LEARNING SHARED AND INDIVIDUAL STRUCTURE IN DYNAMIC NETWORKS WITH DEGREE HETEROGENEITY

Mengxue Li, Rainer von Sachs, Eugen Pircalabelu

LIDAM Discussion Paper ISBA
2026 / 12

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

Learning shared and individual structure in dynamic networks with degree heterogeneity

Mengxue Li, Rainer von Sachs and Eugen Pircalabelu

UCLouvain, Institute of Statistics, Biostatistics and Actuarial Sciences

Abstract

Dynamic networks provide a powerful framework for analysing connectivity data in many fields such as social science, finance or neuroscience (e.g. along neuroimaging studies). In practice, such networks are typically collected from multiple subjects across time and exhibit both temporal dynamics and subject-specific heterogeneity. Ample connectivity networks also contain hub nodes, defined as highly connected nodes that play critical roles. In this work, we propose a mixed-effects dynamic stochastic block model with degree heterogeneity, which simultaneously disentangles the shared connectivity structure from individual variability and recovers the trajectories of hub nodes through time-varying degree parameters. We develop an efficient local approximate estimation procedure, based on a time-localised likelihood approach, provide statistical guarantees such as consistency and asymptotic normality and evaluate its performance through extensive simulations and a case study from the Human Connectome Project focusing on dynamic functional brain connectivity.

Keywords: Replicated time-varying networks; Degree-corrected stochastic block model; Mixed-effects model; Local likelihood estimation; EM algorithm.

1 Introduction

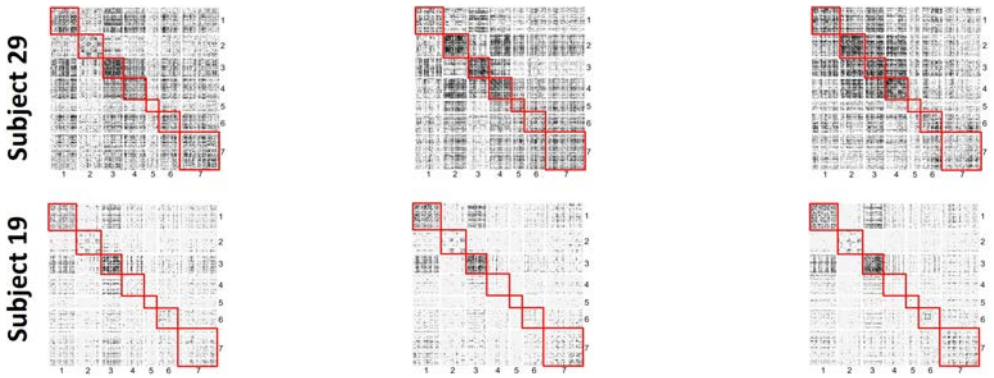
In many neuroimaging studies, the brain is typically partitioned into a set of regions of interest (ROIs), and brain networks are constructed based on the correlation structure among the signals recorded from these regions, where nodes represent ROIs and edges represent functional connections between any pairs of ROIs. Brain parcellation studies, such as Yeo et al. (2011), among many others, have shown that ROIs are often partitioned into functional modules (i.e., *communities*), whose coordinated interactions support complex cognitive processes. At the same time, nodes within communities are usually not equivalent, reflecting *degree heterogeneity* where certain nodes (then called "hubs") display a much higher number of connections than others. Since brain activity is inherently dynamic, evolving in response to external stimuli and cognitive states, both the community connectivity and the degree heterogeneity are likely to change over time. Time-varying networks therefore provide a natural and powerful framework for characterizing such dynamics, and a variety of dynamic network models have been developed, including time-varying stochastic block models (Matias and Miele, 2017; Lei and Lin, 2023), degree-corrected stochastic block models (Li et al., 2026; Agterberg et al., 2025), β-models (Du et al., 2023), latent space models (Zhang et al., 2020; He et al., 2025), and exponential random graph models (Caimo and Gollini, 2020).

In practice, however, the data often consist not of a single time-varying network, but of multiple network sequences collected from multiple subjects or sources under a common experimental condition, which are commonly referred to as *replicated time-varying networks*. In brain connectivity studies, for example, subjects exposed to the same movie stimulus may produce dynamic brain network trajectories that are similar but not identical. On the one hand, one can assume a form of baseline, average connectivity pattern that is shared to some extent by the underlying population in response to the same stimulus (Hasson et al., 2008; Finn and Bandettini, 2021). On the other hand, subject deviations from the average model are quite likely to occur, due to the inherent biological variability at the subject level. Similar data structures occur in gene expression, financial, and social networks, where different individuals, groups, or markets generate multiple network sequences that share a certain common structure while still exhibiting substantial individual heterogeneity.

The central challenge in replicated time-varying networks is to recover a shared connectivity structure while separating it from variation at both the node and subject levels. To the best of our knowledge, existing methods provide only partial solutions to this problem. Dynamic network models capture temporal variation and, in some cases, node-specific heterogeneity, but generally do not distinguish shared dynamics from subject-specific variation. If subject heterogeneity is ignored, treating all network sequences as i.i.d. replicates may confound the individual heterogeneity with the sampling variability, whereas fitting a separate dynamic network model to each subject fails to exploit the shared structure and may therefore lead to findings that are less interpretable and less robust. Multi-subject methods instead focus primarily on variation across observational units. For static networks, Kim et al. (2023) introduced a graph-aware linear mixed-effects model with cell-specific subject effects, while MacDonald et al. (2022) decomposed network structure into shared and individual-specific latent components. For dynamic networks, Zhang et al. (2024) developed a semiparametric network regression relating shared connectivity to subject-specific covariates. Zhang et al. (2020) proposed a mixed-effects SBM for subject-specific departures from a shared community structure. The mixed-effects SBM is closest to our setting, but it assumes homogeneous node propensities within each community. This limitation may confound the hub connectivity with community effects. Estimating node-specific connectivity propensities is also scientifically important since the highly connected nodes play a central role in information integration and transmission in complex networks (Van den Heuvel and Sporns, 2013; Cole et al., 2013). These considerations call for a unified framework that simultaneously incorporates time-varying node heterogeneity, shared community connectivity, and subject-specific variation.

In this paper we propose a time-varying mixed-effects degree-corrected stochastic block model for replicated time-varying networks. The proposed model inherits the flexibility of generalised linear mixed-effects models (GLMM) where we explicitly incorporate the network structure by decomposing the fixed effects into (i) node-specific degree parameters, (ii) community connectivity effects, and capture (iii) subject-to-subject variability via the random effects. Moreover, all model parameters are assumed to be smooth functions of time, thereby providing a more realistic and interpretable representation of the real networks. To capture the smoothness of the underlying functions we base ourselves on a time-localised likelihood approach, for which we develop consistency and asymptotic normality of the estimates of the time-varying model parameters. In our proposed computationally efficient local approximation approach we leverage the similarity of networks observed from the same subject at neighbouring time points leading to more stable estimators than a pointwise approach based on single layers of (often rather sparse) networks would give.

The remainder of the paper is organized as follows. Section 2 introduces the proposed time-varying mixed-effects DCSBM. Section 3 develops the time-localised likelihood estimator, establishes



(a) $t = 8$ (b) $t = 48$ (c) $t = 75$

Figure 1: Representative adjacency matrices for 2 subjects at 3 time points during movie watching, with nodes ordered by community membership.

its theoretical properties, and proposes a local approximate EM algorithm for efficient computation. Comprehensive simulation results are presented in Section 4, followed by a dynamic functional brain connectivity application in Section 5. Section 6 concludes with a brief discussion. The additional details and proofs are provided in the Supplementary Material.

Notations. For a positive integer k, let $\mathbf{1}_k$ be the k-dimensional vector of ones, $\boldsymbol{I}_k$ be the $k \times k$ identity matrix, and $\mathbf{0}$ denote a conformable zero vector or zero matrix determined by the context. For any generic vector or matrix $\boldsymbol{X} = (X_{ij})$, we write $\|\boldsymbol{X}\|_\infty = \max_{i,j} |X_{ij}|$ with a vector viewed as a one-column matrix. Let $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the Euclidean and Frobenius norm. We use $\mathrm{vec}(\cdot)$ for column-wise vectorization, $\mathrm{diag}(\cdot)$ for the diagonal matrix formed from a vector or a matrix, $\mathbb{1}(\cdot)$ for the indicator function, and $\rightsquigarrow$ for the convergence in distribution. For two positive sequences x_T and y_T, we write $x_T = o(y_T)$ if $x_T/y_T \to 0$ as $T \to \infty$, and $x_T = o_p(y_T)$ if $x_T/y_T \to 0$ in probability as $T \to \infty$.

2 Mixed-effects degree-corrected stochastic block models

Suppose that M subjects are observed on a common set of n nodes over T time points. We rescale time to the unit interval $(0, 1]$ and write $u^t = t/T, t = 1, \ldots, T$. For each subject $m \in \{1, \ldots, M\}$ at time point $u^t \in \{1/T, \ldots, 1\}$, the network is recorded by an $n \times n$ symmetric binary adjacency matrix $\boldsymbol{A}_m(u^t) = (A_{m,ij}(u^t))_{1 \le i,j \le n}$ where $A_{m,ij}(u^t) = A_{m,ji}(u^t) = 1$ if there exists a connection between nodes i and j at time point u^t, and $A_{m,ij}(u^t) = A_{m,ji}(u^t) = 0$ otherwise. We model the edges $A_{m,ij}(u^t)$ as independent Bernoulli random variables with

$$\mathbb{E}\left(A_{m,ij}(u^t)\right) = \tau_{m,ij}(u^t),$$

and we refer to $\boldsymbol{\tau}_m(u^t) = (\tau_{m,ij}(u^t))_{1 \le i,j \le n}$ as the $(n \times n)$ connectivity matrix.

In many applications, dynamic networks observed across subjects under the same experimental condition often share substantial similarity in their latent connectivity, while simultaneously exhibiting subject-specific variability. To capture this structure, we decompose the edge probability $\tau_{m,ij}(u^t)$ into (i) a shared term $\Theta_{ij}(u^t)$ capturing the common propensity for an edge to be formed

between nodes i and j, and (ii) a subject-specific component $\omega_m(u^t)$ accounting for the individual variability,

$$\text{logit}(\tau_{m,ij}(u^t)) = \Theta_{ij}(u^t) + \omega_m(u^t), \qquad (\omega_m(u^t))_{m=1,\dots,M} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2(u^t)),$$

where $\text{logit}(x) = \log\left(\frac{x}{1-x}\right)$ for $x \in (0,1)$. Although other monotone invertible link functions, such as the probit link, could be used, we adopt the logit link for its simplicity in modelling connectivity probabilities. Notice that $\omega_m(u^t)$ captures an overall subject-specific shift in connectivity. It is assumed to follow a Gaussian distribution with zero mean and variance $\sigma^2(u^t)$ quantifying the magnitude of variability across subjects over time. This parsimonious specification strikes a reasonable tradeoff between model flexibility and estimability: it accounts for the subject-specific deviations while retaining sufficient pooling across subjects to estimate the shared structure in sparse replicated networks. Figure 1 visually illustrates this observation. We model the shared component $\Theta_{ij}(u^t)$ by a DCSBM structure. The n nodes are partitioned into G latent communities, with $c_i \in \{1, \dots, G\}$ denoting the community membership of node i. Throughout, we assume that the memberships are common across subjects and time points. Let $\boldsymbol{Z}(u^t) = (Z_{gl}(u^t))_{1\le g,l\le G}$ be a $(G \times G)$ symmetric matrix of the community connectivity parameters, where $Z_{gl}(u^t)$ quantifies the propensity for an edge to be formed between community g and l. In addition, each node i is associated with a time-varying degree parameter $\theta_i(u^t)$ to accommodate degree heterogeneity. Thus, we model $\Theta_{ij}(u^t) = \theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t)$ for any pair of nodes (i, j). Under this specification, our model can be expressed in matrix form as

$$\begin{aligned} \boldsymbol{A}_m(u^t) \mid \omega_m(u^t) &\overset{\text{ind}}{\sim} \text{Bernoulli}(\boldsymbol{\tau}_m(u^t)), \\ \text{logit}(\boldsymbol{\tau}_m(u^t)) &= \boldsymbol{\Theta}(u^t) + \omega_m(u^t)\mathbf{1}_n\mathbf{1}_n^\top \quad \text{with} \\ \boldsymbol{\Theta}(u^t) &= \boldsymbol{\theta}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}(u^t)\boldsymbol{C}^\top,\ (\omega_m(u^t))_{m=1,\dots,M} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2(u^t)), \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}(u^t) = (\theta_1(u^t), \dots, \theta_n(u^t))^\top$ and $\boldsymbol{C} \in \{0,1\}^{n\times G}$ is the community membership matrix such that $C_{ig} = 1$ if $c_i = g$ and $C_{ig} = 0$ otherwise. Here, the $\text{logit}(\cdot)$ function is applied elementwise.

Our model can be viewed as a generalised extension of the DCSBM. In particular, when $\omega_m(u^t) = 0$ for all u^t and m, it reduces to a time-varying β-model (see, e.g., Stein et al., 2025) with community covariates. The proposed model extends the mixed-effects SBM of Zhang et al. (2020) by incorporating degree parameters to further accommodate node-level connectivity propensities, which are often substantial in empirical networks (Karrer and Newman, 2011; Agterberg et al., 2025; Li et al., 2026). In brain functional connectivity networks, for example, the excess connectivity of hub regions may be absorbed into the community parameters, thereby confounding the interpretation of community-level connectivity. In contrast, the time-varying degree parameters $\boldsymbol{\theta}(u^t)$ in model (1) enable the proposed model to disentangle degree heterogeneity from the community structure. Moreover, $\theta_i(u^t)$ itself provides an interpretable measure of the node-specific connectivity propensity of node i over time, and may be of independent interest in many studies (e.g., for identifying hub nodes in disease- or task-related analyses; see Van den Heuvel and Sporns, 2013; Cole et al., 2013).

Model (1) suffers from a non-identifiability issue due to the shift invariance of its additive part. A naive solution is to fix an anchor node by setting $\theta_n(u^t) = 0$ for all $u^t \in \{1/T, \dots, 1\}$, as in Du et al. (2023), which, however, makes the interpretation of $\theta_i(u^t)$ depend on an arbitrary choice of the anchor node. Another option is to enforce $\min_{1\le i\le n} \theta_i(u^t) = 0$ for any given time point u^t as in Stein et al. (2025); this constraint is often accompanied by sparsity assumptions on $\boldsymbol{\theta}(u^t)$. Following

the classical identifiability for the DCSBM (Karrer and Newman, 2011), we instead center the degree parameters within each community, i.e. $\sum_{i=1}^{n} \theta_i(u^t)\mathbb{1}(c_i = g) = 0, \forall g = 1, \ldots, G$. Proposition 1 establishes the resulting identifiability.

Proposition 1 (Identifiability)**.** *For any time point $u^t \in \{1/T, \ldots, 1\}$ and given the community membership matrix $\boldsymbol{C}$ with no empty columns, suppose that there exist two sets of parameters $(\boldsymbol{\theta}(u^t), \boldsymbol{Z}(u^t))$ and $(\boldsymbol{\theta}^*(u^t), \boldsymbol{Z}^*(u^t))$ satisfying the conditions $\boldsymbol{C}^\top\boldsymbol{\theta}(u^t) = \mathbf{0}$ and $\boldsymbol{C}^\top\boldsymbol{\theta}^*(u^t) = \mathbf{0}$, then*

$$\boldsymbol{\theta}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}(u^t)\boldsymbol{C}^\top = \boldsymbol{\theta}^*(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}^*(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}^*(u^t)\boldsymbol{C}^\top$$

implies that $\boldsymbol{\theta}(u^t) = \boldsymbol{\theta}^*(u^t)$ and $\boldsymbol{Z}(u^t) = \boldsymbol{Z}^*(u^t)$.

3 Estimation

We focus on the case with known community memberships, which arises in many applications where prior module information is available; for example, brain regions are often assigned to established functional systems (Yeo et al., 2011; Zhang et al., 2020). In this setting, we develop a time-localised likelihood estimator, establish its theoretical properties, and propose a local approximate EM algorithm for computation. For the situation of unknown memberships, in the Appendix we propose a pre-processing step via a residual joint spectral embedding method based on an auxiliary likelihood that can be used prior to the developments of the setting treated here.

3.1 Local likelihood estimation

For each time point u^t, define $\boldsymbol{\xi}(u^t) = \left(\boldsymbol{\theta}(u^t)^\top, \mathrm{vec}(\boldsymbol{Z}(u^t))^\top, \sigma^2(u^t)\right)^\top$ as the parameter vector, where $\boldsymbol{\theta}(u^t) \in \mathbb{R}^n$, $\boldsymbol{Z}(u^t) \in \mathbb{R}^{G\times G}$, and $\sigma^2(u^t) \in \mathbb{R}_+$. Due to the identifiability constraint on the degree parameters, for a given community membership matrix $\boldsymbol{C}$, we define the constrained parameter space as

$$\Xi_{\boldsymbol{C}} = \left\{ \left(\boldsymbol{\theta}^\top, \mathrm{vec}(\boldsymbol{Z})^\top, \sigma^2\right)^\top : \boldsymbol{\theta} \in \mathbb{R}^n, \boldsymbol{C}^\top\boldsymbol{\theta} = \mathbf{0}, \boldsymbol{Z} \in \mathbb{R}^{G\times G}, \boldsymbol{Z}^\top = \boldsymbol{Z}, \sigma^2 \in \mathbb{R}_+ \right\}.$$

Let $\boldsymbol{A}(u^t) = \{\boldsymbol{A}_m(u^t)\}_{1\le m\le M}$ denote the corresponding collection of observed adjacency matrices for the M subjects. The contribution to the marginal log-likelihood of subject m, obtained by integrating out the random effects in model (1), is

$$\begin{aligned}\ell_m(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^t)) = \log \int_{\mathbb{R}} \prod_{1\le i<j\le n} & \frac{\exp\left(A_{m,ij}(u^t) \times (\Theta_{ij}(u^t) + \omega_m(u^t))\right)}{1 + \exp\left(\Theta_{ij}(u^t) + \omega_m(u^t)\right)} \\ & \times \phi(\omega_m(u^t); 0, \sigma^2(u^t))\, \mathrm{d}\,\omega_m(u^t)\end{aligned}$$

where $\phi(\cdot; 0, \sigma^2(u^t))$ denotes the normal density with mean zero and variance $\sigma^2(u^t)$. The marginal log-likelihood is therefore $L(\boldsymbol{\xi}(u^t); \boldsymbol{A}(u^t)) = \sum_{m=1}^{M} \ell_m(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^t))$, and the pointwise maximum likelihood estimator is obtained as

$$\hat{\boldsymbol{\xi}}(u^t) = \underset{\boldsymbol{\xi}(u^t)\in\Xi_{\boldsymbol{C}}}{\arg\max}\ L(\boldsymbol{\xi}(u^t); \boldsymbol{A}(u^t)).$$

Although model (1) can be formulated as a GLMM, it involves a coupled structure in which each dyad (i, j) depends jointly on $\theta_i(u^t)$ and $\theta_j(u^t)$. Such structures are well known to pose substantial computational challenges even in classical dyadic data models (see, e.g., Graham 2017), since the dimension of the fixed effects depends on n, the number of nodes. As a result, standard GLMM implementations (e.g., `glmer()` in the `lme4` package in `R`, or others alike) are not directly applicable to our model. More importantly, as is typical with real data, the observed graphs are sparse, making the pointwise maximum likelihood estimator unstable because the effective information at a single time point is limited relative to the dimension of $\boldsymbol{\Theta}(u^t)$. In principle, one may reduce this instability by increasing the number of subjects M; in practice, however, achieving a sufficiently large M is difficult in some domains, such as rare-disease studies or applications with costly data acquisition, limiting the practical applicability of this approach.

The above considerations motivate a local likelihood approach (Fan et al., 1998) that leverages the similarity of networks observed from the same subject at neighboring time points to reduce estimation variance, thereby leading to a more stable and accurate procedure for estimation. To this end, let $\boldsymbol{A} = \{\boldsymbol{A}_m(u^t)\}_{\substack{1\le m\le M\\ 1\le t\le T}}$ be the collection of all observed networks. We construct the local (kernel-weighted) log-likelihood at a given time point u^t, denoted by $\tilde{L}_h(\boldsymbol{\xi}(u^t); \boldsymbol{A})$, as

$$\tilde{L}_h(\boldsymbol{\xi}(u^t); \boldsymbol{A}) = \sum_{s=1}^{T} L(\boldsymbol{\xi}(u^t); \boldsymbol{A}(u^s))\, K_h(u^s - u^t)$$

where $h \in (0, 1]$ is a bandwidth parameter and $K_h(u) = h^{-1}K(u/h)$ with a kernel function $K(\cdot)$ that is even, supported on $[-1, 1]$, and satisfies $\int_{-1}^{1} K(u)\, du = 1$. The corresponding smooth estimators at time point u^t are obtained by solving the following optimisation programme

$$\tilde{\boldsymbol{\xi}}_h(u^t) = \underset{\boldsymbol{\xi}(u^t)\in \Xi_{\boldsymbol{C}}}{\arg\max}\ \tilde{L}_h(\boldsymbol{\xi}(u^t); \boldsymbol{A}). \tag{2}$$

We next study the asymptotic behaviour of the smooth estimators $\tilde{\boldsymbol{\xi}}_h(u)$ in (2). Under the identifiability constraints in Proposition 1, $\tilde{\boldsymbol{\xi}}_h(u)$ is a constrained local M-estimator. The following analysis extends the constrained M-estimation theory (Geyer, 1994; Moore et al., 2008) to a local likelihood setting. Let $\dim(\cdot)$ denote the dimension of a vector or a matrix. Since the constraints in $\Xi_{\boldsymbol{C}}$ are linear except for the interior condition $\sigma^2 > 0$, the tangent space at an interior point is

$$\Xi_{\boldsymbol{C}}^{\tan} = \left\{ (\boldsymbol{v}_\theta^\top, \mathrm{vec}(\boldsymbol{v}_Z)^\top, v_\sigma)^\top \in \mathbb{R}^{\dim(\boldsymbol{\xi}(u))} : \boldsymbol{C}^\top \boldsymbol{v}_\theta = \boldsymbol{0},\ \boldsymbol{v}_Z^\top = \boldsymbol{v}_Z \right\}.$$

Because every community is nonempty, $\boldsymbol{C}$ has full column rank G. Therefore, $\dim(\Xi_{\boldsymbol{C}}^{\tan}) = n - G + \frac{G(G+1)}{2} + 1$. Let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{\dim(\boldsymbol{\xi}(u))\times \dim(\Xi_{\boldsymbol{C}}^{\tan})}$ be any matrix with orthonormal columns spanning $\Xi_{\boldsymbol{C}}^{\tan}$, that is, $\boldsymbol{\mathcal{P}}^\top \boldsymbol{\mathcal{P}} = \boldsymbol{I}_{\dim(\Xi_{\boldsymbol{C}}^{\tan})}$ and $\mathrm{col}(\boldsymbol{\mathcal{P}}) = \Xi_{\boldsymbol{C}}^{\tan}$. For each subject m, let

$$S_{\ell,m}(\boldsymbol{\xi}(u^t)) = \frac{\partial \ell_m(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m)}{\partial \boldsymbol{\xi}(u^t)}, \quad H_{\ell,m}(\boldsymbol{\xi}(u^t)) = \frac{\partial^2 \ell_m(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m)}{\partial \boldsymbol{\xi}(u^t) \partial \boldsymbol{\xi}(u^t)^\top},$$

denote the score vector and Hessian matrix of the marginal log-likelihood contribution, and thus the relevant quantities of their projections onto $\Xi_{\boldsymbol{C}}^{\tan}$ are $\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t))$ and $\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\xi}(u^t))\boldsymbol{\mathcal{P}}$, respectively. For the purpose of theoretical analysis, we need the following assumptions to hold uniformly for $u \in (0, 1]$.

Assumption 1. (i) Assume that n and G are fixed. The membership matrix $\boldsymbol{C} \in \mathbb{R}^{n\times G}$ is known and has no empty column. The asymptotic regime is $MT \to \infty$ with bandwidth $h \to 0$ satisfying $MTh \to \infty$.

(ii) There exist a large $a > 0$ and $0 < \underline{\sigma}^2 \leq \overline{\sigma}^2 < \infty$ such that true parameter $\boldsymbol{\xi}(u)$ belongs to a compact subset

$$\left\{ \left(\boldsymbol{\theta}^\top, \mathrm{vec}(\boldsymbol{Z})^\top, \sigma^2 \right)^\top \in \Xi_{\boldsymbol{C}} : \|\boldsymbol{\theta}\|_\infty \leq a,\ \|\boldsymbol{Z}\|_\infty \leq a,\ \underline{\sigma}^2 \leq \sigma^2 \leq \overline{\sigma}^2 \right\}.$$

(iii) For each component $\xi_k(u)$ of $\boldsymbol{\xi}(u)$, $\xi_k(\cdot) \in C^2[0,1], \forall k = 1, \ldots, \dim(\boldsymbol{\xi}(u))$.

(iv) The kernel function $K(u)$ is bounded, symmetric, supported on $[-1,1]$, and satisfies $\int K(x)\,dx = 1$, $\int x^2 K(x)\,dx < \infty$ and $\int K^2(x)\,dx < \infty$.

(v) For each m, the marginal log-likelihood contribution $\ell_m(\boldsymbol{\xi}(u); \boldsymbol{A}_m(u))$ is three times continuously differentiable with respect to $\boldsymbol{\xi}(u)$. Its second derivatives are uniformly bounded over the parameter space, and its third derivatives exist.

(vi) There exist positive definite matrices $\boldsymbol{J}(u)$ and $\boldsymbol{\Upsilon}(u)$ such that

$$\lim_{M\to\infty} -M^{-1} \sum_{m=1}^{M} \boldsymbol{\mathcal{P}}^\top \mathbb{E}\left(H_{\ell,m}(\boldsymbol{\xi}(u)) \right) \boldsymbol{\mathcal{P}} = \boldsymbol{J}(u),$$

$$\lim_{M\to\infty} M^{-1} \sum_{m=1}^{M} \mathrm{Var}\left(\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u)) \right) = \boldsymbol{\Upsilon}(u).$$

Assumption 1 (i) fixes the network size and the number of communities. Whereas the second statement of this assumption is standard in network statistics, we note that treating in this theoretical work networks of fixed size n is complementary to the body (Li, von Sachs, and Pircalabelu (2026), Zhao et al. (2012)) of existing theory for estimation of network *structures* under the paradigm of asymptotically changing node sparsity. We, however, address a different problem, namely that of parameter estimation in a multi-subject setting. The nonempty column ensures that $\boldsymbol{C}$ has full column rank, so that the constraint $\boldsymbol{C}^\top \boldsymbol{\theta} = \boldsymbol{0}$ imposes G independent restrictions. Assumption 1 (ii) is a common boundedness condition for M-estimation under network models, as in Zhang et al. (2020); Li et al. (2023); He et al. (2025). It excludes degenerate regions of the parameter space and keeps the derivatives of the likelihood under control. Assumption 1 (iii)–(iv) are the usual conditions for nonparametric regression. In particular, the smoothness condition permits a second-order expansion around the target time point u, with leading smoothing bias of order h^2. This requirement is analogous to the smoothness assumptions used in Fan et al. (1998), Li et al. (2026). From a modeling perspective, it formalizes the gradual temporal evolution of the model parameters under a continuous experimental stimulus, which differs fundamentally from the change-point framework designed for abrupt structural breaks. Assumption 1 (v) gives the differentiability conditions needed for a Taylor expansion of the constrained likelihood score. The bounded second derivatives control the local curvature, while the third derivative condition makes the Taylor remainder negligible. Finally, Assumption 1 (vi) requires the projected Hessian and score covariance to have positive definite limits, ensuring nondegenerate curvature and score variation along the admissible tangent directions.

Theorem 1 (Consistency)**.** *Suppose that Assumption 1 holds. For any fixed $u \in (0,1]$, consider a sequence of grid points $u^t = t/T$, $t = 1, \ldots, T$ such that $u^t \to u$ as $T \to \infty$, then the smooth estimator $\tilde{\boldsymbol{\xi}}_h(u^t)$ satisfies $\tilde{\boldsymbol{\xi}}_h(u^t) \xrightarrow{p} \boldsymbol{\xi}(u)$. Moreover, at any given time point u^t, it holds that $\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u)\|_2 = O_p(h^2 + (MTh)^{-1/2})$.*

Theorem 1 establishes the pointwise convergence rate of the proposed smooth estimator. The term h^2 is the smoothing bias, while the second is the stochastic error associated with the effective local sample size MTh. The result shows the efficiency gain from local smoothing, as the stochastic error is reduced from the pointwise order $M^{-1/2}$, obtained by fitting a GLMM at a single time point, to $(MTh)^{-1/2}$. The resulting bias–variance tradeoff is analogous to that of kernel-smoothed estimators studied for the time-varying β-model (Du et al., 2023), the DCSBM model of Li, von Sachs, and Pircalabelu (2026), and the network vector autoregressive model (Li et al., 2026). Balancing the squared bias with the variance, the optimal bandwidth is therefore of order $(MT)^{-1/5}$, and hence the pointwise convergence rate for $\tilde{\boldsymbol{\xi}}_h(u^t)$ is of order $(MT)^{-2/5}$, provided that $Th \to \infty$. In applications, since the true curve, however, is unknown, the bandwidth is selected by maximizing the leave-one-out cross-validated likelihood criterion described in Section S4 of the Supplementary Material.

Theorem 2 (Asymptotic normality)**.** *Suppose that Assumption 1 holds. Further assume that the bandwidth satisfies $MTh^5 = O(1)$. For any fixed $u \in (0,1]$, consider a sequence of grid points $u^t = t/T$, $t = 1, \ldots, T$ such that $u^t \to u$ as $T \to \infty$, then,*

$$\sqrt{MTh}\left[\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t) - \frac{\nu_{21}}{2}h^2\boldsymbol{\mathcal{P}}\boldsymbol{J}^{-1}(u^t)\boldsymbol{a}(u^t)\right] \rightsquigarrow \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\xi(u)),$$

where $\boldsymbol{a}(u^t)$ is a deterministic vector determined by the second derivative of the true curve $\boldsymbol{\xi}(u)$ around u^t, $\nu_{xy} = \int u^x K^y(u)\,\mathrm{d}\,u$ and $\boldsymbol{\Sigma}_\xi(u) = \nu_{02}\boldsymbol{\mathcal{P}}\boldsymbol{J}^{-1}(u)\boldsymbol{\Upsilon}(u)\boldsymbol{J}^{-1}(u)\boldsymbol{\mathcal{P}}^\top$.

Theorem 2 provides a second-order expansion for the smooth estimator $\tilde{\boldsymbol{\xi}}_h(u^t)$, after projecting onto the tangent space to eliminate the non-identifiable directions. We now give several comments on this result. First, the bias term in Theorem 2 is of order $O(\sqrt{MTh^5})$. If an undersmoothing condition, $\sqrt{MTh}\,h^2 \to 0$, is imposed, then the bias vanishes asymptotically. Otherwise, a bias correction is required for constructing confidence intervals and hypothesis tests. The covariance matrix in Theorem 2 has the sandwich form for a local M-estimator under constraints. Here $\boldsymbol{J}(u)$ represents the limiting negative curvature of the local marginal likelihood, whereas $\boldsymbol{\Upsilon}(u)$ is the limiting covariance of the projected score contribution.

Note that Theorems 1 and 2 hold for either fixed or diverging M, as long as $h \to 0$, $T \to \infty$ such that $Th \to \infty$ for Theorem 1, with the additional condition $Th^5 = O(1)$ for Theorem 2. In the fixed-M case, the matrices $\boldsymbol{J}(u)$ and $\boldsymbol{\Upsilon}(u)$ are understood as their fixed-M counterparts, $\boldsymbol{J}_M(u) = -M^{-1}\sum_{m=1}^M \boldsymbol{\mathcal{P}}^\top \mathbb{E}\left(H_{\ell,m}(\boldsymbol{\xi}(u); \boldsymbol{A}(u))\right)\boldsymbol{\mathcal{P}}$, and $\boldsymbol{\Upsilon}_M(u) = M^{-1}\sum_{m=1}^M \mathrm{Var}\left(\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u); \boldsymbol{A}(u))\right)$, and accordingly, the conditions on positive definiteness in Assumption 1(vi) are imposed on $\boldsymbol{J}_M(u)$ and $\boldsymbol{\Upsilon}_M(u)$.

Remark 1. The tangent space formulation in Theorem 2 is algebraically equivalent to a Lagrange multiplier adjustment of the local likelihood expansion. Consider first the degree component under the constraint $\boldsymbol{C}^\top\boldsymbol{\theta}(u) = \mathbf{0}$. Let generically $S(\boldsymbol{\theta}(u))$ and $H(\boldsymbol{\theta}(u))$ be the score vector and Hessian matrix in a local quadratic expansion of a given likelihood function. Any feasible perturbation $\boldsymbol{\delta}_\theta$ satisfies $\boldsymbol{C}^\top\boldsymbol{\delta}_\theta = \mathbf{0}$ due to the imposed constraint. Since $\boldsymbol{C}$ has full column rank, if $\boldsymbol{\mathcal{P}}_\theta$ is an

orthonormal basis of the null space $\ker(\boldsymbol{C}^\top) = \{\boldsymbol{v} : \boldsymbol{C}^\top \boldsymbol{v} = \boldsymbol{0}\}$, then every feasible perturbation has the representation $\boldsymbol{\delta}_\theta = \boldsymbol{\mathcal{P}}_\theta \boldsymbol{v}_\theta$. Solving the projected first-order equation for the quadratic approximation gives

$$\boldsymbol{\delta}_\theta = -\boldsymbol{\mathcal{P}}_\theta (\boldsymbol{\mathcal{P}}_\theta^\top H(\boldsymbol{\theta}(u)) \boldsymbol{\mathcal{P}}_\theta)^{-1} \boldsymbol{\mathcal{P}}_\theta^\top S(\boldsymbol{\theta}(u)),$$

provided that the projected Hessian is nonsingular. This is equivalent to the KKT equations $H(\boldsymbol{\theta}(u))\boldsymbol{\delta}_\theta + S(\boldsymbol{\theta}(u)) + \boldsymbol{C}\boldsymbol{\lambda} = \boldsymbol{0}$, and $\boldsymbol{C}^\top \boldsymbol{\delta}_\theta = \boldsymbol{0}$. Premultiplication by $\boldsymbol{\mathcal{P}}_\theta^\top$ eliminates the multiplier term. Conversely, if the projected equation holds, then $H(\boldsymbol{\theta}(u))\boldsymbol{\delta}_\theta + S(\boldsymbol{\theta}(u))$ is orthogonal to $\ker(\boldsymbol{C}^\top)$, and hence belongs to $\mathrm{col}(\boldsymbol{C})$, so that a Lagrange multiplier exists. Thus, the tangent space formulation in Theorem 2 is algebraically equivalent to a Lagrange multiplier adjustment.

Theorem 2 applies the same argument to the full parameter vector. The matrix $\boldsymbol{\mathcal{P}}$ is an orthonormal basis of $\Xi_{\boldsymbol{C}}^{\tan}$, and the projected score and Hessian are precisely of the form $\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t))$ and $\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\xi}(u^t))\boldsymbol{\mathcal{P}}$. Thus the Taylor expansion of the marginal log-likelihood is performed only along feasible tangent directions. This is the geometric principle underlying constrained likelihood theory and tangent space projection in (semi)-parametric inference; see, for example, Geyer (1994), Bickel et al. (1993), and Moore et al. (2008). Closely related, Li et al. (2023) use a Lagrange-adjusted Hessian construction to enforce the identifiability restrictions. This equivalence motivates the constrained one-step update in Section 3.2.

Theorem 2 establishes joint asymptotic normality for the full parameter vector at a fixed time point. As discussed above, node-level degree heterogeneity is itself scientifically meaningful in many applications; however, its time-varying trajectory is rarely isolated as a formal asymptotic target in the literature. We therefore state below the limiting distribution for an individual degree parameter. For any fixed node $i \in \{1, \ldots, n\}$, let $\boldsymbol{e}_i$ be the fixed selection vector such that $\theta_i(u) = \boldsymbol{e}_i^\top \boldsymbol{\xi}(u)$. Define the bias-corrected version of $\tilde{\theta}_{i,h}(u)$ by $\tilde{\theta}_{i,h}^{\mathrm{bc}}(u^t) = \tilde{\theta}_{i,h}(u^t) - \frac{\nu_{21}}{2} h^2 \boldsymbol{e}_i^\top \boldsymbol{\mathcal{P}} \boldsymbol{J}^{-1}(u^t) \boldsymbol{a}(u^t)$.

Corollary 1. *Suppose that the conditions of Theorem 2 hold. For any fixed node $i \in \{1, \ldots, n\}$ and any sequence of grid points $u^t = t/T$ such that $u^t \to u \in (0, 1]$,*

$$\sqrt{MTh}\left[\tilde{\theta}_{i,h}^{\mathrm{bc}}(u^t) - \theta_i(u^t)\right] \rightsquigarrow \mathcal{N}\left(0, \Sigma_{\theta,i}(u)\right),$$

where $\Sigma_{\theta,i}(u) = \nu_{02} \boldsymbol{e}_i^\top \boldsymbol{\mathcal{P}} \boldsymbol{J}^{-1}(u) \boldsymbol{\Upsilon}(u) \boldsymbol{J}^{-1}(u) \boldsymbol{\mathcal{P}}^\top \boldsymbol{e}_i$.

Analogous individual limits for other components of $\boldsymbol{\xi}(u)$ follow immediately by replacing $\boldsymbol{e}_i$ with the corresponding selection vector.

3.2 Local approximate EM

Programme (2) can, in principle, be optimized by integrating out the random effects and maximizing the resulting local marginal log-likelihood. This strategy, however, is computationally demanding in our framework since the degree parameters $\boldsymbol{\theta}(u^t)$ constitute a crossed fixed-effect term whose dimension grows with n, and therefore leads to a high-dimensional, coupled optimisation problem under the identifiability constraints in Proposition 1. Instead, we treat $\omega_m(u^t)$ as missing data and proceed within an EM-type framework (Dempster et al., 1977) to optimize (2). A key obstacle is that, under the logit link in (1), the posterior distribution of $\omega_m(u^t)$ is nonconjugate. Hence, the evaluation of conditional moments in the E step is analytically intractable and might be unstable, particularly for sparse networks. We therefore propose a local approximate EM that adopts two complementary approximations to address these challenges.

For each target time point u^t, the complete-data log-likelihood contribution from subject m is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^t), \omega_m(u^t)) = &\sum_{1\le i<j\le n} A_{m,ij}(u^t)\left[\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)+\omega_m(u^t)\right] \\ &- \sum_{1\le i<j\le n} \log\left[1+\exp(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)+\omega_m(u^t))\right] \\ &- \frac{\omega_m^2(u^t)}{2\sigma^2(u^t)} - \frac{\log(\sigma^2(u^t))}{2} - \frac{\log(2\pi)}{2}. \end{aligned} \tag{3}$$

Under Assumption 1 (iii), we adopt a local polynomial approximation to $\boldsymbol{\xi}(u)$ in a neighborhood of u^t, paralleling the construction of the local marginal criterion. This leads to the local complete-data log-likelihood,

$$\tilde{\mathcal{L}}_h(\boldsymbol{\xi}(u^t); \boldsymbol{A}, \boldsymbol{\omega}) = \sum_{s=1}^{T}\sum_{m=1}^{M} \mathcal{L}(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^s), \omega_m(u^s))\, K_h(u^s-u^t) \tag{4}$$

With the above local criterion, our LA-EM algorithm can be formulated as follows:

LA-E step. Let $\tilde{\boldsymbol{\xi}}_h^{(r)} = \{\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t)\}_{1\le t\le T}$ be the parameter estimates from the (r)-th iteration, where we write $\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t) = (\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t),\, \tilde{\boldsymbol{Z}}_h^{(r)}(u^t),\, \tilde{\sigma}_h^{2,(r)}(u^t))^\top$. The local Q-function at time point u^t is then defined as,

$$\begin{aligned} Q_h(\boldsymbol{\xi}(u^t)|\tilde{\boldsymbol{\xi}}_h^{(r)}) &= \mathbb{E}\left(\tilde{\mathcal{L}}_h(\boldsymbol{\xi}(u^t); \boldsymbol{A}, \boldsymbol{\omega})\big|\boldsymbol{A}, \tilde{\boldsymbol{\xi}}_h^{(r)}\right) \\ &= \sum_{s=1}^{T}\sum_{m=1}^{M} K_h(u^s-u^t) \int_{\mathbb{R}} \mathcal{L}(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^s), \omega_m(u^s)) \\ &\quad \times f_{\tilde{\boldsymbol{\xi}}_h^{(r)}}(\omega_m(u^s) \mid \boldsymbol{A}_m(u^s))\, \mathrm{d}\,\omega_m(u^s), \end{aligned} \tag{5}$$

where $f_{\tilde{\boldsymbol{\xi}}_h^{(r)}}(\omega_m(u^s) \mid \boldsymbol{A}_m(u^s))$ denotes the posterior density of $\omega_m(u^s)$ under the current estimates $\tilde{\boldsymbol{\xi}}_h^{(r)}$. Note that this posterior does not have a closed-form expression due to the logit link. We therefore consider a Laplace approximation in this work. For any generic $u^t \in \{1/T, \ldots, 1\}$, define the posterior mode and local curvature by

$$\hat{\omega}_m(u^t) = \arg\max_{\omega_m(u^t)\in\mathbb{R}} \mathcal{L}\big(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t); \boldsymbol{A}_m(u^t), \omega_m(u^t)\big), \tag{6}$$

$$\hat{v}_m(u^t) = -\frac{\partial^2}{\partial\omega_m^2(u^t)} \mathcal{L}\big(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t); \boldsymbol{A}_m(u^t), \omega_m(u^t)\big)\Big|_{\omega_m(u^t)=\hat{\omega}_m(u^t)}. \tag{7}$$

For fixed $\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t)$, the objective in (6) is strictly concave with respect to $\omega_m(u^t)$. Hence the posterior mode $\hat{\omega}_m(u^t)$ is unique, and it can be computed by Newton iterations, which typically converge fast in practice. A second-order Taylor expansion of $\mathcal{L}(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t); \boldsymbol{A}_m(u^t), \omega_m(u^t))$ around $\hat{\omega}_m(u^t)$ yields

$$\begin{aligned}\mathcal{L}\big(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t); \boldsymbol{A}_m(u^t), \omega_m(u^t)\big) \approx &\mathcal{L}\big(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^t); \boldsymbol{A}_m(u^t), \hat{\omega}_m(u^t)\big) \\ &- \frac{1}{2}\hat{v}_m^{(r)}(u^t)\big(\omega_m(u^t)-\hat{\omega}_m(u^t)\big)^2. \end{aligned}$$

Consequently, the marginal probability for subject m can be rewritten as

$$\begin{aligned}\mathbb{P}(\boldsymbol{A}_m(u^s)) &= \int_{\mathbb{R}} \exp\left(\mathcal{L}(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^s); \boldsymbol{A}_m(u^s), \omega_m(u^s))\right) \mathrm{d}\,\omega_m(u^s) \\ &\approx \sqrt{\frac{2\pi}{\hat{v}_m(u^s)}} \exp\left(\mathcal{L}(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^s); \boldsymbol{A}_m(u^s), \hat{\omega}_m(u^s))\right) ,\end{aligned}$$

which leads to the Laplace approximation of the posterior as,

$$\begin{aligned}f_{\tilde{\boldsymbol{\xi}}_h^{(r)}}(\omega_m(u^s) \mid \boldsymbol{A}_m(u^s)) &= \frac{\exp(\mathcal{L}(\tilde{\boldsymbol{\xi}}_h^{(r)}(u^s); \boldsymbol{A}_m(u^s), \omega_m(u^s)))}{\mathbb{P}(\boldsymbol{A}_m(u^s))} \\ &\approx \phi\left(\omega_m(u^s); \hat{\omega}_m(u^s), \hat{v}_m^{-1}(u^s)\right).\end{aligned} \tag{8}$$

Substituting this approximation into (5), and dropping constants, the local Q-function can be approximated by $Q_{h,1} + Q_{h,2} + Q_{h,3}$ where

$$Q_{h,1} = \sum_{s=1}^{T} K_h(u^s - u^t) \left(\sum_{i=1}^{n} d_i(u^s)\, \theta_i(u^t) + \frac{1}{2} \sum_{g,l=1}^{G} O_{gl}(u^s)\, Z_{gl}(u^t) \right) \tag{9}$$

$$\begin{aligned}Q_{h,2} = &- \sum_{m=1}^{M} \sum_{1 \le i < j \le n} \sum_{s=1}^{T} K_h(u^s - u^t) \int_{\mathbb{R}} \log\left[1 + \exp\left(\Theta_{ij}(u^t) + \omega_m(u^s)\right)\right] \\ &\times \phi\left(\omega_m(u^s); \hat{\omega}_m(u^s),\, \hat{v}_m^{-1}(u^s)\right) \mathrm{d}\,\omega_m(u^s),\end{aligned} \tag{10}$$

$$Q_{h,3} = -\frac{1}{2} \sum_{s=1}^{T} K_h(u^s - u^t) \left[M \log(\sigma^2(u^t)) + \frac{\sum_{m=1}^{M} \hat{\omega}_m^2(u^s) + \hat{v}_m^{-1}(u^s)}{\sigma^2(u^t)} \right], \tag{11}$$

where $d_i(u^t) = \sum_{m=1}^{M} \sum_{j=1}^{n} A_{m,ij}(u^t)$ and $O_{gl}(u^t) = \sum_{m=1}^{M} \sum_{i,j=1}^{n} A_{m,ij}(u^t) \mathbb{1}(c_i = g, c_j = l)$.

LA-M step. (1) Update of the variance. Maximizing $Q_{h,3}$ with respect to $\sigma^2(u^t)$ gives the closed-form update at the $(r+1)$-th iteration as

$$\tilde{\sigma}_h^{2,(r+1)}(u^t) = \frac{\sum_{m=1}^{M} \sum_{s=1}^{T} K_h(u^s - u^t) \left(\hat{\omega}_m^2(u^s) + \hat{v}_m^{-1}(u^s)\right)}{M \sum_{s=1}^{T} K_h(u^s - u^t)} , \tag{12}$$

recalling that $\hat{\omega}_m^2(u^s)$ and $\hat{v}_m(u^s)$ have been calculated in the r-th iteration.

(2) Update of the fixed effects. Since the maximization of $Q_{h,1} + Q_{h,2}$ is subject to identifiability constraints, and has no closed-form solution, we propose a constrained one-step update based on the local Q-function (Fan and Chen, 1999). Fix $\boldsymbol{Z}(u^t)$ at its current estimate $\tilde{\boldsymbol{Z}}_h^{(r)}(u^t)$ and consider $Q_{h,1} + Q_{h,2}$ as a function of $\boldsymbol{\theta}(u^t)$ only. For notational convenience, define $Q_h(\boldsymbol{\theta}(u^t)) = Q_{h,1} + Q_{h,2}$ and let $S_{Q,h}(\boldsymbol{\theta}(u^t))$ and $H_{Q,h}(\boldsymbol{\theta}(u^t))$ denote the gradient vector and second derivative matrix of $Q_h(\boldsymbol{\theta}(u^t))$ with respect to $\boldsymbol{\theta}(u^t)$; see the Supplementary Material for their explicit expressions. We then apply a Taylor expansion of $Q_h(\boldsymbol{\theta}(u^t))$ around $\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)$,

$$\begin{aligned}Q_h(\boldsymbol{\theta}(u^t)) =& Q_h(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)) + S_{Q,h}(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t))^\top \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t) \\ &+ \frac{1}{2} \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t)^\top H_{Q,h}(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)) \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t) + O(\|\boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t)\|_2^3),\end{aligned}$$

where $\boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t) = \boldsymbol{\theta}(u^t) - \tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)$. To enforce the identifiability constraint, we next introduce the Lagrange multipliers $\boldsymbol{\lambda}(u^t) \in \mathbb{R}^G$ and set first-order conditions, which yields the KKT system

$$\begin{pmatrix} H_{Q,h}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) & \boldsymbol{C} \\ \boldsymbol{C}^\top & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t) \\ \boldsymbol{\lambda}(u^t) \end{pmatrix} = - \begin{pmatrix} S_{Q,h}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) \\ \mathbf{0} \end{pmatrix},$$

where on the right-hand side expression, the second block is zero since $\boldsymbol{C}^\top \tilde{\boldsymbol{\theta}}_h^{(r)}(u^t) = \mathbf{0}$. Solving the KKT system gives

$$\begin{aligned} \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t) = &- H_{Q,h}^{-1}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) S_{Q,h}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) + H_{Q,h}^{-1}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) \\ &\times \boldsymbol{C}\Big(\boldsymbol{C}^\top H_{Q,h}^{-1}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big)\boldsymbol{C}\Big)^{-1} \boldsymbol{C}^\top H_{Q,h}^{-1}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big) S_{Q,h}\big(\tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)\big). \end{aligned}$$

Thus, the constrained one-step estimator of $\boldsymbol{\theta}(u^t)$ is given by $\tilde{\boldsymbol{\theta}}_h^{(r+1)}(u^t) = \tilde{\boldsymbol{\theta}}_h^{(r)}(u^t) + \boldsymbol{\delta}_{\boldsymbol{\theta}}(u^t)$. Analogously, fixing $\boldsymbol{\theta}(u^t) = \tilde{\boldsymbol{\theta}}_h^{(r+1)}(u^t)$, $\tilde{\boldsymbol{Z}}_h^{(r+1)}(u^t) = \tilde{\boldsymbol{Z}}_h^{(r)}(u^t) + \boldsymbol{\delta}_{\boldsymbol{Z}}(u^t)$, where $\boldsymbol{\delta}_{\boldsymbol{Z}}(u^t) = -H_{Q,h}^{-1}(\tilde{\boldsymbol{Z}}_h^{(r)}(u^t)) S_{Q,h}(\tilde{\boldsymbol{Z}}_h^{(r)}(u^t))$. Here $S_{Q,h}\big(\boldsymbol{Z}(u^t)\big)$ and $H_{Q,h}\big(\boldsymbol{Z}(u^t)\big)$ are the gradient vector and Hessian matrix of $Q_{h,1} + Q_{h,2}$ with respect to $\boldsymbol{Z}(u^t)$, respectively.

The initialization procedure and full LA-EM procedure are given in Section S5 of the Supplementary Material.

4 Simulation studies

This section shows the finite-sample performance of the proposed procedure. As discussed previously, in many applications, the number of subjects M is often limited by participant availability and data acquisition costs. For instance, the application in Section 5 analyzes 34 young subjects from a publicly available longitudinal dataset. In such settings, one cannot simply rely on having a large number of replicates to obtain stable estimates. We therefore set $M = 20$ and $n = 50$, and study the asymptotic behaviour in Theorems 1 and 2 with varying $T \in \{15, 30, 60, 120\}$. We use throughout the Epanechnikov kernel and select the bandwidth by the leave-one-out cross-validated likelihood criterion from the Supplementary Material.

4.1 Estimation error

Data-generating process. We generate undirected binary networks from model (1) for $M = 20$ subjects, with $n = 50$ nodes observed over T time points. The nodes are assigned independently to three communities with probabilities $(0.2, 0.4, 0.4)^\top$.

- Scenario 1 (Baseline with smooth trajectories). For the degree parameters, nodes within each community are split into two groups of nearly equal size, with degree parameters given by $f_{\boldsymbol{\theta}}(u^t) = 0.8 + 0.1\sin(2\pi u^t)$ for the first group and $-f_{\boldsymbol{\theta}}(u^t)$ for the second group. The community connectivity parameters are set to $Z_{gl}(u^t) = -1.2 + 0.2\sin(2\pi u^t - 1)$ if $g = l$, and $Z_{gl}(u^t) = -2 + 0.2\sin(2\pi u^t + 1)$ if $g \neq l$. The variance curve is given by $\sigma^2(u^t) = (\sigma_0 + 0.05\sin(2\pi u^t))^2$ with $\sigma_0 = 0.5$.
- Scenario 2 (No degree heterogeneity). We keep $\boldsymbol{Z}(u^t)$ and $\sigma^2(u^t)$ the same as in Scenario 1, but set the degree parameters to zero, i.e., $\theta_i(u^t) = 0$, $\forall\, i = 1, \ldots, n,\ \ t = 1, \ldots, T$. This reduces the model to a mixed-effects SBM with homogeneous nodes.

- Scenario 3 (High variability across subjects). We keep $\boldsymbol{\theta}(u^t)$ and $\boldsymbol{Z}(u^t)$ the same as in Scenario 1, while increasing σ_0 to 1. This setting therefore presents a substantially larger dispersion across subjects over time than in Scenario 1.

- Scenario 4 (Piecewise constant trajectories). The variance function is as in Scenario 1, whereas $\boldsymbol{\theta}(u^t)$ and $\boldsymbol{Z}(u^t)$ are assumed to be piecewise constant with abrupt changes. We sample five change points $2 \leq \eta_1 < \cdots < \eta_5 \leq T$ and set $\eta_0 = 1$, $\eta_6 = T + 1$. Setting $Z_{gl}^{(0)} = -1.2$ for $g = l$ and $Z_{gl}^{(0)} = -2$ for $g \neq l$, we define,

$$\boldsymbol{Z}(u^t) = \boldsymbol{Z}_0 + \Delta_{\kappa,1}^{(Z)} \mathbf{1}_G + \Delta_{\kappa,2}^{(Z)} (\mathbf{1}_G \mathbf{1}_G^\top - \mathbf{1}_G) \qquad \Delta_{\kappa,1}^{(Z)}, \Delta_{\kappa,2}^{(Z)} \overset{\text{i.i.d.}}{\sim} \text{Unif}(-0.3, 0.3),$$

for $t = \eta_\kappa, \ldots, \eta_{\kappa+1} - 1$ with $\kappa = 1, \ldots, 5$. The degree trajectory is generated analogously, with $f_{\boldsymbol{\theta}}^{(0)} = 0.8$ and independent jumps $\Delta_\kappa^{(\boldsymbol{\theta})} \overset{\text{i.i.d.}}{\sim} \text{Unif}(-0.3, 0.3)$. We then center $\boldsymbol{\theta}(u^t)$ within communities as in Scenario 1 to enforce the identifiability constraint. This scenario deliberately violates the smoothness assumption and evaluates the robustness of our method under model misspecification.

Methods. We compare our proposed method, denoted by ME-DCSBM$_{\text{LA-EM}}$, with three benchmarks. The pointwise method ME-DCSBM$_{\text{PW}}$ fits model (1) separately at each time point by a standard EM algorithm, and therefore does not exploit temporal smoothness. The second competitor is the time-varying mixed-effects SBM proposed by Zhang et al. (2020), referred to as ME-SBM, which accounts for the subject-specific variability but does not include the degree heterogeneity. The final competing approach we consider is a time-varying β-model with community structure, denoted by βBM, which extends the framework of Stein et al. (2025). It assumes that networks $\{\boldsymbol{A}_m(u^t)\}_{1 \leq m \leq M}$ observed from multiple subjects are i.i.d. replicates,

$$(\boldsymbol{A}_m(u^t))_{m=1,\ldots,M} \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(\boldsymbol{\tau}(u^t)), \quad \forall m \in \{1, \ldots, M\},$$

with $\text{logit}(\boldsymbol{\tau}(u^t)) = \boldsymbol{\theta}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n \boldsymbol{\theta}(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}(u^t)\boldsymbol{C}^\top$. The same identifiability constraints as for our procedure are imposed on the βBM. Here, the membership matrix $\boldsymbol{C}$ defines the design matrix of covariates, and the community connectivity matrix $\boldsymbol{Z}(u^t)$ plays the role of the matrix of regression coefficients. Estimation of the above βBM is carried out through a Newton-Raphson algorithm as described in the Supplementary Material.

Let $\tilde{\boldsymbol{\theta}}(u^t), \tilde{\boldsymbol{Z}}(u^t)$ and $\tilde{\sigma}^2(u^t)$ generally denote the estimators obtained from different methods. The estimation errors here are evaluated by the averaged squared errors across time. That is, $\text{Err}_{\boldsymbol{\theta}} = \frac{1}{nT} \sum_{t=1}^T \|\tilde{\boldsymbol{\theta}}(u^t) - \boldsymbol{\theta}(u^t)\|_2^2$, $\text{Err}_{\boldsymbol{Z}} = \frac{1}{G^2 T} \sum_{t=1}^T \|\tilde{\boldsymbol{Z}}(u^t) - \boldsymbol{Z}(u^t)\|_F^2$ and $\text{Err}_{\sigma^2} = \frac{1}{T} \sum_{t=1}^T \left(\tilde{\sigma}^2(u^t) - \sigma^2(u^t)\right)^2$. All figures in this section present the mean results over 50 replications.

Results. Tables 1-3 report estimation errors (in units of 10^{-3}) for all parameters. Across all settings, we observe that the accuracy of both our method and the ME-SBM typically increases with the growing number of time points T, providing finite-sample evidence supporting the consistency result in Theorem 1. This is expected since both methods incorporate temporal aggregation, which increases the local effective sample size and improves recovery of trajectories. In contrast, the pointwise method could not benefit from the larger T since its effective sample size remains essentially fixed across time (primarily relying on M and n). In the presence of degree heterogeneity in Scenarios 1, 3 and 4, the proposed method yields smaller errors for both $\boldsymbol{\theta}(u^t)$ and $\boldsymbol{Z}(u^t)$ than ME-SBM. This observation provides empirical support for the argument that, when degree heterogeneity is present but ignored, node-specific degree propensities may confound the estimation of the

Table 1: Estimation errors ($\mathrm{Err}_{\boldsymbol{\theta}}$, 10^{-3}) with standard deviations in parentheses.

Scenario	Method	$T = 15$	$T = 30$	$T = 60$	$T = 120$
1	ME-DCSBM$_{\text{LA-EM}}$	**2.09 (0.26)**	**1.21 (0.12)**	**0.70 (0.06)**	**0.41 (0.03)**
	ME-DCSBM$_{\text{PW}}$	8.43 (0.51)	8.39 (0.30)	8.47 (0.24)	8.47 (0.18)
	ME-SBM	–	–	–	–
	βBM	8.73 (0.49)	8.74 (0.32)	8.82 (0.24)	8.84 (0.17)
2	ME-DCSBM$_{\text{LA-EM}}$	**1.45 (0.24)**	**0.86 (0.10)**	**0.48 (0.06)**	**0.28 (0.02)**
	ME-DCSBM$_{\text{PW}}$	7.45 (0.45)	7.48 (0.31)	7.53 (0.24)	7.52 (0.15)
	ME-SBM	–	–	–	–
	βBM	6.99 (0.42)	7.02 (0.29)	7.07 (0.23)	7.06 (0.14)
3	ME-DCSBM$_{\text{LA-EM}}$	**2.08 (0.22)**	**1.16 (0.12)**	**0.70 (0.06)**	**0.41 (0.03)**
	ME-DCSBM$_{\text{PW}}$	8.34 (0.61)	8.35 (0.33)	8.45 (0.25)	8.41 (0.21)
	ME-SBM	–	–	–	–
	βBM	17.70 (1.87)	17.85 (1.37)	17.92 (1.01)	18.10 (0.62)
4	ME-DCSBM$_{\text{LA-EM}}$	**4.22 (0.26)**	**3.00 (0.17)**	**2.04 (0.07)**	**1.67 (0.04)**
	ME-DCSBM$_{\text{PW}}$	7.99 (0.50)	7.97 (0.28)	8.39 (0.24)	8.15 (0.15)
	ME-SBM	–	–	–	–
	βBM	8.20 (0.47)	7.98 (0.28)	8.76 (0.26)	8.47 (0.16)

Table 2: Estimation errors ($\mathrm{Err}_{\boldsymbol{Z}}$, 10^{-3}) with standard deviations in parentheses.

Scenario	Method	$T = 15$	$T = 30$	$T = 60$	$T = 120$
1	ME-DCSBM$_{\text{LA-EM}}$	**4.36 (1.91)**	**2.36 (1.13)**	**1.38 (0.68)**	**0.79 (0.32)**
	ME-DCSBM$_{\text{PW}}$	15.97 (4.35)	15.81 (3.34)	15.44 (2.33)	15.27 (1.58)
	ME-SBM	120.53 (6.18)	115.55 (10.91)	113.18 (7.02)	113.38 (5.16)
	βBM	20.45 (6.18)	19.86 (4.53)	19.20 (2.73)	19.45 (1.93)
2	ME-DCSBM$_{\text{LA-EM}}$	**6.15 (2.27)**	**4.38 (1.73)**	**3.55 (1.12)**	**2.93 (0.69)**
	ME-DCSBM$_{\text{PW}}$	15.86 (4.36)	15.60 (3.31)	15.35 (2.15)	15.19 (1.52)
	ME-SBM	14.72 (3.18)	10.67 (1.87)	8.99 (0.75)	7.83 (0.57)
	βBM	21.38 (6.56)	20.50 (4.68)	19.85 (2.82)	20.07 (1.94)
3	ME-DCSBM$_{\text{LA-EM}}$	**12.93 (6.76)**	**7.07 (3.49)**	**4.55 (2.40)**	**2.80 (1.12)**
	ME-DCSBM$_{\text{PW}}$	55.55 (16.83)	53.49 (13.14)	52.76 (8.43)	52.51 (5.88)
	ME-SBM	119.98 (30.66)	111.45 (20.38)	106.49 (13.42)	106.01 (9.48)
	βBM	109.22 (29.58)	104.20 (20.48)	102.55 (13.74)	97.75 (4.98)
4	ME-DCSBM$_{\text{LA-EM}}$	**8.13 (2.32)**	**5.59 (1.09)**	**3.99 (0.93)**	**3.41 (0.41)**
	ME-DCSBM$_{\text{PW}}$	15.92 (4.19)	15.41 (3.44)	15.26 (2.25)	15.15 (1.58)
	ME-SBM	95.30 (15.40)	64.95 (8.84)	116.18 (8.05)	98.47 (5.26)
	βBM	20.18 (6.12)	19.62 (4.79)	19.22 (2.81)	19.30 (1.94)

community connectivity parameters. Scenario 3 yields the largest estimation errors for all methods. The degradation is, however, most pronounced for the βBM, especially for $\boldsymbol{Z}(u^t)$, whose errors are markedly larger than those of the competing methods. This behaviour is reasonable given the fact that the βBM regards the subject-specific networks as i.i.d. replicates and thus misattributes the subject variability to sampling noise, which in turn yields considerable bias and variance in $\tilde{\boldsymbol{Z}}(u^t)$. Furthermore, the pointwise ME-DCSBM method also performs poorly in Scenario 3. Intuitively, when the variability is large, the signal-to-noise ratio for the shared trajectories decreases, thus it is harder to disentangle this effect from individual fluctuations even when such heterogeneity is modelled. Interestingly, the kernel-smoothing step provides one possible way of increasing the local effective information to mitigate this difficulty. Finally, in Scenario 4, the proposed method remains competitive under piecewise-constant trajectories, provided that the change in magnitude is moderate.

Table 3: Estimation errors (Err_{σ^2}, 10^{-3}) with standard deviations in parentheses.

Scenario	Method	$T = 15$	$T = 30$	$T = 60$	$T = 120$
1	ME-DCSBM$_{\text{LA-EM}}$	**1.58 (0.98)**	**0.83 (0.50)**	**0.48 (0.27)**	**0.33 (0.18)**
	ME-DCSBM$_{\text{PW}}$	6.76 (2.40)	6.74 (1.79)	6.66 (1.22)	6.81 (1.03)
	ME-SBM	11.42 (2.08)	11.24 (1.52)	10.98 (1.27)	10.84 (0.91)
	βBM	–	–	–	–
2	ME-DCSBM$_{\text{LA-EM}}$	**1.47 (0.99)**	**0.83 (0.48)**	**0.50 (0.30)**	**0.35 (0.19)**
	ME-DCSBM$_{\text{PW}}$	6.82 (2.53)	6.69 (1.74)	6.67 (1.16)	6.79 (0.97)
	ME-SBM	6.84 (2.54)	6.70 (1.74)	6.68 (1.17)	6.81 (0.98)
	βBM	–	–	–	–
3	ME-DCSBM$_{\text{LA-EM}}$	**22.17 (14.66)**	**11.52 (5.85)**	**7.40 (4.06)**	**4.97 (2.69)**
	ME-DCSBM$_{\text{PW}}$	102.31 (38.06)	101.37 (25.41)	101.48 (17.67)	102.14 (14.89)
	ME-SBM	165.78 (29.31)	164.36 (23.09)	161.29 (18.53)	158.35 (13.85)
	βBM	–	–	–	–
4	ME-DCSBM$_{\text{LA-EM}}$	**1.65 (1.22)**	**0.88 (0.49)**	**0.87 (0.35)**	**0.37 (0.18)**
	ME-DCSBM$_{\text{PW}}$	6.85 (2.63)	6.68 (1.63)	6.71 (1.20)	6.77 (1.03)
	ME-SBM	10.53 (2.08)	8.92 (1.54)	11.41 (1.31)	10.49 (0.96)
	βBM	–	–	–	–

4.2 Asymptotic normality

We empirically evaluate the asymptotic normality of the proposed estimators. For simplicity, we focus on the baseline setting in Scenario 1 with fixed $n = 50, M = 20, G = 3$, while varying $T \in \{60, 120, 240\}$. Inference is performed at the particular time point $u^t = 1/2$. All results in this section are based on 200 replications. In the implementation, we first estimate the asymptotic variance $\boldsymbol{\Upsilon}(u)$ by a kernel-smoothed plug-in estimator over time,

$$\hat{\boldsymbol{\Upsilon}}(u^t) = \frac{1}{MT}\sum_{s=1}^{T} K_h(u^s - u^t) \sum_{m=1}^{M} \boldsymbol{\mathcal{P}}^\top \hat{S}_{\ell,m}(\hat{\boldsymbol{\xi}}(u^t); \boldsymbol{A}_m(u^s))\hat{S}_{\ell,m}^\top(\hat{\boldsymbol{\xi}}(u^t); \boldsymbol{A}_m(u^s))\boldsymbol{\mathcal{P}} \ ,$$

where $\boldsymbol{\mathcal{P}}$ is one orthonormal basis for the identifiable tangent space. Since the restrictions are linear, we compute $\boldsymbol{\mathcal{P}}$ from the QR decomposition of community membership matrix $\boldsymbol{C}$. Similarly, the projected information matrix is estimated as

$$\hat{\boldsymbol{J}}(u^t) = -\frac{1}{MT}\sum_{s=1}^{T} K_h(u^s - u^t) \sum_{m=1}^{M} \boldsymbol{\mathcal{P}}^\top \hat{H}_{\ell,m}(\hat{\boldsymbol{\xi}}(u^t); \boldsymbol{A}_m(u^s))\boldsymbol{\mathcal{P}}.$$

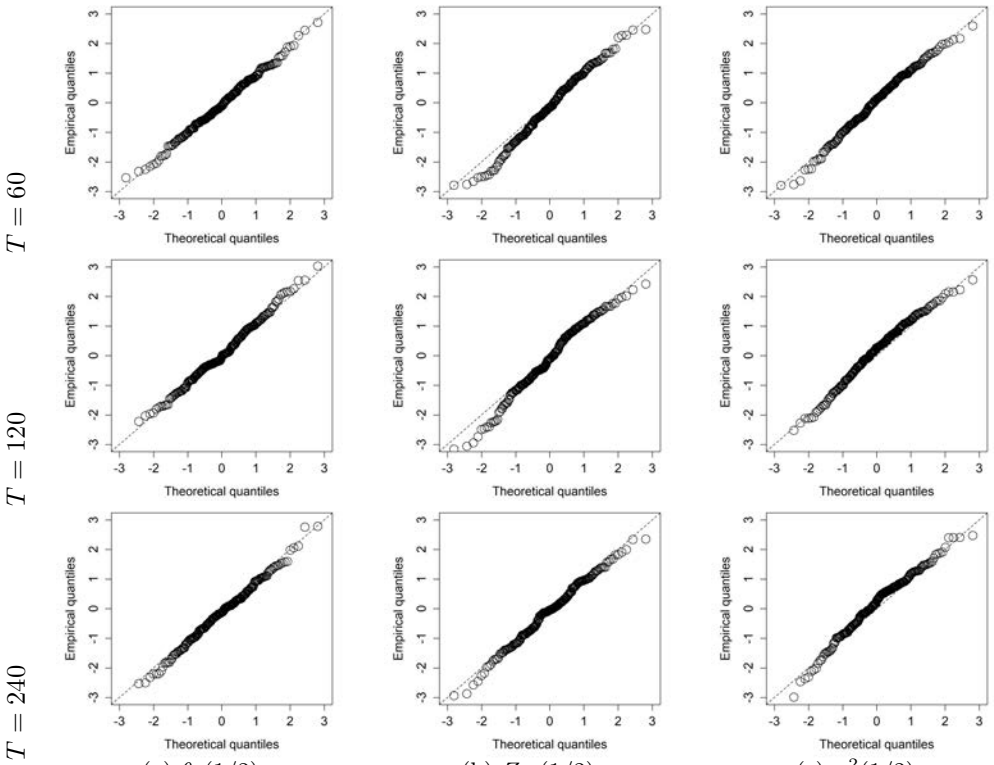
The plug-in covariance estimator is $\hat{\boldsymbol{\Sigma}}_\xi(u) = \nu_{02}\boldsymbol{\mathcal{P}}\hat{\boldsymbol{J}}^{-1}(u^t)\hat{\boldsymbol{\Upsilon}}(u^t)\hat{\boldsymbol{J}}^{-1}(u^t)\boldsymbol{\mathcal{P}}^\top$. Let $\hat{B}(u^t) = \frac{\nu_{21}h^2}{2}\boldsymbol{\mathcal{P}}\hat{\boldsymbol{J}}^{-1}(u^t)\hat{\boldsymbol{a}}(u^t)$ be the estimated leading bias. For any $k = 1, \ldots, \dim(\boldsymbol{\xi}(u^t))$, the nominal 95% pointwise confidence interval for $\xi_{h,k}(u^t)$ is

$$\left[\tilde{\xi}_{h,k}(u^t) - \hat{B}_k(u^t) - 1.96\sqrt{\frac{\hat{\boldsymbol{\Sigma}}_{\xi,kk}(u^t)}{MTh}},\ \tilde{\xi}_{h,k}(u^t) - \hat{B}_k(u^t) + 1.96\sqrt{\frac{\hat{\boldsymbol{\Sigma}}_{\xi,kk}(u^t)}{MTh}}\right].$$

Table 4 reports the average empirical coverage rates for $\boldsymbol{\theta}(u^t), \boldsymbol{Z}(u^t)$, and $\sigma^2(u^t)$ over 200 replications. We observe that as T increases, the empirical coverage rates gradually come closer to 95%. We further plot Q-Q-plots for selected parameters, using $\tilde{\theta}_1(u^t)$, $\tilde{Z}_{11}(u^t)$ and $\tilde{\sigma}^2(u^t)$ as

Table 4: Empirical coverage rates of 95% pointwise confidence intervals under the baseline setting with fixed $n = 50$, $M = 20$, and $u^t = 1/2$.

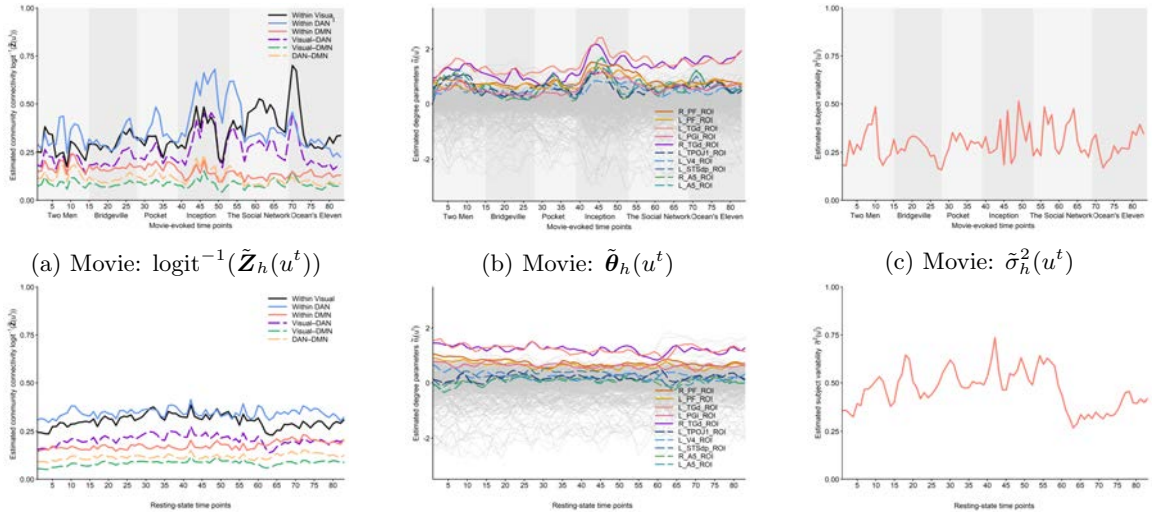
T	h	$\boldsymbol{\theta}(1/2)$	$\boldsymbol{Z}(1/2)$	$\sigma^2(1/2)$
60	$h = 0.155$	0.940	0.924	0.930
120	$h = 0.135$	0.942	0.938	0.935
240	$h = 0.118$	0.936	0.943	0.935



(a) $\theta_1(1/2)$ (b) $Z_{11}(1/2)$ (c) $\sigma^2(1/2)$

Figure 2: Q-Q-plots for selected parameters at $u^t = 1/2$ with $T \in \{60, 120, 240\}$.

examples. The results show that as T increases, the empirical distributions also gradually converge to the standard normal distribution, which provides the empirical support for the asymptotic normality of our estimators as established in Theorem 2.



(a) Movie: $\text{logit}^{-1}(\tilde{\boldsymbol{Z}}_h(u^t))$ (b) Movie: $\tilde{\boldsymbol{\theta}}_h(u^t)$ (c) Movie: $\tilde{\sigma}_h^2(u^t)$

(d) Rest: $\text{logit}^{-1}(\tilde{\boldsymbol{Z}}_h(u^t))$ (e) Rest: $\tilde{\boldsymbol{\theta}}_h(u^t)$ (f) Rest: $\tilde{\sigma}_h^2(u^t)$

Figure 3: Estimated parameter trajectories under movie watching and resting state.

5 Application to functional brain connectivity during movie watching

We apply our proposed method to the dynamic functional connectivity networks derived from a publicly available Human Connectome Project dataset, and mainly focus on $M = 34$ young adults (age 22–30) who completed the full movie-watching protocol. For comparison, we also include a resting-state segment of matched length from the same subjects. By construction, we have $T = 83$ undirected binary networks on $n = 360$ ROIs from the Glasser atlas (Glasser et al., 2016), grouped into $G = 7$ functional modules using the Yeo parcellation (Yeo et al., 2011). We fit the proposed time-varying mixed-effects DCSBM using an Epanechnikov kernel, with $h_{\text{opt}} = 0.03$ as selected using the leave-one-out cross-validated likelihood. Figure 3 presents the results. The data descriptions, network construction procedures, implementation details, and sensitivity analyses are provided in the Supplementary Material.

Community connectivity. For interpretation purposes, we transform the estimated community connectivity matrices to the probability scale. Compared with the rest condition, the probability-scale estimates suggest more pronounced temporal variation during movie watching, particularly of the within-connectivity for visual and DAN modules, as well as for visual–DAN coupling. This pattern is consistent with prior evidence on stimulus-driven functional connectivity during naturalistic viewing (Tian et al., 2021; Kringelbach et al., 2023). The resting-state curves are comparatively stable.

Regional and subject-level heterogeneity. The degree parameters reveal time-varying region-specific heterogeneity. Their trajectories allow us to identify ROIs whose connectivity propensity is enhanced or attenuated in response to the stimulus, thereby revealing stimulus-evoked regional reconfiguration. The estimated variance is generally smaller during movie watching than during rest, indicating weaker between-subject variability under naturalistic stimulation. This find-

ing agrees with prior studies showing that movie stimulation induces a more structured and more strongly shared functional connectivity organization across subjects (Hasson et al., 2008; Finn and Bandettini, 2021; Kringelbach et al., 2023).

Further results on subject-specific effects, individualized reconstruction errors, and out-of-sample discrimination results, are provided in the Supplementary Material.

6 Discussion

Motivated by functional brain connectivity, in this paper we suggest a model for modelling and estimating time-varying connectivity in (frequently sparsely observed) stochastic networks that are governed by both dynamic node heterogeneity and heterogeneity across subject-replicates. The key idea is to augment a dynamic DCSBM with random effects and to estimate its time-varying components through a time-localised likelihood, using a local approximate EM algorithm. This yields a unified estimator for the shared community connectivity, time-varying degree heterogeneity, and subject variability, for which we establish consistency and asymptotic normality.

Our work opens the line for addressing various subsequent statistical questions. The asymptotic normality developed here can be further used for testing node-heterogeneity: in the context of brain connectivity this would give more insights to identify the most important ROIs. Another alley would be to embed our analysis into the context of a classification problem where the interest lies in identifying underlying states or stimuli per subject.

Appendix Residualized joint spectral embedding

In the main text, we develop the estimation procedure conditional on known community memberships, motivated by applications where the prior module information is available, such as brain connectivity, financial networks, and certain social networks. This appendix considers the extension to unknown node memberships, and introduces a residualized joint spectral embedding procedure as a pre-processing step before applying the LA-EM algorithm. The idea is that the leading nuisance components are first removed, after which the common community structure is recovered from the principal eigenspace of an aggregated residual matrix.

For each time point u^t, we consider a community-free auxiliary model,

$$\operatorname{logit}(\boldsymbol{\tau}_m^{\text{aux}}(u^t)) = \boldsymbol{\theta}^{\text{aux}}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}^{\text{aux}}(u^t)^\top + \omega_m^{\text{aux}}(u^t)\mathbf{1}_n\mathbf{1}_n^\top,$$

where $\boldsymbol{\tau}_m^{\text{aux}}(u^t) = (\tau_{m,ij}^{\text{aux}}(u^t))_{1\le i,j\le n}$. Here, $\boldsymbol{\theta}^{\text{aux}}(u^t)$ and $\omega_m^{\text{aux}}(u^t)$ represent the node and subject effects under the auxiliary model. To identify this decomposition, we impose that $\sum_{m=1}^M \omega_m^{\text{aux}}(u^t) = 0$. Let $\hat{\tau}_{m,ij}^{\text{aux}}(u^t)$ denote the fitted edge probability from the auxiliary model. We then define the estimated residual matrix $\hat{\boldsymbol{R}}_m(u^t) = (\hat{R}_{m,ij}(u^t))_{1\le i,j\le n}$ by setting $\hat{R}_{m,ii}(u^t) = 0$ and, for any $i \neq j$,

$$\hat{R}_{m,ij}(u^t) = \frac{A_{m,ij}(u^t) - \hat{\tau}_{m,ij}^{\text{aux}}(u^t)}{\left[\hat{\tau}_{m,ij}^{\text{aux}}(u^t)(1-\hat{\tau}_{m,ij}^{\text{aux}}(u^t))\right]^{1/2}}.$$

Note that the standardization step stabilizes the variance across edges with different baseline probabilities, which in turn reduces the influence of heterogeneous edge variances in the residual matrices and yields a more stable spectral embedding. The common community structure is then extracted

from the residual inner-product $\sum_{m=1}^{M}\sum_{t=1}^{T}\hat{\boldsymbol{R}}_m(u^t)\hat{\boldsymbol{R}}_m(u^t)^\top$. Finally, we compute its eigen decomposition. Let $\hat{\boldsymbol{U}} \in \mathbb{R}^{n\times G}$ be the matrix whose columns are the leading G eigenvectors of the inner-product matrix, with the i-th row $\hat{\boldsymbol{U}}_{i\cdot}$ representing the spectral embedding of node i. We normalize the rows of $\hat{\boldsymbol{U}}$ by defining the matrix $\hat{\boldsymbol{U}}^* \in \mathbb{R}^{n\times G}$ with rows, $\hat{\boldsymbol{U}}^*_{i\cdot} = \frac{\hat{\boldsymbol{U}}_{i\cdot}}{\|\hat{\boldsymbol{U}}_{i\cdot}\|_2}$, $i = 1, \ldots, n$, whenever $\|\hat{\boldsymbol{U}}_{i\cdot}\|_2 > 0$. Applying K-means with G clusters to the rows of $\hat{\boldsymbol{U}}^*$, we obtain the estimated community memberships matrix $\hat{\boldsymbol{C}}$.

Several comments are in order. Notice that when $\boldsymbol{C}$ is unknown, community memberships are identifiable only up to permutation. We additionally assume that, for each u^t, the community connectivity matrix $\boldsymbol{Z}(u^t)$ has no two identical columns. Together with the constraints in Proposition 1, this ensures identifiability of the model with unknown memberships up to label permutation. Second, the auxiliary model is community-free by construction, and hence, it does not impose a community partition on the data. Instead, it provides a baseline description for non-community components of the network, so that residual matrices encode both random noise and any systematic structure not captured by the auxiliary model. If the latent communities manifest themselves through systematic residual connectivity patterns, such information is preserved in $\hat{\boldsymbol{R}}_m(u^t)$. The residual inner-product matrix aggregates weak community signals over both subjects and time points. When the networks share a common community structure, the signal components accumulate in $\sum_{m=1}^{M}\sum_{t=1}^{T}\hat{\boldsymbol{R}}_m(u^t)\hat{\boldsymbol{R}}_m(u^t)^\top$, whereas random fluctuations may be partially averaged out. This follows the same signal-aggregation principle of multilayer networks (**?**Agterberg et al., 2025; Li et al., 2026), but the aggregation here is performed after removing the nuisance heterogeneity through the auxiliary model. The simulation results where the communities are estimated by this approach of residual joint spectral embedding are provided in Section S6 of the Supplementary Material.

Supplementary Materials

The Supplementary Material contains the proofs, bandwidth-selection procedure, algorithmic details, additional simulation results, and further application details, including pre-processing, sensitivity analyses and additional results.

Acknowledgements

Mengxue Li would like to express her sincere gratitude to Hernando Ombao and Laura Symul for their insightful discussions, constructive comments, and invaluable suggestions. She also gratefully acknowledges the financial support provided by the China Scholarship Council.

References

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

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1993). *Efficient and adaptive estimation*

for semiparametric models. Johns Hopkins Series in the Mathematical Sciences. Baltimore, MD: Johns Hopkins University Press.

Caimo, A. and I. Gollini (2020). A multilayer exponential random graph modelling approach for weighted networks. *Computational Statistics & Data Analysis 142*, 106825.

Cole, M. W., J. R. Reynolds, J. D. Power, G. Repovs, A. Anticevic, and T. S. Braver (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience 16*(9), 1348–1355.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B: Statistical Methodology 39*(1), 1–22.

Du, Y., L. Qu, T. Yan, and Y. Zhang (2023). Time-varying β-model for dynamic directed networks. *Scandinavian Journal of Statistics 50*(4), 1687–1715.

Fan, J. and J. Chen (1999). One-step local quasi-likelihood estimation. *Journal of the Royal Statistical Society Series B: Statistical Methodology 61*(4), 927–943.

Fan, J., M. Farmen, and I. Gijbels (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society Series B: Statistical Methodology 60*(3), 591–608.

Finn, E. S. and P. A. Bandettini (2021). Movie-watching outperforms rest for functional connectivity-based prediction of behavior. *NeuroImage 235*, 117963.

Geyer, C. J. (1994). On the asymptotics of constrained M-estimation. *The Annals of Statistics 22*(4), 1993 – 2010.

Glasser, M. F., T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature 536*(7615), 171–178.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Hasson, U., O. Furman, D. Clark, Y. Dudai, and L. Davachi (2008). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron 57*(3), 452–462.

He, Y., J. Sun, Y. Tian, Z. Ying, and Y. Feng (2025). Semiparametric modeling and analysis for longitudinal network data. *The Annals of Statistics 53*(4), 1406–1430.

Karrer, B. and M. E. J. Newman (2011). Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics 83*(1), 016107.

Kim, Y., D. Kessler, and E. Levina (2023). Graph-aware modeling of brain connectivity networks. *The Annals of Applied Statistics 17*(3), 2095–2117.

Kringelbach, M. L., Y. S. Perl, E. Tagliazucchi, and G. Deco (2023). Toward naturalistic neuroscience: Mechanisms underlying the flattening of brain hierarchy in movie-watching compared to rest and task. *Science Advances 9*(2), eade6049.

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

Li, D., B. Peng, S. Tang, and W. Wu (2026). Estimation of grouped time-varying network vector autoregressive models. *The Annals of Statistics 54*(2), 621–646.

Li, J., S. Wu, C. Cui, G. Xu, and J. Zhu (2023). Statistical inference on latent space models for network data. *arXiv preprint arXiv:2312.06605*.

Li, M., R. von Sachs, and E. Pircalabelu (2026). Time-varying degree-corrected stochastic block models. *Scandinavian Journal of Statistics 53*(3), in press.

MacDonald, P. W., E. Levina, and J. Zhu (2022). Latent space models for multiplex networks with shared structure. *Biometrika 109*(3), 683–706.

Matias, C. and V. Miele (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society Series B: Statistical Methodology 79*(4), 1119–1141.

Moore, T. J., B. M. Sadler, and R. J. Kozick (2008). Maximum-likelihood estimation, the Cramér–Rao bound, and the method of scoring with parameter constraints. *IEEE Transactions on Signal Processing 56*(3), 895–908.

Stein, S., R. Feng, and C. Leng (2025). A sparse β regression model for network analysis. *Journal of the American Statistical Association 120*(550), 1281–1293.

Tian, L., M. Ye, C. Chen, X. Cao, and T. Shen (2021). Consistency of functional connectivity across different movies. *NeuroImage 233*, 117926.

Van den Heuvel, M. P. and O. Sporns (2013). Network hubs in the human brain. *Trends in Cognitive Sciences 17*(12), 683–696.

Yeo, B. T., F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology 106*(3), 1125–1165.

Zhang, J., W. W. Sun, and L. Li (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *Journal of the American Statistical Association 115*(532), 2022–2036.

Zhang, M., B. Cai, W. Dai, D. Kong, H. Zhao, and J. Zhang (2024). Learning brain connectivity in social cognition with dynamic network regression. *The Annals of Applied Statistics 18*(4), 3405–3424.

Zhang, X., S. Xue, and J. Zhu (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning*, pp. 11288–11297. PMLR.

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

Supplementary Material for "Learning shared and individual structure in dynamic networks with degree heterogeneity"

Mengxue Li, Rainer von Sachs and Eugen Pircalabelu

UCLouvain, Institute of Statistics, Biostatistics and Actuarial Sciences

Notations. Throughout the paper, for a positive integer k, let $\mathbf{1}_k$ be the k-dimensional vector of ones, $\boldsymbol{I}_k$ be the $k \times k$ identity matrix, and $\mathbf{0}$ denote a conformable zero vector or zero matrix determined by the context. For a vector $\boldsymbol{X} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$, write $\|\boldsymbol{X}\|_k = (\sum_{i=1}^p |X_i|^k)^{1/k}$ for $k \geq 1$. For any generic vector or matrix $\boldsymbol{X} = (X_{ij})$, write $\|\boldsymbol{X}\|_\infty = \max_{i,j} |X_{ij}|$ with a vector viewed as a one-column matrix. Let $\|\cdot\|_2$, $\|\cdot\|_F$ and $\|\cdot\|_{\mathrm{op}}$ denote the Euclidean, Frobenius and operator norm, respectively. For any generic square matrix $\boldsymbol{X}$, let $\lambda_{\min}(\cdot)$, $\lambda_{\max}(\cdot)$ and $\lambda_k(\cdot)$ denote its minimal, maximal and k-th largest eigenvalues. We use $\mathrm{vec}(\cdot)$ for column-wise vectorization, $\mathrm{diag}(\cdot)$ for the diagonal matrix formed from a vector or the diagonal of a matrix, $\mathbb{1}(\cdot)$ for the indicator function, and $\rightsquigarrow$ for the convergence in distribution, respectively. For two positive sequences x_T and y_T, write $x_T = o(y_T)$ if $x_T/y_T \to 0$ as $T \to \infty$, $x_T = o_p(y_T)$ if $x_T/y_T \xrightarrow{p} 0$ as $T \to \infty$, and $x_T \asymp y_T$ if there exist constants $0 < a < b < \infty$ such that $a \leq x_T/y_T \leq b$ for all sufficiently large T.

S1. Proofs for Section 2

Proof of Proposition 1. Fix an arbitrary time point $u^t \in \{1/T, \ldots, 1\}$. Suppose that there exist two sets of parameters $(\boldsymbol{\theta}(u^t), \boldsymbol{Z}(u^t))$ and $(\boldsymbol{\theta}^*(u^t), \boldsymbol{Z}^*(u^t))$ satisfying the equation as follows,

$$\boldsymbol{\theta}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}(u^t)\boldsymbol{C}^\top = \boldsymbol{\theta}^*(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\boldsymbol{\theta}^*(u^t)^\top + \boldsymbol{C}\boldsymbol{Z}^*(u^t)\boldsymbol{C}^\top. \tag{S1.1}$$

Define

$$\Delta\boldsymbol{\theta}(u^t) = \boldsymbol{\theta}(u^t) - \boldsymbol{\theta}^*(u^t), \qquad \Delta\boldsymbol{Z}(u^t) = \boldsymbol{Z}(u^t) - \boldsymbol{Z}^*(u^t).$$

Then (S1.1) is equivalent to

$$\Delta\boldsymbol{\theta}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\Delta\boldsymbol{\theta}(u^t)^\top + \boldsymbol{C}\Delta\boldsymbol{Z}(u^t)\boldsymbol{C}^\top = \mathbf{0}. \tag{S1.2}$$

Moreover, since both parameter sets satisfy the identifiability constraint that $\boldsymbol{C}^\top\boldsymbol{\theta}(u^t) = \boldsymbol{C}^\top\boldsymbol{\theta}^*(u^t) = \mathbf{0}$, we have

$$\boldsymbol{C}^\top\Delta\boldsymbol{\theta}(u^t) = \mathbf{0}. \tag{S1.3}$$

Left-multiplying both sides of (S1.2) by $\boldsymbol{C}^\top$, and right-multiplying by $\boldsymbol{C}$ yields

$$\boldsymbol{C}^\top\Delta\boldsymbol{\theta}(u^t)\mathbf{1}_n^\top\boldsymbol{C} + \boldsymbol{C}^\top\mathbf{1}_n\Delta\boldsymbol{\theta}(u^t)^\top\boldsymbol{C} + \boldsymbol{C}^\top\boldsymbol{C}\,\Delta\boldsymbol{Z}(u^t)\,\boldsymbol{C}^\top\boldsymbol{C} = \mathbf{0}.$$

By (S1.3), we note that the first and second terms vanish, and we thus have that

$$\boldsymbol{C}^\top \boldsymbol{C}\, \Delta \boldsymbol{Z}(u^t)\, \boldsymbol{C}^\top \boldsymbol{C} = \boldsymbol{0}.$$

Since $\boldsymbol{C} \in \{0,1\}^{n \times G}$ is the community membership matrix where $C_{ig} = 1$ if $c_i = g$ and $C_{ig} = 0$ otherwise, we observe that $\boldsymbol{C}^\top \boldsymbol{C}$ is a diagonal matrix with all positive entries $n_g = \sum_{i=1}^n \mathbb{1}(c_i = g) > 0$ for any $g = 1, \ldots, G$, i.e.,

$$\boldsymbol{C}^\top \boldsymbol{C} = \mathrm{diag}(n_1, \ldots, n_G).$$

Therefore, the term, $\boldsymbol{C}^\top \boldsymbol{C}$, is invertible, implying that $\Delta \boldsymbol{Z}(u^t) = \boldsymbol{0}$, i.e., $\boldsymbol{Z}(u^t) = \boldsymbol{Z}^*(u^t)$.

Since $\Delta \boldsymbol{Z}(u^t) = \boldsymbol{0}$, the last term on the left-hand side of (S1.2) disappears. Right-multiplying both sides by $\boldsymbol{1}_n$ gives

$$\Delta \boldsymbol{\theta}(u^t) \boldsymbol{1}_n^\top \boldsymbol{1}_n + \boldsymbol{1}_n \Delta \boldsymbol{\theta}(u^t)^\top \boldsymbol{1}_n = \boldsymbol{0},$$

where the left-hand side can be rewritten as

$$\begin{pmatrix} n\Delta\theta_1(u^t) \\ n\Delta\theta_2(u^t) \\ \vdots \\ n\Delta\theta_n(u^t) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^n \Delta\theta_i(u^t) \\ \sum_{i=1}^n \Delta\theta_i(u^t) \\ \vdots \\ \sum_{i=1}^n \Delta\theta_i(u^t) \end{pmatrix} = \begin{pmatrix} n\Delta\theta_1(u^t) + \sum_{i=1}^n \Delta\theta_i(u^t) \\ n\Delta\theta_2(u^t) + \sum_{i=1}^n \Delta\theta_i(u^t) \\ \vdots \\ n\Delta\theta_n(u^t) + \sum_{i=1}^n \Delta\theta_i(u^t) \end{pmatrix}.$$

Therefore, it holds that $\Delta\theta_1(u^t) = \ldots = \Delta\theta_n(u^t) = -\frac{\sum_{i=1}^n \Delta\theta_i(u^t)}{n}$, which implies that all entries of $\Delta\boldsymbol{\theta}(u^t)$ are zero. □

S2. Proofs for Section 3

S2.1 Definitions

Before proving the main results, we first recall some definitions that will be used throughout the proofs. For any $u \in (0,1]$, $\boldsymbol{\xi}(u) = \left(\boldsymbol{\theta}(u)^\top, \mathrm{vec}(\boldsymbol{Z}(u))^\top, \sigma^2(u)\right)^\top$ denotes the true parameter curve, and, at each fixed time point $u^t = t/T, \forall t = 1, \ldots, T$, the true parameter $\boldsymbol{\xi}(u^t)$ belongs to the constrained parameter space

$$\Xi_{\boldsymbol{C}} = \left\{ \left(\boldsymbol{\theta}^\top, \mathrm{vec}(\boldsymbol{Z})^\top, \sigma^2\right)^\top : \boldsymbol{\theta} \in \mathbb{R}^n,\ \boldsymbol{C}^\top \boldsymbol{\theta} = \boldsymbol{0},\ \boldsymbol{Z} \in \mathbb{R}^{G \times G},\ \boldsymbol{Z}^\top = \boldsymbol{Z},\ \sigma^2 \in \mathbb{R}_+ \right\}.$$

Recall in Section 3.1 of the main text that the marginal log-likelihood contribution of subject m is defined as

$$\begin{aligned}\ell_m(\boldsymbol{\xi}(u^t); \boldsymbol{A}_m(u^t)) = \log \int_{\mathbb{R}} \prod_{1 \le i < j \le n} & \frac{\exp\left(A_{m,ij}(u^t) \times (\Theta_{ij}(u^t) + \omega_m(u^t))\right)}{1 + \exp\left(\Theta_{ij}(u^t) + \omega_m(u^t)\right)} \\ & \times \phi(\omega_m(u^t); 0, \sigma^2(u^t))\, \mathrm{d}\, \omega_m(u^t).\end{aligned}$$

Let $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$ be a candidate parameter at a given time point u^t. We consider the scaled version of the time-localised likelihood criterion in (2) as,

$$\begin{aligned}\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) &= \frac{1}{MT} \sum_{s=1}^{T} L(\boldsymbol{\zeta}(u^t); \boldsymbol{A}(u^s)) K_h(u^s - u^t) \\ &= \frac{1}{MT} \sum_{s=1}^{T} \sum_{m=1}^{M} \ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s)) K_h(u^s - u^t).\end{aligned} \tag{S2.1}$$

Thus, the smooth estimator in (2) of the main text is equivalently defined as

$$\tilde{\boldsymbol{\xi}}_h(u^t) = \arg \max_{\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}} \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) = \arg \max_{\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}} \tilde{L}_h(\boldsymbol{\zeta}(u^t); \boldsymbol{A}).$$

We further define the population criterion associated with the marginal log-likelihood at a given time point u^t as

$$\mathbb{L}(\boldsymbol{\zeta}(u^t)) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}_{\boldsymbol{\xi}(u^t)}[\ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^t)],$$

where the expectation is under the true parameter $\boldsymbol{\xi}(u^t)$.

Since the parameters are subject to the linear constraints in Proposition 1, the corresponding tangent space at any interior point of $\Xi_{\boldsymbol{C}}$ can be expressed as

$$\Xi_{\boldsymbol{C}}^{\tan} = \left\{ \left(\boldsymbol{v}_\theta^\top, \mathrm{vec}(\boldsymbol{v}_Z)^\top, v_\sigma\right)^\top : \boldsymbol{C}^\top \boldsymbol{v}_\theta = \boldsymbol{0},\ \boldsymbol{v}_Z^\top = \boldsymbol{v}_Z \right\}.$$

Let $\boldsymbol{\mathcal{P}} \in \mathbb{R}^{\dim(\boldsymbol{\xi}(u)) \times \dim(\Xi_{\boldsymbol{C}}^{\tan})}$ be a matrix with orthonormal columns spanning $\Xi_{\boldsymbol{C}}^{\tan}$, that is,

$$\boldsymbol{\mathcal{P}}^\top \boldsymbol{\mathcal{P}} = \boldsymbol{I}_{\dim(\Xi_{\boldsymbol{C}}^{\tan})}, \qquad \mathrm{col}(\boldsymbol{\mathcal{P}}) = \Xi_{\boldsymbol{C}}^{\tan}.$$

Hence, any feasible parameter vector $\boldsymbol{\zeta}(u^t)$ in a sufficiently small neighbourhood of $\boldsymbol{\xi}(u^t)$ admits the unique representation

$$\boldsymbol{\zeta}(u^t) = \boldsymbol{\xi}(u^t) + \boldsymbol{\mathcal{P}}\boldsymbol{\vartheta}(u^t),$$

where $\boldsymbol{\vartheta}(u^t) \in \left\{\boldsymbol{\vartheta}(u^t) \in \mathbb{R}^{\dim(\Xi_{\boldsymbol{C}}^{\tan})} : \boldsymbol{\xi}(u^t) + \boldsymbol{\mathcal{P}}\boldsymbol{\vartheta}(u^t) \in \Xi_{\boldsymbol{C}}^{\tan}\right\}$, denotes the intrinsic free coordinate under the constraints at the given time point u^t. Thus, locally, the constrained maximization problem in (S2.1) is equivalent to maximizing over the free coordinate $\boldsymbol{\vartheta}(u^t)$. In other words, the analysis of the original constrained estimator reduces to that of the intrinsic coordinate estimator. On the event that $\tilde{\boldsymbol{\xi}}_h(u^t)$ belongs to a sufficiently small neighbourhood of $\boldsymbol{\xi}(u^t)$, there exists a unique $\tilde{\boldsymbol{\vartheta}}_h(u^t) \in \mathbb{R}^{\dim(\Xi_{\boldsymbol{C}}^{\tan})}$ satisfying

$$\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t) = \boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\vartheta}}_h(u^t), \qquad \left\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\right\|_2 = \left\|\tilde{\boldsymbol{\vartheta}}_h(u^t)\right\|_2,$$

where the last equality follows from $\boldsymbol{\mathcal{P}}^\top\boldsymbol{\mathcal{P}} = \boldsymbol{I}_{\dim(\Xi_{\boldsymbol{C}}^{\tan})}$. Hence, once the estimator is localised in the neighbourhood of its true value, the desired rate for $\tilde{\boldsymbol{\xi}}_h(u^t)$ follows directly from the corresponding rate for $\tilde{\boldsymbol{\vartheta}}_h(u^t)$.

S2.2 Preliminary lemmas

Throughout, $u^t = t/T$ denotes a target grid point satisfying $u^t \to u \in (0,1]$. All localised criteria below are evaluated at this same target point u^t. We now state the following lemmas.

Lemma S1. *Suppose that Assumption 1 holds. If $h \to 0$, $MT \to \infty$, and $MTh \to \infty$, then there exists a sequence $\epsilon_{M,T,h} \to 0$ such that*

$$\mathbb{P}\left(\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left|\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}(\boldsymbol{\zeta}(u^t))\right| < \epsilon_{M,T,h}\right) \to 1.$$

For a generic parameter value $\boldsymbol{\zeta}(u^t)$, define the score vector and Hessian matrix of the scaled local criterion $\mathbb{L}_{M,T,h}$ as

$$S_{M,T,h}(\boldsymbol{\zeta}(u^t)) = \frac{\partial \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))}{\partial \boldsymbol{\zeta}(u^t)}, \qquad H_{M,T,h}(\boldsymbol{\zeta}(u^t)) = \frac{\partial^2 \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))}{\partial \boldsymbol{\zeta}(u^t)\,\partial \boldsymbol{\zeta}(u^t)^\top},$$

respectively. We next establish the central limit theorem for the projected score evaluated at the true parameter value, which provides the stochastic expansion used to prove Theorem 2 in the main text.

Lemma S2. *Suppose that Assumption 1 holds. For any fixed $u \in (0,1)$, consider a sequence of grid points $u^t = t/T, t = 1,\dots,T$ such that $u^t \to u$ as $T \to \infty$. If $h \to 0$ and $MT \to \infty$ such that $MTh \to \infty, MTh^5 = O(1)$, then it holds that*

$$\sqrt{MTh}\left[\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t)) - \frac{\nu_{21}}{2}h^2\boldsymbol{a}(u^t)\right] \rightsquigarrow N(0, \nu_{02}\boldsymbol{\Upsilon}(u)),$$

where $\boldsymbol{a}(u^t)$ is a deterministic vector determined by the second derivative of the true curve $\boldsymbol{\xi}(u)$ around u^t, $\nu_{xy} = \int u^x K^y(u)\,\mathrm{d}u$, and $\boldsymbol{\Upsilon}(u)$, as defined in Assumption 1, is the limit of the score covariance along the sequence $u^t \to u$.

Lemma S3. *Define $r_{M,T,h} = h^2 + (MTh)^{-1/2}$. Suppose that Assumption 1 holds. If $h \to 0$, $MT \to \infty$, and $MTh \to \infty$, then, for any $\epsilon > 0$, there exists a finite constant $a_\epsilon > 0$ such that, for each fixed u^t,*

$$\liminf_{MT\to\infty} \mathbb{P}\left(\sup_{\|\boldsymbol{\alpha}\|_2=1} \mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}\right) < \mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t)) \right) \geq 1-\epsilon, \tag{S2.2}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^{\dim(\Xi_{\boldsymbol{C}}^{\tan})}$.

Proof of Lemma S1. Decompose

$$\begin{aligned}
&\left|\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}(\boldsymbol{\zeta}(u^t))\right| \\
&\leq \left|\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{E}[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))]\right| + \left|\mathbb{E}[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))] - \mathbb{L}(\boldsymbol{\zeta}(u^t))\right| \\
&= |I_1(\boldsymbol{\zeta}(u^t))| + |I_2(\boldsymbol{\zeta}(u^t))|,
\end{aligned} \tag{S2.3}$$

where terms $I_1(\boldsymbol{\zeta}(u^t)) = \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{E}[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))]$ and $I_2(\boldsymbol{\zeta}(u^t)) = \mathbb{E}[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))] - \mathbb{L}(\boldsymbol{\zeta}(u^t))$ correspond to the stochastic error and the deterministic smoothing bias, respectively.
Step 1. Control of the stochastic term $I_1(\boldsymbol{\zeta}(u^t))$. Note that $K_h(u) = \frac{1}{h}K(u/h)$. For any fixed $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$, we define

$$X_{m,s}(\boldsymbol{\zeta}(u^t)) = \ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s)) - \mathbb{E}[\ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s))].$$

Then

$$I_1(\boldsymbol{\zeta}(u^t)) = \frac{1}{MTh}\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right) X_{m,s}(\boldsymbol{\zeta}(u^t)).$$

By Assumption 1, the parameter space is compact, n is fixed and the log-likelihood $\ell_m(\cdot)$ is continuously differentiable. Since the sample space of $\boldsymbol{A}_m(u^s)$ is also finite for fixed n, compactness and continuity imply that,

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \sup_{m\in\{1,\ldots,M\}} \sup_{s\in\{1,\ldots,T\}} \left|\ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s))\right| < \infty.$$

This implies that there exist some finite constants $a_X > 0$ such that

$$\left|X_{m,s}(\boldsymbol{\zeta}(u^t))\right| \leq a_X, \quad \mathrm{Var}(X_{m,s}(\boldsymbol{\zeta}(u^t))) \leq a_X^2, \tag{S2.4}$$

uniformly over m, s and t. Therefore, Bernstein's inequality gives that, for every $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}\left(\left|I_1(\boldsymbol{\zeta}(u^t))\right| > \epsilon\right) &= \mathbb{P}\left(\left|\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right) X_{m,s}(\boldsymbol{\zeta}(u^t))\right| > MTh\epsilon\right) \\
&\leq 2\exp\left(-\frac{(MTh\epsilon)^2/2}{MTh(a_X')^2 + a_X' MTh\epsilon/3}\right) \\
&= 2\exp\left(-\frac{3MTh(\epsilon)^2}{6(a_X')^2 + 2a_X'\epsilon}\right)
\end{aligned} \tag{S2.5}$$

due to the fact that

$$\begin{aligned}
&\mathrm{Var}\left(\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right) X_{m,s}(\boldsymbol{\zeta}(u^t))\right) \\
&= \sum_{m=1}^{M}\sum_{s=1}^{T} K^2\left(\frac{u^s-u^t}{h}\right) \mathrm{Var}(X_{m,s}(\boldsymbol{\zeta}(u^t))) \\
&\leq MTh(a_X^{'})^2,
\end{aligned}$$

with some constant $a_X^2 \leq (a_X^{'})^2$.

We now make the bound uniform over $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$. For each fixed $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$, the Bernstein bound in (S2.5) implies that $I_1(\boldsymbol{\zeta}(u^t)) = o_p(1)$. To establish uniform convergence, we first derive deterministic bounds for generic arguments. Let $A \in \{0,1\}^{n\times n}$ and $\boldsymbol{\zeta} \in \Xi_{\boldsymbol{C}}$ denote generic deterministic arguments. Since $\boldsymbol{A}(u^s) \in \{0,1\}^{n\times n}$ for every m and s, any bound that holds uniformly over A and $\boldsymbol{\zeta}$, also applies, in particular, to $A = \boldsymbol{A}_m(u^s)$ and $\boldsymbol{\zeta} = \boldsymbol{\zeta}(u^t)$. Define

$$F_m(A) = 2 \sup_{\boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} |\ell_m(\boldsymbol{\zeta}; A)|.$$

By the compactness of $\Xi_{\boldsymbol{C}}$,the continuity of $\ell_m(\boldsymbol{\zeta}; A)$ and $S_{\ell,m}(\boldsymbol{\zeta}; A)$ in $\boldsymbol{\zeta}$, and the fact that n is fixed, there exist constants $a_\ell, a_\ell^* < \infty$ such that

$$\begin{aligned}
\sup_{m, A\in\{0,1\}^{n\times n}} F_m(A) &\leq a_\ell, \\
\sup_{m, A\in\{0,1\}^{n\times n},\, \boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} \|S_{\ell,m}(\boldsymbol{\zeta}; A)\|_2 &\leq a_\ell^*.
\end{aligned}$$

It follows by the triangle inequality and Jensen's inequality,

$$\begin{aligned}
&\sup_{m, A\in\{0,1\}^{n\times n},\, \boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} |\ell_m(\boldsymbol{\zeta}; A) - \mathbb{E}(\ell_m(\boldsymbol{\zeta}; A_m(u^s)))| \\
&\leq \sup_{m, A\in\{0,1\}^{n\times n},\, \boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} |\ell_m(\boldsymbol{\zeta}; A)| + \sup_{m,\, \boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} |\mathbb{E}(\ell_m(\boldsymbol{\zeta}; A_m(u^s)))| \\
&\leq 2 \sup_{m, A\in\{0,1\}^{n\times n},\, \boldsymbol{\zeta}\in\Xi_{\boldsymbol{C}}} |\ell_m(\boldsymbol{\zeta}; A)| \\
&\leq \sup_{m, A\in\{0,1\}^{n\times n}} F_m(A) \leq a_\ell.
\end{aligned}$$

Hence the class $\{\ell_m(\boldsymbol{\zeta}; A) - \mathbb{E}(\ell_m(\boldsymbol{\zeta}; A_m(u^s))) : \boldsymbol{\zeta} \in \Xi_{\boldsymbol{C}}\}$ is uniformly bounded. Moreover, for any $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \Xi_{\boldsymbol{C}}$, the mean value theorem gives

$$|\ell_m(\boldsymbol{\zeta}_1; A) - \ell_m(\boldsymbol{\zeta}_2; A)| \leq a_\ell^* \|\boldsymbol{\zeta}_1 - \boldsymbol{\zeta}_2\|_2,$$

uniformly over m and $A \in \{0,1\}^{n\times n}$. As stated in Assumption 1, the kernel function $K(u)$ is bounded and compactly supported on $[-1,1]$, and the effective local sample size satisfies $MTh \to \infty$. Hence, by using the uniform law of large numbers, we obtain that

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left|\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{E}[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))]\right| \xrightarrow{p} 0,$$

which means that

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left|I_1(\boldsymbol{\zeta}(u^t))\right| = o_p(1). \tag{S2.6}$$

Step 2. Control of the smoothing bias $I_2(\boldsymbol{\zeta}(u^t))$. It remains to bound the smoothing bias. For any $m = 1, \ldots, M$ and $s = 1, \ldots, T$, we consider

$$\boldsymbol{A}_m(u^s) = \boldsymbol{\tau}_m(u^s) + \boldsymbol{\varepsilon}_m(u^s),$$

where $\boldsymbol{\tau}_m(u^s)$ is determined by the true parameter $\boldsymbol{\xi}(u^s)$ and $\boldsymbol{\varepsilon}_m(u^s)$ denotes the centered error matrix with zero mean. Hence, the expectation of the marginal log-likelihood can be considered as a function of two quantities, (1) the candidate parameter $\boldsymbol{\zeta}(u^t)$ and (2) the true parameter $\boldsymbol{\xi}(u^s)$ that generates the data at time u^s. To do this, we define

$$\begin{aligned}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^s) &= \mathbb{E}_{\boldsymbol{\xi}(u^s)}\left[\ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s))\right] \\ &= \mathbb{E}\left[\ell_m(\boldsymbol{\zeta}(u^t); \boldsymbol{\tau}_m(u^s) + \boldsymbol{\varepsilon}_m(u^s))\right].\end{aligned}$$

With this notation, we can rewrite the bias term $I_2(\boldsymbol{\zeta}(u^t))$ as

$$\begin{aligned}I_2(\boldsymbol{\zeta}(u^t)) &= \mathbb{E}\left[\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))\right] - \mathbb{L}(\boldsymbol{\zeta}(u^t)) \\ &= \frac{1}{MTh}\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s - u^t}{h}\right)\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^s) - \frac{1}{M}\sum_{m=1}^{M}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^t).\end{aligned} \tag{S2.7}$$

By Assumption 1, the true parameter curve $\boldsymbol{\xi}(\cdot)$ is twice continuously differentiable and remains in a compact subset of the parameter space. Moreover, the marginal log-likelihood $\ell_m(\boldsymbol{\xi}(u); \boldsymbol{A}(u))$ is three times continuously differentiable in the model parameters, with uniformly bounded second derivatives over the compact parameter space. Hence, its first and second derivatives are uniformly bounded by continuity and compactness. Since n is fixed, the sample space of $\boldsymbol{A}_m$ is finite, and thus the corresponding marginal probabilities also have uniformly bounded first and second derivatives. It follows that, uniformly over $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$ and m, the function $\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)$ is twice continuously differentiable in u, with uniformly bounded first and second derivatives, i.e.,

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}},\, u\in[0,1],\, m}\left|\frac{\partial^x}{\partial u^x}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)\right| < \infty, \qquad x = 1, 2. \tag{S2.8}$$

A second-order Taylor expansion around u^t gives that

$$\begin{aligned}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^s) &= \ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^t) + \frac{\partial}{\partial u}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)\Big|_{u=u^t}(u^s - u^t) \\ &\quad + \frac{1}{2}\times\frac{\partial^2}{\partial u^2}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)\Big|_{u=u^{t^*}}(u^s - u^t)^2 \\ &= \ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^t) + a_{1,m,t}(u^s - u^t) + \frac{a_{2,m,t^*}}{2}(u^s - u^t)^2,\end{aligned} \tag{S2.9}$$

where u^{t^*} lies between u^s and u^t but may vary from term to term. Here

$$a_{1,m,t} = \frac{\partial}{\partial u}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)\Big|_{u=u^t}, \quad a_{2,m,t^*} = \frac{\partial^2}{\partial u^2}\ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u)\Big|_{u=u^{t^*}}.$$

Substituting the Taylor expansion into $I_2(\boldsymbol{\zeta}(u^t))$ yields,

$$\begin{aligned} I_2(\boldsymbol{\zeta}(u^t)) = & \frac{1}{M}\sum_{m=1}^{M} \ell_m^{\mathbb{E}}(\boldsymbol{\zeta}(u^t), u^t) \left\{ \frac{1}{Th}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right) - 1 \right\} \\ & + \frac{1}{MTh}\sum_{m=1}^{M} a_{1,m,t} \sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right)(u^s-u^t) \\ & + \frac{1}{2MTh}\sum_{m=1}^{M} a_{2,m,t^*} \sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right)(u^s-u^t)^2. \end{aligned}$$

Since $K(u)$ is symmetric, compactly supported on $[-1,1]$, and integrates to one, using the regular time grid, we have that

$$\begin{aligned} I_2(\boldsymbol{\zeta}(u^t)) &= \frac{1}{2MTh}\sum_{m=1}^{M}\sum_{s=1}^{T} a_{2,m,t^*} K\left(\frac{u^s-u^t}{h}\right)(u^s-u^t)^2 \\ &= \frac{1}{2M}\sum_{m=1}^{M} a_{2,m,t}\nu_{21}h^2 + o(h^2) \end{aligned}$$

where the remainder $o(h^2)$ accounts for replacing a_{2,m,t^*} by $a_{2,m,t}$, since $a_{2,m,t^*} = a_{2,m,t} + o(1)$ as $u^{t^*} \to u^t$. Thus, as $h \to 0$, (S2.8) implies that,

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left| I_2(\boldsymbol{\zeta}(u^t)) \right| = o(1).$$

Combining this bound with the uniform bound in (S2.6) gives

$$\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left| \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}(\boldsymbol{\zeta}(u^t)) \right| = o_p(1),$$

which completes the proof of Lemma S1. □

Proof of Lemma S2. For notational simplicity, we define the projected score of the function $\ell_m\big(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s)\big)$ by

$$S_{\ell,m}(\boldsymbol{\zeta}(u^t)) = \frac{\partial \ell_m\big(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s)\big)}{\partial \boldsymbol{\zeta}(u^t)}.$$

By the definition of $S_{M,T,h}\big(\boldsymbol{\zeta}(u^t)\big)$, we have

$$\begin{aligned} \boldsymbol{\mathcal{P}}^\top S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big) &= \frac{1}{MT}\sum_{s=1}^{T}\sum_{m=1}^{M} K_h(u^s-u^t)\boldsymbol{\mathcal{P}}^\top \left.\frac{\partial \ell_m\big(\boldsymbol{\zeta}(u^t); \boldsymbol{A}_m(u^s)\big)}{\partial \boldsymbol{\zeta}(u^t)}\right|_{\boldsymbol{\zeta}(u^t)=\boldsymbol{\xi}(u^t)} \\ &= \frac{1}{MT}\sum_{s=1}^{T}\sum_{m=1}^{M} K_h(u^s-u^t)\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t)). \end{aligned}$$

Therefore,

$$\begin{aligned} & \sqrt{MTh}\left\{ \boldsymbol{\mathcal{P}}^\top S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big) - \mathbb{E}\left[\boldsymbol{\mathcal{P}}^\top S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big) \right] \right\} \\ & = \frac{1}{\sqrt{MTh}}\sum_{s=1}^{T}\sum_{m=1}^{M} K\left(\frac{u^s-u^t}{h}\right)\left[\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t)) - \mathbb{E}_{\boldsymbol{\xi}(u^s)}\left(\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t)) \right) \right]. \end{aligned}$$

We now use the Cramér–Wold theorem to show the results. For an arbitrary fixed vector $\boldsymbol{\alpha}^* \in \mathbb{R}^{\dim(\Xi_C^{\tan})}$, we consider the linear combination

$$\sqrt{MTh}\,\boldsymbol{\alpha}^{*\top}\left\{\boldsymbol{\mathcal{P}}^\top S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big) - \mathbb{E}\left[\boldsymbol{\mathcal{P}}^\top S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big)\right]\right\}.$$

By Assumption 1 (ii) and Assumption 1 (v), since n is fixed, it is clear that the sample space of each $\boldsymbol{A}_m(u^s)$ is finite and the true parameter curve stays in a compact subset. Hence, the projected score is uniformly bounded over s, t and m. That is, there exists a constant $a_S > 0$ such that

$$\sup_{m,s,t}\left\|\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t))\right\|_2 \le a_S.$$

Define

$$Y_{m,s}(\boldsymbol{\xi}(u^t)) = \frac{1}{\sqrt{MTh}} K\left(\frac{u^s - u^t}{h}\right)\boldsymbol{\alpha}^{*\top}\left[\boldsymbol{\mathcal{P}}^\top S_{\ell,m}\big(\boldsymbol{\xi}(u^t)\big) - \mathbb{E}_{\boldsymbol{\xi}(u^s)}\left(\boldsymbol{\mathcal{P}}^\top S_{\ell,m}\big(\boldsymbol{\xi}(u^t)\big)\right)\right]. \quad \text{(S2.10)}$$

Since $K(u)$ is bounded, it follows that $\sup_{m,s}|Y_{m,s}(\boldsymbol{\xi}(u^t))| \le \frac{a_Y}{\sqrt{MTh}} \to 0$ for some constant $a_Y > 0$ if $MTh \to \infty$. This implies that for any fixed $\epsilon > 0$, there exists a sufficiently large MT such that the event $\{|Y_{m,s}(\boldsymbol{\xi}(u^t))| > \epsilon\}$ is empty, uniformly over m and s. Therefore, for any $\epsilon > 0$, the Lindeberg condition holds, i.e.,

$$\sum_{s=1}^{T}\sum_{m=1}^{M}\mathbb{E}\left[Y_{m,s}^2(\boldsymbol{\xi}(u^t))\mathbb{1}(|Y_{m,s}(\boldsymbol{\xi}(u^t))| > \epsilon)\right] \to 0.$$

It remains to compute the limiting variance. Since

$$\begin{aligned}
&\sum_{s=1}^{T}\sum_{m=1}^{M}\mathrm{Var}_{\boldsymbol{\xi}(u^s)}[Y_{m,s}(\boldsymbol{\xi}(u^t))] \\
&= \frac{1}{Th}\sum_{s=1}^{T}K^2\left(\frac{u^s-u^t}{h}\right)\frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}_{\boldsymbol{\xi}(u^s)}\left[\boldsymbol{\alpha}^{*\top}\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t))\right] \\
&= \frac{1}{Th}\sum_{s=1}^{T}K^2\left(\frac{u^s-u^t}{h}\right)\left\{\frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}_{\boldsymbol{\xi}(u)}\left[\boldsymbol{\alpha}^{*\top}\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u))\right] + o(1)\right\},
\end{aligned}$$

where the last equality holds since $\sup_{\{s:u^s-u^t\le h\}}\|\boldsymbol{\xi}(u^s)-\boldsymbol{\xi}(u)\|_2 = o(1)$ as $T\to\infty$ and $h\to 0$ such that $u^t \to u$. By Assumption 1 (vi), we obtain that

$$\frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}_{\boldsymbol{\xi}(u)}\left[\boldsymbol{\alpha}^{*\top}\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u))\right] = \boldsymbol{\alpha}^{*\top}\boldsymbol{\Upsilon}(u)\boldsymbol{\alpha}^* + o(1).$$

Thus, we have that

$$\begin{aligned}
&\sum_{s=1}^{T}\sum_{m=1}^{M}\mathrm{Var}\left[Y_{m,s}(\boldsymbol{\xi}(u^t))\right] \\
&= \frac{1}{Th}\sum_{s=1}^{T}K^2\left(\frac{u^s-u^t}{h}\right)\left\{\frac{1}{M}\sum_{m=1}^{M}\mathrm{Var}_{\boldsymbol{\xi}(u^s)}\left[\boldsymbol{\alpha}^{*\top}\boldsymbol{\mathcal{P}}^\top S_{\ell,m}(\boldsymbol{\xi}(u^t))\right]\right\} \\
&= \int K^2(u)\,\mathrm{d}u\left[\boldsymbol{\alpha}^{*\top}\boldsymbol{\Upsilon}(u)\boldsymbol{\alpha}^* + o(1)\right] \to \nu_{02}\,\boldsymbol{\alpha}^{*\top}\boldsymbol{\Upsilon}(u)\boldsymbol{\alpha}^*.
\end{aligned}$$

Together with the Lindeberg condition verified above, the Lindeberg-Feller central limit theorem yields

$$\sum_{s=1}^{T}\sum_{m=1}^{M} Y_{m,s}(\boldsymbol{\xi}(u^t)) \rightsquigarrow \mathcal{N}\left(0, \nu_{02}\,\boldsymbol{\alpha}^{*\top}\boldsymbol{\Upsilon}(u)\boldsymbol{\alpha}^{*}\right).$$

Since $\boldsymbol{\alpha}^{*} \in \mathbb{R}^{\dim(\Xi_{\mathcal{C}}^{\tan})}$ is arbitrary, combining the definition of $Y_{m,s}(\boldsymbol{\xi}(u^t))$ in (S2.10), the Cramér-Wold theorem implies that

$$\sqrt{MTh}\left\{\boldsymbol{\mathcal{P}}^{\top} S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big) - \mathbb{E}\left[\boldsymbol{\mathcal{P}}^{\top} S_{M,T,h}\big(\boldsymbol{\xi}(u^t)\big)\right]\right\} \rightsquigarrow \mathcal{N}\left(\mathbf{0}, \nu_{02}\,\boldsymbol{\Upsilon}(u)\right).$$

We now treat the expectation. Similarly to (S2.7), we have that

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\mathcal{P}}^{\top} S_{M,T,h}(\boldsymbol{\xi}(u^t))] &= \frac{1}{MTh}\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right)\boldsymbol{\mathcal{P}}^{\top}\left[\frac{\partial}{\partial \boldsymbol{\xi}(u^t)}\ell_m^{\mathbb{E}}(\boldsymbol{\xi}(u^t), u^s)\right] \\
&= \frac{1}{MTh}\sum_{m=1}^{M}\sum_{s=1}^{T} K\left(\frac{u^s-u^t}{h}\right)\boldsymbol{\mathcal{P}}^{\top} S_{\ell,m}^{\mathbb{E}}(u^s), \qquad \text{(S2.11)}
\end{aligned}$$

where $S_{\ell,m}^{\mathbb{E}}(u^s) = \frac{\partial}{\partial \boldsymbol{\xi}(u^t)}\ell_m^{\mathbb{E}}(\boldsymbol{\xi}(u^t), u^s)$. Under the smoothness condition in Assumption 1, the Taylor expansion of $S_{\ell,m}^{\mathbb{E}}(u^s)$ around u^t provides that

$$\begin{aligned}
S_{\ell,m}^{\mathbb{E}}(u^s) = S_{\ell,m}^{\mathbb{E}}(u^t) &+ \frac{\partial}{\partial u} S_{\ell,m}^{\mathbb{E}}(u)\Big|_{u=u^t}(u^s-u^t) \\
&+ \frac{1}{2}\,\frac{\partial^2}{\partial u^2} S_{\ell,m}^{\mathbb{E}}(u)\Big|_{u=u^{t^*}}(u^s-u^t)^2,
\end{aligned}$$

where u^{t^*} lies between u^s and u^t but may vary from term to term. Notice that

$$\frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\mathcal{P}}^{\top} S_{\ell,m}^{\mathbb{E}}(u^t) = \mathbf{0}$$

holds since $\boldsymbol{\xi}(u^t)$ satisfies the first-order condition for the constrained population criterion at time point u^t. Define $\boldsymbol{a}(u^t) = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\mathcal{P}}^{\top}\,\frac{\partial^2}{\partial u^2} S_{\ell,m}^{\mathbb{E}}(u)\Big|_{u=u^t}$. Substituting the Taylor expansion into (S2.11) therefore yields

$$\mathbb{E}[\boldsymbol{\mathcal{P}}^{\top} S_{M,T,h}(\boldsymbol{\xi}(u^t))] = \frac{\nu_{21}}{2}h^2\boldsymbol{a}(u^t) + o(h^2), \qquad \text{(S2.12)}$$

where the term involving the first-order derivative is negligible given the symmetry of the kernel function $K(u)$. □

Proof of Lemma S3. We first apply the Taylor expansion to $\mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}\right)$ around $\boldsymbol{\xi}(u^t)$, which implies that

$$\begin{aligned}
&\mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}\right) - \mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t)\right) \\
&= a_\epsilon r_{M,T,h}\boldsymbol{\alpha}^{\top}\boldsymbol{\mathcal{P}}^{\top} S_{M,T,h}(\boldsymbol{\xi}(u^t)) + \frac{a_\epsilon^2 r_{M,T,h}^2}{2}\boldsymbol{\alpha}^{\top}\boldsymbol{\mathcal{P}}^{\top} H_{M,T,h}(\bar{\boldsymbol{\zeta}}(u^t))\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}, \qquad \text{(S2.13)}
\end{aligned}$$

where $\bar{\boldsymbol{\zeta}}(u^t)$ lies on the line segment joining $\boldsymbol{\xi}(u^t)$ and $\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}$. To prove (S2.2), we need to control the projected score and the projected Hessian in (S2.13).

Step 1. Control of the score term. By the mean and variance calculations used in the proof of Lemma S2, it follows that

$$\left\|\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t))\right\|_2 = O_p(r_{M,T,h}).$$

Hence, for every $\epsilon > 0$, there exists a constant $a_\epsilon^* > 0$ such that $\left\|\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t))\right\|_2 \leq a_\epsilon^* r_{M,T,h}$ holds with probability at least $1-\epsilon$. By the Cauchy-Schwarz inequality and $\|\boldsymbol{\alpha}\|_2 = 1$, it holds that,

$$\begin{aligned}
&a_\epsilon r_{M,T,h}\boldsymbol{\alpha}^\top\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t)) \\
&\leq a_\epsilon r_{M,T,h}\|\boldsymbol{\alpha}\|_2\left\|\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t))\right\|_2 \\
&\leq a_\epsilon^* a_\epsilon r_{M,T,h}^2. \qquad \text{(S2.14)}
\end{aligned}$$

Step 2. Control of the Hessian term. By Assumption 1, the projected population information matrix

$$\boldsymbol{J}(u) = -\lim_{M\to\infty}\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\xi}(u))\boldsymbol{\mathcal{P}}]$$

is positive definite. Hence, $\lambda_{\min}(\boldsymbol{J}(u)) > 0$. It suffices to show that the projected empirical Hessian is uniformly close to its population counterpart in a neighbourhood of $\boldsymbol{\xi}(u^t)$. That is,

$$\left\|-\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\zeta}}(u^t))\boldsymbol{\mathcal{P}} - \boldsymbol{J}(u)\right\|_{\text{op}} = o_p(1),$$

as $u^t \to u$. For any fixed sufficiently small $\delta > 0$, by the same argument as in Step 1 of Lemma S1, applied componentwise to the Hessian entries, and by the boundedness and uniform Lipschitz continuity of the Hessian matrix, we have

$$\sup_{\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\leq\delta}\left\|\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}} - \mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}}]\right\|_{\text{op}} = o_p(1). \qquad \text{(S2.15)}$$

Moreover, applying the same Taylor expansion in the time argument u, as in Step 2 of Lemma S1, to each entry of the population Hessian matrix $\mathbb{E}[H_{\ell,m}(\boldsymbol{\zeta}(u^t)]$, we obtain that

$$\sup_{\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\leq\delta}\left\|\mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}}] - \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}}]\right\|_{\text{op}} = O(h^2). \qquad \text{(S2.16)}$$

By the continuity of the population Hessian matrix in $\boldsymbol{\zeta}(u^t)$, as $\delta \to 0$,

$$\sup_{\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\leq\delta}\left\|\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}}] - \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}[\boldsymbol{\mathcal{P}}^\top H_{\ell,m}(\boldsymbol{\xi}(u^t))\boldsymbol{\mathcal{P}}]\right\|_{\text{op}} = o(1). \qquad \text{(S2.17)}$$

Combining (S2.15)-(S2.17) and using the definition of $\boldsymbol{J}(u)$ gives that

$$\sup_{\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\leq\delta}\left\|-\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\boldsymbol{\zeta}(u^t))\boldsymbol{\mathcal{P}} - \boldsymbol{J}(u)\right\|_{\text{op}} = o_p(1).$$

Since $\bar{\boldsymbol{\zeta}}(u^t)$ lies on the line segment joining $\boldsymbol{\xi}(u^t)$ and $\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}$, and since $\|\boldsymbol{\alpha}\|_2 = 1$, we have

$$\|\bar{\boldsymbol{\zeta}}(u^t) - \boldsymbol{\xi}(u^t)\|_2 \leq a_\epsilon r_{M,T,h}\|\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}\|_2 \leq a_\epsilon r_{M,T,h}.$$

Since $r_{M,T,h} \to 0$ as $h \to 0$ and $MTh \to \infty$, it follows that, for any fixed sufficiently small $\delta > 0$, $\bar{\boldsymbol{\zeta}}(u^t)$ belongs to the neighbourhood $\{\boldsymbol{\zeta}(u^t) : \|\boldsymbol{\zeta}(u^t) - \boldsymbol{\xi}(u^t)\|_2 \leq \delta\}$ for all sufficiently large MT. Therefore,

$$\left\| -\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\zeta}}(u^t))\boldsymbol{\mathcal{P}} - \boldsymbol{J}(u) \right\|_{\text{op}} = o_p(1), \tag{S2.18}$$

which means that, with probability tending to one, it follows that

$$\lambda_{\min}(-\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\zeta}}(u^t))\boldsymbol{\mathcal{P}}) \geq \frac{1}{2}\lambda_{\min}(\boldsymbol{J}(u)).$$

Thus, uniformly over $\{\boldsymbol{\alpha} : \|\boldsymbol{\alpha}\|_2 = 1\}$, it follows that

$$\boldsymbol{\alpha}^\top\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\zeta}}(u^t))\boldsymbol{\mathcal{P}}\boldsymbol{\alpha} \leq -\frac{1}{2}\lambda_{\min}(\boldsymbol{J}(u)). \tag{S2.19}$$

Substituting (S2.14) and (S2.19) into (S2.13), with probability tending to one,

$$\begin{aligned}&\mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t) + a_\epsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\alpha}\right) - \mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t)\right)\\ &\leq a_\epsilon^* a_\epsilon r_{M,T,h}^2 - \frac{a_\epsilon^2}{4} r_{M,T,h}^2 \lambda_{\min}(\boldsymbol{J}(u))\end{aligned} \tag{S2.20}$$

holds uniformly over $\{\boldsymbol{\alpha} : \|\boldsymbol{\alpha}\|_2 = 1\}$. Setting $a_\epsilon > 4a^*/\lambda_{\min}(\boldsymbol{J}(u))$, then the right-hand side of (S2.20) is strictly negative. □

S2.3 Proof of Theorem 1: Consistency

Proof of Theorem 1. By the definition of the smooth estimator in (2), the latter is equivalent to

$$\tilde{\boldsymbol{\xi}}_h(u^t) = \underset{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}}{\arg\max}\, \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)), \quad u^t \in \{1/T, \ldots, 1\}. \tag{S2.21}$$

To show the consistency of $\tilde{\boldsymbol{\xi}}_h(u^t)$, we first identify the uniqueness of the maximizer of the population criterion. For model (1), the difference between the population marginal log-likelihoods evaluated at a candidate parameter $\boldsymbol{\zeta}(u^t)$ and the true parameter $\boldsymbol{\xi}(u^t)$ is equal to the negative Kullback–Leibler divergence between the corresponding marginal distributions. More precisely, for any $\boldsymbol{\zeta}(u^t) \in \Xi_{\boldsymbol{C}}$,

$$\mathbb{L}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}(\boldsymbol{\xi}(u^t)) = -\frac{1}{M}\sum_{m=1}^{M} \text{KL}\left(\mathbb{P}_{\boldsymbol{\xi}(u^t),m} | \mathbb{P}_{\boldsymbol{\zeta}(u^t),m}\right) \leq 0,$$

where $\mathbb{P}_{\boldsymbol{\xi}(u^t),m}$ denotes the true marginal distribution of $\boldsymbol{A}_m(u^t)$ after integrating out the random effect $\omega_m(u^t)$, and $\mathbb{P}_{\boldsymbol{\zeta}(u^t),m}$ is the corresponding marginal distribution under the candidate parameter. Here, equality holds only when the two marginal distributions coincide. Under Proposition 1 and Assumption 1, this implies

$$\mathbb{L}(\boldsymbol{\zeta}(u^t)) = \mathbb{L}(\boldsymbol{\xi}(u^t)) \quad \Longleftrightarrow \quad \boldsymbol{\zeta}(u^t) = \boldsymbol{\xi}(u^t).$$

Thus, $\boldsymbol{\xi}(u^t)$ is the unique maximizer of the population criterion.

Next, we show the convergence of $\tilde{\boldsymbol{\xi}}_h(u^t)$. Since $\Xi_{\boldsymbol{C}}$ is compact and $\mathbb{L}(\boldsymbol{\zeta}(u^t))$ is continuous, the uniqueness of the maximizer implies that there always exists $\epsilon_\delta > 0$ such that

$$\max_{\{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}:\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\geq\delta\}} \mathbb{L}(\boldsymbol{\zeta}(u^t)) \leq \mathbb{L}(\boldsymbol{\xi}(u^t)) - 2\epsilon_\delta. \tag{S2.22}$$

As defined in (S2.21), $\tilde{\boldsymbol{\xi}}_h(u^t)$ is the maximizer of $\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))$ over $\Xi_{\boldsymbol{C}}$, which means

$$\begin{aligned}
&\mathbb{P}\left(\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\|_2 \geq \delta\right) \\
&\leq \mathbb{P}\left(\max_{\{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}:\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\geq\delta\}} \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) \geq \mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t))\right).
\end{aligned} \tag{S2.23}$$

Using (S2.22), the probability on the right-hand side of (S2.23) is bounded by

$$\begin{aligned}
&\mathbb{P}\left(\max_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}:\|\boldsymbol{\zeta}(u^t)-\boldsymbol{\xi}(u^t)\|_2\geq\delta} \mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t)) \geq 0\right) \\
&\leq \mathbb{P}\left(\sup_{\boldsymbol{\zeta}(u^t)\in\Xi_{\boldsymbol{C}}} \left|\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t)) - \mathbb{L}(\boldsymbol{\zeta}(u^t))\right| > \epsilon_\delta\right) \\
&\quad + \mathbb{P}\left(\left|\mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t)) - \mathbb{L}(\boldsymbol{\xi}(u^t))\right| > \epsilon_\delta\right).
\end{aligned} \tag{S2.24}$$

By Lemma S1, both terms on the right-hand side of (S2.24) converge to zero. Thus,

$$\mathbb{P}\left(\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\|_2 \geq \delta\right) \longrightarrow 0, \tag{S2.25}$$

which means that the smooth estimator $\tilde{\boldsymbol{\xi}}_h(u^t)$ converges to its true value $\boldsymbol{\xi}(u^t)$ in probability.

It remains to derive the rate. Since $\tilde{\boldsymbol{\xi}}_h(u^t)$ is in an arbitrarily small neighbourhood of $\boldsymbol{\xi}(u^t)$ with probability tending to one, and also since the constraints defining $\Xi_{\boldsymbol{C}}$ are linear in this neighbourhood, there exists a vector $\tilde{\boldsymbol{\vartheta}}(u^t) \in \mathbb{R}^{\dim(\Xi_{\boldsymbol{C}}^{\tan})}$ such that

$$\tilde{\boldsymbol{\xi}}_h(u^t) = \boldsymbol{\xi}(u^t) + \boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\vartheta}}(u^t), \quad \|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\|_2 = \|\tilde{\boldsymbol{\vartheta}}(u^t)\|_2.$$

By Lemma S3, it holds that, for any $\epsilon > 0$, there exists a sufficiently large constant $a_\epsilon > 0$ such that

$$\mathbb{P}\left[\sup_{\|\boldsymbol{\vartheta}\|_2=1} \mathbb{L}_{M,T,h}\left(\boldsymbol{\xi}(u^t) + a_\varepsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\vartheta}\right) < \mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t))\right] \geq 1 - \varepsilon$$

for all sufficiently large M and T. Moreover, the Hessian expansion used in Lemma S3 and Assumption 1 (vi) imply that, in the local neighbourhood of $\boldsymbol{\xi}(u^t)$, the projected sample criterion is strictly concave with probability tending to one. Hence, if $\|\tilde{\boldsymbol{\vartheta}}(u^t)\|_2 > a_\varepsilon r_{M,T,h}$, then the line segment joining $\boldsymbol{\xi}(u^t)$ and $\tilde{\boldsymbol{\xi}}_h(u^t)$ intersects the boundary $\{\boldsymbol{\xi}(u^t) + a_\varepsilon r_{M,T,h}\boldsymbol{\mathcal{P}}\boldsymbol{\vartheta} : \|\boldsymbol{\vartheta}\|_2 = 1\}$. By concavity of $\mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t) + \boldsymbol{\mathcal{P}}\boldsymbol{\vartheta})$ and the definition of $\tilde{\boldsymbol{\xi}}_h(u^t)$ in (S2.21), the criterion value at this boundary point would have to be at least $\mathbb{L}_{M,T,h}(\boldsymbol{\xi}(u^t))$, contradicting Lemma S3. Therefore, it holds that, for all sufficiently large M and T,

$$\mathbb{P}\left(\|\tilde{\boldsymbol{\vartheta}}(u^t)\|_2 \leq a_\varepsilon r_{M,T,h}\right) \geq 1 - \varepsilon.$$

Finally, using the orthonormality of $\boldsymbol{\mathcal{P}}$, we have that, at any given time point u^t,

$$\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\|_2 = O_p\left(h^2 + (MTh)^{-1/2}\right).$$

Moreover, by the smoothness assumption, if $u^t \to u$ as $T \to \infty$, it holds that

$$\|\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u)\|_2 = O_p\left(h^2 + (MTh)^{-1/2}\right).$$

□

S2.4 Proof of Theorem 2: Asymptotic normality

Proof of Theorem 2. By the consistency in Theorem 1, it is clear that $\tilde{\boldsymbol{\xi}}_h(u^t)$ lies in a shrinking neighbourhood of $\boldsymbol{\xi}(u^t)$ with probability tending to one. Hence, on this event, it admits the unique representation

$$\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t) = \boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\vartheta}}_h(u^t),$$

where $\tilde{\boldsymbol{\vartheta}}_h(u^t) = O_p(r_{M,T,h})$. Since $\tilde{\boldsymbol{\xi}}_h(u^t)$ is the maximizer of the constrained criterion $\mathbb{L}_{M,T,h}(\boldsymbol{\zeta}(u^t))$, it satisfies the projected first-order condition,

$$\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\tilde{\boldsymbol{\xi}}_h(u^t)) = \mathbf{0}. \tag{S2.26}$$

Applying Taylor expansion to the left-hand side of (S2.26) around $\boldsymbol{\xi}(u^t)$ yields

$$\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t)) + \boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\xi}}_h(u^t))\boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\vartheta}}_h(u^t) = \mathbf{0},$$

where $\bar{\boldsymbol{\xi}}_h(u^t)$ lies on the line segment joining $\tilde{\boldsymbol{\xi}}_h(u^t)$ and $\boldsymbol{\xi}(u^t)$. By Theorem 1, we have that $\|\bar{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t)\|_2 = O_p(r_{M,T,h})$. It then follows from (S2.18) that

$$\boldsymbol{\mathcal{P}}^\top H_{M,T,h}(\bar{\boldsymbol{\xi}}_h(u^t))\boldsymbol{\mathcal{P}} = -\boldsymbol{J}(u^t) + o_p(1).$$

Consequently,

$$\tilde{\boldsymbol{\vartheta}}_h(u^t) = \boldsymbol{J}^{-1}(u^t)\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t)) + o_p(r_{M,T,h}). \tag{S2.27}$$

Recall from Lemma S2 that

$$\sqrt{MTh}\left[\boldsymbol{\mathcal{P}}^\top S_{M,T,h}(\boldsymbol{\xi}(u^t)) - \frac{\nu_{21}}{2}h^2\boldsymbol{a}(u^t)\right] \rightsquigarrow \mathcal{N}(0, \nu_{02}\boldsymbol{\Upsilon}(u)).$$

Moreover, under the conditions in Theorem 2,

$$\sqrt{MTh}\, o_p(r_{M,T,h}) = o_p(1).$$

Since $u^t \to u$ and $\boldsymbol{J}(\cdot)$ is continuous, by Slutsky's theorem, we have

$$\sqrt{MTh}\left[\tilde{\boldsymbol{\vartheta}}_h(u^t) - \frac{\nu_{21}}{2}h^2\boldsymbol{J}^{-1}(u^t)\boldsymbol{a}(u^t)\right] \rightsquigarrow \mathcal{N}(\mathbf{0}, \nu_{02}\boldsymbol{J}^{-1}(u)\boldsymbol{\Upsilon}(u)\boldsymbol{J}^{-1}(u)).$$

Finally, transforming back from the intrinsic coordinates to the original constrained parameter space through $\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t) = \boldsymbol{\mathcal{P}}\tilde{\boldsymbol{\vartheta}}_h(u^t)$ gives the following CLT,

$$\sqrt{MTh}\left[\tilde{\boldsymbol{\xi}}_h(u^t) - \boldsymbol{\xi}(u^t) - \frac{\nu_{21}}{2}h^2\boldsymbol{\mathcal{P}}\boldsymbol{J}^{-1}(u^t)\boldsymbol{a}(u^t)\right] \rightsquigarrow \mathcal{N}(\mathbf{0}, \nu_{02}\boldsymbol{\mathcal{P}}\boldsymbol{J}^{-1}(u)\boldsymbol{\Upsilon}(u)\boldsymbol{J}^{-1}(u)\boldsymbol{\mathcal{P}}^\top).$$

□

S3. Derivative expressions

S3.1 Derivatives for the time-varying mixed-effects DCSBM

In this section, we provide the preliminary derivations for the proposed LA-EM algorithm developed in Section 3, including analytical expressions for the gradient vector and second derivative matrix of the local Q-function with respect to the model parameters. Recall that the terms in the approximation to the local Q-function that depend on $\boldsymbol{\theta}(u^t)$ and $\boldsymbol{Z}(u^t)$ are given by $Q_{h,1}$ and $Q_{h,2}$, where

$$
\begin{aligned}
Q_{h,1} &= \sum_{s=1}^{T} K_h(u^s - u^t) \left(\sum_{i=1}^{n} d_i(u^s)\, \theta_i(u^t) + \frac{1}{2} \sum_{g,l=1}^{G} O_{gl}(u^s)\, Z_{gl}(u^t) \right) \\
Q_{h,2} &= -\sum_{m=1}^{M} \sum_{1 \le i < j \le n} \sum_{s=1}^{T} K_h(u^s - u^t) \int_{\mathbb{R}} \log\left[1 + \exp\left(\Theta_{ij}(u^t) + \omega_m(u^s)\right)\right] \\
&\quad \times \phi\left(\omega_m(u^s); \hat{\omega}_m(u^s),\, \hat{v}_m^{-1}(u^s)\right) \mathrm{d}\,\omega_m(u^s),
\end{aligned}
$$

with $\Theta_{ij}(u^t) = \theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t)$. To simplify notation, for any $i, j = 1, \ldots, n$, we define the following two quantities,

$$
\begin{aligned}
\mathbb{M}_{ij,m}(u^s, u^t) &= \int_{\mathbb{R}} \frac{\exp(\theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t) + \omega_m(u^s))}{1 + \exp(\theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t) + \omega_m(u^s))} \\
&\quad \times \phi(\omega_m(u^s); \hat{\omega}_m(u^s), \hat{v}_m^{-1}(u^s))\, \mathrm{d}\,\omega_m(u^s), \\
\mathbb{V}_{ij,m}(u^s, u^t) &= \int_{\mathbb{R}} \frac{\exp(\theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t) + \omega_m(u^s))}{[1 + \exp(\theta_i(u^t) + \theta_j(u^t) + Z_{c_i c_j}(u^t) + \omega_m(u^s))]^2} \\
&\quad \times \phi(\omega_m(u^s); \hat{\omega}_m(u^s), \hat{v}_m^{-1}(u^s))\, \mathrm{d}\,\omega_m(u^s)\,.
\end{aligned}
$$

With $\boldsymbol{Z}(u^t)$ fixed at its current estimate $\tilde{\boldsymbol{Z}}_h^{(r)}(u^t)$, we substitute it into $\mathbb{M}_{ij,m}(u^s, u^t)$ and $\mathbb{V}_{ij,m}(u^s, u^t)$, and, with a slight abuse of notation, we still denote the resulting quantities by $\mathbb{M}_{ij,m}(u^s, u^t)$ and $\mathbb{V}_{ij,m}(u^s, u^t)$. Writing $Q_h(\boldsymbol{\theta}(u^t)) = Q_{h,1} + Q_{h,2}$ as a function of $\boldsymbol{\theta}(u^t)$, the gradient vector $S_{Q,h}(\boldsymbol{\theta}(u^t))$ is defined as

$$
S_{Q,h}(\boldsymbol{\theta}(u^t)) = \left(S_{Q,h,1}(\boldsymbol{\theta}(u^t)), \ldots, S_{Q,h,n}(\boldsymbol{\theta}(u^t))\right)^\top.
$$

By the chain rule, the i-th component of $S_{Q,h}(\boldsymbol{\theta}(u^t))$ can be expressed as

$$
\begin{aligned}
S_{Q,h,i}(\boldsymbol{\theta}(u^t)) &= \frac{\partial Q_{h,1}}{\partial \theta_i(u^t)} + \frac{\partial Q_{h,2}}{\partial \theta_i(u^t)} \\
&= \sum_{s=1}^{T} K_h(u^s - u^t) \left[d_i(u^s) - \sum_{m=1}^{M} \sum_{j \neq i} \mathbb{M}_{ij,m}(u^s, u^t) \right].
\end{aligned} \tag{S3.1}
$$

Moreover, the entries of the second derivative matrix

$$
H_{Q,h}(\boldsymbol{\theta}(u^t)) = \left(H_{Q,h,ij}(\boldsymbol{\theta}(u^t))\right)_{1 \le i,j \le n},
$$

satisfy

$$H_{Q,h,ij}(\boldsymbol{\theta}(u^t)) = \begin{cases} -\sum_{m=1}^{M}\sum_{s=1}^{T} K_h(u^s-u^t)\sum_{k\neq i}\mathbb{V}_{ik,m}(u^s,u^t), & \text{if} \quad i=j, \\ -\sum\limits_{m=1}^{M}\sum\limits_{s=1}^{T} K_h(u^s-u^t)\mathbb{V}_{ij,m}(u^s,u^t), & \text{if} \quad i\neq j. \end{cases} \tag{S3.2}$$

Similarly, for derivatives with respect to $\boldsymbol{Z}(u^t)$, $\mathbb{M}_{ij,m}(u^s,u^t)$ and $\mathbb{V}_{ij,m}(u^s,u^t)$ are understood to be evaluated at $\boldsymbol{\theta}(u^t) = \tilde{\boldsymbol{\theta}}_h^{(r)}(u^t)$, again with a slight abuse of notation. Since $\boldsymbol{Z}(u^t)$ is a $G\times G$ matrix, we consider the vectorization of $\boldsymbol{Z}(u^t)$, denoted by $\mathrm{vec}(\boldsymbol{Z}(u^t))$. Accordingly, we define the gradient vector with respect to $\mathrm{vec}(\boldsymbol{Z}(u^t))$ by

$$S_{Q,h}(\boldsymbol{Z}(u^t)) = \nabla_{\mathrm{vec}(\boldsymbol{Z}(u^t))}(Q_{h,1}+Q_{h,2}) \in \mathbb{R}^{G^2\times 1}.$$

Its component associated with $Z_{gl}(u^t)$ is then given by

$$S_{Q,h,gl}(\boldsymbol{Z}(u^t)) = \frac{\partial Q_{h,1}}{\partial Z_{gl}(u^t)} + \frac{\partial Q_{h,2}}{\partial Z_{gl}(u^t)}.$$

Hence, for $g,l=1,\ldots,G$, we have that

$$S_{Q,h,gl}(\boldsymbol{Z}(u^t)) = \frac{1}{2}\sum_{s=1}^{T} K_h(u^s-u^t)\left[O_{gl}(u^s) - \sum_{m=1}^{M}\sum_{\substack{i,j=1,\\ j\neq i}}^{n} \mathbb{1}(c_i=g,c_j=l)\mathbb{M}_{ij,m}(u^s,u^t)\right]. \tag{S3.3}$$

Regarding the second derivative, observe that

$$\begin{aligned}\frac{\partial^2(Q_{h1}+Q_{h2})}{\partial Z_{gl}(u^t)\,\partial Z_{g'l'}(u^t)} = -\frac{1}{2}\sum_{s=1}^{T} K_h(u^s-u^t)\sum_{m=1}^{M}\sum_{\substack{i,j=1,\\ j\neq i}}^{n}\mathbb{1}(c_i=g,c_j=l) \\ \times\,\mathbb{1}(c_i=g',c_j=l')\,\mathbb{V}_{ij,m}(u^s,u^t).\end{aligned}$$

Unless $(g,l)=(g',l')$, the indicator product degenerates to zero. Therefore, under the full vectorization of $\boldsymbol{Z}(u^t)$, the second derivative matrix $H_{Q,h}(\boldsymbol{Z}(u^t))$ is a diagonal matrix where the diagonal entries are given by

$$\frac{\partial^2 Q_h}{\partial Z_{gl}^2(u^t)} = -\frac{1}{2}\sum_{s=1}^{T} K_h(u^s-u^t)\sum_{m=1}^{M}\sum_{\substack{i,j=1,\\ j\neq i}}^{n}\mathbb{1}(c_i=g,c_j=l)\mathbb{V}_{ij,m}(u^s,u^t), \quad g,l=1,\ldots,G. \tag{S3.4}$$

S3.2 Derivatives for the time-varying βBM

Recall the model setup in Section 4. To distinguish this model from our model (1), we write $\boldsymbol{\tau}^{\beta\mathrm{BM}}(u^t)$ for its edge probability matrix at time point u^t. Under the βBM, the edge probability $\tau_{ij}^{\beta\mathrm{BM}}(u^t)$ is shared across subjects and is modeled as

$$\tau_{ij}^{\beta\mathrm{BM}}(u^t) = \frac{\exp(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t))}{1+\exp(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t))}, \qquad i,j=1,\ldots,n.$$

Recalling that $\boldsymbol{A}(u^t) = \{\boldsymbol{A}_m(u^t)\}_{1\le m\le M}$, the log-likelihood at time point u^t is

$$
\begin{aligned}
&\mathcal{L}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t), \boldsymbol{Z}(u^t); \boldsymbol{A}(u^t)\big) \\
=&\sum_{m=1}^{M}\sum_{1\le i<j\le n} A_{m,ij}(u^t)\left[\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)\right] \\
&-\sum_{m=1}^{M}\sum_{1\le i<j\le n}\log\left[1+\exp\big(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)\big)\right] \\
=&\sum_{i=1}^{n}\left(\sum_{m=1}^{M}\sum_{j=1}^{n}A_{m,ij}(u^t)\right)\theta_i(u^t)+\frac{1}{2}\sum_{g,l=1}^{G}\left(\sum_{m=1}^{M}\sum_{i,j=1}^{n}A_{m,ij}(u^t)\right)\mathbb{1}(c_i=g,c_j=l)Z_{gl}(u^t) \\
&-\sum_{m=1}^{M}\sum_{1\le i<j\le n}\log\left[1+\exp\big(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)\big)\right] \\
=&\sum_{i=1}^{n}d_i(u^t)\theta_i(u^t)+\frac{1}{2}\sum_{g,l=1}^{G}O_{gl}(u^t)Z_{gl}(u^t) \\
&-\sum_{m=1}^{M}\sum_{1\le i<j\le n}\log\left[1+\exp\big(\theta_i(u^t)+\theta_j(u^t)+Z_{c_ic_j}(u^t)\big)\right].
\end{aligned}
$$

To present the Newton-Raphson updates in Algorithm 2, we next derive below the score vector and Hessian matrices with respect to $\boldsymbol{\theta}(u^t)$ and $\boldsymbol{Z}(u^t)$.

Letting $S^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big) = \Big(S_1^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big),\ldots,S_n^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big)\Big)^\top$ be the score vector, the i-th component is defined by

$$
S_i^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big) = \frac{\partial\mathcal{L}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t),\boldsymbol{Z}(u^t);\boldsymbol{A}(u^t)\big)}{\partial\theta_i(u^t)} = d_i(u^t) - M\sum_{j\neq i}\tau_{ij}^{\beta\mathrm{BM}}(u^t). \tag{S3.5}
$$

Moreover, we write

$$
H^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big) = \Big(H_{ij}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big)\Big)_{1\le i,j\le n},
$$

as the Hessian matrix where the entries satisfy

$$
\begin{aligned}
H_{ij}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t)\big) &= \frac{\partial^2\mathcal{L}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t),\boldsymbol{Z}(u^t);\boldsymbol{A}(u^t)\big)}{\partial\theta_i(u^t)\,\partial\theta_j(u^t)} \\
&= \begin{cases} -M\sum_{k\neq i}\tau_{ik}^{\beta\mathrm{BM}}(u^t)\big[1-\tau_{ik}^{\beta\mathrm{BM}}(u^t)\big], & \text{if } i=j, \\ -M\tau_{ij}^{\beta\mathrm{BM}}(u^t)\big[1-\tau_{ij}^{\beta\mathrm{BM}}(u^t)\big], & \text{if } i\neq j. \end{cases}
\end{aligned} \tag{S3.6}
$$

We next give the derivatives with respect to $\boldsymbol{Z}(u^t)$. Again, we consider the vectorization of $\boldsymbol{Z}(u^t)$. Analogously to (S3.3) and (S3.4), the score vector is

$$
S^{\beta\mathrm{BM}}(\boldsymbol{Z}(u^t)) = \nabla_{\mathrm{vec}(\boldsymbol{Z}(u^t))}\mathcal{L}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t),\boldsymbol{Z}(u^t);\boldsymbol{A}(u^t)\big) \in \mathbb{R}^{G^2\times 1}
$$

where

$$S_{gl}^{\beta\mathrm{BM}}(\boldsymbol{Z}(u^t)) = \frac{1}{2}\left[O_{gl}(u^t) - M\sum_{\substack{i,j=1,\\ j\neq i}}^{n} \mathbb{1}(c_i = g, c_j = l)\tau_{ij}^{\beta\mathrm{BM}}(u^t)\right]. \tag{S3.7}$$

And the Hessian matrix $H^{\beta\mathrm{BM}}(\boldsymbol{Z}(u^t))$ has a diagonal form where the diagonal entries for any pair (g, l) are

$$\frac{\partial^2 \mathcal{L}^{\beta\mathrm{BM}}\big(\boldsymbol{\theta}(u^t), \boldsymbol{Z}(u^t); \boldsymbol{A}(u^t)\big)}{\partial Z_{gl}^2(u^t)} = -\frac{M}{2}\sum_{\substack{i,j=1,\\ j\neq i}}^{n} \mathbb{1}(c_i = g, c_j = l)\tau_{ij}^{\beta\mathrm{BM}}(u^t)[1 - \tau_{ij}^{\beta\mathrm{BM}}(u^t)]. \tag{S3.8}$$

S4. Bandwidth selection

As in most nonparametric procedures, the choice of bandwidth h is both important and nontrivial in practice. To this end, we proposed a leave-one-out cross-validation strategy that yields a data-adaptive bandwidth selection. Such methods are widely used for selecting tuning parameters in many fields (Härdle et al., 1988), and also work well in our simulation and data analysis. For computational convenience, we prefer selecting a common bandwidth h for all time points when a smoothing assumption is imposed on the parameter trajectories as in Assumption 1.

Let $\mathcal{H} = \{h_1, \ldots, h_L\}$ be a grid of candidate bandwidths. Fixing a target time point $u^t \in \{1/T, \ldots, 1\}$, let $\mathcal{T}^{(-t)} = \{1, \ldots, T\} \setminus \{t\}$ denote the training index set, and let $\mathcal{A}^{(-t)} = \{\boldsymbol{A}_m(u^{t'}) : m = 1, \ldots, M, t' \in \mathcal{T}^{(-t)}\}$ be the corresponding training sample consisting of all networks excluding the networks observed at time u^t. To implement leave-one-out fitting, we exclude all contributions associated with the time point u^t from the local criterion in (4), that is, we replace $\sum_{s=1}^{T}$ with $\sum_{s \in \mathcal{T}^{(-t)}}$ in both the LAE-step and the LAM-step. Given any candidate bandwidth $h \in \mathcal{H}$, fitting model (1) to the training sample $\mathcal{A}^{(-t)}$ yields the leave-one-out estimator at time point u^t, denoted by

$$\tilde{\boldsymbol{\xi}}_h^{(-t)}(u^t) = \left(\tilde{\boldsymbol{\theta}}_h^{(-t)}(u^t),\, \tilde{\boldsymbol{Z}}_h^{(-t)}(u^t),\, \tilde{\sigma}_h^{2,(-t)}(u^t)\right).$$

Based on these estimators, we can evaluate the predictive performance of the candidate h via the predictive marginal log-likelihood of each subject-specific observation $\boldsymbol{A}_m(u^t)$. Aggregating these contributions over all subjects and time points leads to the cross-validation score

$$\mathrm{CV}(h) = \frac{1}{MT} \sum_{t=1}^{T} \sum_{m=1}^{M} \log \mathbb{P}(\boldsymbol{A}_m(u^t);\, \tilde{\boldsymbol{\xi}}_h^{(-t)}(u^t)),$$

and thus, the selected bandwidth is defined as

$$\hat{h}_{\mathrm{opt}} \;=\; \arg\max_{h \in \mathcal{H}} \mathrm{CV}(h).$$

In practice, this marginal likelihood, which entails only a one-dimensional integral, can be efficiently evaluated by using the Laplace approximation or the (adaptive) Gauss-Hermite quadrature methods.

S5. Algorithmic details

In Section 3.2 of the main text, we introduced the local approximate EM (LA-EM) algorithm for computing the proposed (time-localised likelihood) smooth estimator. This section provides additional implementation details and pseudocode.

S5.1 Initialization

Considering that EM-type algorithms may be sensitive to initialization values, we construct a two-step initialization procedure that first estimates the fixed effects from a simplified model and then initializes the variance via a one-step update.

Step 1 (Fixed-effects initialization). We first suppress the subject-specific random intercepts by imposing $\omega_m(u^t) = 0, \forall m = 1, \ldots, M,\ \forall u^t = 1/T, \ldots, 1$. For each subject m, we aggregate the observed networks across time and form a single adjacency matrix $\bar{A}_m = (\bar{A}_{m,ij})_{1 \le i,j \le n}$ by majority vote, namely,

$$\bar{A}_{m,ij} = \mathbb{1}\left(\sum_{t=1}^{T} A_{m,ij}(u^t) \ge \frac{T}{2}\right), \qquad 1 \le i < j \le n,$$

with $\bar{A}_{m,ii} = 0$ and $\bar{A}_{m,ij} = \bar{A}_{m,ji}$. We then fit a βBM to $\{\bar{A}_m\}_{1 \le m \le M}$ and apply the Newton-Raphson algorithm in Algorithm 2 to obtain estimates $\tilde{\boldsymbol{\theta}}^0$ and $\tilde{\boldsymbol{Z}}^0$, which are used as initial values for all time points. Since this stage is used solely to provide a stable warm start for the subsequent LA-EM iterations, we cap the Newton-Raphson algorithm at three iterations; in our simulation studies, this is sufficient.

Step 2 (Variance initialization). Given $\tilde{\boldsymbol{\theta}}^{(0)}$ and $\tilde{\boldsymbol{Z}}^{(0)}$ from Step 1, we can compute the posterior mode $\hat{\omega}_m(u^t)$ and the local curvature $\hat{v}_m(u^t)$ for each m and u^t, according to (6) and (7). Substituting these quantities into (12) gives,

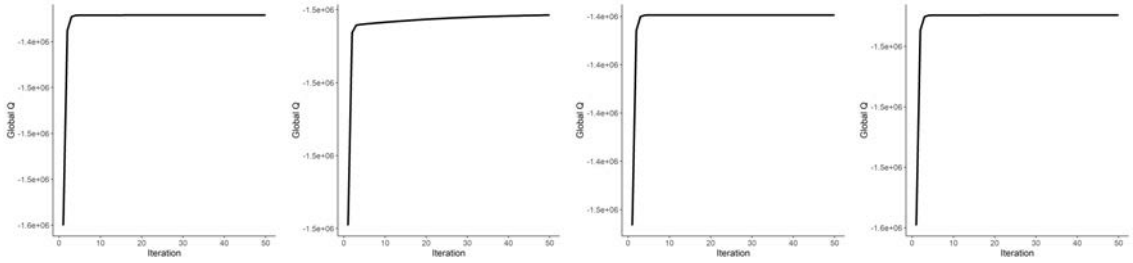
$$\tilde{\sigma}^{2,(0)}(u^t) = \frac{1}{M} \sum_{m=1}^{M} \left\{\hat{\omega}_m^2(u^t) + \hat{v}_m^{-1}(u^t)\right\},$$

which can be considered as the initial value at time point u^t.

S5.2 Convergence behaviour

In this section, we study the empirical convergence behaviour of the proposed LA-EM algorithm under the four simulation scenarios we considered in the main text. For each iteration, we compute the global approximate objective by summing the local Q-functions over all time points, denoted by 'Global Q'. Figure S5.1 reports representative trajectories of Global Q across all scenarios with $T = 120$. It shows that the global objective stabilizes within a few iterations.

S5.3 Pseudocode



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3 (d) Scenario 4

Figure S5.1: Convergence behaviour of the proposed LA-EM algorithm across iterations for the four simulation scenarios in the main text with $T = 120$.

Algorithm 1 Newton–Raphson algorithm for computing the posterior mode $\hat{\omega}_m(u^t)$

Input: Subject-specific adjacency matrix $\boldsymbol{A}_m(u^t)$, parameter estimate $\Theta(u^t)$, variance estimate $\sigma^2(u^t)$, convergence threshold ϵ, step size η, maximum iteration number R_ω.
Initialize: $\omega^{(0)} = 0$, $r = 0$, and $\Delta_\omega^{(0)} = \infty$.

1. **while** $r < R_\omega$, $\Delta_\omega^{(r)} > \epsilon$ **do**
2. Compute the first derivative:

$$S(\omega^{(r)}) = \sum_{1 \le i < j \le n} \left[A_{m,ij}(u^t) - \frac{\exp(\Theta_{ij}(u^t) + \omega^{(r)})}{1 + \exp(\Theta_{ij}(u^t) + \omega^{(r)})} \right] - \frac{\omega^{(r)}}{\sigma^2(u^t)}.$$

3. Compute the second derivative:

$$H(\omega^{(r)}) = - \sum_{1 \le i < j \le n} \frac{\exp(\Theta_{ij}(u^t) + \omega^{(r)})}{\left[1 + \exp(\Theta_{ij}(u^t) + \omega^{(r)})\right]^2} - \frac{1}{\sigma^2(u^t)}$$

4. Update $\omega^{(r+1)} = \omega^{(r)} - \eta S(\omega^{(r)})(H(\omega^{(r)}))^{-1}$.
5. Compute the difference $\Delta_\omega^{(r+1)} = |\omega^{r+1} - \omega^r|$.
6. Set $r = r + 1$.
7. **end while**
8. **return** $\hat{\omega}_m(u^t) = \omega^{(r+1)}$.

Algorithm 2 Newton-Raphson algorithm for the βBM

Input: Adjacency matrices $\{\boldsymbol{A}_m(u^t)\}_{\substack{1\le m\le M\\ 1\le t\le T}}$, community memberships $\boldsymbol{C}$, convergence thresholds ϵ_p and ϵ_l, step sizes η_θ and η_Z, maximum iterations R.

Initialize: $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}^{(0)}(u^1), \ldots, \boldsymbol{\theta}^{(0)}(u^T))^\top$, $\boldsymbol{Z}^{(0)} = (\boldsymbol{Z}^0(u^1), \ldots, \boldsymbol{Z}^{(0)}(u^T))$, $\mathcal{L}^{(0)}(u^1) = \ldots = \mathcal{L}^{(0)}(u^T) = -\infty$, $\Delta_p^{(0)} = \Delta_{\mathcal{L}}^{(0)} = +\infty$, $r = 0$.

1. **for** $t = 1, \ldots, T$ **do**
2. **while** $r < R$, $\Delta_p^{(r)} > \epsilon_p$, $\Delta_{\mathcal{L}}^{(r)} > \epsilon_l$ **do**
3. Compute $S^{\beta\text{BM}}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)$ based on (S3.5) and $H^{\beta\text{BM}}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)$ based on (S3.6).
4. Compute

$$\begin{aligned}\delta_{\boldsymbol{\theta}}(u^t) = &-(H^{\beta\text{BM}})^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big) S^{\beta\text{BM}}\big(\boldsymbol{\theta}^{(r)}(u^t)\big) + (H^{\beta\text{BM}})^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)\boldsymbol{C}\\ &\times \Big(\boldsymbol{C}^\top (H^{\beta\text{BM}})^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)\boldsymbol{C}\Big)^{-1}\boldsymbol{C}^\top (H^{\beta\text{BM}})^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big) S^{\beta\text{BM}}\big(\boldsymbol{\theta}^{(r)}(u^t)\big).\end{aligned}$$

5. Update $\boldsymbol{\theta}^{(r+1)}(u^t) = \boldsymbol{\theta}^{(r)}(u^t) + \eta_{\boldsymbol{\theta}}\, \delta_{\boldsymbol{\theta}}(u^t)$.
6. Compute $S^{\beta\text{BM}}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$ based on (S3.7) and $H^{\beta\text{BM}}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$ based on (S3.8).
7. Compute $\delta_{\boldsymbol{Z}}(u^t) = -(H^{\beta\text{BM}})^{-1}\big(\boldsymbol{Z}^{(r)}(u^t)\big) S^{\beta\text{BM}}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$.
8. Update $\boldsymbol{Z}^{(r+1)}(u^t) = \boldsymbol{Z}^{(r)}(u^t) + \eta_{\boldsymbol{Z}}\, \delta_{\boldsymbol{Z}}(u^t)$.
9. Compute the log-likelihood:

$$\begin{aligned}\mathcal{L}^{\beta\text{BM},(r+1)}(u^t) = \sum_{m=1}^{M} \sum_{1\le i<j\le n} \Big\{ &A_{m,ij}(u^t)\Big[\theta_i^{(r+1)}(u^t) + \theta_j^{(r+1)}(u^t) + Z_{c_i c_j}^{(r+1)}(u^t)\Big]\\ &- \log\Big[1 + \exp\big(\theta_i^{(r+1)}(u^t) + \theta_j^{(r+1)}(u^t) + Z_{c_i c_j}^{(r+1)}(u^t)\big)\Big]\Big\}.\end{aligned}$$

10. Compute the relative difference:

$$\Delta_p^{(r+1)} = \frac{\max\limits_{1\le t\le T}\Big\{\|\boldsymbol{\theta}^{(r+1)}(u^t) - \boldsymbol{\theta}^{(r)}(u^t)\|_\infty, \|\boldsymbol{Z}^{(r+1)}(u^t) - \boldsymbol{Z}^{(r)}(u^t)\|_\infty\Big\}}{1 + \max\limits_{1\le t\le T}\Big\{\|\boldsymbol{\theta}^{(r)}(u^t)\|_\infty, \|\boldsymbol{Z}^{(r)}(u^t)\|_\infty\Big\}}.$$

11. Compute the objective difference $\Delta_{\mathcal{L}}^{(r+1)} = |\mathcal{L}^{\beta\text{BM},(r+1)}(u^t) - \mathcal{L}^{\beta\text{BM},(r)}(u^t)|$.
12. Set $r = r + 1$.
13. **end while**
14. **end for**
15. **return** $\hat{\boldsymbol{\theta}} = (\boldsymbol{\theta}^{(r+1)}(u^1), \ldots, \boldsymbol{\theta}^{(r+1)}(u^T))^\top$ and $\hat{\boldsymbol{Z}} = (\boldsymbol{Z}^{(r+1)}(u^1), \ldots, \boldsymbol{Z}^{(r+1)}(u^T))$.

Algorithm 3 Local Approximate EM

Input: Adjacency matrices $\{\boldsymbol{A}_m(u^t)\}_{\substack{1\le m\le M\\ 1\le t\le T}}$, community memberships $\boldsymbol{C}$, bandwidth h, kernel $K_h(\cdot)$, convergence thresholds ϵ_p and ϵ_l, step sizes η_θ and η_Z, maximum iterations R.
Initialize: $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}^{(0)}(u^1), \ldots, \boldsymbol{\theta}^{(0)}(u^T))^\top$, $\boldsymbol{Z}^{(0)} = (\boldsymbol{Z}^0(u^1), \ldots, \boldsymbol{Z}^{(0)}(u^T))$, $\sigma^{2,(0)} = (\sigma^{2,(0)}(u^1), \ldots, \sigma^{2,(0)}(u^T))^\top$, $Q^{(0)}(u^1) = \ldots = Q^{(0)}(u^T) = -\infty$, $\Delta_p^{(0)} = \Delta_Q^{(0)} = +\infty$, $r = 0$.

1. **for** $t = 1, \ldots, T$ **do**
2. **while** $r < R$, $\Delta_p^{(r)} > \epsilon_p$, $\Delta_Q^{(r)} > \epsilon_l$ **do**

LAE-step:

3. **for** $s = 1, \ldots, T$ **do**
4. **for** $m = 1, \ldots, M$ **do**
5. Update $\hat{\omega}_m(u^s)$ based on Algorithm 1, and $\hat{v}_m(u^s)$ based on (7).
6. **end for**
7. **end for**

LAM-step:

8. Update $\sigma_h^{2,(r+1)}(u^t) = \frac{\sum_{m=1}^M \sum_{s=1}^T K_h(u^s - u^t)\big(\hat{\omega}_m^2(u^s) + \hat{v}_m^{-1}(u^s)\big)}{M \sum_{s=1}^T K_h(u^s - u^t)}$.
9. Compute $S_{Q,h}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)$ from (S3.1) and $H_{Q,h}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)$from (S3.2).
10. Compute

$$\begin{aligned}\delta_{\boldsymbol{\theta}}(u^t) = &- H_{Q,h}^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)\, S_{Q,h}\big(\boldsymbol{\theta}^{(r)}(u^t)\big) + H_{Q,h}^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)\boldsymbol{C} \\ &\times \Big(\boldsymbol{C}^\top H_{Q,h}^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big)\boldsymbol{C}\Big)^{-1} \boldsymbol{C}^\top H_{Q,h}^{-1}\big(\boldsymbol{\theta}^{(r)}(u^t)\big) S_{Q,h}\big(\boldsymbol{\theta}^{(r)}(u^t)\big).\end{aligned}$$

11. Update $\boldsymbol{\theta}^{(r+1)}(u^t) = \boldsymbol{\theta}^{(r)}(u^t) + \eta_\theta\, \delta_{\boldsymbol{\theta}}(u^t)$.
12. Compute $S_{Q,h}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$ from (S3.3) and $H_{Q,h}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$ from (S3.4).
13. Compute $\delta_{\boldsymbol{Z}}(u^t) = -H_{Q,h}^{-1}\big(\boldsymbol{Z}^{(r)}(u^t)\big)\, S_{Q,h}\big(\boldsymbol{Z}^{(r)}(u^t)\big)$.
14. Update $\boldsymbol{Z}^{(r+1)}(u^t) = \boldsymbol{Z}^{(r)}(u^t) + \eta_Z\, \delta_{\boldsymbol{Z}}(u^t)$.
15. Compute $Q^{(r+1)}(u^t) = Q_{h,1} + Q_{h,2} + Q_{h,3}$.
16. Compute the relative difference

$$\Delta_p^{(r+1)} = \frac{\max\Big\{\|\boldsymbol{\theta}^{(r+1)}(u^t) - \boldsymbol{\theta}^{(r)}(u^t)\|_\infty, \|\boldsymbol{Z}^{(r+1)}(u^t) - \boldsymbol{Z}^{(r)}(u^t)\|_\infty, |\sigma^{2,(r+1)}(u^t) - \sigma^{2,(r)}(u^t)|\Big\}}{1 + \max\Big\{\|\boldsymbol{\theta}^{(r)}(u^t)\|_\infty, \|\boldsymbol{Z}^{(r)}(u^t)\|_\infty, |\sigma^{2,(r)}(u^t)|\Big\}}.$$

17. Compute the objective difference $\Delta_Q^{(r+1)} = |Q(r+1)(u^t) - Q(r)(u^t)|$.
18. Set $r = r + 1$.
19. **end while**
20. **end for**
21. **return** $\tilde{\boldsymbol{\theta}}_h = (\boldsymbol{\theta}^{(r+1)}(u^1), \ldots, \boldsymbol{\theta}^{(r+1)}(u^T))^\top$, $\tilde{\boldsymbol{Z}}_h = (\boldsymbol{Z}^{(r+1)}(u^1), \ldots, \boldsymbol{Z}^{(r+1)}(u^T))$, $\tilde{\sigma}_h^2(u^t) = (\sigma^{2,(r+1)}(u^1), \ldots, \sigma^{2,(r+1)}(u^T))^\top$.

S6. Results for the residualized joint spectral embedding

In the Appendix of the main paper, we introduced a residualized joint spectral embedding (RJSE) method for estimating unknown community memberships. The resulting RJSE procedure is summarized in Algorithm 4. We now evaluate its performance under all simulation settings we considered in Section 5 of the main text. Since the data-generating paradigm includes subject-specific random effects, we construct a coarse-grained benchmark based on the works in Agterberg et al. (2025) and Lei and Lin (2023). Specifically, we define

$$\bar{A}^*(u^t) = M^{-1} \sum_{m=1}^{M} A_m(u^t), \qquad t = 1, \ldots, T,$$

and arrange the averaged networks according to their time order. We then apply the degree-corrected multiple adjacency spectral embedding method (DC-MASE) of Agterberg et al. (2025) and the debiased sum-of-squares method (De-SoS) of Lei and Lin (2023) to the sequence $\{\bar{A}^*(u^t)\}_{t=1,\ldots,T}$. Here, the averaging step preserves the time-varying network structure targeted by these benchmarks, while partially attenuating subject-specific fluctuations, and hence provides a reasonable benchmark for comparison. To evaluate the detection accuracy, we consider the misclassification rate after optimal label permutation as the performance metric, defined as,

$$ME(\hat{\boldsymbol{C}}) = \frac{1}{2n} \min_{\Pi \in \Pi_G} \|\hat{C}\Pi - C\|_1,$$

where $\hat{\boldsymbol{C}}$ denotes the estimated community memberships, and Π_G is the set of all $G \times G$ permutation matrices. All figures in this section present the mean results over 50 independent trials.

The results are summarized in Table S6.1. Overall, we observe that our proposed residualized joint spectral embedding method achieves exact community recovery in all configurations. This suggests that the auxiliary model successfully accounts for node-level and subject-level nuisance variation, thereby disentangling the common community structure from these nuisance components. Among the two benchmark methods, DC-MASE generally outperforms De-SoS, especially in the presence of node heterogeneity. This is consistent with the fact that DC-MASE incorporates the degree corrections through row-wise spherical normalization of the spectral embeddings, whereas De-SoS does not explicitly account for such effects.

Across scenarios, the two benchmark methods also perform well when both sources of heterogeneity are relatively mild. They achieve zero error in Scenario 2, where the nodes are homogeneous and $\sigma_0 = 0.5$, and remain competitive in Scenarios 1 and 4, where the node heterogeneity is present while σ_0 remains the same. This suggests that the simple averaging over subjects can partially mitigate the effects of heterogeneity. However, their performance deteriorates as σ_0 increases. As in Scenario 3, DC-MASE and De-SoS exhibit large misclassification errors across all values of T. This behavior is expected, since a large σ_0 will introduce stronger distortions in the averaged networks, which may perturb the leading eigenspace and rotate it away from the community subspace. As a result, the coarse-grained methods fail to recover the true communities. Accordingly, plugging their estimated memberships into the parameter estimation procedure in Section 3 is expected to yield less accurate estimates than those obtained using our proposed method.

Algorithm 4 Residualized joint spectral embedding

Input: Adjacency matrices $\{\boldsymbol{A}_m(u^t)\}_{\substack{1\le m\le M\\ 1\le t\le T}}$, number of communities G.

Output: Estimated community membership matrix $\hat{\boldsymbol{C}}$.

1. Fit the community-free auxiliary model and compute the fitted edge probability matrices $\{\hat{\boldsymbol{\tau}}_m^{\text{aux}}(u^t)\}_{\substack{1\le m\le M\\ 1\le t\le T}}$.
2. **for** $m = 1, \ldots, M$ and $t = 1, \ldots, T$ **do**
3. Construct the standardized residual matrix $\hat{\boldsymbol{R}}_m(u^t)$ with

$$\hat{R}_{m,ij}(u^t) = \begin{cases} 0 & i = j \\ \frac{A_{m,ij}(u^t) - \hat{\tau}^{\text{aux}}_{m,ij}(u^t)}{\left[\hat{\tau}^{\text{aux}}_{m,ij}(u^t)(1-\hat{\tau}^{\text{aux}}_{m,ij}(u^t))\right]^{1/2}} & i \neq j. \end{cases}$$

4. **end for**
5. Construct the residual joint matrix $\sum_{t=1}^{T}\sum_{m=1}^{M} \hat{\boldsymbol{R}}_m(u^t)\hat{\boldsymbol{R}}_m(u^t)^\top$.
6. Let $\hat{\boldsymbol{U}}$ be the matrix formed by the leading G eigenvectors of $\sum_{t=1}^{T}\sum_{m=1}^{M} \hat{\boldsymbol{R}}_m(u^t)\hat{\boldsymbol{R}}_m(u^t)^\top$.
7. Normalize the rows of $\hat{\boldsymbol{U}}$ by defining $\hat{\boldsymbol{U}}^*_{i\cdot} = \frac{\hat{\boldsymbol{U}}_{i\cdot}}{\|\hat{\boldsymbol{U}}_{i\cdot}\|_2}$ whenever $\|\hat{\boldsymbol{U}}_{i\cdot}\|_2 > 0$.
8. Apply K-means with G clusters to the rows of $\hat{\boldsymbol{U}}^*$.
9. **return** $\hat{\boldsymbol{C}}$.

Table S6.1: Average misclassification errors for community membership estimation over 50 Monte Carlo trials under the four scenarios of the main text.

Scenario	**Method**	$T = 15$	$T = 30$	$T = 60$	$T = 120$
1	RJSE	0.00	0.00	0.00	0.00
	DC-MASE	0.14	0.08	0.02	0.02
	De-SoS	0.14	0.14	0.12	0.11
2	RJSE	0.00	0.00	0.00	0.00
	DC-MASE	0.00	0.00	0.00	0.00
	De-SoS	0.00	0.00	0.00	0.00
3	RJSE	0.00	0.00	0.00	0.00
	DC-MASE	0.25	0.25	0.26	0.30
	De-SoS	0.29	0.31	0.37	0.36
4	RJSE	0.00	0.00	0.00	0.00
	DC-MASE	0.21	0.00	0.00	0.00
	De-SoS	0.09	0.00	0.04	0.00

S7. Additional details for dynamic functional connectivity during movie watching

S7.1 Data description and preprocessing

We apply our proposed method to the dynamic functional connectivity networks derived from a publicly available Human Connectome Project (HCP) dataset, using minimally preprocessed 7T cortical-surface fMRI data. Our primary analysis focuses on $M = 34$ young adults (age 22–30) who completed the full movie-watching protocol. In the part of the experiment considered here, participants passively viewed a series of audiovisual clips in two movie-watching scans: one comprising clips from independent films and the other comprising clips from Hollywood films. Any two consecutive clips are separated by 20-second rest periods, during which the word "REST" is displayed on a dark background. To reduce transient onset effects and to avoid bias from non-stimulus periods, we discard all rest blocks and remove the first 20 seconds at the beginning of each clip. In addition, several very short test clips are excluded to ensure that each retained clip provides sufficient temporal support for reliable dynamic connectivity estimation. After these preprocessing steps, the final movie-watching dataset comprises six clips: *Two Men* (225 seconds), *Bridgeville* (202 seconds), *Pockets* (169 seconds), *Inception* (208 seconds), *The Social Network* (240 seconds), and *Ocean's Eleven* (230 seconds), totaling 1274 seconds per subject.

The fMRI data are parcellated into $n = 360$ cortical regions of interest using the Glasser HCP-MMP atlas (Glasser et al., 2016). We adopt this parcellation because it is natively aligned with the HCP dataset and the CIFTI format (Ji et al., 2019; Gruskin and Patel, 2022). Furthermore, the $n = 360$ ROIs are partitioned into $G = 7$ functional modules (communities) from Yeo et al. (2011): (1) Visual, (2) Somatomotor network (SMN), (3) Dorsal attention network (DAN), (4) Ventral attention network (VAN), (5) Limbic, (6) Frontoparietal network (FPN), and (7) Default mode network (DMN). For each subject, we construct a sequence of Pearson correlation matrices among the $n = 360$ ROIs using a sliding-window procedure with window length 30 and step size 15, yielding a total of $T = 83$ connectivity snapshots per subject. Each correlation matrix is then converted into an undirected binary adjacency matrix by hard thresholding at 0.55 in absolute value. By construction, the resulting networks have an average density of approximately 0.12. The same construction was applied to both the movie-watching and resting-state data to ensure that the two conditions were comparable.

S7.2 Additional analysis of estimated parameters

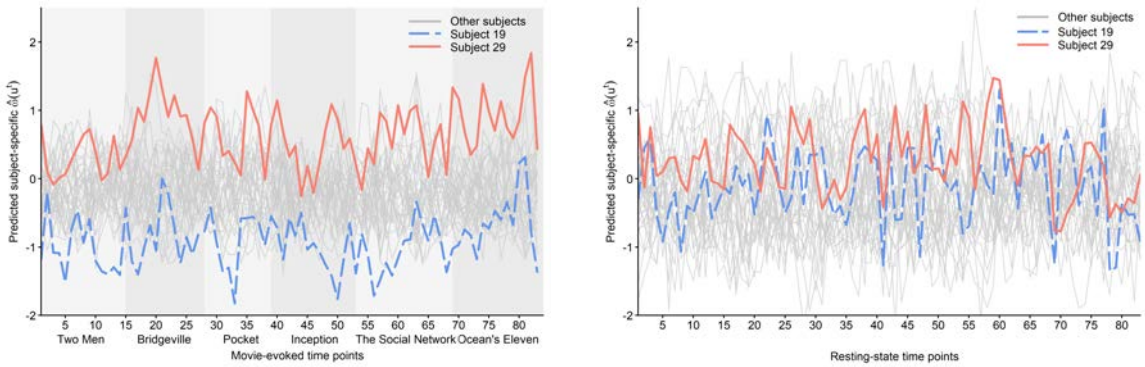
Persistent and movie-driven ROIs. As mentioned in Section 2 in the main text, we interpret the degree-correction parameters $\boldsymbol{\theta}(u^t)$ as the time-varying degree propensities where a larger positive value of $\theta_i(u^t)$ indicates that the i-th ROI has a stronger tendency to form connections. In a sense, an ROI with a larger value of $\theta_i(u^t)$ tends to be more active in the brain network. Motivated by this interpretation, we compute, for each ROI $i, i = 1, \ldots, 360$, the frequency with which its corresponding $\theta_i(u^t)$ appears among the top 20 values under the movie-watching and resting-state conditions, and use these frequencies to identify two types of active regions: (i) persistent ROIs, that remain highly active, with high frequencies of being ranked among the top 20 nodes, across both movie-watching and resting-state conditions, (ii) movie-driven (event-driven) ROIs, which are frequently ranked among the top 20 nodes during movie watching but appear much less prominently under the resting-state condition. The top 5 persistent regions, along with the top 5 movie-driven re-

Table S7.2: Top persistent and movie-driven ROIs based on the frequency with which $\tilde{\theta}_i(u^t)$ is ranked among the top 20 degree parameters across time.

Type	ROI	Module	Movie frequency	Rest frequency
Persistent	R PF ROI	VAN	83	83
Persistent	L PF ROI	VAN	83	83
Persistent	L TGd ROI	Limbic	82	83
Persistent	L PGi ROI	DMN	79	83
Persistent	R TGd ROI	Limbic	71	83
Movie-driven	L TPOJ1 ROI	SMN	75	2
Movie-driven	L V4 ROI	Visual	75	9
Movie-driven	L STSdp ROI	DMN	70	16
Movie-driven	R A5 ROI	SMN	65	0
Movie-driven	L A5 ROI	SMN	65	0

gions, are summarized in Table S7.2, and their estimated temporal behaviour is presented in Figure 3 (b) and (e) of the main text. Notably, the persistent ROIs are mainly located in the higher-order association cortex, including the inferior parietal lobule and the anterior temporal cortex. The former is involved in attentional reorienting and large-scale information integration, whereas the latter is typically associated with semantic processing and social conceptual representation. These regions may therefore play an important role in supporting fundamental brain function and integrating information. Interestingly, a comparison of panels (b) and (e) in Figure 3 shows that the movie-driven ROIs display clear temporal fluctuations during movie watching, whereas their trajectories, under the resting-state condition, are relatively flat and remain close to zero. These differences may be associated with the anatomical locations of these regions and the functional roles they are known to subserve. L_V4_ROI is particularly noteworthy, as it is closely associated with color and shape processing, object representation, selective attention, and the extraction of object information from complex visual scenes (Pasupathy et al., 2020). L_TPOJ1_ROI, located in the temporo-parieto-occipital junction, shows connectivity with auditory areas (e.g., L_A5_ROI), visual regions involved in motion, face, and word processing (e.g., L_V4_ROI), somatosensory regions, and inferior frontal language areas. Rolls et al. (2023) show that, these regions are well positioned to be active during movie watching, which involves naturalistic stimuli rich in speech, action, scene dynamics, and narrative information.

Subject-specific random effects. To measure the subject-specific random shifts, we exploit the fitted posterior distribution and compute, for each subject m and time point u^t, the maximum a posteriori estimator (MAP) of the random effects $\omega_m(u^t)$ by solving the optimization problem in (6) of the main text. It is worth noting that, under the Gaussian prior of $\omega_m(u^t)$, this MAP estimator coincides with the empirical Bayes predictor (i.e., the posterior mode). Figure S7.2 displays the predicted trajectories of $\hat{\omega}_m(u^t)$ across movie-evoked time points. We observe that some subjects (e.g., Subjects 19 and 29) deviate substantially from the population mean, not only in the magnitude of the deviations but also in their temporal evolution. This suggests a substantial individual heterogeneity in dynamic connectivity, and further indicates that such effects are not merely static offsets but may evolve with the movie stimulus and its changing cognitive demands.



(a) $\hat{\boldsymbol{\omega}}(u^t)$ under the movie-watching condition

(b) $\hat{\boldsymbol{\omega}}(u^t)$ under the resting-state condition

Figure S7.2: Predicted trajectories of $\hat{\omega}_m(u^t)$. Colored curves indicate two representative subjects, Subject 29 (solid) and Subject 19 (dashed); gray curves correspond to the remaining subjects.

Moreover, we provide the time-varying distributions of $\hat{\omega}_m(u^t)$ for the six movies in Figure S7.3.

To highlight even more the importance of accurately predicting random effects, we construct another metric of evaluation that focuses on comparing the observed adjacency matrix with the probability estimates produced by our fitted model. Given $\hat{\omega}_m(u^t)$, we can obtain the subject-specific edge probability matrix

$$\tilde{\boldsymbol{\tau}}_m(u^t) = \text{logit}^{-1}\left(\tilde{\boldsymbol{\theta}}(u^t)\mathbf{1}_n^\top + \mathbf{1}_n\tilde{\boldsymbol{\theta}}(u^t)^\top + \boldsymbol{C}\tilde{\boldsymbol{Z}}(u^t)\boldsymbol{C}^\top + \hat{\omega}_m(u^t)\mathbf{1}_n\mathbf{1}_n^\top\right),$$

which enables individualized network reconstruction. To assess predictive accuracy, we consider the average squared Frobenius loss between the observed $\boldsymbol{A}_m(u^t)$ and the estimated $\tilde{\boldsymbol{\tau}}_m(u^t)$. Specifically, we compute

$$\hat{\text{Err}}_{\text{LA-EM}} = \frac{1}{n^2MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\|\boldsymbol{A}_m(u^t) - \tilde{\boldsymbol{\tau}}_m(u^t)\|_F^2.$$

As a benchmark, we fit a βBM with community covariates (as in our simulation studies), which produces only a common estimate, $\tilde{\boldsymbol{\tau}}(u^t)$, shared across subjects. We then compute an analogous error, denoted by $\hat{\text{Err}}_{\beta\text{BM}}$. Empirically, $\hat{\text{Err}}_{\text{LA-EM}}$ is 0.09 for the movie-watching data and 0.10 for the resting-state data, whereas the error $\hat{\text{Err}}_{\beta\text{BM}}$ equals 0.10 and 0.11 for the movie-watching and the resting-state fMRI, representing a relative reduction of approximately 10%. This improvement suggests that incorporating subject-specific random shifts can yield more accurate probability predictions, which are essential for the analyses of individual differences.

Out-of-sample discrimination between movie watching and rest. Although model (1) is not formulated for classification, one could construct a simple discriminant method between movie watching and resting state using the predictive distribution for each observed network given the out-of-sample estimates of parameters from all methods in Section 4 of the main text. We implement five-fold cross-validation with subject-wise splits. In each fold, we first estimate the model parameters separately using the movie-watching and resting-state training data, and then compute the assigned

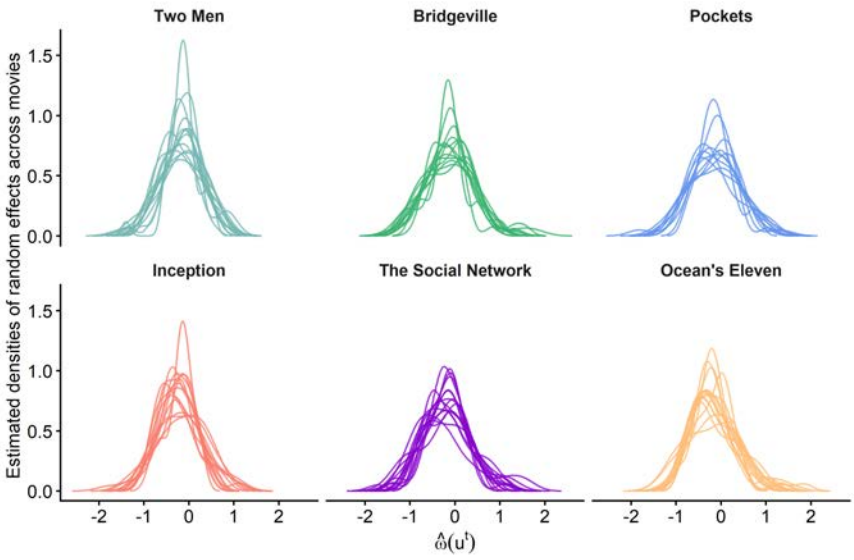


Figure S7.3: Estimated densities of subject-specific random effects, $\boldsymbol{\omega}(u^t)$, across movie clips.

to the class with the larger log likelihood scores for all testing samples under the two fitted models. Each observation in the testing set is assigned to the class that attains the larger scores. Here, the misclassification error (denoted by ME) is defined as the total proportion of observations whose predicted class differs from the true class. Intuitively, comparing ME across methods provides a quantitative measure of their ability to capture stimulus-evoked network dynamics. The procedure is provided as follows.

1. **Subject splitting.** Randomly partition the 34 subjects into $K = 5$ folds $\mathcal{M}_1, \ldots, \mathcal{M}_5$ of (approximately) equal size. To avoid information leakage, we assign all observations (movie-watching and resting-state) from the same subject to the same fold.

2. **For each fold $k = 1, \ldots, 5$, we perform the following steps:**

   (a) **Estimation from the training sets.** We fit the model on the movie training set, namely $\{A_m^{\text{mov}}(t) : m \in \mathcal{M}_{k'}, \forall k' \neq k\}$, to obtain movie-related model parameters, denoted by

   $$\tilde{\boldsymbol{\xi}}_{\text{mov}}(u^t) = (\tilde{\boldsymbol{\theta}}_{\text{mov}}(u^t), \tilde{\boldsymbol{Z}}_{\text{mov}}(u^t), \tilde{\sigma}^2_{\text{mov}}(u^t)),$$

   for $u^t \in \{1/T, \ldots, 1\}$. Similarly, calculate the resting-state parameters

   $$\tilde{\boldsymbol{\xi}}_{\text{rest}}(u^t) = (\tilde{\boldsymbol{\theta}}_{\text{rest}}(u^t), \tilde{\boldsymbol{Z}}_{\text{rest}}(u^t), \tilde{\sigma}^2_{\text{rest}}(u^t)),$$

   on set $\{A_m^{\text{rest}}(t) : m \in \mathcal{M}_{k'}, \forall k' \neq k\}$.

   (b) **Score calculation on the testing sets.** For each $m \in \mathcal{M}_k$ and $u^t \in \{1/T, \ldots, 1\}$, we compute the log likelihood score of each observation $A_m(u^t)$ under each condition (movie watching and resting state),

   $$\text{Score}_*(m, u^t) = \log \mathbb{P}(\boldsymbol{A}_m(u^t); \tilde{\boldsymbol{\xi}}_*(u^t)), \quad \text{with} \quad * \in \{\text{mov}, \text{rest}\}.$$

(c) **Classification.** Define the log-likelihood difference as

$$\Delta(m, u^t) = \text{Score}_{\text{mov}}(m, u^t) - \text{Score}_{\text{rest}}(m, u^t).$$

We consider two types of classifications: (i) Pointwise classification, which predicts "Movie" at time point u^t if $\Delta(m, u^t) > 0$ and "Rest" otherwise; and (ii) Global classification, which predicts "Movie" for subject m if $\sum_{t=1}^{T} \Delta(m, u^t) > 0$ and "Rest" otherwise.

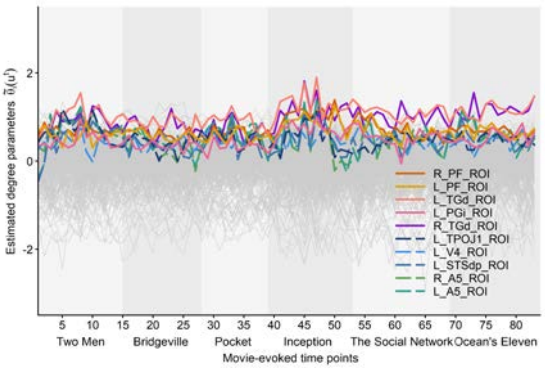
The pointwise and global misclassification errors are reported in Table S7.3. Overall, the proposed method outperforms the competing methods.

Table S7.3: Cross-validated misclassification errors (ME) for discriminating movie watching from resting state for four methods. Pointwise ME denotes the overall proportion of misclassified observations across subjects and time points, whereas Global ME denotes the proportion of misclassified subjects.

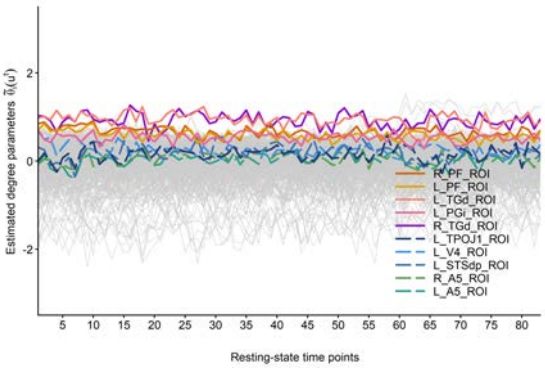
	ME-DCSBM$_{\text{LA-EM}}$	ME-DCSBM$_{\text{PW}}$	ME-SBM	βBM
Pointwise ME	0.16	0.16	0.31	0.22
Global ME	0.01	0.03	0.10	0.06

S7.3 Sensitivity analysis

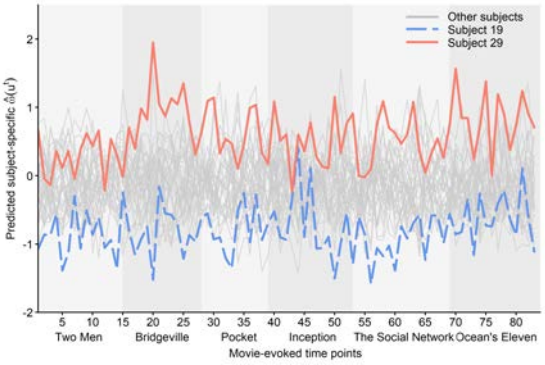
In constructing dynamic connectivity networks, both the window and step sizes may influence the resulting network estimates. If the window size is too large, the constructed network might mix signals from different periods, and thus, it might reduce the ability to capture the dynamic features of the brain functional connectivity. If the window length is too short, however, the amount of information available within each window may be insufficient, leading to greater sensitivity to noise. Similarly, a large step size will reduce the number of the constructed networks, whereas a small step size may increase the noise in networks. As such, we check the sensitivity of our proposed method to these choices in this section. To avoid confounding the effect of window length with that of step size, we fix the step size at 15 and consider window sizes of 20 and 40. Figures S7.4 and S7.5 report the estimated model parameters under these two settings. We observe that the resulting estimates are very similar to those reported in the main text, further suggesting that the proposed method is robust to the considered choices of window length.



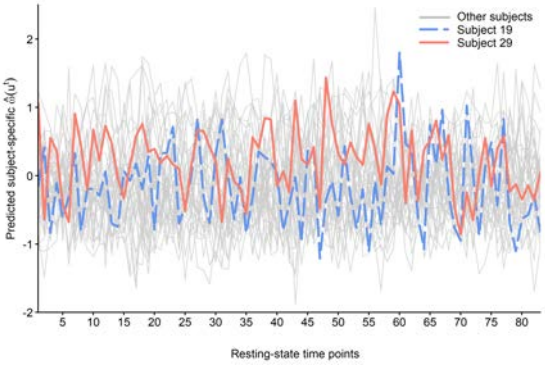
(a) $\tilde{\boldsymbol{\theta}}(u^t)$ under the movie-watching condition



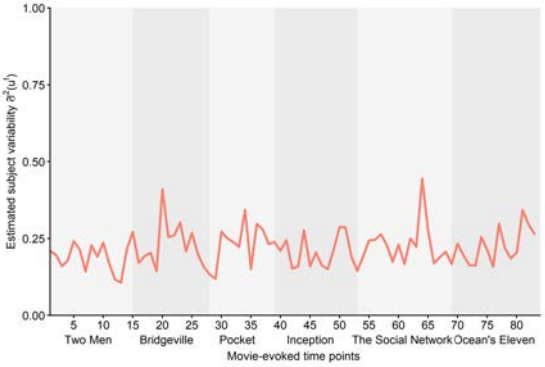
(b) $\tilde{\boldsymbol{\theta}}(u^t)$ under the resting-state condition



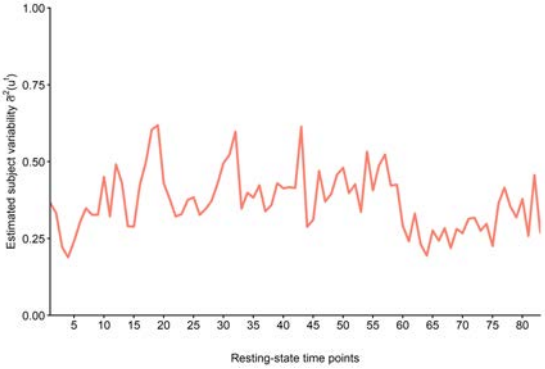
(c) $\hat{\boldsymbol{\omega}}(u^t)$ under the movie-watching condition



(d) $\hat{\boldsymbol{\omega}}(u^t)$ under the resting-state condition

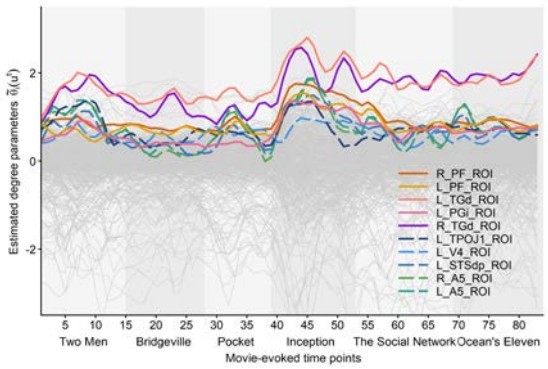


(e) $\tilde{\sigma}^2(u^t)$ under the movie-watching condition

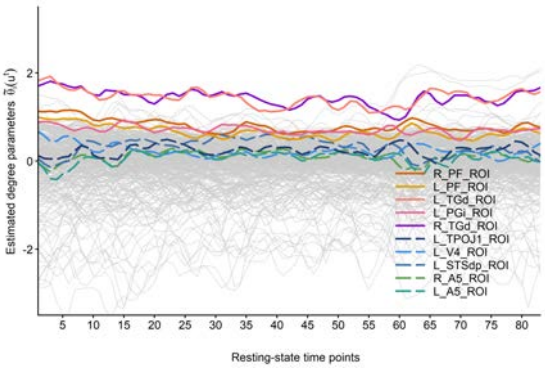


(f) $\tilde{\sigma}^2(u^t)$ under the resting-state condition

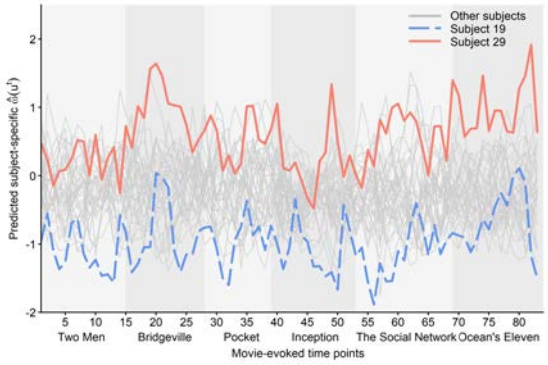
Figure S7.4: Estimated trajectories of model parameters under the alternative window length of 20.



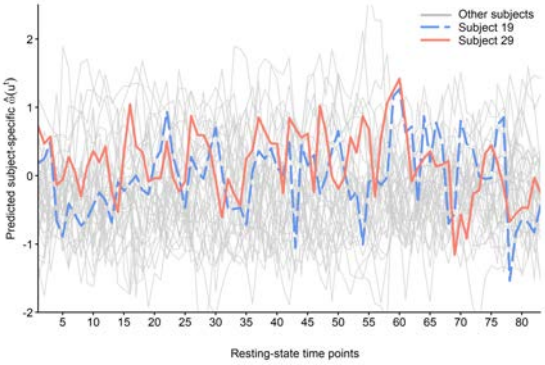
(a) $\tilde{\boldsymbol{\theta}}(u^t)$ under the movie-watching condition



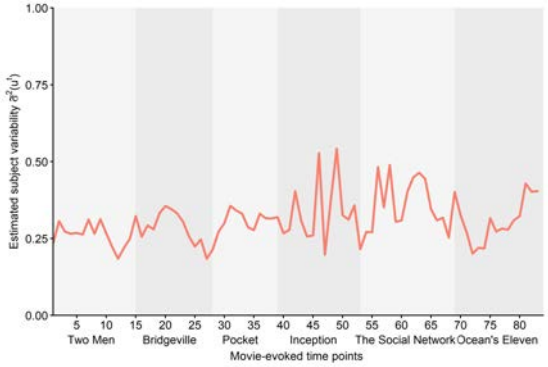
(b) $\tilde{\boldsymbol{\theta}}(u^t)$ under the resting-state condition



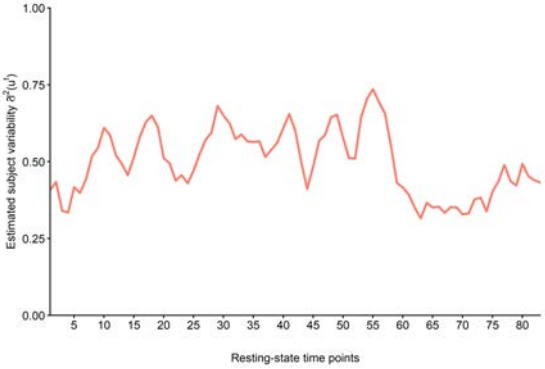
(c) $\hat{\boldsymbol{\omega}}(u^t)$ under the movie-watching condition



(d) $\hat{\boldsymbol{\omega}}(u^t)$ under the resting-state condition



(e) $\tilde{\sigma}^2(u^t)$ under the movie-watching condition



(f) $\tilde{\sigma}^2(u^t)$ under the resting-state condition

Figure S7.5: Estimated trajectories of model parameters under the alternative window length of 40.

References

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

Glasser, M. F., T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature 536*(7615), 171–178.

Gruskin, D. C. and G. H. Patel (2022). Brain connectivity at rest predicts individual differences in normative activity during movie watching. *NeuroImage 253*, 119100.

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

Ji, J. L., M. Spronk, K. Kulkarni, G. Repovš, A. Anticevic, and M. W. Cole (2019). Mapping the human brain's cortical-subcortical functional network organization. *NeuroImage 185*, 35–57.

Lei, J. and K. Z. Lin (2023). Bias-adjusted spectral clustering in multi-layer stochastic block models. *Journal of the American Statistical Association 118*(544), 2433–2445.

Pasupathy, A., D. V. Popovkina, and T. Kim (2020). Visual functions of primate area v4. *Annual Review of Vision Science 6*(1), 363–385.

Rolls, E. T., J. P. Rauschecker, G. Deco, C.-C. Huang, and J. Feng (2023). Auditory cortical connectivity in humans. *Cerebral Cortex 33*(10), 6207–6227.

Yeo, T., F. M. Krienen, J. Sepulcre, M. R. Sabuncu, D. Lashkari, M. Hollinshead, J. L. Roffman, J. W. Smoller, L. Zöllei, J. R. Polimeni, et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology 106*(3), 1125–1165.