK MODELS

Mengxue Li, Rainer von Sachs, Eugen Pircalabelu

LIDAM Discussion Paper ISBA
2024 / 14

UCLouvain

ISBA | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

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.html

Time-varying degree-corrected stochastic block models

Mengxue Li Rainer von Sachs Eugen Pircalabelu

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

September 9, 2024

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 smoothed 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 et al., 2018; Samdin et al., 2019).

A classical model for the network displaying a community structure is the stochastic block model (SBM) introduced in Holland et al. (1983) and developed by many authors (Bickel and Chen, 2009; Zhao et al., 2012; Abbe, 2018). 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 practice. 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 et al., 2013; Wang et al., 2023), variational expectation-maximization (Bickel et al., 2013; Wang and Bickel, 2017), information-theoretic (Gao et al., 2018), and spectral clustering (Rohe et al., 2011; Qin and Rohe, 2013; Jin, 2015), 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 (or a multilayer network) with n nodes and T time points. In recent years, numerous studies have sought to extend community detection procedures to dynamic networks. Huang et al. (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 the communities. Bhattacharyya and Chatterjee (2020) and Agterberg et al. (2022) 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. A final line of work is based on probabilistic generative models. Matias and Miele (2017) and Riverain et al. (2023) applied Markov chain method to estimate the dynamic community structures which has proven to be intractable for investigating the theoretical properties of the estimators. Li et al. (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. In particular, we allow all nodes of the network to change freely their communities between any two time points without extra constraints. Moreover, as a novelty compared to the existing reviewed literature, we accompany our approach with asymptotic theory on rates of consistency of our 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-smoothed estimators of edge probabilities to be treated 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 rates of convergence essentially for two times differentiable curves (over

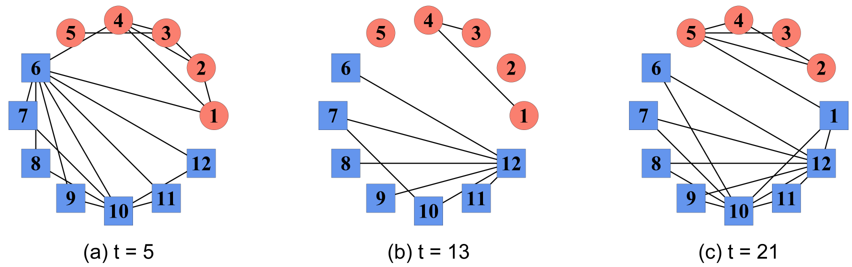


Figure 1: A toy example of a dynamic subnetwork for the Travian dataset with $n = 12$ nodes at three time points $t = 5, 13, 21$. The red circles and blue squares denote $K = 2$ distinct communities.

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 using 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 smoothed 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 on 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 only focus 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 the 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 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

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

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. Zhao et al. (2012); Gao et al. (2018)). 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 behaviour 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 et al., 2020; Lei and Lin, 2023) and modularity (Clauset et al., 2004; Bickel and Chen, 2009; 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 $k \neq l$, and $O_{kk}(\boldsymbol{e}^t)$ is twice the number of edges within the community without self-loops in networks. The quantity $O_k(\boldsymbol{e}^t)$ denotes the total number of marginal connections from nodes in community k to 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, the log-likelihood function under the given parameters $\boldsymbol{c}^t, \boldsymbol{\theta}^t, \boldsymbol{Z}^t$ is,

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

which yields the closed form MLEs of θ_i^t and $Z_{c_i^t c_j^t}^t$, given $\boldsymbol{c}^t$, as follows,

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

Plugging in these MLEs and ignoring constants, we can obtain a function depending only on the community assignments with the likelihood already maximized given the vector $\boldsymbol{c}^t$. Therefore, for any generic candidate $\boldsymbol{e}^t$, the criterion for community detection is given by

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

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.

Roughly speaking, the estimation based on equation (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 equation (4) leads to strongly consistent estimators for the communities as long as $\lambda_n / \log n \to \infty$ as

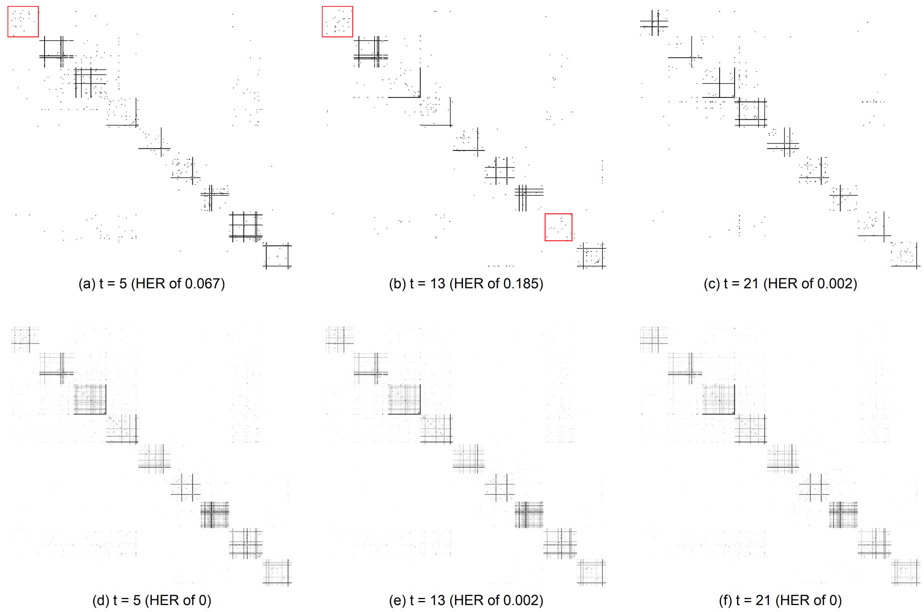


Figure 2: The original adjacency matrices (top row: (a)-(c)) and the smoothed adjacency matrices (bottom row: (d)-(f)) with a bandwidth of 0.47 are presented for three time points $t = 5, 13, 21$ in the Travian dataset among 411 players assigned to 9 communities. The black and white segments represent the presence ($A_{ij}^t = 1$) and absence ($A_{ij}^t = 0$) of an edge, respectively. And 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.

$n \to \infty$. 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 across these three original 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 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 equation (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 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 (18)) based on smoothed adjacency matrices are substantially lower than those using criterion (4) separately for each individual network.

3 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 $t/T, t = 1, \ldots, T$ of the unit interval in 'rescaled' time $u \in [0, 1]$. More specifically, recall the edge probability $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t)$ defined in equation (2) which is our target at time point t, now to be linked to a smooth nonparametric curve - which we will call $\tau_{ij}(u)$ - in rescaled time $u = t/T$. That is, for any pair of nodes (i, j), we suppose that there exists such a curve $\tau_{ij}(u), u \in [0, 1]$, such that

$$\tau_{ij}(t/T) \ = \ \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A_{ij}^t | u = t/T). \tag{5}$$

Observe that $\tau_{ij}(t/T)$ denotes the probability of an edge connecting nodes i and j at time point t. Consequently, the edge A_{ij}^t can be viewed as an empirical, hence noisy, version of $\tau_{ij}(t/T)$, that is

$$A_{ij}^t = \tau_{ij}(t/T) + \varepsilon_{ij}^t \tag{6}$$

where we make the assumption of independent 'errors' ε_{ij}^t associated with the observed probability for an edge between nodes i and j (see later on in Assumption 2). Note that, comparing model (6) with equation (2) we observe that the quantities $\boldsymbol{\theta}^t$ and $\boldsymbol{W}^t$ of the right-hand side of (2) will also be connected to 'target' versions defined in rescaled time u, inheriting as such the smoothness properties of $\tau_{ij}(u)$ in time. We refrain however from explicitly stating those as it will be sufficient to picture the product of the right-hand side of (2) to be a smooth function in time as a whole.

As we have assumed that $\tau_{ij}(t/T)$ is not only a smooth function 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 in 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 $\tau_{ij}(t/T)$ by averaging the observations A_{ij}^t over neighbors in time t for any $i, j \in \{1, \ldots, n\}$. Let $U(t/T) = [t/T - h, t/T + h]$ be a neighborhood of time point $t \in \{1, \ldots, T\}$ for some bandwidth $h \in (0, 1]$, then for any $i, j \in \{1, \ldots, n\}$, a basic kernel estimator of $\tau_{ij}(t/T)$ can be written as

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

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 $t \in \{1, \ldots, T\}$, we replace the adjacency matrix $\boldsymbol{A}^t$ with the kernel estimator $\hat{\boldsymbol{\tau}}_h(t/T) = (\hat{\tau}_{ij,h}(t/T))_{n \times n}$. Correspondingly, for any vector $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$, we define the smoothed counting variables as follows,

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

$$\hat{O}_{k,h}(\boldsymbol{e}^t) = \sum_{l=1}^{K} \hat{O}_{kl,h}(\boldsymbol{e}^t), \quad \hat{d}_{i,h}^t = \sum_{j=1}^{n} \hat{\tau}_{ij,h}(t/T).$$

As well, given $\boldsymbol{c}^t$, we can obtain the smoothed MLEs of θ_i^t and $Z_{c_i^t c_j^t}^t$ from (3) as,

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

where $\hat{O}_{c_i^t c_j^t,h}(\boldsymbol{c}^t) = \sum_{s,m=1}^{n} \hat{\tau}_{sm,h}(t/T) \mathbf{1}_{\{c_s^t = c_i^t, c_m^t = c_j^t\}}$, and $\hat{O}_{c_i^t,h}(\boldsymbol{c}^t) = \sum_{s,j=1}^{n} \hat{\tau}_{sj,h}(t/T) \mathbf{1}_{\{c_s^t = c_i^t\}}$. Plugging in these smoothed

MLEs, the final estimator for $\tau_{ij}(t/T)$ at time point t is written as

$$\tilde{\tau}_{ij,h}(t/T) = \hat{\theta}_{i,h}^t \hat{\theta}_{j,h}^t \hat{Z}_{c_i^t c_j^t,h}^t = \frac{\hat{d}_{i,h}^t \hat{d}_{j,h}^t \hat{O}_{c_i^t c_j^t,h}(\boldsymbol{c}^t)}{\hat{O}_{c_i^t,h}(\boldsymbol{c}^t) \hat{O}_{c_j^t,h}(\boldsymbol{c}^t)}, \tag{10}$$

where $\tilde{\boldsymbol{\tau}}_h(t/T) = (\tilde{\tau}_{ij,h}(t/T))_{n\times n}$ replaces the estimated edge probability matrices. 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 $\hat{O}_h(\boldsymbol{e}^t) = (\hat{O}_{kl,h}(\boldsymbol{e}^t))_{K\times K}$.

Remark 2. The kernel estimator $\hat{\tau}_{ij,h}(t/T)$ 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 $Q\left(\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(O(\boldsymbol{e}^t)\right)\right)$ by smoothing as well. It has 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 $\boldsymbol{\tau}(t/T) = (\tau_{ij}(t/T))_{n\times n}$, serving as the target linked to nonparametric techniques, encapsulates all parameters comprehensively. The goal is to choose a bandwidth where the final estimator $\tilde{\boldsymbol{\tau}}_h(t/T)$ is close to $\boldsymbol{\tau}(t/T)$ for any time point $t \in \{1,\dots,T\}$.

To facilitate our theoretical development, we need first to determine the error of $\tilde{\boldsymbol{\tau}}_h(t/T)$. For any matrices $\boldsymbol{R}, \boldsymbol{R}^{'} \in \mathbb{R}^{n\times n}$, define $d(\boldsymbol{R}, \boldsymbol{R}^{'})$ as the scaled mean-squared error (MSE) between two matrices $\boldsymbol{R}$ and $\boldsymbol{R}^{'}$, as

$$d(\boldsymbol{R}, \boldsymbol{R}^{'}) = \mathbb{E}\left(\frac{1}{n^2}\|\boldsymbol{R} - \boldsymbol{R}^{'}\|^2\right) \tag{12}$$

where $\|\cdot\|^2$ denotes the sum of squares for all entries of a vector or matrix. Next, we require the following assumptions to state an upper bound on the error of the final estimator $\tilde{\boldsymbol{\tau}}_h(t/T)$.

Assumption 1 (Edge probability functions)**.** *Uniformly over $i, j \in \{1,\dots,n\}$, assume that the probability functions $\tau_{ij}(u)$ have continuous second derivatives and bounded third derivatives for all $u \in (0,1]$. Furthermore, as all $\tau_{ij}(u)$ are proportional to ρ_n, we note that also all derivatives are proportional to ρ_n across all time points u.*

Assumption 2 (Errors)**.** *Assume that the error ε_{ij}^t is a random variable with zero mean such that its variance satisfies $\sigma_{ij}^2(u) = \tau_{ij}(u)(1-\tau_{ij}(u))$, and that $\varepsilon_{ij}^t \perp \varepsilon_{i'j'}^{t'}$ holds for any $i' \neq i, j' \neq j \in \{1,\dots,n\}$ and any $t' \neq t \in \{1,\dots,T\}$.*

Assumption 3 (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 $\boldsymbol{c}^t \in \{1,\dots,K\}^n$ such that $n_a(\boldsymbol{c}^t)$ is a positive integer for any community $a \in \{1,\dots,K\}$ and any time point $t \in \{1,\dots,T\}$.*

Assumption 4 (Identifiability of parameters)**.** *For any time point $t \in \{1,\dots,T\}$ and any community $a \in \{1,\dots,K\}$, the degree parameter vector $\boldsymbol{\theta}^t$ satisfies $\sum_{i=1}^n \theta_i^t \mathbf{1}_{\{c_i^t=a\}} = n_a(\boldsymbol{c}^t)$. Moreover, to avoid severely unbalanced degrees, we assume that $\sup_{i\in\{1,\dots,n\}} \theta_i^t = O(1)$ for any time point t.*

Assumption 1 describes the temporal evolution of the network. Following the assumptions in previous works (Pensky and Zhang, 2019; Yang and Jie, 2020; Keriven and Vaiter, 2022; Lin and Lei, 2024), we posit that nodes are gradually switching their community across time, which implies a certain smoothness in $\tau_{ij}(u)$ across time. However, note that in contrast to the smoothness assumptions in Pensky and Zhang (2019); Keriven and Vaiter (2022), which allow at most $s \ll n$ nodes to change their communities, all nodes in our context are permitted to change their communities between any two time points with a relatively weak constraint (i.e., $s = n$). Moreover, we observe that $\tau_{ij}(u)$ is proportional to ρ_n considering its definition. Applying Taylor expansion, it is immediate that all derivatives are also dependent on ρ_n. Assumption 2 is natural for nonparametric curve estimation. However note the specific form of the variance of ε_{ij}^t coming from our model where A_{ij}^t is assumed to be a Bernoulli random variable with variance $\tau_{ij}(t/T)\left(1-\tau_{ij}(t/T)\right)$. Investigating the number K of communities, though interesting in itself, it is not the primary focus of this work. Thus, as stated in Assumption 3, K is assumed to be fixed and not varying throughout time. Finally, 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.

Remark 3. [Additional remark for Assumption 1.] Note that assuming a certain number of derivatives in time in Assumption 1 is merely for convenience to keep the presentation of the proofs manageable (in particular, without changing any results, we could give up the assumption of the existence of a third derivative in time). Qualitatively, the nature of the results would also not change if we assumed less regularity in time - resulting of course into slower rates of convergence for the bias and the MSE of our considered kernel estimators (completely analogously to classical nonparametric curve estimation). However this would not jeopardise 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 by a more refined kernel as to be found in the classical literature.

Throughout this work, we define $a_n \sim b_n$ for any two sequences a_n and b_n if the quantity a_n/b_n tends to a non-zero 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 our first result now.

Our first proposition provides a bound on the scaled mean-squared error of $\tilde{\boldsymbol{\tau}}_h(t/T)$.

Proposition 1 (MSE for scaled $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T)$)**.** *Under Assumptions 1-4 and given the true community structure $\boldsymbol{c}^t$ for each time point $t \in \{1,\ldots,T\}$, we have that*

$$d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T), \rho_n^{-1}\boldsymbol{\tau}(t/T)) \le C_1 h^4 + \frac{C_2}{2n\kappa_n\rho_n Th} \tag{13}$$

for constants C_1 and C_2 depending on $\theta_i^t, W_{c_i^t c_j^t}^t$ and $\tau_{ij}^{''}(t/T)$ over $i,j \in \{1,\ldots,n\}$ and $t \in \{1,\ldots,T\}$.

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 $\tilde{\tau}_{ij,h}(t/T)$ is a function of the smoothed counting variables $\hat{d}_{i,h}^t, \hat{O}_{kl,h}(\boldsymbol{e}^t)$ and $\hat{O}_{k,h}(\boldsymbol{e}^t)$ as defined in (8). Lemma A.2 (which itself is based on Lemma A.1 on bias and variance of our Step 1 - smoother $\hat{\tau}_{ij,h}(t/T)$) provides typical nonparametric rates of convergence of these kernel smoothed counting variables to their corresponding 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 $\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_k(\boldsymbol{e}^t))$, local in time t. Then, the essential step of the proof is the treatment of a Taylor expansion of $\tilde{\tau}_{ij,h}(t/T)$ being expressed as a (five-dimensional) function of the smoothed counting variables about its population analog $\tau_{ij}(t/T)$ being expressed as the same function of the corresponding population quantities - see equation (A.10) in Appendix A. This eventually leads to the desired rates of convergence of bias, variance and mean-squared error of the scaled $\tilde{\boldsymbol{\tau}}_h(t/T)$. □

We next give some insights into why we need to consider the scaled error. Observe that, since $\tau_{ij}(t/T)$ is proportional to ρ_n, if we considered the non-scaled error of $\tilde{\boldsymbol{\tau}}_h(t/T)$, this 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.

We observe in the upper bound for the MSE of $\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T)$ in (13) that in contrast to the first term, i.e. the typical bias term in nonparametrically estimating curves with two continuous derivatives, depending only on the bandwidth h, the second, i.e. the variance, is inversely proportional to the 'effective local sample size' $n\kappa_n\rho_n Th$. For this one, we can distinguish two scenarios: If the community sizes are balanced, i.e. if $\kappa_n \sim 1$ because $n_{\min} \asymp n/K$, the variance depends only on the effective local sample size of the network $n\rho_n Th$ needing to tend to infinity. 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 tend to infinity 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 error of $\tilde{\boldsymbol{\tau}}_h(t/T)$ across all time points. Proposition 1 yields the following result for the optimal bandwidth h_{opt}.

Corollary 1 (Bandwidth selection). *Under the conditions of Proposition 1, the optimal bandwidth h_{opt} minimizing the asymptotically leading term of the global mean-squared error $T^{-1}\sum_{t=1}^{T} d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T), \rho_n^{-1}\boldsymbol{\tau}(t/T))$ is of order*

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

Observe that Corollary 1 states the asymptotic behaviour of h_{opt}. It is intuitive that the bandwidth h_{opt} will decrease when the number of nodes n or the network sparsity 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}.

Remark 4. In practice, the selection procedure of h_{opt} is challenging because the true value of $\tau_{ij}(t/T)$ 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

$$\hat{\tau}_{ij,h,-t}(t/T) = \frac{\sum_{t' \in U(t/T)\setminus\{t\}} A_{ij}^{t'}}{|\{A_{ij}^{t'} : t' \in U(t/T) \setminus \{t\}\}|}.$$

We then substitute $\hat{\tau}_{ij,h,-t}(t/T)$ for $\hat{\tau}_{ij,h}(t/T)$ in Step 2. Following the definition in (10), we construct the leave-one-out estimator $\tilde{\tau}_{ij,h,-t}(t/T)$ with

$$\tilde{\tau}_{ij,h,-t}(t/T) = \frac{\hat{d}_{i,h,-t}^t \hat{d}_{j,h,-t}^t \hat{O}_{c_i^t c_j^t,h,-t}(\boldsymbol{c}^t)}{\hat{O}_{c_i^t,h,-t}(\boldsymbol{c}^t)\hat{O}_{c_j^t,h,-t}(\boldsymbol{c}^t)},$$

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

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

4 Consistency of kernel smoothed estimators

Recall that in Step 2, the smoothed estimator of θ_i^t for any $i \in \{1, \dots, n\}$ is given by (9), as well as the smoothed profile-likelihood criterion $Q(\hat{O}_h(\boldsymbol{e}^t))$ provided in (11). We now investigate theoretical properties for the smoothed 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 $t \in \{1, \dots, T\}$ is

$$\hat{\boldsymbol{c}}^t = \arg\max_{\boldsymbol{e}^t} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right). \tag{14}$$

We first investigate the consistency in estimating the node degree parameters. With slight abuse of the same notation $d(\cdot, \cdot)$, as already used in equation (12), we define the scaled mean-squared error now between any two vectors $\boldsymbol{\theta}, \boldsymbol{\theta}^{'} \in \mathbb{R}^n$ as follows,

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}^{'}) = \mathbb{E}\left(\frac{1}{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}^{'}\|^2\right).$$

Theorem 1 (Consistency of degree parameters)**.** *Suppose that Assumptions 1-4 are holding, and consider the true community structure $\boldsymbol{c}^t$ for each time point $t \in \{1, \dots, T\}$ be given. Let $\hat{\boldsymbol{\theta}}_h^t = \left(\hat{\theta}_{1,h}^t, \dots, \hat{\theta}_{n,h}^t\right)^\top$ be the smoothed estimators of the degree parameters where $\hat{\theta}_{i,h}^t$ is defined in (9). Provided that $\rho_n, h \to 0$ as $nT \to \infty$ such that $n\rho_n Th \to \infty$, the estimator $\hat{\boldsymbol{\theta}}_h^t$ is mean-squared error consistent for $\boldsymbol{\theta}^t$ at any time point $t \in \{1, \dots, T\}$, that is*

$$d(\hat{\boldsymbol{\theta}}_h^t, \boldsymbol{\theta}^t) \leq C_3 h^4 + \frac{C_4}{2n\rho_n Th} \tag{15}$$

for constants C_3 and C_4 depending on $\theta_i^t, W_{c_i^t c_j^t}^t$ and $\tau_{ij}^{''}(t/T)$ over $i, j \in \{1, \dots, n\}$ and $t \in \{1, \dots, T\}$.

Proof. The main technical component of this proof is similar to the proof of Proposition 1 following the idea that $\hat{\theta}_{i,h}^t$, defined in (9) 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 about its population limit delivers the aforementioned rates of convergence via those of Lemma A.2. For details, we refer to Appendix B. □

Theorem 1 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 1, note that no additional assumptions of balancedness in community sizes is necessary (unlike for Proposition 1, where the upper bound for $\tilde{\boldsymbol{\tau}}_h(t/T)$ depends on κ_n).

Remark 5. The results of Proposition 1 and Theorem 1 are shown conditionally on the given true community assignments $\boldsymbol{c}^t$. However, the upper bounds in (13) and (15) still work when replacing $\boldsymbol{c}^t$ by the estimator $\hat{\boldsymbol{c}}^t$ if one can prove that $\hat{\boldsymbol{c}}^t$ converges to $\boldsymbol{c}^t$ almost surely.

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 1 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 $\mathcal{N}_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) = \{1 \leq i \leq n : c_i^t = a, e_i^t = k\}$ 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 satisfies 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 scaled $O_{kl}(\boldsymbol{e}^t)$ where

$$H_{kl}(\boldsymbol{e}^t) = \mathbb{E}_{\boldsymbol{c}^t, \boldsymbol{\theta}^t}\left(\frac{O_{kl}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \sum_{a,b=1}^K W_{ab}^t \times \frac{\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n} \times \frac{\Lambda_{b(l)}(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n}, \tag{16}$$

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

Similarly to Zhao et al. (2012) and Sengupta and Chen (2018), we also define the population version of the criterion based on the original observation $\boldsymbol{A}^t$ 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 1 (Uniform convergence of $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$)**.** *Under Assumptions 1-4, if $\rho_n, h \to 0$ and both n and T diverging such that $n\rho_n T h \to \infty$, then over $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$, 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 $t \in \{1, \ldots, T\}$, that is, there exists $\epsilon_n^t \to 0$ such that*

$$\mathbb{P}\left(\max_{\boldsymbol{e}^t}\left|Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(H(\boldsymbol{e}^t)\right)\right| < \epsilon_n^t\right) \to 1, \text{ as } \rho_n, h \to 0, n\rho_n T h \to \infty. \tag{17}$$

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

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

$$\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) = \min_{\pi(\boldsymbol{e}^t)} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{e_i^t \neq c_i^t\}}, \tag{18}$$

where $\pi(\boldsymbol{e}^t)$ denotes 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. If one can show that $\hat{\boldsymbol{c}}^t \to \boldsymbol{c}^t$, 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 a unique maximizer of $Q\left(H(\boldsymbol{e}^t)\right)$.

Assumption 5 (Detectability of community structures)**.** *For each time point $t \in \{1, \ldots, T\}$, we assume that the community connectivity matrix $\boldsymbol{W}^t$ does not have two identical columns.*

Lemma 2 (Uniqueness of the maximizer of $Q(H(\boldsymbol{e}^t))$). *Under Assumptions 3-5, $Q(H(\boldsymbol{e}^t))$ is uniquely maximized at the true labels $\boldsymbol{c}^t$, that is, for any generic vector $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$,*

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

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

$$M'\theta_{\inf}\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \leq Q\left(H(\boldsymbol{c}^t)\right) - Q\left(H(\boldsymbol{e}^t)\right) \leq M\theta_{\sup}\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \tag{19}$$

with $\theta_{\inf} = \inf_{\substack{i=1,\ldots,n,\\ t=1,\ldots,T}} \theta_i^t$ and $\theta_{\sup} = \sup_{\substack{i=1,\ldots,n,\\ t=1,\ldots,T}} \theta_i^t$.

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

This Lemma also gives the lower and upper bounds for the population optimality gap for incorrect community assignments. Combining Lemmas 1 and 2, we have the following theorem.

Theorem 2 (Consistency of community detection). *Under Assumptions 1-5, if $\rho_n, h \to 0$ such that $n\rho_n Th \to \infty$ as both n and T diverge, the community estimators, as defined in (14), are weakly consistent for any time point $t \in \{1, \ldots, T\}$, that is*

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

and 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 2. In other words, Theorem 2 shows the effectiveness of pooling information across the networks to enhance the performance of community detection. 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 2 still guarantee a strongly consistent community estimation requiring that T diverges such that $(Th)/\log n \to \infty$. 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 simulations. Therefore, in this paper, using criterion (11) prevails in sparse and semi-sparse networks where $\lambda_n < \log n$.

5 Numerical simulations

In this section, we explore the performance of our method constructed in Section 3 for recovering the community structures 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. Consider $T = 100$ networks and assign n nodes to $K = 2$ communities. Assume that community label c_i^t of each node follows a Markov chain with initial probabilities ($\mathbb{P}(c_i^1 = 1) = p, \mathbb{P}(c_i^1 = 2) = 1 - p$) and a state transition matrix $\begin{pmatrix} \gamma & 1-\gamma \\ 1-\gamma & \gamma \end{pmatrix}$. Roughly speaking, each node will steadily alter the community assignment with the probability $(1-\gamma)$ over time. Next let all the diagonal elements of $\boldsymbol{Z}^t$ follow the function $\rho_n \times W_{\text{in}}(t/T)$ while all the off-diagonal elements follow the function $\rho_n \times W_{\text{out}}(t/T)$ with $W_{\text{in}}(x) = 1.3 + 0.1\sin(2\pi x - 0.1\pi)$ and $W_{\text{out}}(x) = 0.7 - 0.1\sin(2\pi x)$. 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 $f(t/T) = \theta_0 + 0.1\sin(2\pi(t/T))$ if $\theta_0 > 1$, and $f(t/T) = 1$ if $\theta_0 = 1$. The other half of the nodes are chosen accordingly to satisfy the identifiability constraint in Assumption 4 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 the model (1). Based on the aforementioned data-generating process, we focus on four scenarios, each corresponding to the impact of different parameters: n, ρ_n (network sizes), γ (community-switching stability), θ_0 (heterogeneity in node degrees), and p (imbalance in community sizes) on their performance. Beyond these four scenarios, we explored the robustness of our approach to the smoothness condition on edge probabilities ($\tau_{ij}(t/T)$). Numerical results are provided in Appendix C.

Methods. We compare our proposed time-varying DCSBM (TVDCSBM) with other methods designed for dynamic networks. One natural approach involves applying the static DCSBM separately to each adjacency matrix $\boldsymbol{A}^t$ (SDCSBM). This method relies solely on the current network for estimation. Another candidate considered is the dynamic SBM (DSBM) developed by Matias and Miele (2017), as this method can identify changes in communities using Markov chains. To evaluate the clustering accuracy, we consider the HER, defined in equation (18), as the measure of the agreement between the estimated and true communities. This index ranges from 0 to 1, with lower values indicating better agreement. In our dynamic framework, we calculate the average value across all time points to compare overall performance across methods. The figures in this section all present the mean results over 100 replications.

Scenario 1: Network sizes. We consider $n = 50, 100, 200$ corresponding respectively to small, medium and large networks, and then vary ρ_n from 0.05 to 0.2. The results are presented in Figure 3 (a)-(c). On one hand, intuitively, our method significantly improved the detection performance compared to other methods, particularly when the network size or sparsity is limited (as shown in Figure 3 (a) and (b)). This result highlights the importance of aggregating information across networks when a single network (e.g., a semi-sparse or sparse network) may not provide sufficient information for accurate recovery. Conversely, SDCSBM and DSBM perform poorly for the network with the limited size or sparsity. On the other hand, we find that the advantage of TVDCSBM is dissipating with growing ρ_n in large networks. When $\rho_n = 0.2$, the results of SDCSBM are close to those of TVDCSBM, whereas DSBM exhibits the best performance. This result is sensible because there is sufficient information within each network to accurately recover communities in such scenarios, lessening the necessity of aggregation across networks. Furthermore, when the information within a network is substantial, the Markov chain can recover the abrupt state transition in time-varying communities of nodes. However, DSBM is sensitive to the heterogeneity of node degrees owing to the standard SBM. We will further discuss this later for Scenario 3.

Scenario 2: Community-switching stability. In this part, we assign $\gamma = 0.9, 0.95, 1$. These three cases correspond respectively to low, medium and high community-switching stability. Notice that when $\gamma = 1$, all nodes share a common community structure for all time points. In Figure 3 (d), we observe that our method, TVDCSBM, is able to accurately recover the community structure for all ρ_n with all average HERs below 0.05. In comparison, DSBM demonstrates slightly better performance only in denser networks and SDCSBM performs much worse for all networks. As γ decreases, the performance of both TVDCSBM and DSBM deteriorates, but DSBM shows a rapid decrease in performance. With low community stability, the nodes are more likely to change their communities from time point to time point. This would introduce additional noise and makes it challenging for these dynamic models to effectively aggregate information from other networks. Furthermore, as inferred from Corollary 1, a smaller value of ρ_n will lead to a larger h in order to gain more information. At the same time, this also means that more noise might be added into the estimation process. Consequently, it is reasonable to expect that TVDCSBM may not perform as well in sparser networks at $\gamma = 0.9$. Interestingly, SDCSBM

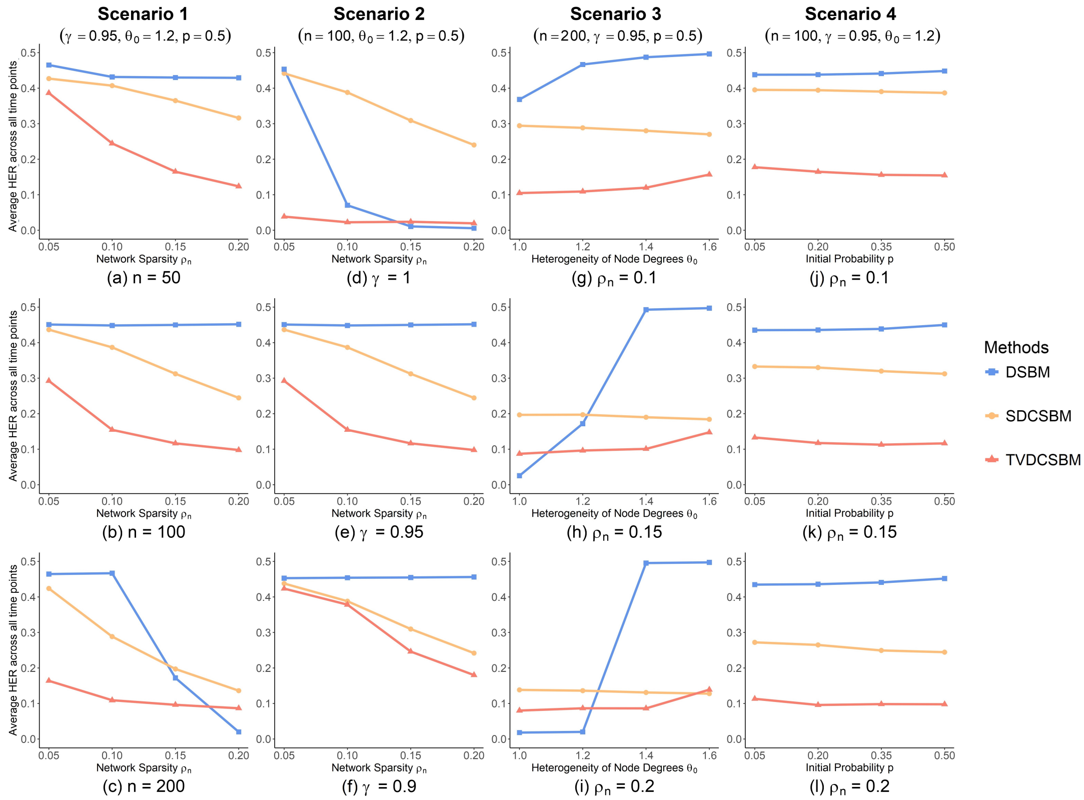


Figure 3: Average Hamming error rates (HER) for three methods: DSBM (blue squares), SDCSBM (orange circles) and TVDCSBM (red triangles) across four scenarios.

has relatively stable results. It is however intuitive that SDCSBM is not affected by the time-varying structure, as it relies solely on current information. Overall, TVDCSBM still slightly outperforms the other two approaches.

Scenario 3: Heterogeneity in node degrees. As shown in Figure 1, there exist hub nodes in real-world data.Therefore, in this scenario, we evaluate the robustness of the methods to heterogeneity of node degrees. θ_0 varies from 1 to 1.6, where $\theta_0 = 1$ represents the standard SBM. As θ_0 increases, the differences in node degrees within the same community also increase. Figure 3 (g)-(i) illustrate the results for three different settings of ρ_n. It is obvious that DSBM displays the highest sensitivity to θ_0 with its HER sharply increasing as the heterogeneity goes up, even within denser networks at $\rho_n = 0.2$. While TVDCSBM is also sensitive to θ_0, its performance is overall superior. It can also identify the community structure of the dynamic networks under the assumption of a standard SBM. Roughly speaking, TVDCSBM has a better robustness to the heterogeneity of node degrees compared to DSBM. In contrast, SDCSBM shows a relatively stable performance across varying levels of heterogeneity in node degrees.

Scenario 4: Unbalanced community sizes. The preceding simulations all involve cases with balanced community sizes. In this study, we vary p from 0.05 to 0.5 where the community sizes tend to be more balanced as p approaches 0.5. Across Figure 3 (j)-(l), observe that the HERs of TVDCSBM and SDCSBM slightly decrease with growing p, whereas the HERs of DSBM rise. We further see that the discrepancies for TVDCSBM in detection results caused by the imbalance in community sizes are gradually diminishing as ρ_n grows. This supports the results of Proposition 1. For the unbalanced case, the network needs to be sufficiently dense to guarantee the identification of connectivity within a community of minimal size. In conclusion, TVDCSBM also performs better than the other two methods across various settings of the network sparsity as well as the imbalance in community sizes.

Table 1: Summary statistics of node degrees in the Workplace dataset

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 Applications

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

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 1 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 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 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 0.155, while DSBM exhibits relatively lower performance and has an HER of 0.355. Applying the static DCSBM to each yields the worst performance (the HER of 0.41). 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 the unbalanced community sizes.

6.2 Travian game dataset analysis

Travian is the 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. 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.

[1]http://www.sociopatterns.org/datasets/contacts-in-a-workplace
[2]http://ial.eecs.ucf.edu/travian.php

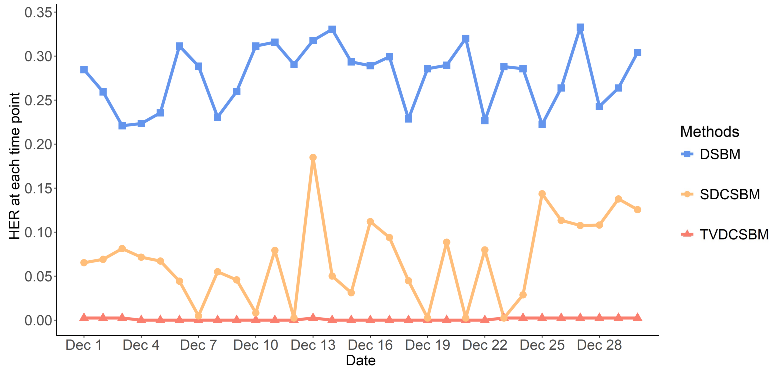


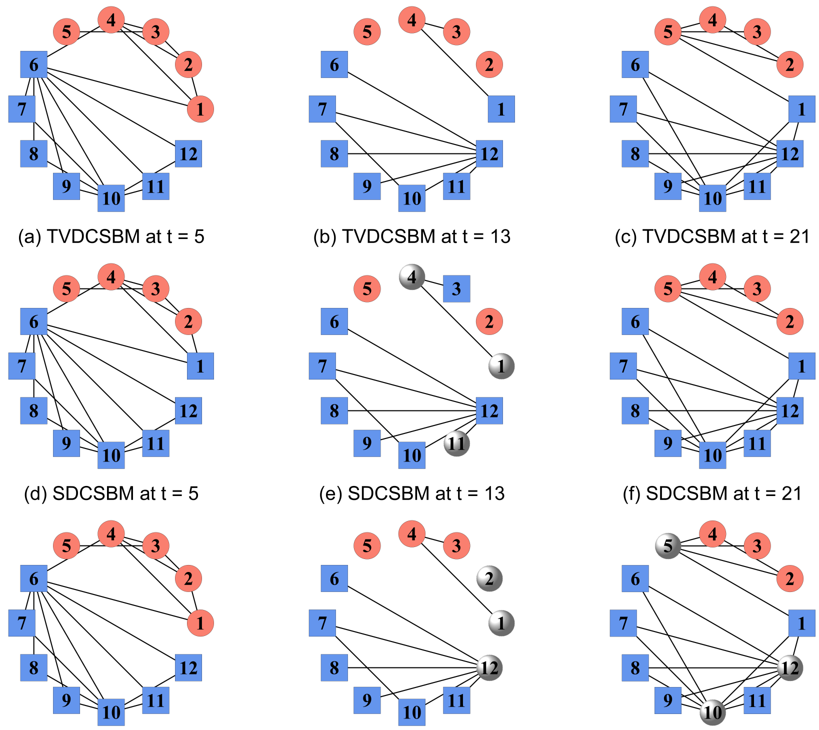
Figure 4: Community detection results for the Travian dataset at each time point for three methods: DSBM (blue squares), SDCSBM (orange circles) and TVDCSBM (red triangles).

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 be always 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 the 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 balanced community sizes ranging from 40 to 51 nodes in each community.

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

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 the 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 partitions found by TVDCSBM align remarkably well with 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



(a) TVDCSBM at t = 5 (b) TVDCSBM at t = 13 (c) TVDCSBM at t = 21

(d) SDCSBM at t = 5 (e) SDCSBM at t = 13 (f) SDCSBM at t = 21

(g) DSBM at t = 5 (h) DSBM at t = 13 (i) DSBM at t = 21

Figure 5: Results of the subnetwork in Travian dataset for three time points. The red circles and blue squares denote two distinct communities, while the gray spheres indicate the nodes misclassified into communities other than those depicted by the clusters of red circles and blue squares.

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 consistent 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 et al. (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 et al. (2019), or the use of node covariate information for improved community structure estimation.

References

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

Agterberg, J., Lubberts, Z., and Arroyo, J. (2022). Joint spectral clustering in multilayer degree-corrected stochastic blockmodels. arXiv preprint arXiv:2212.05053.

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

Bhattacharyya, S. and Chatterjee, S. (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, pages 92–103. Springer.

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

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

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

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

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

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

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

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

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

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

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

Keriven, N. and Vaiter, S. (2022). Sparse and smooth: improved guarantees for spectral clustering in the dynamic stochastic block model. Electronic Journal of Statistics, 16(1):1330–1366.

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

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

Li, D., Yuan, Y., Zhang, X., and Qu, A. (2024). Joint modeling of change-point identification and dependent dynamic community detection. Statistica Sinica Preprint No:SS-202100182.

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

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

Matias, C. and Miele, V. (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 Zhang, T. (2019). Spectral clustering in the dynamic stochastic block model. Electronic Journal of Statistics, 13(1):678–709.

Qin, T. and Rohe, K. (2013). Regularized spectral clustering under the degree-corrected stochastic blockmodel. In Advances in neural information processing systems, volume 26 of NIPS'13, pages 3120–3128, Red Hook, NY, USA. Curran Associates Inc.

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

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

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

Samdin, S. B., Ting, C.-M., and Ombao, H. (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), pages 1483–1487.

Sengupta, S. and Chen, Y. (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 Esseen, C.-G. (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., Zhang, J., Liu, B., Zhu, J., and Guo, J. (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 Bickel, P. J. (2017). Likelihood-based model selection for stochastic block models. The Annals of Statistics, 45(2):500–528.

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

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

8 Appendix

A Proofs of results in Section 3

Lemma A.1 (Bias and variance for $\hat{\tau}_{ij,h}(t/T)$). *Under Assumptions 1 and 2, the kernel-based estimator $\hat{\tau}_{ij,h}(t/T)$, as defined in (7), satisfies*

$$\begin{aligned} &\textit{Bias:} \qquad \mathbb{E}(\hat{\tau}_{ij,h}(t/T)) - \tau_{ij}(t/T) = \frac{\tau_{ij}^{''}(t/T)C^*}{12}h^2 + O(\rho_n h^3), \\ &\textit{Variance:} \quad Var(\hat{\tau}_{ij,h}(t/T)) = \frac{\sigma_{ij}^2(t/T)}{2Th} + O\left(\frac{\rho_n}{T}\right), \end{aligned}$$

where $2 < C^* \leq 6$ is a constant and $\sigma_{ij}^2(t/T) = \tau_{ij}(t/T)(1-\tau_{ij}(t/T))$.

This lemma serves as a preparatory step for subsequent proofs.

Proof. In Step 1, define an even number $2L$ to be the number of observations in the neighborhood $U(t/T)$, that is $L = \frac{1}{2}\left|\left\{A_{ij}^{t'}, t' \in U(t/T)\right\}\right| = Th$. Then $U(t/T)$ becomes $[(t-L)/T, (t+L)/T]$ and $\hat{\tau}_{ij,h}(t/T)$ can be redefined as $\hat{\tau}_{ij,h}(t/T) = \frac{1}{2L}\sum_{t'=t-L}^{t+L} A_{ij}^{t'}$. Combining this with equation (6), we obtain that

$$\hat{\tau}_{ij,h}(t/T) = \frac{1}{2L}\sum_{t'=t-L}^{t+L}\left[\tau_{ij}(t'/T) + \varepsilon_{ij}^{t'}\right] = \frac{1}{2L}\sum_{t'=t-L}^{t+L}\tau_{ij}(t'/T) + \frac{1}{2L}\sum_{t'=t-L}^{t+L}\varepsilon_{ij}^{t'}. \tag{A.1}$$

Based on Assumption 2, we know that the error $\varepsilon_{ij}^{t'}$ has zero mean, and then the expectation of $\hat{\tau}_{ij,h}(t/T)$ becomes

$$\mathbb{E}\left(\hat{\tau}_{ij,h}(t/T)\right) = \frac{1}{2L}\sum_{t'=t-L}^{t+L}\tau_{ij}(t'/T). \tag{A.2}$$

Note that Assumption 1 gives that the third derivative of $\tau_{ij}(t'/T)$ is bounded and proportional to ρ_n. For any $t'/T \in U(t/T)$, using a Taylor expansion of $\tau_{ij}(t'/T)$ at point t/T, we have that

$$\tau_{ij}(t'/T) = \tau_{ij}(t/T) + \tau_{ij}^{'}(t/T)\left(\frac{t'-t}{T}\right) + \frac{\tau_{ij}^{''}(t/T)}{2}\left(\frac{t'-t}{T}\right)^2 + O\left(\rho_n\left|\frac{t'-t}{T}\right|^3\right) \tag{A.3}$$

where $\tau_{ij}^{'}(t/T)$ and $\tau_{ij}^{''}(t/T)$ are the first and second derivatives evaluated in t/T. Next plugging (A.3) into (A.2) implies

$$\begin{aligned} \mathbb{E}\left(\hat{\tau}_{ij,h}(t/T)\right) =& \tau_{ij}(t/T) + \frac{\tau_{ij}^{'}(t/T)}{2L} \times \frac{\sum_{t'=t-L}^{t+L}(t'-t)}{T} + \frac{\tau_{ij}^{''}(t/T)}{4L} \\ &\times \frac{\sum_{t'=t-L}^{t+L}(t'-t)^2}{T^2} + O\left(\frac{\rho_n\sum_{t'=t-L}^{t+L}|t'-t|^3}{T^3 L}\right). \end{aligned}$$

Notice that $\sum_{t'=t-L}^{t+L}(t'-t) = 0$ in the second term. For the third and the remainder term, since we have that $\sum_{t'=t-L}^{t+L}(t'-t)^2 = \frac{1}{3}(2L^3+3L^2+L)$ and $\sum_{t'=t-L}^{t+L}|t'-t|^3 \leq 2L^4$ as $|t'-t| \leq L$, the bias can be written as

$$\begin{aligned} \mathbb{E}\left(\hat{\tau}_{ij,h}(t/T)\right) - \tau_{ij}(t/T) &= \frac{\tau_{ij}^{''}(t/T)}{12} \times \left(\frac{L}{T}\right)^2\left(2 + \frac{3}{L} + \frac{1}{L^2}\right) + O\left(\rho_n\left(\frac{L}{T}\right)^3\right) \\ &= \frac{\tau_{ij}^{''}(t/T)C^*}{12}h^2 + O\left(\rho_n h^3\right) \end{aligned}$$

for some constant $2 < C^* \le 6$ since the term $\left(2 + \frac{3}{L} + \frac{1}{L^2}\right) \in (2, 6]$ as $L \ge 1$.

We next consider the variance of $\hat{\tau}_{ij,h}(t/T)$. As Assumption 2 states the independence of errors $\varepsilon_{ij}^{t'}$ over time points $t' \in \{1, 2 \ldots, T\}$, equation (A.1) implies that

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

For any $t'/T \in U(t/T)$, we use a Taylor expansion of $\sigma_{ij}^2(t'/T)$ at point (t/T),

$$\sigma_{ij}^2(t'/T) = \sigma_{ij}^2(t/T) + O\left(\sigma_{ij}^{2\,'}(\xi_t)\left|\frac{t'-t}{T}\right|\right) = \sigma_{ij}^2(t/T) + O\left(\rho_n\left|\frac{t'-t}{T}\right|\right),$$

where $\sigma_{ij}^{2\,'}(\xi_t)$ represents the first derivative of $\sigma_{ij}^2(t/T)$ evaluated in $\xi_t = T^{-1}\left[t + \alpha'(t - t')\right]$ with $\alpha' \in (0, 1)$. Based on Assumptions 1 and 2, we conclude that $\sigma_{ij}^{2\,'}(\xi_t) = \tau_{ij}'(\xi_t) - 2\tau_{ij}(\xi_t)\tau_{ij}'(\xi_t)$ is proportional to ρ_n. Therefore, the variance of $\hat{\tau}_{ij,h}(t/T)$ takes the form

$$\mathrm{Var}(\hat{\tau}_{ij,h}(t/T)) = \frac{1}{4L^2}\sum_{t'=t-L}^{t+L}\sigma_{ij}^2(t/T) + O\left(\frac{\rho_n\sum_{t'=t-L}^{t+L}|t'-t|}{TL^2}\right) = \frac{\sigma_{ij}^2(t/T)}{2Th} + O\left(\frac{\rho_n}{T}\right).$$

□

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, $n^2\rho_n$. 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)**.** *Assume that for any time point $t \in \{1, \ldots, T\}$, both limits $C^t = \lim_{n\to\infty}\frac{\sum_{i,j=1}^n \tau_{ij}''(t/T)C^*}{n^2\rho_n}$ and $V^t = \lim_{n\to\infty}\frac{\sum_{i,j=1}^n \sigma_{ij}^2(t/T)}{n^2\rho_n}$ are constants. Under Assumptions 1-4, and given true community assignments $\mathbf{c}^t$, the smoothed counting variables, defined in (8), satisfy that*

$$\mathbb{E}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) = \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right) + C_{d_i}^t h^2 + O(h^3), \tag{A.4}$$

$$Var\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) = \frac{V_{d_i}^t}{2n\rho_n Th} + O\left(\frac{1}{n\rho_n T}\right) \tag{A.5}$$

for some constants $C_{d_i}^t = \frac{\sum_{j=1}^n \tau_{ij}''(t/T)C^*}{12n\rho_n}, V_{d_i}^t = \frac{\sum_{j=1}^n \sigma_{ij}^2(t/T)}{n\rho_n}$.

$$\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = H_{kl}(\boldsymbol{e}^t) + C_{ij}^t\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3\right), \tag{A.6}$$

$$Var\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{V_{ij}^t\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{2n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_n T}\right) \tag{A.7}$$

for any vector $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$ and some constants $C_{ij}^t = \frac{\sum_{i,j=1}^n \tau_{ij}''(t/T)C^*\mathbf{1}_{\{e_i^t=k, e_j^t=l\}}}{12n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)\rho_n}, V_{ij}^t = \frac{\sum_{i,j=1}^n \sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k, e_j^t=l\}}}{n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)\rho_n}$.

$$\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = H_k(\boldsymbol{e}^t) + C_i^t\tilde{\kappa}_k(\boldsymbol{e}^t)h^2 + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)h^3\right), \tag{A.8}$$

$$Var\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{V_i^t\tilde{\kappa}_k(\boldsymbol{e}^t)}{2n^2\rho_n Th} + O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_n T}\right) \tag{A.9}$$

for any vector $\boldsymbol{e}^t \in \{1,\ldots,K\}^n$ *and some constants* $C_i^t = \frac{\sum_{i,j=1}^n \tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k\}}}{12nn_k(\boldsymbol{e}^t)\rho_n}, V_i^t = \frac{\sum_{i,j=1}^n \sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k\}}}{nn_k(\boldsymbol{e}^t)\rho_n}$. *Furthermore, if there exist* $\rho_n, h \to 0$ *and* $nT \to \infty$ *such that* $n\rho_n Th \to \infty$, *then for any* $i \in \{1,\ldots,n\}, k,l \in \{1,\ldots,K\}$ *and any vector* $\boldsymbol{e}^t \in \{1,\ldots,K\}^n$,

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

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 $\hat{\tau}_{ij,h}(t/T)$'s are independent variables over all i and j as A_{ij}^t's are independent given $\boldsymbol{\theta}^t, \boldsymbol{c}^t$. Relying on the results of Lemma A.1, it holds

$$\begin{aligned}
\mathbb{E}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) &= \frac{1}{n\rho_n}\sum_{j=1}^n \mathbb{E}(\hat{\tau}_{ij,h}(t/T)) = \frac{1}{n\rho_n}\sum_{j=1}^n\left[\tau_{ij}(t/T) + \frac{\tau_{ij}^{''}(t/T)C^*}{12}h^2 + O(\rho_n h^3)\right] \\
&= \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right) + \frac{\sum_{j=1}^n \tau_{ij}^{''}(t/T)C^*}{12n\rho_n}h^2 + O(h^3) = \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right) + C_{d_i}^t h^2 + O(h^3),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right) &= \frac{1}{n^2\rho_n^2}\sum_{j=1}^n \mathrm{Var}(\hat{\tau}_{ij,h}(t/T)) = \frac{1}{n^2\rho_n^2}\sum_{j=1}^n\left[\frac{\sigma_{ij}^2(t/T)}{2Th} + O\left(\frac{\rho_n}{T}\right)\right] \\
&= \frac{\sum_{j=1}^n \sigma_{ij}^2(t/T)}{n\rho_n} \times \frac{1}{2n\rho_n Th} + O\left(\frac{1}{n\rho_n T}\right) = \frac{V_{d_i}^t}{2n\rho_n Th} + O\left(\frac{1}{n\rho_n T}\right),
\end{aligned}$$

where $C_{d_i}^t = \frac{\sum_{j=1}^n \tau_{ij}^{''}(t/T)C^*}{12n\rho_n}$ is proportional to C^t and $V_{d_i}^t = \frac{\sum_{j=1}^n \sigma_{ij}^2(t/T)}{n\rho_n}$ is proportional to V^t. Given that the limits C^t and V^t are finite, it is not difficult to show that $C_{d_i}^t$ and $V_{d_i}^t$ are also finite. Similarly, the expectations and variances of $\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}$ and $\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}$ can be written as follows,

$$\begin{aligned}
&\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^2\rho_n}\sum_{i,j=1}^n \mathbb{E}(\hat{\tau}_{ij,h}(t/T))\mathbf{1}_{\{e_i^t=k.e_j^t=l\}} \\
&= \frac{1}{n^2\rho_n}\sum_{i,j=1}^n\left[\tau_{ij}(t/T) + \frac{\tau_{ij}^{''}(t/T)C^*}{12}h^2 + O(\rho_n h^3)\right]\mathbf{1}_{\{e_i^t=k.e_j^t=l\}} \\
&= \frac{1}{n^2}\sum_{a,b=1}^K \frac{Z_{ab}^t}{\rho_n}\sum_{i=1}^n \theta_i^t\mathbf{1}_{\{e_i^t=k,c_i^t=a\}}\sum_{j=1}^n \theta_j^t\mathbf{1}_{\{e_j^t=l,c_j^t=b\}} + \frac{\sum_{i,j=1}^n \tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}}{12n^2\rho_n}h^2 + \frac{\sum_{i,j=1}^n \mathbf{1}_{\{e_i^t=k,e_j^t=l\}}O(h^3)}{n^2} \\
&= \frac{\sum_{a,b=1}^K W_{ab}^t\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n^2} + \frac{\sum_{i,j=1}^n \tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}}{12n^2\rho_n}h^2 + \frac{\sum_{i,j=1}^n \mathbf{1}_{\{e_i^t=k,e_j^t=l\}}O(h^3)}{n^2} \\
&= H_{kl}(\boldsymbol{e}^t) + C_{ij}^t\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3\right),
\end{aligned}$$

and

$$\operatorname{Var}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^{n}\operatorname{Var}(\hat{\tau}_{ij,h}(t/T))\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^{n}\left[\frac{\sigma_{ij}^2(t/T)}{2Th}+O\left(\frac{\rho_n}{T}\right)\right]\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}$$
$$= \frac{\sum_{i,j=1}^{n}\sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}}{n^2\rho_n}\times\frac{1}{2n^2\rho_nTh}+O\left(\frac{n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)}{n^4\rho_nT}\right)$$
$$= \frac{V_{ij}^t\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{2n^2\rho_nTh}+O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_nT}\right)$$

with constants $C_{ij}^t = \frac{\sum_{i,j=1}^{n}\tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}}{12n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)\rho_n}$ and $V_{ij}^t = \frac{\sum_{i,j=1}^{n}\sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k,e_j^t=l\}}}{n_k(\boldsymbol{e}^t)n_l(\boldsymbol{e}^t)\rho_n}$.

$$\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right) = \frac{1}{n^2\rho_n}\sum_{i,j=1}^{n}\mathbb{E}(\hat{\tau}_{ij,h}(t/T))\mathbf{1}_{\{e_i^t=k\}} = \frac{1}{n^2\rho_n}\sum_{i,j=1}^{n}\left[\tau_{ij}(t/T)+\frac{\tau_{ij}^{''}(t/T)C^*}{12}h^2+O(\rho_nh^3)\right]\mathbf{1}_{\{e_i^t=k\}}$$
$$= \frac{1}{n^2}\sum_{a,b=1}^{K}\frac{Z_{ab}^t}{\rho_n}\sum_{i=1}^{n}\theta_i^t\mathbf{1}_{\{e_i^t=k,c_i^t=a\}}\sum_{j=1}^{n}\theta_j^t\mathbf{1}_{\{c_j^t=b\}}+\frac{\sum_{i,j=1}^{n}\tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k\}}}{12n^2\rho_n}h^2+\frac{\sum_{i,j=1}^{n}\mathbf{1}_{\{e_i^t=k\}}O(h^3)}{n^2}$$
$$= \frac{\sum_{a,b=1}^{K}W_{ab}^tn_b(\boldsymbol{c}^t)\Lambda_{a(k)}(\boldsymbol{c}^t,\boldsymbol{e}^t)}{n^2}+\frac{\sum_{i,j=1}^{n}\tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k\}}}{12n^2\rho_n}h^2+\frac{\sum_{i,j=1}^{n}\mathbf{1}_{\{e_i^t=k\}}O(h^3)}{n^2}$$
$$= H_k(\boldsymbol{e}^t)+C_i^t\tilde{\kappa}_k(\boldsymbol{e}^t)h^2+O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)h^3\right),$$
$$\operatorname{Var}\left(\frac{\hat{O}_{k,h}(\boldsymbol{c}^t)}{n^2\rho_n}\right) = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^{n}\operatorname{Var}(\hat{\tau}_{ij,h}(t/T))\mathbf{1}_{\{e_i^t=k\}} = \frac{1}{n^4\rho_n^2}\sum_{i,j=1}^{n}\left[\frac{\sigma_{ij}^2(t/T)}{2Th}+O\left(\frac{\rho_n}{T}\right)\right]\mathbf{1}_{\{e_i^t=k\}}$$
$$= \frac{\sum_{i,j=1}^{n}\sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k\}}}{n^2\rho_n}\times\frac{1}{2n^2\rho_nTh}+O\left(\frac{n_k(\boldsymbol{e}^t)}{n^3\rho_nT}\right) = \frac{V_i^t\tilde{\kappa}_k(\boldsymbol{e}^t)}{2n^2\rho_nTh}+O\left(\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_nT}\right)$$

for constants $C_i^t = \frac{\sum_{i,j=1}^{n}\tau_{ij}^{''}(t/T)C^*\mathbf{1}_{\{e_i^t=k\}}}{12nn_k(\boldsymbol{e}^t)\rho_n}$ and $V_i^t = \frac{\sum_{i,j=1}^{n}\sigma_{ij}^2(t/T)\mathbf{1}_{\{e_i^t=k\}}}{nn_k(\boldsymbol{e}^t)\rho_n}$. Hence, the results of equations (A.6) through (A.9) are obtained. By the mean-variance decomposition, the MSE of each smoothed counting variable tends to 0 as

$$\mathbb{E}\left|\frac{\hat{d}_{i,h}^t}{n\rho_n}-\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right)\right|^2 = \left[\mathbb{E}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right)-\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d_i^t}{n\rho_n}\right)\right]^2+\operatorname{Var}\left(\frac{\hat{d}_{i,h}^t}{n\rho_n}\right)$$
$$= O\left(h^4+\frac{1}{n\rho_nTh}\right)\to 0,$$
$$\mathbb{E}\left|\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}-H_{kl}(\boldsymbol{e}^t)\right|^2 = \left[\mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right)-H_{kl}(\boldsymbol{e}^t)\right]^2+\operatorname{Var}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right)$$
$$= O\left(\tilde{\kappa}_k^2(\boldsymbol{e}^t)\tilde{\kappa}_l^2(\boldsymbol{e}^t)h^4+\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)}{n^2\rho_nTh}\right)\to 0,$$
$$\mathbb{E}\left|\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}-H_k(\boldsymbol{e}^t)\right|^2 = \left[\mathbb{E}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right)-H_k(\boldsymbol{e}^t)\right]^2+\operatorname{Var}\left(\frac{\hat{O}_{k,h}(\boldsymbol{e}^t)}{n^2\rho_n}\right)$$
$$= O\left(\tilde{\kappa}_k^2(\boldsymbol{e}^t)h^4+\frac{\tilde{\kappa}_k(\boldsymbol{e}^t)}{n^2\rho_nTh}\right)\to 0$$

as $\rho_n, h\to 0$ and $nT\to\infty$ such that $n\rho_nTh\to\infty$. 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 $\tilde{\tau}_{ij,h}(t/T)$ for any $i,j \in \{1,\ldots,n\}$. Define

$$\boldsymbol{u} = \left(\frac{\hat{d}^t_{i,h}}{n\rho_n}, \frac{\hat{d}^t_{j,h}}{n\rho_n}, \frac{\hat{O}_{c^t_i c^t_j,h}(\boldsymbol{c}^t)}{n^2\rho_n}, \frac{\hat{O}_{c^t_i,h}(\boldsymbol{c}^t)}{n^2\rho_n}, \frac{\hat{O}_{c^t_j,h}(\boldsymbol{c}^t)}{n^2\rho_n} \right)^\top,$$

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

$$\tilde{\tau}_{ij,h}(t/T) = f(\boldsymbol{u}) = \frac{\rho_n u_1 u_2 u_3}{u_4 u_5}.$$

Correspondingly, let

$$\boldsymbol{\nu} = \left(\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d^t_i}{n\rho_n}\right), \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}\left(\frac{d^t_j}{n\rho_n}\right), H_{c^t_i c^t_j}(\boldsymbol{c}^t), H_{c^t_i}(\boldsymbol{c}^t), H_{c^t_j}(\boldsymbol{c}^t) \right)^\top,$$

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

Remark 7. Let $\kappa_{c^t_i} = n_{c^t_i}(\boldsymbol{c}^t)/n$ 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 tends to infinity, the scaled terms $(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\nu_3$, $\kappa^{-1}_{c^t_i}\nu_4$ and $\kappa^{-1}_{c^t_j}\nu_5$ are bounded below and above by non-zero constants. It is natural to allow the lower and upper bound of $\boldsymbol{\theta}^t$ and $\boldsymbol{W}^t$ satisfy $C_{\inf} = \left(\inf_{\substack{i\in\{1,\ldots,n\},\\ t\in\{1,\ldots,T\}}} \theta^t_i\right)^2 \left(\inf_{\substack{i,j\in\{1,\ldots,n\},\\ t\in\{1,\ldots,T\}}} W^t_{c^t_i c^t_j}\right)$ and $C_{\sup} = \left(\sup_{\substack{i\in\{1,\ldots,n\},\\ t\in\{1,\ldots,T\}}} \theta^t_i\right)^2 \left(\sup_{\substack{i,j\in\{1,\ldots,n\},\\ t\in\{1,\ldots,T\}}} W^t_{c^t_i c^t_j}\right)$ are positive constants independent of n such that $C_{\inf}\rho_n \le \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A^t_{ij}) \le C_{\sup}\rho_n$ over $i,j \in \{1,\ldots,n\}$. It is easy to see that

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

$$\kappa_{c^t_i}\kappa_{c^t_j} C_{\inf} \le H_{c^t_i c^t_j}(\boldsymbol{c}^t) = \frac{\sum_{i,j=1}^n \mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(A^t_{ij})\mathbf{1}_{\{e^t_i=k, e^t_j=l\}}}{n^2\rho_n} \le \kappa_{c^t_i}\kappa_{c^t_j} C_{\sup},$$

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

which means that ν_1, ν_2 and the scaled terms $(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\nu_3, \kappa^{-1}_{c^t_i}\nu_4, \kappa^{-1}_{c^t_j}\nu_5$ are bounded below and above by non-zero constants.

To simplify the notations, define the index set $\mathcal{N}^t_a = \{1 \le i \le n : c^t_i = a\}$ of nodes whose true community is a. With this definition, it is clear that the identifiability of θ^t_i in Assumption 4 is equivalent to $\sum_{s\in\mathcal{N}^t_{c^t_i}} \theta^t_s = n_{c^t_i}(\boldsymbol{c}^t)$. By a Taylor expansion around the point $\boldsymbol{\nu}$, we have that

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

where $\nabla f(\boldsymbol{\xi})^\top (\boldsymbol{u} - \boldsymbol{\nu})$ is the Taylor remainder term. The term $\nabla f(\boldsymbol{\xi})$ denotes the vector of first derivatives of the function $f(\boldsymbol{u})$ evaluated in $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_5)^\top$ where $\xi_m = u_m + \alpha(u_m - \nu_m)$ 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 $\tau_{ij}(t/T)$. Given the community labels $\boldsymbol{c}^t$ and the identifiability of $\boldsymbol{\theta}^t$ via Assumption 4, we have that

$$
\begin{aligned}
f(\boldsymbol{\nu}) &= \frac{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_i^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(d_j^t)\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t c_j^t}(\boldsymbol{c}^t))}{\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_i^t}(\boldsymbol{c}^t))\mathbb{E}_{\boldsymbol{c}^t,\boldsymbol{\theta}^t}(O_{c_j^t}(\boldsymbol{c}^t))} \\
&= \frac{\left(\sum_{j=1}^n \theta_i^t\theta_j^t Z_{c_i^t c_j^t}^t\right)\left(\sum_{i=1}^n \theta_i^t\theta_j^t Z_{c_i^t c_j^t}^t\right)\left(\sum_{s\in\mathcal{N}_{c_i^t}^t, m\in\mathcal{N}_{c_j^t}^t} \theta_s^t\theta_m^t Z_{c_i^t c_j^t}^t\right)}{\left(\sum_{s\in\mathcal{N}_{c_i^t}^t}\sum_{j=1}^n \theta_s^t\theta_j^t Z_{c_i^t c_j^t}^t\right)\left(\sum_{m\in\mathcal{N}_{c_j^t}^t}\sum_{i=1}^n \theta_m^t\theta_i^t Z_{c_i^t c_j^t}^t\right)} \\
&= \frac{\theta_i^t\left(\sum_{j=1}^n \theta_j^t Z_{c_i^t c_j^t}^t\right)\times\theta_j^t\left(\sum_{i=1}^n \theta_i^t Z_{c_i^t c_j^t}^t\right)\times\left(\sum_{s\in\mathcal{N}_{c_i^t}^t}\theta_s^t\right)\left(\sum_{m\in\mathcal{N}_{c_j^t}^t}\theta_m^t\right)Z_{c_i^t c_j^t}^t}{\left(\sum_{s\in\mathcal{N}_{c_i^t}^t}\theta_s^t\right)\left(\sum_{j=1}^n \theta_j^t Z_{c_i^t c_j^t}^t\right)\left(\sum_{m\in\mathcal{N}_{c_j^t}^t}\theta_m^t\right)\left(\sum_{i=1}^n \theta_i^t Z_{c_i^t c_j^t}^t\right)} \\
&= \theta_i^t\theta_j^t Z_{c_i^t c_j^t}^t = \tau_{ij}(t/T). \qquad \text{(A.10)}
\end{aligned}
$$

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

$$
R_{f1} = \frac{\rho_n\xi_2\xi_3}{\xi_4\xi_5}(u_1-\nu_1), \quad R_{f2} = \frac{\rho_n\xi_1\xi_3}{\xi_4\xi_5}(u_2-\nu_2), \quad R_{f3}\frac{\rho_n\xi_1\xi_2}{\xi_4\xi_5}(u_3-\nu_3),
$$
$$
R_{f4} = \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4^2\xi_5}(u_4-\nu_4), \quad R_{f5} = \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4\xi_5^2}(u_5-\nu_5).
$$

Combining it with (A.10), the Taylor series for $\tilde{\tau}_{ij,h}(t/T)$ becomes

$$
\tilde{\tau}_{ij,h}(t/T) = \tau_{ij}(t/T) + R_{f1} + R_{f2} + R_{f3} - R_{f4} - R_{f5}. \qquad \text{(A.11)}
$$

To show the asymptotic behaviour of this remainder term, we need to show the convergence of each part R_{fm}. Since R_{f2} has the same behaviour as R_{f1} and as well R_{f5} has the same behaviour 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 $m \in \{1,\ldots,5\}$, u_m converges to its population version ν_m in quadratic mean which implies that $\xi_m \in (u_m,\nu_m)$ also converges to ν_m as $h,\rho_n \to 0$ and $nT\to\infty$ such that $n\rho_n Th\to\infty$. 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

$$
(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\xi_3 \xrightarrow{P} (\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3, \qquad \frac{1}{\kappa_{c_i^t}^{-1}\xi_4} \xrightarrow{P} \frac{1}{\kappa_{c_i^t}^{-1}\nu_4}, \qquad \text{and} \qquad \frac{1}{\kappa_{c_j^t}^{-1}\xi_5} \xrightarrow{P} \frac{1}{\kappa_{c_j^t}^{-1}\nu_5} \qquad \text{(A.12)}
$$

as $\rho_n, h\to 0$ and $n\rho_n Th\to\infty$, which implies that

$$
\begin{aligned}
R_{f1} = \frac{\rho_n\xi_2\xi_3}{\xi_4\xi_5}(u_1-\nu_1) &= \frac{\xi_2\times(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\xi_3}{\left(\kappa_{c_i^t}^{-1}\xi_4\right)\left(\kappa_{c_j^t}^{-1}\xi_5\right)}\times\rho_n(u_1-\nu_1) \\
&= \left[\frac{\nu_2\times(\kappa_{c_i^t}\kappa_{c_j^t})^{-1}\nu_3}{\left(\kappa_{c_i^t}^{-1}\nu_4\right)\left(\kappa_{c_j^t}^{-1}\nu_5\right)} + o_p(1)\right]\times\rho_n(u_1-\nu_1) \\
&= M_{\nabla f_1}\rho_n(u_1-\nu_1) + o_p(\rho_n(u_1-\nu_1)), \qquad \text{(A.13)}
\end{aligned}
$$

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.13).

By equation (A.4), we have that

$$\begin{aligned}\mathbb{E}(R_{f1}) &= M_{\nabla f_1}\rho_n \mathbb{E}\left(u_1-\nu_1\right)+o(\rho_n\mathbb{E}(u_1-\nu_1))\\ &= M_{\nabla f_1}\rho_n C^t_{d_i}h^2+O(\rho_n h^3)+o(\rho_n h^2)=M_{\nabla f_1}C^t_{d_i}\rho_n h^2+o(\rho_n h^2),\end{aligned} \tag{A.14}$$

and also substituting the variance (A.5), we have that

$$\begin{aligned}\mathrm{Var}(R_{f1}) &= M^2_{\nabla f_1}\rho_n^2 \mathrm{Var}\left(u_1-\nu_1\right)+o(\rho_n^2\mathrm{Var}(u_1-\nu_1))\\ &= M^2_{\nabla f_1}\rho_n^2\times\frac{V^t_{d_i}}{2n\rho_n Th}+O\left(\rho_n^2\times\frac{1}{n\rho_n T}\right)+o\left(\rho_n^2\times\frac{1}{n\rho_n Th}\right)\\ &= \frac{M^2_{\nabla f_1}V^t_{d_i}\rho_n}{2nTh}+o\left(\frac{\rho_n}{nTh}\right).\end{aligned} \tag{A.15}$$

Step 2: R_{f3}. We use next the same technique to handle R_{f3}. Notice that by expression (A.12) and Slutsky's Theorem, it is easy to show that

$$\begin{aligned}R_{f3} &= \frac{\rho_n\xi_1\xi_2}{\xi_4\xi_5}\left(u_3-\nu_3\right)=\frac{\xi_1\xi_2}{\left(\kappa^{-1}_{c^t_i}\xi_4\right)\left(\kappa^{-1}_{c^t_j}\xi_5\right)}\times(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n(u_3-\nu_3)\\ &= \left[\frac{\nu_1\nu_2}{\left(\kappa^{-1}_{c^t_i}\nu_4\right)\left(\kappa^{-1}_{c^t_j}\nu_5\right)}+o_p(1)\right]\times(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n(u_3-\nu_3)\\ &= M_{\nabla f_3}(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n(u_3-\nu_3)+o_p\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n(u_3-\nu_3)\right)\end{aligned} \tag{A.16}$$

with a constant $M_{\nabla f_3}=\frac{\nu_1\nu_2}{\left(\kappa^{-1}_{c^t_i}\nu_4\right)\left(\kappa^{-1}_{c^t_j}\nu_5\right)}$. Following (A.16), (A.6) and (A.7), we have that

$$\begin{aligned}\mathbb{E}(R_{f3}) &= M_{\nabla f_3}(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n\mathbb{E}\left(u_3-\nu_3\right)+o\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n\mathbb{E}(u_3-\nu_3)\right)\\ &=M_{\nabla f_3}(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n\times C^t_{ij}\kappa_{c^t_i}\kappa_{c^t_j}h^2+O\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n\times\kappa_{c^t_i}\kappa_{c^t_j}h^3\right)+o\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\rho_n\times\kappa_{c^t_i}\kappa_{c^t_j}h^2\right)\\ &=M_{\nabla f_3}C^t_{ij}\rho_n h^2+o(\rho_n h^2),\end{aligned} \tag{A.17}$$

$$\begin{aligned}\mathrm{Var}(R_{f3}) &= M^2_{\nabla f_3}(\kappa_{c^t_i}\kappa_{c^t_j})^{-2}\rho_n^2\mathrm{Var}\left(u_3-\nu_3\right)+o\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-2}\rho_n^2\mathrm{Var}(u_3-\nu_3)\right)\\ &=M^2_{\nabla f_3}(\kappa_{c^t_i}\kappa_{c^t_j})^{-2}\rho_n^2\times\frac{V^t_{ij}\kappa_{c^t_i}\kappa_{c^t_j}}{2n^2\rho_n Th}+O\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-2}\rho_n^2\times\frac{\kappa_{c^t_i}\kappa_{c^t_j}}{n^2\rho_n T}\right)+o\left((\kappa_{c^t_i}\kappa_{c^t_j})^{-2}\rho_n^2\times\frac{\kappa_{c^t_i}\kappa_{c^t_j}}{n^2\rho_n Th}\right)\\ &=\frac{M^2_{\nabla f_3}V^t_{ij}\rho_n}{2n^2\kappa_{c^t_i}\kappa_{c^t_j}Th}+o\left(\frac{\rho_n}{n^2\kappa_{c^t_i}\kappa_{c^t_j}Th}\right).\end{aligned} \tag{A.18}$$

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

$$\begin{aligned}R_{f4} &= \frac{\rho_n\xi_1\xi_2\xi_3}{\xi_4^2\xi_5}\left(u_4-\nu_4\right)=\frac{\xi_1\xi_2\times(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\xi_3}{\left(\kappa^{-2}_{c^t_i}\xi_4^2\right)\left(\kappa^{-1}_{c^t_j}\xi_5\right)}\times\kappa^{-1}_{c^t_i}\rho_n(u_4-\nu_4)\\ &= \left[\frac{\nu_1\nu_2\times(\kappa_{c^t_i}\kappa_{c^t_j})^{-1}\nu_3}{\left(\kappa^{-2}_{c^t_i}\nu_4^2\right)\left(\kappa^{-1}_{c^t_j}\nu_5\right)}+o_p(1)\right]\times\kappa^{-1}_{c^t_i}\rho_n(u_4-\nu_4)\\ &= M_{\nabla f_4}\kappa^{-1}_{c^t_i}\rho_n(u_4-\nu_4)+o_p\left(\kappa^{-1}_{c^t_i}\rho_n(u_4-\nu_4)\right)\end{aligned} \tag{A.19}$$

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.19), (A.8) and (A.9), we calculate that

$$
\begin{aligned}
\mathbb{E}(R_{f4}) &= M_{\nabla f_4} \kappa_{c_i^t}^{-1} \rho_n \mathbb{E}\left(u_4 - \nu_4\right) + o\left(\kappa_{c_i^t}^{-1} \rho_n \mathbb{E}(u_4 - \nu_4)\right) \\
&= M_{\nabla f_4} \kappa_{c_i^t}^{-1} \rho_n \times C_i^t \kappa_{c_i^t} h^2 + O\left(\kappa_{c_i^t}^{-1} \rho_n \times \kappa_{c_i^t} h^3\right) + o\left(\kappa_{c_i^t}^{-1} \rho_n \times \kappa_{c_i^t} h^2\right) \\
&= M_{\nabla f_4} C_i^t \rho_n h^2 + o(\rho_n h^2), \qquad (A.20) \\
\mathrm{Var}(R_{f4}) &= M_{\nabla f_4}^2 \kappa_{c_i^t}^{-2} \rho_n^2 \mathrm{Var}\left(u_4 - \nu_4\right) + o\left(\kappa_{c_i^t}^{-2} \rho_n^2 \mathrm{Var}(u_4 - \nu_4)\right) \\
&= M_{\nabla f_4}^2 \kappa_{c_i^t}^{-2} \rho_n^2 \times \frac{V_i^t \kappa_{c_i^t}}{2n^2 \rho_n T h} + O\left(\kappa_{c_i^t}^{-2} \rho_n^2 \times \frac{\kappa_{c_i^t}}{n^2 \rho_n T}\right) + o\left(\kappa_{c_i^t}^{-2} \rho_n^2 \times \frac{\kappa_{c_i^t}}{n^2 \rho_n T h}\right) \\
&= \frac{M_{\nabla f_4}^2 V_i^t \rho_n}{2n^2 \kappa_{c_i^t} T h} + o\left(\frac{\rho_n}{n^2 \kappa_{c_i^t} T h}\right). \qquad (A.21)
\end{aligned}
$$

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,

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

Therefore, taking the expectation on both sides of equation (A.11), and then combining (A.14), (A.17) and (A.20), the bias for $\tilde{\tau}_{ij,h}(t/T)$ is determined as

$$
\begin{aligned}
&\mathbb{E}(\tilde{\tau}_{ij,h}(t/T)) - \tau_{ij}(t/T) = \mathbb{E}(R_{f1}) + \mathbb{E}(R_{f2}) + \mathbb{E}(R_{f3}) - \mathbb{E}(R_{f4}) - \mathbb{E}(R_{f5}) \\
&= \left(M_{\nabla f_1} C_{d_i}^t + M_{\nabla f_2} C_{d_j}^t + M_{\nabla f_3} C_{ij}^t - M_{\nabla f_4} C_i^t - M_{\nabla f_5} C_j^t\right) \rho_n h^2 + o(\rho_n h^2) \\
&= \tau_{ij}(t/T) + C_1 \rho_n h^2 + o(\rho_n h^2), \qquad (A.22)
\end{aligned}
$$

where C_1 is a constant related to $\theta_i^t, W_{c_i^t c_j^t}^t$ and $\tau_{ij}^{''}(t/T)$ for all $i, j \in \{1, \ldots, n\}$.

We next proceed to investigate the variance for $\tilde{\tau}_{ij,h}(t/T)$. Using equation (A.11), for any $i \neq j \in \{1, \ldots, n\}$, the variance can be expressed as,

$$
\begin{aligned}
\mathrm{Var}(\tilde{\tau}_{ij,h}(t/T)) &= \mathrm{Var}\left(R_{f1} + R_{f2} + R_{f3} - R_{f4} - R_{f5}\right) \\
&\leq \mathbb{E}\left(\sum_{m=1}^{5} |R_{fm} - \mathbb{E}(R_{fm})|\right)^2 \\
&\leq 5 \sum_{m=1}^{5} \mathbb{E}\left(R_{fm} - \mathbb{E}(R_{fm})\right)^2 = 5 \sum_{m=1}^{5} \mathrm{Var}\left(R_{fm}\right). \qquad (A.23)
\end{aligned}
$$

Note that the C_r inequality states that $\left(\sum_{i=1}^{n} |x_i|\right)^r \leq C_r \sum_{i=1}^{n} |x_i|^r$ 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.23). Next combining (A.15), (A.18) and (A.21), inequality (A.23) implies that

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

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 $\kappa_n \leq \kappa_{c_i^t}, \forall i \in \{1,\dots,n\}$. Additionally, due to $n\kappa_n \leq n^2\kappa_n^2 \leq n^2\kappa_n$ and $n\kappa_n \leq n$, all the terms $O\left(\frac{\rho_n}{n^2\kappa_n Th}\right), O\left(\frac{\rho_n}{n^2\kappa_n^2 Th}\right)$ and $O\left(\frac{\rho_n}{nTh}\right)$ are of smaller or equivalent order compared to $O\left(\frac{\rho_n}{n\kappa_n Th}\right)$ for any κ_n. Then the variance for $\tilde{\tau}_{ij,h}(t/T)$ becomes

$$\mathrm{Var}(\tilde{\tau}_{ij,h}(t/T)) \leq \frac{C_2\rho_n}{2n\kappa_n Th} + O\left(\frac{\rho_n}{n\kappa_n T}\right), \tag{A.24}$$

where C_2 is a constant related to $\theta_i^t, W_{c_i^t c_j^t}^t$ for all $i,j \in \{1,\dots,n\}$. To investigate the essential issue of the impact of network sparsity on parameter estimations, we consider the scaled quantity $(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T))$ since $\tau_{ij}(t/T)$ is proportional to ρ_n. Recalling the definition of $d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T), \rho_n^{-1}\boldsymbol{\tau}(t/T))$, we have that

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

where $\Delta_{1,h}^t = \frac{1}{n^2\rho_n^2}\sum_{i,j=1}^n \mathbb{E}\left[\tilde{\tau}_{ij,h}(t/T) - \mathbb{E}(\tilde{\tau}_{ij,h}(t/T))\right]^2, \Delta_{2,h}^t = \frac{1}{n^2\rho_n^2}\sum_{i,j=1}^n \left[\mathbb{E}\left(\tilde{\tau}_{ij,h}(t/T)\right) - \tau_{ij}(t/T)\right]^2$. Note that $\Delta_{1,h}^t$ and $\Delta_{2,h}^t$ represent the variance and the squared bias, respectively. Combining (A.22) and (A.24), we obtain that the leading term of $d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T), \rho_n^{-1}\boldsymbol{\tau}(t/T))$ is

$$d(\rho_n^{-1}\tilde{\boldsymbol{\tau}}_h(t/T), \rho_n^{-1}\boldsymbol{\tau}(t/T)) \leq C_1 h^4 + \frac{C_2}{2n\kappa_n\rho_n Th}.$$

□

B Proofs of results in Section 4

Proof of Theorem 1. This proof is similar to the proof of Proposition 1. As $\hat{\theta}_{i,h}^t$, defined in (3) can be seen as a function of $u_1 = \frac{\hat{d}_{i,h}^t}{n\rho_n}$ and $u_4 = \frac{\hat{O}_{c_i^t,h}(\boldsymbol{c}^t)}{n^2\rho_n}$, we have that

$$\hat{\theta}_{i,h}^t = g(u_1, u_4) = \frac{n_{c_i^t}(\boldsymbol{c}^t)u_1}{nu_4} = \frac{u_1}{\kappa_{c_i^t}^{-1}u_4}.$$

Since we have that

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

equality which holds given the $\boldsymbol{c}^t$ and by the assumption $\sum_{s\in\mathcal{N}_{c_i^t}^t}\theta_s^t=n_{c_i^t}(\boldsymbol{c}^t)$, we use again a Taylor expansion of $\hat{\theta}_{i,h}^t$ around the point $(\nu_1,\nu_4)^\top$ implying that

$$\begin{aligned}\hat{\theta}_{i,h}^t=&g(\nu_1,\nu_4)+\nabla g(\xi_1',\xi_4')^\top\begin{pmatrix}u_1-\nu_1\\u_4-\nu_4\end{pmatrix}\\=&\theta_i^t+\frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}(u_1-\nu_1)-\frac{\xi_1'}{\kappa_{c_i^t}^{-1}(\xi_4')^2}(u_4-\nu_4)\\=&\theta_i^t+R_{g1}-R_{g4}\end{aligned}\tag{B.1}$$

where the term $\nabla g(\xi_1',\xi_4')^\top\binom{u_1-\nu_1}{u_4-\nu_4}$ 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 $g(u_1,u_4)$ evaluated in $(\xi_1',\xi_4')^\top$ with $\xi_m'=u_m+\alpha'(u_m-\nu_m)$ for some $\alpha'\in(0,1)$ and any $m\in\{1,4\}$. We then rewrite the remainder term as $(R_{g1}-R_{g4})$ where

$$R_{g1}=\frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}(u_1-\nu_1),\quad R_{g4}=\frac{\xi_1'}{\kappa_{c_i^t}^{-1}(\xi_4')^2}(u_4-\nu_4).$$

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

Given that $\xi_m'\in(u_m,\nu_m),\forall m\in\{1,4\}$, we conclude from Lemma A.2 that $(\xi_1',\xi_4')^\top$ converges to $(\nu_1,\nu_4)^\top$ in quadratic mean as $\rho_n,h\to0$ and $n\rho_nTh\to\infty$. Considering that $\kappa_{c_i^t}^{-1}\nu_4$ is a non-zero constant, as stated in Remark 7, we have that

$$\frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}\xrightarrow{P}\frac{1}{\kappa_{c_i^t}^{-1}\nu_4},\quad\frac{1}{\kappa_{c_i^t}^{-2}(\xi_4')^2}\xrightarrow{P}\frac{1}{\kappa_{c_i^t}^{-2}\nu_4^2}\quad\text{as }\rho_n,h\to0,\ n\rho_nTh\to\infty.\tag{B.2}$$

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

$$\begin{aligned}R_{g1}&=\frac{1}{\kappa_{c_i^t}^{-1}\xi_4'}(u_1-\nu_1)=\left[M_{\nabla g_1}+o_p(1)\right](u_1-\nu_1)\\&=M_{\nabla g_1}(u_1-\nu_1)+o_p(u_1-\nu_1).\end{aligned}$$

Taking the expectation on both sides and plugging in the expectation from (A.4), we obtain that

$$\mathbb{E}(R_{g1})=M_{\nabla g_1}\mathbb{E}(u_1-\nu_1)+o\left(\mathbb{E}(u_1-\nu_1)\right)=M_{\nabla g_1}C_{d_i}^th^2+o\left(h^2\right),\tag{B.3}$$

furthermore, by (A.5),

$$\begin{aligned}\mathrm{Var}(R_{g1})&=M_{\nabla g_1}^2\mathrm{Var}(u_1-\nu_1)+o\left(\mathrm{Var}(u_1-\nu_1)\right)\\&=\frac{M_{\nabla g_1}^2V_{d_i}^t}{2n\rho_nTh}+O\left(\frac{1}{n\rho_nT}\right)+o\left(\frac{1}{n\rho_nTh}\right)\\&=\frac{M_{\nabla g_1}^2V_{d_i}^t}{2n\rho_nTh}+o\left(\frac{1}{n\rho_nTh}\right).\end{aligned}\tag{B.4}$$

As for R_{g4}, applying Slutsky's Theorem once again, expression (B.2) yields that

$$
\begin{aligned}
R_{g4} &= \frac{\xi_1'}{\kappa_{c_i^t}^{-1}(\xi_4')^2}(u_4-\nu_4) = \frac{\xi_1'}{\kappa_{c_i^t}^{-2}(\xi_4')^2} \times \kappa_{c_i^t}^{-1}(u_4-\nu_4) \\
&= \left[\frac{\nu_1}{\kappa_{c_i^t}^{-2}\nu_4^2} + o_p(1)\right] \times \kappa_{c_i^t}^{-1}(u_4-\nu_4) \\
&= M_{\nabla g_4}\kappa_{c_i^t}^{-1}(u_4-\nu_4) + o_p\left(\kappa_{c_i^t}^{-1}(u_4-\nu_4)\right),
\end{aligned}
$$

where $M_{\nabla g_4} = \frac{\nu_1}{\kappa_{c_i^t}^{-2}\nu_4^2}$. Then, by (A.8) and (A.9), the expectation and variance can be obtained as follows,

$$
\begin{aligned}
\mathbb{E}(R_{g4}) &= M_{\nabla g_4}\kappa_{c_i^t}^{-1}\mathbb{E}(u_4-\nu_4) + o\left(\kappa_{c_i^t}^{-1}\mathbb{E}(u_4-\nu_4)\right) \\
&= M_{\nabla g_4}\kappa_{c_i^t}^{-1} \times C_i^t\kappa_{c_i^t}h^2 + O\left(\kappa_{c_i^t}^{-1} \times \kappa_{c_i^t}h^3\right) + o\left(\kappa_{c_i^t}^{-1} \times \kappa_{c_i^t}h^2\right) \\
&= M_{\nabla g_4}C_i^t h^2 + o\left(h^2\right),
\end{aligned} \tag{B.5}
$$

and

$$
\begin{aligned}
\mathrm{Var}(R_{g4}) &= M_{\nabla g_4}^2\kappa_{c_i^t}^{-2}\mathrm{Var}(u_4-\nu_4) + o\left(\kappa_{c_i^t}^{-2}\mathrm{Var}(u_4-\nu_4)\right) \\
&= M_{\nabla g_4}^2\kappa_{c_i^t}^{-2} \times \frac{V_i^t\kappa_{c_i^t}}{2n^2\rho_n Th} + O\left(\kappa_{c_i^t}^{-2} \times \frac{\kappa_{c_i^t}}{n^2\rho_n T}\right) + o\left(\kappa_{c_i^t}^{-2} \times \frac{\kappa_{c_i^t}}{n^2\rho_n Th}\right) \\
&= \frac{M_{\nabla g_4}^2 V_i^t}{2n^2\kappa_{c_i^t}\rho_n Th} + o\left(\frac{1}{n^2\kappa_{c_i^t}\rho_n Th}\right).
\end{aligned} \tag{B.6}
$$

Taking expectation on both sides in equation (B.1) and combining equations (B.3) and (B.5), we have that

$$
\begin{aligned}
\mathbb{E}(\hat{\theta}_{i,h}^t) &= \theta_i^t + \mathbb{E}(R_{g1}) - \mathbb{E}(R_{g4}) = \theta_i^t + \left[M_{\nabla g_1}C_{d_i}^t - M_{\nabla g_4}C_i^t\right]h^2 + o\left(h^2\right) \\
&= \theta_i^t + C_3h^2 + o(h^2)
\end{aligned}
$$

with a constant C_3 related to $\theta_i^t, W_{c_i^t c_j^t}^t$ and $\tau_{ij}^{''}(t/T)$ for all $i,j \in \{1,\ldots,n\}$, which implies that the bias of $\hat{\theta}_{i,h}^t$ is as follows,

$$
\mathbb{E}(\hat{\theta}_{i,h}^t) - \theta_i^t = C_3h^2 + o(h^2).
$$

Since the C_r inequality states that $\mathrm{Var}(x_1+x_2) \leq 2\big(\mathrm{Var}(x_1)+\mathrm{Var}(x_2)\big)$ as $r=2$ and $n=2$, from (B.1), we conclude that

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

and then substitute (B.4) and (B.6) to get that

$$
\begin{aligned}
\mathrm{Var}(\hat{\theta}_{i,h}^t) &\leq \frac{2M_{\nabla g_1}^2 V_{d_i}^t}{2n\rho_n Th} + \frac{2M_{\nabla g_4}^2 V_i^t}{2n^2\kappa_{c_i^t}\rho_n Th} + o\left(\frac{1}{n\rho_n Th}\right) + o\left(\frac{1}{n^2\kappa_{c_i^t}\rho_n Th}\right) \\
&= \frac{C_4}{2n\rho_n Th} + o\left(\frac{1}{n\rho_n Th}\right),
\end{aligned}
$$

where C_4 is a constant related to $\theta_i^t, W_{c_i^t c_j^t}^t$ for all $i,j \in \{1,\ldots,n\}$. Note that the last equality holds due to $n \leq n^2\kappa_{c_i^t}$ for any $\kappa_{c_i^t}$. We now turn back to the distance $d(\hat{\boldsymbol{\theta}}_h^t, \boldsymbol{\theta}^t)$, which by the bias and variance of $\hat{\theta}_{i,h}^t$, satisfies

$$
\begin{aligned}
d(\hat{\boldsymbol{\theta}}_h^t, \boldsymbol{\theta}^t) &= \mathbb{E}\left(\frac{1}{n}\|\hat{\boldsymbol{\theta}}_h^t - \boldsymbol{\theta}^t\|^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n (\hat{\theta}_{i,h}^t - \theta_i^t)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\hat{\theta}_{i,h}^t - \mathbb{E}(\hat{\theta}_{i,h}^t) + \mathbb{E}(\hat{\theta}_{i,h}^t) - \theta_i^t\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^n \left[\mathbb{E}\left(\hat{\theta}_{i,h}^t - \mathbb{E}(\hat{\theta}_{i,h}^t)\right)^2 + \left(\mathbb{E}(\hat{\theta}_{i,h}^t) - \theta_i^t\right)^2\right] \\
&\le C_3 h^4 + \frac{C_4}{2n\rho_n Th}.
\end{aligned}
$$

$\square$

Proof of Lemma 1. For any generic matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, let $\|\boldsymbol{A}\|_\infty = \max_{i,j=1,\ldots,n}|A_{ij}|$ 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

$$
\begin{aligned}
\left|Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(H(\boldsymbol{e}^t)\right)\right| &\le L\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty \\
&\le L\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty + L\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t)\right\|_\infty
\end{aligned} \tag{B.7}
$$

where L are the constants occurring in the Lipschitz continuity of $Q(\cdot)$. To show (17), we need to establish uniform convergences of both l_∞ norms over $\boldsymbol{e}^t \in \{1,\ldots,K\}^n$.

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 $\hat{\tau}_{ij,h}(t/T)$ and $\hat{O}_{kl,h}(\boldsymbol{e}^t)$ and letting $S_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^n \sum_{t'\in U(t/T)} A_{ij}^{t'}\mathbf{1}_{\{e_i^t=k, e_j^t=l\}}$, 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_{ij}^t's are independent given the $\boldsymbol{\theta}^t$ and $\boldsymbol{c}^t$ for all time points and $|A_{ij}^t| \le 1$. For any fixed $\varepsilon > 0$, letting $B = 2$, $w = 2n^2\rho_n Th\varepsilon$ and $\alpha = \frac{3\varepsilon^2}{3C_{\sup}+2\varepsilon}$, then we have that

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

as $\mathrm{Var}(S_{kl}(\boldsymbol{e}^t)) = \sum_{i,j=1}^n \sum_{t'\in U(t/T)} \mathrm{Var}(A_{ij}^{t'})\mathbf{1}_{\{e_i^t=k, e_j^t=l\}} \le 2C_{\sup}n^2\rho_n Th$ for a constant $C_{\sup} = \left(\sup_{i,t}\theta_i^t\right)^2\left(\sup_{i,j,t} W_{c_i^t c_j^t}^t\right)$. 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 $\varepsilon > 0$,

$$
\mathbb{P}\left(\max_{\boldsymbol{e}^t}\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty \ge \varepsilon\right) \le 2K^{n+2}\exp\left\{-\alpha n^2\rho_n Th\right\} = 2K^2\left[K\exp\left\{-\alpha n\rho_n Th\right\}\right]^n \to 0, \tag{B.9}
$$

where the last convergence follows when $n\rho_n Th \to \infty$. This means that the first term of (B.7) converges to 0 in probability if $n\rho_n Th \to \infty$. To inspect the second term of (B.7), by equality A.6 we have that

$$\begin{aligned}
&\left\| \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - H(\boldsymbol{e}^t) \right\|_\infty \le \left\| \mathbb{E}\left(\frac{\hat{O}_{kl,h}(\boldsymbol{e}^t)}{n^2\rho_n} \right) - H_{kl}(\boldsymbol{e}^t) \right\|_1 \\
&\le \sum_{k,l=1}^{K} \left[\left| C_{ij}^t \right| \tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + O\left(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3\right) \right] = \left| C_{ij}^t \right| h^2 + O(h^3) \to 0, \qquad \text{(B.10)}
\end{aligned}$$

as $h \to 0$. Combining (B.9) and (B.10) implies that the left-hand side of (B.7) converges to 0 in probability if $\rho_n, h \to 0$ and $nT \to \infty$ such that $n\rho_n Th \to \infty$. $\square$

Proof of Lemma 2. Recalling the definitions of $H_{kl}(\boldsymbol{c}^t)$ and $H_k(\boldsymbol{c}^t)$, we have that

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

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 $\boldsymbol{e}^t \in \{1, \ldots, K\}^n$,

$$\begin{aligned}
Q\left(H(\boldsymbol{e}^t)\right) &= \sum_{k,l=1}^{K} \left[\frac{\sum_{a,b=1}^K \Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)W_{ab}^t}{n^2} \times \log G_{kl}^t \right. \\
&\qquad \left. - \frac{\sum_{a,s=1}^K n_s(\boldsymbol{c}^t)W_{as}^t \Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n^2} \times \frac{\sum_{b,s'=1}^K n_{s'}(\boldsymbol{c}^t)W_{bs'}^t \Lambda_{b(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n^2} \times G_{kl}^t \right] \\
&= \frac{1}{n^2}\sum_{k,l=1}^{K}\sum_{a,b=1}^{K} \Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) \left[W_{ab}^t \log G_{kl}^t - \frac{W_{ab}^t}{B_{ab}^t} \times G_{kl}^t \right] \\
&\le \frac{1}{n^2}\sum_{k,l=1}^{K}\sum_{a,b=1}^{K} \Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)\Lambda_{b(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) \left[W_{ab}^t \log B_{ab}^t - W_{ab}^t \right] \\
&= \sum_{a,b=1}^{K} \left\{ \frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2} \log\left[\frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2} \times \left(\frac{n_a(\boldsymbol{c}^t)\sum_{s=1}^K n_s(\boldsymbol{c}^t)W_{as}^t}{n^2} \right)^{-1} \right. \right. \\
&\qquad \left. \left. \times \left(\frac{n_b(\boldsymbol{c}^t)\sum_{s'=1}^K n_{s'}(\boldsymbol{c}^t)W_{bs'}^t}{n^2} \right)^{-1} \right] - \frac{n_a(\boldsymbol{c}^t)n_b(\boldsymbol{c}^t)W_{ab}^t}{n^2} \right\} \\
&= \sum_{a,b=1}^{K} \left[H_{ab}(\boldsymbol{c}^t) \log \frac{H_{ab}(\boldsymbol{c}^t)}{H_a(\boldsymbol{c}^t)H_b(\boldsymbol{c}^t)} - H_{ab}(\boldsymbol{c}^t) \right] = Q\left(H(\boldsymbol{c}^t)\right), \qquad \text{(B.11)}
\end{aligned}$$

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 $Q(H(\boldsymbol{e}^t)) = Q(H(\boldsymbol{c}^t))$ 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 $Q(H(\boldsymbol{e}^t))$. 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 $\mathcal{N}_{a(k)}^t$ must be a non-empty set, indicating $n_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t) > 0$. Thus, the equality in (B.11) gives that $G_{kl}^t = B_{ab}^t$ 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 $\boldsymbol{B}^t = (B_{ab}^t)_{K \times K}$ and $\boldsymbol{G}^t = (G_{kl}^t)_{K \times K}$ are constructed by $\boldsymbol{W}^t$ and $H(\boldsymbol{e}^t)$, both matrices $\boldsymbol{B}^t, \boldsymbol{G}^t$ are symmetric. Under Assumption 5, we can further conclude that $\boldsymbol{B}^t$ does not have two identical columns. Finally, combining the above conditions with Assumption 3, 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 $Q\left(H(\boldsymbol{c}^t)\right) - Q\left(H(\boldsymbol{e}^t)\right)$. Let $\|\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)\|_1 = \sum_{a,k=1}^K |\Lambda_{a(k)}(\boldsymbol{c}^t, \boldsymbol{e}^t)|$. Note that by Assumptions (b) and (c) of Zhao et al. (2012), $Q(\cdot)$ 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

$$M' \left\| \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t)}{n} - \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n} \right\|_1 \le Q\left(H(\boldsymbol{c}^t)\right) - Q\left(H(\boldsymbol{e}^t)\right) \le M \left\| \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t)}{n} - \frac{\Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)}{n} \right\|_1 \tag{B.12}$$

since, from equation (16), $H(\boldsymbol{e}^t)$ 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,

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

from which we can conclude that

$$2\theta_{\inf} \sum_{i=1}^n \mathbf{1}_{\{e_i^t \ne c_i^t\}} \le \left\|\Lambda(\boldsymbol{c}^t, \boldsymbol{c}^t) - \Lambda(\boldsymbol{c}^t, \boldsymbol{e}^t)\right\|_1 \le 2\theta_{\sup} \sum_{i=1}^n \mathbf{1}_{\{e_i^t \ne c_i^t\}} \tag{B.13}$$

and after plugging it into (B.12) inequality (19) follows. □

Lemma B.1. *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,*

$$\mathbb{P}\left(\max_{\{\boldsymbol{e}^t : n\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \le m_{\text{mis}}\}} \left\| \frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n} - \frac{\mathbb{E}\left(\hat{O}_h(\boldsymbol{e}^t)\right)}{n^2\rho_n} + \frac{\mathbb{E}\left(\hat{O}_h(\boldsymbol{c}^t)\right)}{n^2\rho_n} \right\|_\infty > \frac{Cm_{\text{mis}}}{n} \right) \to 0. \tag{B.14}$$

Proof. The technique of proving this proof is similar to Lemma A.1 of Zhao et al. (2012). Define P_1 as the probability on the left-hand side of (B.14). Letting $S_{kl}(\boldsymbol{e}^t) = \sum_{i,j=1}^n \sum_{t' \in U(t/T)} A_{ij}^{t'} \mathbf{1}_{\{e_i^t = k, e_j^t = l\}}$, 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 without loss of generality if we let $n\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) = m_{\text{mis}}$ and $e_1^t \ne c_1^t, \ldots, e_{m_{\text{mis}}}^t \ne c_{m_{\text{mis}}}^t, e_{m_{\text{mis}}+1}^t = c_{m_{\text{mis}}+1}^t, \ldots, e_n^t = c_n^t$,

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

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

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

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

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

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

Since we set $n\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) \le m_{\text{mis}}$, there exist $\binom{n}{m_{\text{mis}}}K^{m_{\text{mis}}}$ possible assignments of $\boldsymbol{e}^t$ and K^2 elements in the matrix $\hat{O}_h(\boldsymbol{e}^t)$. Let $\varepsilon = (Cm_{\text{mis}})/n$. Using the union bound, if $C \le 6C_{\text{sup}}$,

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

where the last convergence follows if $(n\rho_n Th)/\log n \to \infty$. Otherwise, if $C > 6C_{\text{sup}}$,

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

as $(n\rho_n Th)/\log n \to \infty$. Consequently, the convergence (B.14) holds for any constant C if $(n\rho_n Th)/\log n \to \infty$. □

Proof of Theorem 2. **Step 1:** we first show the weak consistency. Let $\iota(\boldsymbol{e}^t, \boldsymbol{c}^t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{e_i^t \ne c_i^t\}}$. Equation (14) states that $\hat{\boldsymbol{c}}^t$ is the maximizer of $Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right)$. Thus, in order to show the weak consistency, we only need to show that there exists $\delta_n^t \to 0$ such that

$$\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\ge\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) > Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right)\right) \to 0. \tag{B.15}$$

By the continuity and the uniqueness of $Q(H(\boldsymbol{e}^t))$ in Lemma 2, 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 $\rho_n, h \to 0, n\rho_n Th \to \infty$ and where the last convergence holds by equation (17) in Lemma 1.

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.15) we analyze the behaviour 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

$$
\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:1\leq n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}} Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) - Q\left(\frac{\hat{O}_h(\boldsymbol{c}^t)}{n^2\rho_n}\right) > 0\right) \to 0. \tag{B.16}
$$

Since $Q(\cdot)$ is a continuous function, we can establish from (A.6) that $Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)$ converges to $Q\left(H(\boldsymbol{e}^t)\right)$ for any $\boldsymbol{e}^t \in \{1,\dots,K\}^n$ as $h \to 0$ and $n\rho_n Th \to \infty$. Also note that by Lemma 2 $Q\left(H(\boldsymbol{e}^t)\right) \leq Q\left(H(\boldsymbol{c}^t)\right)$, 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 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

$$
\begin{aligned}
&Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\\
&= -\mathrm{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)\right)^\top \mathrm{vec}\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) + o\left(\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_1\right)\\
&\geq M_1\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_1 + o\left(\left\|\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right\|_1\right)
\end{aligned} \tag{B.17}
$$

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 the 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.6) the vanishing bias between $\frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n}$ and $H_{kl}(\boldsymbol{e}^t)$ is given by

$$C_{ij}^t\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^2 + O(\tilde{\kappa}_k(\boldsymbol{e}^t)\tilde{\kappa}_l(\boldsymbol{e}^t)h^3)$$

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

$$\begin{aligned}
&\frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{e}^t))}{n^2\rho_n} - H_{kl}(\boldsymbol{e}^t) - \frac{\mathbb{E}(\hat{O}_{kl,h}(\boldsymbol{c}^t))}{n^2\rho_n} + H_{kl}(\boldsymbol{c}^t)\\
&= \frac{h^2}{12n^2\rho_n}\sum_{i=1}^{m_{\text{mis}}}\tau_{ii}^{''}(t/T)C^*\left(\mathbf{1}_{\{e_i^t=k,e_i^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_i^t=l\}}\right) + \frac{O(h^3)}{n^2}\sum_{i=1}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k,e_i^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_i^t=l\}}\right)\\
&\quad + \frac{h^2}{6n^2\rho_n}\sum_{i<j}^{m_{\text{mis}}}\tau_{ij}^{''}(t/T)C^*\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right) + \frac{O(h^3)}{n^2}\sum_{i<j}^{m_{\text{mis}}}\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\quad + \frac{h^2}{6n^2\rho_n}\sum_{i=1}^{m_{\text{mis}}}\sum_{j=m_{\text{mis}}+1}^{n}\tau_{ij}^{''}(t/T)C^*\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&\quad + \frac{O(h^3)}{n^2}\sum_{i=1}^{m_{\text{mis}}}\sum_{j=m_{\text{mis}}+1}^{n}\tau_{ij}^{''}(t/T)C^*\left(\mathbf{1}_{\{e_i^t=k,e_j^t=l\}} - \mathbf{1}_{\{c_i^t=k,c_j^t=l\}}\right)\\
&= o\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right),
\end{aligned}$$

since $h \to 0$ and $\tau_{ij}^{''}(t/T)$ is proportional to ρ_n for any pair of nodes (i,j) as stated in Assumption 1. As such, $\|R_1\|_1 = K^2\|R_1\|_\infty = o\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right)$ given that K is fixed. Considering R_2, as, from equation (16), $H(\boldsymbol{e}^t)$ 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.13). By the (reverse) triangle inequality we have that $\left|\|R_2\|_1 - \|R_1\|_1\right| \le \|R_1 - R_2\|_1 \le \|R_1\|_1 + \|R_2\|_1$ implying that $\|R_1 - R_2\|_1 \sim \iota(\boldsymbol{e}^t,\boldsymbol{c}^t)$. Therefore, plugging this quantity into (B.17) we conclude that there exists a positive constant M'' such that

$$Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right) - Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) \ge M''\iota(\boldsymbol{e}^t,\boldsymbol{c}^t) \tag{B.18}$$

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 (1.1) in Bickel et al. (2015). For any generic vector $\boldsymbol{e}^t \in \{1,\dots,K\}^n$, we use a Taylor expansion again 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

$$Q\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n}\right) = Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) + \operatorname{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\right)^\top \operatorname{vec}\left(\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right) + O_p\left(\left\|\frac{\hat{O}_h(\boldsymbol{e}^t)}{n^2\rho_n} - \frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right\|_\infty^2\right) \tag{B.19}$$

where $\operatorname{vec}\left(\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)\right)$ represents the vector of the 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

$$\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)=\nabla Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)+O\left(\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\right). \tag{B.20}$$

Employing the Taylor series in (B.19) 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.20), 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)+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)\\
&\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)+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+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\\
&\quad+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

$$\begin{aligned}
&\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:1\le n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}}\sum_{m=1}^4 C_mI_m+Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{e}^t))}{n^2\rho_n}\right)-Q\left(\frac{\mathbb{E}(\hat{O}_h(\boldsymbol{c}^t))}{n^2\rho_n}\right)>0\right)\\
&\le\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:1\le n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)<n\delta_n^t\}}\sum_{m=1}^4 C_mI_m-M''\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)>0\right)\\
&\le\sum_{m=1}^3\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}}I_m>\frac{M''m_{\text{mis}}}{4C_mn}\right)+\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(I_4>\frac{M''m_{\text{mis}}}{4C_4n}\right),
\end{aligned}$$

where the first inequality holds when plugging in the bound from (B.18). To show (B.16), it is enough to show that $\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}}I_m>\frac{M''m_{\text{mis}}}{4C_mn}\right)$ for any $m=1,2,3$, and $\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(I_4>\frac{M''m_{\text{mis}}}{4C_4n}\right)$ converge to 0.

For I_1: Let $C=\frac{M''}{4C_1}$ and $(n\rho_nTh)/\log n\to\infty$, then Lemma B.1 yields that

$$\begin{aligned}
\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)=m_{\text{mis}}\}}I_1>\frac{M''m_{\text{mis}}}{4C_1n}\right)&\le\sum_{m_{\text{mis}}=1}^n\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}}I_1>\frac{M''m_{\text{mis}}}{4C_1n}\right)\\
&\le n\max_{m_{\text{mis}}=1,\ldots,n}\mathbb{P}\left(\max_{\{\boldsymbol{e}^t:n\iota(\boldsymbol{e}^t,\boldsymbol{c}^t)\le m_{\text{mis}}\}}I_1>\frac{M''m_{\text{mis}}}{4C_1n}\right)\to 0
\end{aligned}$$

due to the exponential decay.

For I_2: Let $\varepsilon = \frac{M''}{4C_2}$ and apply the bound in (B.9), 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 $\alpha = \frac{3\varepsilon^2}{3C_{\sup}+2\varepsilon}$. Then applying the union bound, we conclude from inequality (B.8) that

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

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

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

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.8) implies that

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

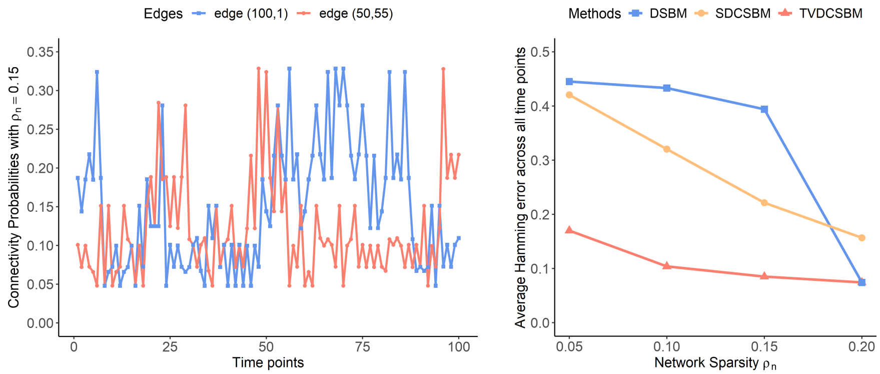
the convergence $\sum_{m_{\text{mis}}=1}^{n} \mathbb{P}\left(I_4 > \frac{M'' m_{\text{mis}}}{4C_4 n} \right) \to 0$ follows due to the exponential decay as $(n\rho_n Th)/\log n \to \infty$. □

C Additional simulation results

In this section, we investigate the robustness of our proposed procedure to the smoothness assumption for edge probabilities $\tau_{ij}(t/T)$. We consider a different data-generating process where we give up the smooth functions imposed on the community connectivity parameter $\boldsymbol{Z}^t$ and the degree parameter $\boldsymbol{\theta}^t$. More specifically, the parameters $\boldsymbol{Z}^t$ and $\boldsymbol{\theta}^t$ are constrained to jump between two configurations where

$$\begin{aligned}
\boldsymbol{Z}^t &= \rho_n \begin{pmatrix} 1.3 & 0.7 \\ 0.7 & 1.3 \end{pmatrix} \mathbf{1}_{\{t \text{ is odd number}\}} + \rho_n \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.3 \end{pmatrix} \mathbf{1}_{\{t \text{ is even number}\}}, \\
\theta_0^t &= 1.2 \times \mathbf{1}_{\{t \text{ is odd number}\}} + 0.8 \times \mathbf{1}_{\{t \text{ is even number}\}}.
\end{aligned}$$

To ensure the identifiability of $\boldsymbol{\theta}^t$, half of the nodes for each community are randomly chosen to have at a given time point t the same value θ_0^t for the degree parameters. The other half of the nodes is set accordingly to satisfy the constraint in Assumption 4 at each time point. We then use the same method described in Section 5 to generate the time-varying community structures. Lastly, we generate $T = 100$ networks with $n = 100$ nodes and $K = 2$ disjoint communities according to the specified parameters. In Figure 6 (a), we plot the true curves of the varying generative probabilities of two edges across time, obviously showing significant fluctuations.



(a) (b)

Figure 6: The left panel (a) illustrates the true values of edge probabilities over time: the blue squares depict the edge probability between nodes 100 and 1 (i.e., $\tau_{100,1}(t/T)$) whereas the red circles denote the edge probability between nodes 50 and 55 (i.e., $\tau_{50,55}(t/T)$). The right panel (b) presents the average Hamming error rates for three methods: DSBM (blue squares), SDCSBM (orange circles) and TVDCSBM (red triangles)

The results for all community detection competitors are presented in Figure 6 (b). We observe that our proposed method, TVDCSBM, demonstrates robustness to the smoothness of edge probabilities. Even when the probabilities of edges are not stationary over time, TVDCSBM is still able to estimate the varying community structures overall better than the competing methods. Moreover, the advantage of TVDCSBM becomes particularly pronounced when the network is sparse. As discussed in Remark 3, violating moderately the smoothness assumption, would not severely jeopardise the applicability of our approach in practice.