

SEMIPARAMETRIC ESTIMATION FOR DYNAMIC NETWORKS WITH SHIFTED CONNECTING INTENSITIES

BY ZITONG ZHANG^a AND SHIZHE CHEN^b

Department of Statistics, University of California Davis, ^aaztzhang@ucdavis.edu, ^bbszdchen@ucdavis.edu

Neural circuits are of paramount importance in the nervous system, as they are the essential infrastructure in guiding animal behavior. However, modeling the development of neural circuits poses significant challenges due to inherent properties of the development process. First, the neural circuit development process is transient, where the course of development can only be observed once. Second, despite potentially sharing similar underlying mechanisms for development, neural circuits from different subjects possess distinct sets of neurons, which limits the sharing of information across subjects. Third, neurons have diverse, unobserved activation times, which may obscure the analysis of neural activities. In light of these challenges, this study presents a novel approach aimed at clustering neurons based on their connecting behaviors while accommodating disparities at the neuron level. To this end, we propose a dynamic stochastic block model that accommodates unknown time shifts. We establish the conditions that guarantee the identifiability of cluster memberships of nodes and representative connecting intensities across clusters. Using methods for shape invariant models, we propose computationally efficient semiparametric estimation procedures to simultaneously estimate time shifts, cluster memberships, and connecting intensities. We illustrate the performance of the proposed procedures via extensive simulation experiments. We further apply the proposed method on a motor circuit development data from zebrafish to reveal distinct roles of neurons and identify representative connecting behaviors.

1. Introduction. Neural circuits are fundamental components of the nervous system that play a crucial role in instructing animal behavior. Understanding the neural circuit development is beneficial for understanding how the nervous system works, identifying the causes of nervous system disorders, and developing new treatments for nervous system disorders. However, the study of neural circuit development has been limited by difficulty in data acquisition until recently. Modern technologies have enabled the recording of individual neural activities over extended durations, thereby facilitated the exploration of neural circuit development.

Using novel technologies, Wan et al. (2019) recorded the development of a neural circuit in the spine of zebrafish during embryogenesis. The neural circuit development process involves individual neurons forming functional connections, resulting in a highly interconnected network. The process is illustrated in Figure 1(a), where nodes represent neurons and edges represent functional connections between neurons. The dataset on neural circuit development exhibits three inherent properties. First, the neural circuit development process is transient, making it impossible to obtain repeated observations of the same system. Second, neural circuits from different subjects possess distinct sets of neurons, despite potentially sharing similar underlying developmental mechanisms. Third, neurons have different functional maturation progress, resulting in different *activation times* (Wan et al. (2019)). The neural activation times can obscure the similarity in neural connecting behavior. Figure 1(b) depicts two neurons exhibiting seemingly distinct connecting behavior. However, upon adjusting the

Received December 2022; revised December 2023.

Key words and phrases. Dynamic networks, semiparametric estimation, shape invariant models, stochastic block models.

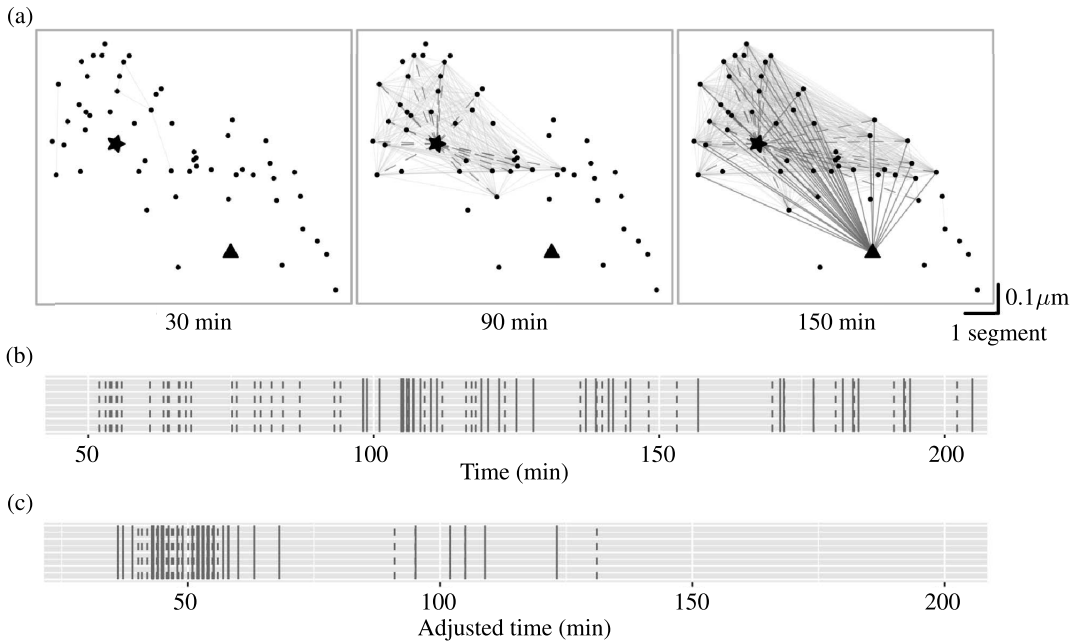


FIG. 1. *Isolated neurons form a densely connected network during embryogenesis of zebrafish. All time stamps indicate the time since the experiment starts. Panel (a) displays neural network snapshots at 30, 90, and 150 minutes, where edges appear progressively over time. In the networks, nodes represent neurons, and edges represent functional connections between neurons. The horizontal axis is the anterior-posterior (AP) axis, and the vertical axis is the left-right (LR) axis of zebrafish (see [Wan et al. \(2019\)](#) for details). Preprocessing steps for obtaining these networks are available in Section S5 in the Supplementary Material ([Zhang and Chen \(2023\)](#)). The star and triangle nodes are two neurons with similar connecting behaviour. Edges associated with the two nodes are displayed as dashed and solid line segments, respectively. The rest of edges are displayed using faint and thin lines. Panel (b) displays connecting behaviour of the two highlighted neurons. Vertical bars represent the time points when edges associated with the two neurons are formed. The line types are consistent with Panel (a). Panel (c) displays the adjusted connecting behaviour of the two highlighted neurons. Each vertical bar has been shifted according to the estimated time shift of the associated edge, as determined by the proposed method. Upon this adjustment, it becomes evident that the connecting behavior of the two neurons exhibits a high degree of similarity.*

connection times based on the estimated neural activation times using our proposed method, Figure 1(c) shows that the two neurons actually share similar connecting behaviors. Additional examples of neuron pairs with similar connecting behavior can be found in Figure S14 in the Supplementary Material ([Zhang and Chen \(2023\)](#)). In light of these inherent properties, our primary objective in this study is to cluster neurons based on their connecting behaviors after accounting for diverse activation times. The resulting clusters are expected to comprise neurons that play similar roles during neural circuit development.

Nonetheless, existing methods in the field of network analysis have limitations in identifying clusters of neurons when varying activation times are present. A canonical approach for clustering nodes in networks is the *dynamic stochastic block models* (see, e.g., [Longepierre and Matias \(2019\)](#), [Matias and Miele \(2017\)](#), [Matias, Rebafka and Villers \(2018\)](#), [Pensky \(2019\)](#), [Xing, Fu and Song \(2010\)](#), [Xu \(2015\)](#), [Xu and Hero \(2014\)](#), [Yang et al. \(2011\)](#)). In essence, the dynamic stochastic block model assumes that nodes in a network fall into distinct clusters, and the probability of an edge forming between a pair of nodes over time, that is, the *connecting intensity*, depends solely on the the cluster memberships of the two nodes. For networks observed at a discrete time grid, the connecting intensities are typically parameterized using Markov chains (see, e.g., [Matias and Miele \(2017\)](#), [Xing, Fu and Song](#)

(2010), Xu (2015), Xu and Hero (2014), Yang et al. (2011)). For networks observed in continuous time, Matias, Rebafka and Villers (2018) propose to model recurrent instantaneous edges (e.g., email exchanges) as point processes, where connecting intensities are estimated nonparametrically. However, a crucial assumption in these models is that nodes within the same cluster share *identical* connecting intensities, which is essential for identifiability and estimation. Consequently, these models are unable to identify neural activation times and, as a result, may overlook the similarity in connecting behavior among neurons.

A class of problems relevant to identifying neural activation times has been studied in the context of *shape invariant models* (see, e.g., Bigot and Gadat (2010), Bigot et al. (2013), Bigot and Gendre (2013), Bontemps and Gadat (2014), Vimond (2010)). These models aim to align a set of curves that are identical up to unknown time shifts. However, the shape invariant models are incapable of handling multiple clusters of curves, where the curves within each cluster are identical up to unknown time shifts.

In this paper we propose a dynamic stochastic block model that allows for unknown node-specific time shifts. We establish the identifiability of the time shifts, the cluster memberships, and the connecting intensities under suitable assumptions in Section 2. We propose semi-parametric estimation procedures that simultaneously estimate model parameters including the time shifts in Section 3, along with computationally efficient algorithms in Section 4. We investigate the performance of the proposed estimation procedures through simulation experiments in Section 5. Finally, we use the proposed model to study neural behaviour during embryogenesis of zebrafish in Section 6.

2. Dynamic stochastic block model. We model the development of neural circuits as a dynamic network. Specifically, a dynamic network is defined as $\{\mathcal{G}(t) : \mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t)), t \in [0, \infty)\}$, where $\mathcal{V} \equiv \{1, 2, \dots, p\}$ denotes a fixed set of nodes across time and $\mathcal{E}(t)$ represents the set of edges at time t , defined as

$$\mathcal{E}(t) \equiv \{(i, j) : i, j \in \mathcal{V}, \text{ nodes } i, j \text{ are connected at time } t\}.$$

In the context of neural circuits, $\mathcal{V}$ corresponds to the set of neurons, and $\mathcal{E}(t)$ denotes the set of functional connections between neurons at time t . The functional connections among neurons are identified by the highly correlated activities of neurons (see illustration in Section S5.1 in the Supplementary Material, Zhang and Chen (2023)). These functional connections have the following properties: (i) no self-connection, that is, $(i, i) \notin \mathcal{E}(t)$ for any $i \in \mathcal{V}$ and $t \geq 0$, (ii) undirected connections, that is, $(i, j) \in \mathcal{E}(t)$ if $(j, i) \in \mathcal{E}(t)$, (iii) no connections at the onset of observation, that is, $\mathcal{E}(0) = \emptyset$, and (iv) persistence, that is, $(i, j) \in \mathcal{E}(t')$ for any $t' \geq t$ if $(i, j) \in \mathcal{E}(t)$. Further exploration of the persistence of connections can be found in Figure S10 in the Supplementary Material (Zhang and Chen (2023)). For simplicity, we assume that the network has stabilized by time T , indicating that $\mathcal{G}(t) = \mathcal{G}(T)$ for any $t > T$. Hereafter, unless otherwise specified, we let (i, j) denote two arbitrary and distinct nodes, that is, $i \neq j$ and $i, j \in \mathcal{V}$.

For any pair of nodes (i, j) , we define the *connecting time* as $t_{i,j} \equiv \inf\{t \in [0, T] : (i, j) \in \mathcal{E}(t)\}$, that is, the time when nodes i and j connect, and we denote $t_{i,j} = \infty$ if $\{t \in [0, T] : (i, j) \in \mathcal{E}(t)\}$ is empty. We define the *connecting process* as $N_{i,j}(t) \equiv \mathbf{I}(t_{i,j} \leq t)$, $t \in [0, T]$, which is known as a point process or a counting process in the literature (see, among others, Daley and Vere-Jones (2003)). The persistence of edges implies that $N_{i,j}(t)$ remains to be unity once an edge is formed. This type of point process is commonly studied in survival analysis as the outcome (see, among others, Kleinbaum and Klein (2012)). We introduce the (marginal) intensity $\lambda_{i,j}(t) \equiv \Pr(dN_{i,j}(t) = 1)/dt$ that we refer to as *connecting intensity*, whose integral $\Lambda_{i,j}(T) \equiv \int_0^T \lambda_{i,j}(t) dt$ is the *connecting probability* between nodes i, j throughout $[0, T]$. In survival analysis the probability $\Lambda_{i,j}(T)$ can be less than one due

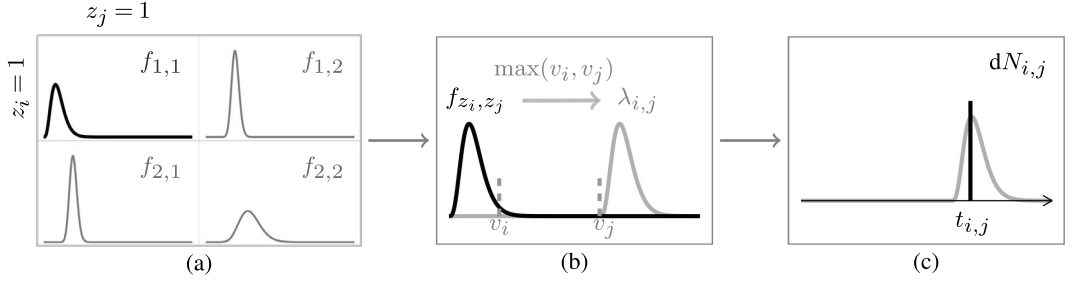


FIG. 2. Graphical representation of our model. Panel (a) shows a matrix of representative connecting intensities $(f_{q,k})_{K \times K}$, where the highlighted curve represents f_{z_i, z_j} . Panel (b) shows the connecting intensity between node i and j after incorporating the time shift, where $\lambda_{i,j} = S^{\max(v_i, v_j)} f_{z_i, z_j}$. Panel (c) shows the observed connecting process, where $\mathbb{E}[dN_{i,j}(t)] = \lambda_{i,j}(t) dt$. Only the connecting process is observed in the data.

to censoring. In our setting we assume that there is no censoring, but $\Lambda_{i,j}(T)$ can still be less than one since the final graph is not necessarily complete. Conditioning on that nodes i and j are connected, the density of the connecting time, or *connecting density*, is $\lambda_{i,j}(t)/\Lambda_{i,j}(T)$.

Following the convention in Matias, Rebafka and Villers (2018) and Pensky (2019), we assume that $\mathcal{V} = \bigcup_{k=1}^K \mathcal{C}_k$, where $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_K$ are disjoint clusters with distinct connecting intensities to be specified in Assumption 1. Nodes within each cluster have identical connecting intensities up to unknown time shifts. To be specific, denote $[K] \equiv \{1, \dots, K\}$, and let $z_i \in [K]$ be the *cluster membership* of node i for $i \in \mathcal{V}$. Then we assume that the connecting intensity between nodes i and j satisfies that

$$(1) \quad \lambda_{i,j}(t) = S^{w_{i,j}} f_{z_i, z_j}(t),$$

where $w_{i,j} \equiv \max(v_i, v_j)$ with $v_i, v_j \in [0, T/2]$ being the *time shifts* of nodes i and j , S^w is the shift operator defined as $S^w x(t) \equiv x(t - w)$, and $f_{z_i, z_j}(\cdot)$ is a nonnegative function with support in $[0, T/2]$ and $\|f_{z_i, z_j}\|_1 \leq 1$. For $q, k \in [K]$, the function $f_{q,k}(\cdot)$ is called the *representative connecting intensity* between any pairs of nodes in clusters q and k . Figure 2 shows a graphical representation of the proposed dynamic stochastic block model with time shifts. Here the bounded supports for the representative connecting intensities and time shifts are put in place to avoid the discussion of censoring, which can be easily relaxed.

The model presented in (1) is tailored to the neural circuit development data. Specifically, $w_{i,j}$ is defined as the time when both neurons i and j are mature enough to establish functional connections. As a result, the proposed model can be applied to identify the clusters of neurons with similar connecting behaviour (i.e., $\mathcal{C}_k$'s), the representative connecting behaviour of neurons (i.e., $f_{q,k}$'s), and activation time of neurons (i.e., v_i 's). Nonetheless, the model in (1) can be easily adapted to other applications with minor changes. For instance, it can be used to analyze information exchanges, such as tweets, emails, or text messages in social networks (see, e.g., Diesner and Carley (2005), Haythornthwaite (1996), Osatuyi (2013)), where nodes represent individuals and edges represent the initiation of information exchange between individuals on a particular topic. In this context, $w_{i,j}$ can be redefined to represent the minimal time shifts of individuals i and j , signifying the time when either individual i or j becomes intrigued by the topic.

The proposed model has close connections with two lines of research in the literature. First of all, the proposed model generalizes the dynamic stochastic block model in Matias, Rebafka and Villers (2018) by allowing unknown time shifts in connecting intensities between pairs of nodes. In fact, if we average out time shifts to treat $\mathbb{E}_{w_{i,j}}[\lambda_{i,j}(\cdot)]$'s as the connecting intensities, the proposed model reduces to the Poisson process stochastic block model (PPSBM) in Matias, Rebafka and Villers (2018). Second, when there is only one cluster, that is, $K = 1$,

the proposed model reduces to a shape invariant model with point processes (Bigot et al. (2013)), where the point processes are assumed to share a common intensity function subject to time shifts. Hence, we call the proposed model as the *shape invariant dynamic stochastic block model*, or SidSBM for short.

Additional conditions are required to ensure the identifiability of parameters in (1). In the case without unknown time shifts, Matias, Rebafka and Villers (2018) establish the identifiability of representative connecting intensities and cluster memberships, under the assumption that all representative connecting intensities are distinct. In the case with one single cluster, Bigot et al. (2013) ensure the identifiability of the common intensity function by assuming a known distribution of time shifts. However, in our setting we need to simultaneously identify representative connecting intensities, cluster memberships, and time shifts. Therefore, additional conditions are required for identifiability, as summarized in Assumption 1 and Proposition 1.

ASSUMPTION 1. Denote true parameters in (1) by $\{v_i^* : i \in \mathcal{V}\}$, $\{f_{q,k}^* : q, k \in [K]\}$ and $\{z_i^* : i \in \mathcal{V}\}$. Let $\mathcal{C}_q^* = \{i \in \mathcal{V} : z_i^* = q\}$ for $q \in [K]$. We assume the true parameters satisfy the following conditions:

(C1) For any two distinct clusters $q, q' \in [K]$, $q \neq q'$, there exists at least one cluster $k \in [K]$ such that $f_{q,k}^* \neq S^w f_{q',k}^*$ for any $w \in \mathbb{R}$.

(C2) For any $q, k \in [K]$, if $f_{q,k}^* \neq 0$, then $\max\{v_i^* : i \in \mathcal{C}_q^*\} > \min\{v_j^* : j \in \mathcal{C}_k^*\}$ and $\min\{v_i^* : i \in \mathcal{C}_q^*\} < \max\{v_j^* : j \in \mathcal{C}_k^*\}$.

(C3) For any $q, k \in [K]$, there exists a sequence of clusters $(q_1, q_2, \dots, q_a) \in [K]^a$ such that $q_1 = q$ and $q_a = k$ and $f_{q_i, q_{i+1}}^* \neq 0$ for $i = 1, \dots, a-1$, where $a \in \mathbb{N}$, $a \geq 2$.

(C4) There exist $q_0, k_0 \in [K]$ such that $\inf\{t \in [0, T] : f_{q_0, k_0}^*(t) > 0\} = 0$.

PROPOSITION 1. Suppose that we observe connecting processes $\{N_{i,j} : i, j \in \mathcal{V}\}$ and that Assumption 1 holds. Then we can verify the following statements:

(P1) The cluster memberships $\{z_i^* : i \in \mathcal{V}\}$ and representative connecting intensities $\{f_{q,k}^* : q, k \in [K]\}$ are identifiable up to a permutation of cluster labels.

(P2) The time shifts $\{v_i^* : i \in \mathcal{V}\}$ are identifiable, except for nodes in the set

(2) $\{i \in \mathcal{V} : v_i^* \leq v_j^* \text{ for any } j \in \mathcal{C}_{z_i^*}^*, \text{ or } v_i^* \leq v_j^* \text{ for any } j \in \mathcal{V} \text{ s.t. } \lambda_{i,j}^* \neq 0\}.$

Assumption 1 lists the technical conditions that ensure the identifiability of the model. Here we briefly explain how conditions in Assumption 1 lead to the identifiability in Proposition 1, and we defer the detailed proof to Section S1 in the Supplementary Material (Zhang and Chen (2023)). First, Condition C1 assumes that any two different clusters are distinct in at least one of their connecting intensities with other clusters, which ensures the identifiability of $\{z_i^* : i \in \mathcal{V}\}$, up to a permutation of cluster labels. Second, Condition C2 requires that the ranges of time shifts overlap for any two clusters with a positive connecting probability. As a result, for any cluster $q \in [K]$, its nonzero connecting intensities $\{f_{q,k}^* : f_{q,k}^* \neq 0, k \in [K]\}$ are identifiable up to a shared time shift. Moreover, Condition C3 grants that, for any pair of nodes, there is a positive probability that these two nodes are directly or indirectly connected. In other words, the network is not separable. Consequently, we can show that the representative connecting intensities of all clusters are identifiable up to a global time shift. Third, Condition C4 pins down the global time shift in a similar way as the zero-sum constraint in Bigot and Gendre (2013). Lastly, the definition of $w_{i,j}$ in (1) implies that there is a set of nodes whose time shifts can not be identified. To see this, consider the case with one cluster. The unidentifiable set (2) reduces to the node with the minimum time shift, which is unidentifiable since the minimum time shift is always masked by other time shifts. We verify in Section 6 that the conditions in Assumption 1 are feasible in real data.

3. Estimation procedure. The estimation of stochastic block models can be performed based on the squared error (Gao, Lu and Zhou (2015), Lei, Chen and Lynch (2020), Pensky (2019)) or the likelihood (Matias, Rebafka and Villers (2018), Pavlović et al. (2020)). We opt to extend the squared-error-based estimation approach to accommodate unknown time shifts, as it allows us to leverage techniques employed in shape invariant models (Bigot and Gendreau (2013), Kneip and Engel (1995)) for fast computing.

In what follows, we represent vectors and matrices in boldface, for example, $\mathbf{f} \equiv (f_{q,k})_{q,k \in [K]}$, $\mathbf{z} \equiv (z_i)_{i \in \mathcal{V}}$, $\mathbf{v} \equiv (v_i)_{i \in \mathcal{V}}$, and let $L_+^2([0, T])$ denote the space of nonnegative squared integrable functions on $[0, T]$. Recall that in our model, we only observe the connecting processes $\{N_{i,j} : i, j \in \mathcal{V}\}$. The optimization problem associated with our model takes the following form:

$$(3) \quad (\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}) = \arg \min_{\Theta_1} \sum_{i,j \in \mathcal{V}, i \neq j} \left\{ \int_0^T S^{w_{i,j}} f_{z_i, z_j}^2(t) dt - 2 \int_0^T S^{w_{i,j}} f_{z_i, z_j}(t) dN_{i,j}(t) \right\},$$

where $w_{i,j} = \max(v_i, v_j)$ and $\Theta_1 \equiv \{(\mathbf{f}, \mathbf{z}, \mathbf{v}) : \mathbf{f} \in \mathcal{B}_1^{K \times K}, \mathbf{z} \in [K]^p, \mathbf{v} \in [0, T]^p\}$ with $\mathcal{B}_1 \equiv \{f \in L_+^2([0, T]) : \|f\|_1 \leq 1\}$. The objective function in (3) is identical to the sum of squared deviations of the observed connecting processes from their intensities up to a constant. To estimate $\mathbf{f}$ nonparametrically, we employ a truncated basis expansion. In particular, we consider the functional space with truncated Fourier bases $\mathcal{B}_{1,\ell_0} \equiv \{f \in \mathcal{B}_1 : \int_0^T \exp\{-i2\pi \ell t/T\} f(t) dt = 0 \text{ for } |\ell| \geq \ell_0\}$. The selection of the tuning parameter ℓ_0 is discussed later in this section.

We notice that replacing the intensities $\{f_{q,k} : q, k \in [K]\}$ with the cumulative intensities $\{F_{q,k} : F_{q,k}(t) \equiv \int_0^t f_{q,k}(s) ds, t \in [0, T], q, k \in [K]\}$ can avoid the frequency truncation. In particular, the optimization problem based on cumulative intensities takes the following form:

$$(4) \quad (\hat{\mathbf{F}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}) = \arg \min_{\Theta_2} \sum_{i,j \in \mathcal{V}, i \neq j} \int_0^T \{S^{w_{i,j}} F_{z_i, z_j}^2(t) - 2S^{w_{i,j}} F_{z_i, z_j}(t) N_{i,j}(t)\} dt,$$

where $w_{i,j} = \max(v_i, v_j)$ and $\Theta_2 \equiv \{(\mathbf{F}, \mathbf{z}, \mathbf{v}) : \mathbf{F} \in \mathcal{B}_2^{K \times K}, \mathbf{z} \in [K]^p, \mathbf{v} \in [0, T]^p\}$ with $\mathcal{B}_2 \equiv \{F \in [0, 1]^{[0, T]} : F \text{ is nondecreasing and right-continuous, } F(0) = 0\}$. The objective function in (4) is almost identical to the one in (3), except for the use of $N_{i,j}$ and F_{z_i, z_j} . In (4) a natural nonparametric estimator of $\mathbf{F}$ is the empirical cumulative intensity, which does not require frequency truncation.

However, the objective functions based on the plain vanilla L^2 -distance in (3) and (4) may tend to overlook the information in the *shapes* of intensities. To see this, consider a simple case where we assign one node to one of K clusters with known intensities and no time shifts. In this case, problem (4) reduces to

$$(5) \quad \hat{z}_1 = \arg \min_{z_1 \in [K]} \sum_{k \in [K]} |C_k^*| \int_0^T |F_{z_1, k}^*(t) - \bar{N}_{1, k}(t)|^2 dt,$$

where $\{C_k^* : k \in [K]\}$ are known clusters for nodes in $\mathcal{V} \setminus \{1\}$, $\{F_{z_1, k}^* : z_1, k \in [K]\}$ are known cumulative intensities across clusters, and $\{\bar{N}_{1, k} : \bar{N}_{1, k} \equiv |C_k^*|^{-1} \sum_{j \in C_k^*} N_{1, j}, k \in [K]\}$ are empirical cumulative intensities. The cumulative intensity, for example, $F_{z_1, k}^*(t)$, can be broken down into two parts: the cumulative distribution $F_{z_1, k}^*(t)/F_{z_1, k}^*(T)$ and the connecting probability $F_{z_1, k}^*(T)$. In a sense the two parts can be seen as the shape and the scale of the cumulative intensity. As shown in Figure 3(a), the minimizer $\hat{z}_1$ of the objective function in (5) might be predominantly decided by scales. Moreover, the problem is exacerbated when time shifts are unknown, as in (4). Figure 3(b) shows that the minimum might be achieved by over estimating the time shifts.

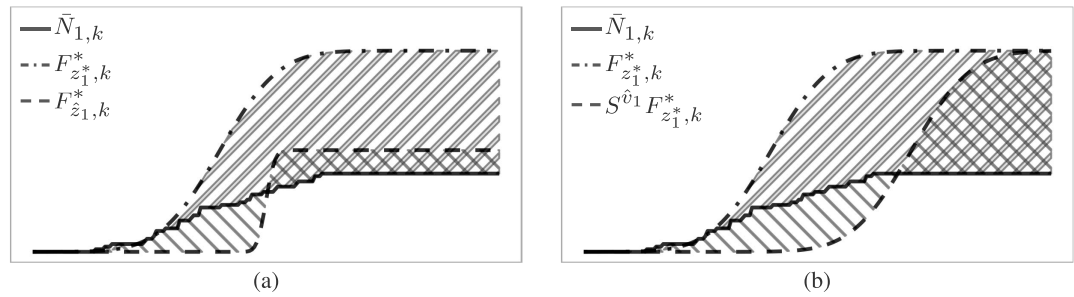


FIG. 3. Deviation of the empirical cumulative intensity from the true cumulative intensity and from the minimizer-induced cumulative intensities. Panel (a) shows that the objective function may select wrong cluster $\hat{z}_1$ due to the similarity in scales, despite the difference in shapes. Panel (b) shows that $\hat{v}_1$, which minimizes the objective function in (4), may shift the cumulative intensity away from the truth.

To address this issue, we propose a new loss function that separates the shape and scale. We modify (4) as

$$\begin{aligned} (\hat{\mathbf{F}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}) &= \arg \min_{\Theta_C} M_C(\mathbf{F}, \mathbf{z}, \mathbf{v}; \gamma) \\ (6) \quad &\equiv \arg \min_{\Theta_C} \left[\sum_{(i,j) \in \mathcal{E}(T)} \frac{1}{T} \int_0^T \left\{ \left| \frac{S^{w_{i,j}} F_{z_i, z_j}(t)}{F_{z_i, z_j}(T)} \right|^2 - 2N_{i,j}(t) \frac{S^{w_{i,j}} F_{z_i, z_j}(t)}{F_{z_i, z_j}(T)} \right\} dt \right. \\ &\quad \left. + \gamma \sum_{i,j \in \mathcal{V}, i \neq j} |N_{i,j}(T) - F_{z_i, z_j}(T)|^2 \right], \end{aligned}$$

where $\gamma \geq 0$ is a user-specified parameter, $\mathcal{E}(T) = \{(i, j) : i, j \in \mathcal{V}, N_{i,j}(T) = 1\}$, $w_{i,j} = \max(v_i, v_j)$, and $\Theta_C \equiv \{(\mathbf{F}, \mathbf{z}, \mathbf{v}) : \mathbf{F} \in \mathcal{B}_2^{K \times K}, \mathbf{z} \in [K]^p, \mathbf{v} \in [0, T]^p, F_{z_i, z_j}(T) > 0 \text{ if } (i, j) \in \mathcal{E}(T)\}$. The objective function in (6) has two components. The first component is the sum of squared deviations related to cumulative connecting distributions (i.e., shapes), and the second component is the sum of squared deviations related to connecting probabilities (i.e., scales). The relative weight of shapes and scales is governed by γ . The selection of γ is discussed later in this section. The estimator in (6) is based on the cumulative connecting intensities; thus, we refer to it as *SidSBM-C*.

In a similar manner, we introduce *SidSBM-P*, an intensity-based estimator that solves the following optimization problem:

$$\begin{aligned} (\hat{\mathbf{f}}, \hat{\mathbf{z}}, \hat{\mathbf{v}}) &= \arg \min_{\Theta_P} M_P(\mathbf{f}, \mathbf{z}, \mathbf{v}; \gamma) \\ (7) \quad &\equiv \arg \min_{\Theta_P} \left[\sum_{(i,j) \in \mathcal{E}(T)} \frac{1}{T} \left\{ \int_0^T \left| \frac{S^{w_{i,j}} f_{z_i, z_j}(t)}{F_{z_i, z_j}(T)} \right|^2 dt - 2 \int_0^T \frac{S^{w_{i,j}} f_{z_i, z_j}(t)}{F_{z_i, z_j}(T)} dN_{i,j}(t) \right\} \right. \\ &\quad \left. + \gamma \sum_{i,j \in \mathcal{V}, i \neq j} |N_{i,j}(T) - F_{z_i, z_j}(T)|^2 \right], \end{aligned}$$

where γ , $\mathcal{E}(T)$, $w_{i,j}$ are defined the same way as in (6), $F_{z_i, z_j}(T) = \int_0^T f_{z_i, z_j}(s) ds$, and $\Theta_P \equiv \{(\mathbf{f}, \mathbf{z}, \mathbf{v}) : \mathbf{f} \in \mathcal{B}_{1, \ell_0}^{K \times K}, \mathbf{z} \in [K]^p, \mathbf{v} \in [0, T]^p, f_{z_i, z_j} \not\equiv 0 \text{ if } (i, j) \in \mathcal{E}(T)\}$.

We adapt the *integrated classification likelihood* (ICL) criterion (Biernacki, Celeux and Govaert (2000), Daudin, Picard and Robin (2008)) to select tuning parameters including the number of clusters K , the relative importance of intensity scales γ , and the frequency truncation parameter ℓ_0 . Integrated classification likelihood is based on a BIC-like approximation

that has been adapted to semiparametric stochastic block models by [Matias, Rebafka and Villers \(2018\)](#). For our model the integrated classification likelihood criterion takes the following form:

$$(8) \quad \begin{aligned} \text{ICL} = & \sum_{(i,j) \in \mathcal{E}_0} \log\{1 - \hat{F}_{\hat{z}_i, \hat{z}_j}(T)\} + \sum_{(i,j) \in \mathcal{E}_1} \log\{\hat{f}_{\hat{z}_i, \hat{z}_j}(t_{i,j})\} + \sum_{k \in [K]} |\hat{\mathcal{C}}_k| \log(p^{-1} |\hat{\mathcal{C}}_k|) \\ & - \frac{1}{2}(K-1) \log p - \frac{1}{4}K(K+1)\ell_0 \log \frac{p(p-1)}{2}, \end{aligned}$$

where $\mathcal{E}_0 \equiv \{(i, j) : i, j \in \mathcal{V}, i < j, N_{i,j}(T) = 0\}$ and $\mathcal{E}_1 \equiv \{(i, j) : i, j \in \mathcal{V}, i < j, N_{i,j}(T) = 1\}$. The sum of the first three terms can be seen as the complete-data log-likelihood, and the last two terms are penalties for cluster memberships and connecting intensities. In the second penalty term, we use ℓ_0 to represent the degrees of freedom in each connecting intensity. For optimization problem (6), we select K and γ by maximizing ICL with $\ell_0 = 2$, where the two degrees of freedom correspond to the shape and the scale. For optimization problem (7), we select K , γ , and ℓ_0 by maximizing ICL. In practice, heuristic methods can be employed for computationally efficient selection of tuning parameters. Further details are provided in Section S4.1 in the Supplementary Material ([Zhang and Chen \(2023\)](#)).

REMARK 1. We can generalize the proposal to networks with recurrent edges, where the edges are events that can happen more than one time. Instances of such networks include, but are not limited to, email exchange networks, payment transaction networks, and bike sharing networks (see, e.g., [Diesner and Carley \(2005\)](#), [Lin, He and Peeta \(2018\)](#), [Loupos, Nathan and Cerf \(2019\)](#)). When the connecting intensities $\mathbf{f}$ have bounded supports, the optimization problems (6) and (7) can be directly applied without modification, and the identifiability results in Section 2 hold. When $\mathbf{f}$ have unbounded supports, identifiability requires additional assumptions on the intensities $\mathbf{f}$. For instance, we can assume that intensities in $\mathbf{f}$ are periodic, and a full period is observed for each intensity in $\mathbf{f}$ (see, e.g., [Bigot et al. \(2013\)](#), [Bigot and Gendre \(2013\)](#)). Furthermore, the optimization problem (7) can be applied directly, when the intensity is periodic.

4. Algorithm. We now present an algorithm for solving the optimization problem (6). The key observation in (6) is that it is straightforward to solve one set of parameters conditioning on the others. The resulting algorithm resembles the classic k-means algorithm that iterates between the centering and clustering steps. In particular, we can write the two steps as follows:

$$(9) \quad (\text{centering step}) \quad \hat{\mathbf{F}}, \hat{\mathbf{v}} = \arg \min_{\mathbf{F}, \mathbf{v}} M_{\mathbf{C}}(\mathbf{F}, \hat{\mathbf{z}}, \mathbf{v}; \gamma),$$

$$(10) \quad (\text{clustering step}) \quad \hat{\mathbf{z}} = \arg \min_{\mathbf{z}} M_{\mathbf{C}}(\hat{\mathbf{F}}, \mathbf{z}, \hat{\mathbf{v}}; \gamma).$$

In the centering step (9), we update $\hat{\mathbf{F}}$ and $\hat{\mathbf{v}}$, conditioning on the current estimates of cluster memberships $\hat{\mathbf{z}}$. Given $\hat{\mathbf{z}}$, the optimization problem (6) reduces to a well-studied problem in the literature of shape invariant models ([Bigot and Gadat \(2010\)](#), [Bigot et al. \(2013\)](#), [Bigot and Gendre \(2013\)](#)). The reduced optimization problem can be decoupled into two parts concerning $\hat{\mathbf{F}}$ and $\hat{\mathbf{v}}$, respectively. In the first part, we have the analytical solution of $\hat{F}_{q,k}$ as

$$(11) \quad \hat{F}_{q,k} = \sum_{i \in \hat{\mathcal{C}}_q, j \in \hat{\mathcal{C}}_k, i \neq j} \frac{S^{-\max(\hat{v}_i, \hat{v}_j)} N_{i,j}}{|\hat{\mathcal{C}}_q| |\hat{\mathcal{C}}_k| - \mathbf{I}(q=k) |\hat{\mathcal{C}}_q|}, \quad q, k \in [K],$$

where $\hat{C}_k \equiv \{i \in \mathcal{V} : \hat{z}_i = k\}$. The estimator in (11) is the empirical cumulative intensity of edges between clusters q and k , where self-edges are eliminated if $q = k$.

In the second part, we solve for $\hat{\mathbf{v}}$ via

$$(12) \quad \hat{\mathbf{v}} = \arg \min_{\mathbf{v}} \sum_{i \in \mathcal{V}} \sum_{k \in [K]} \left\{ |\hat{J}_{i,k}| \bar{N}_{i,k}(T) \frac{1}{T} \int_0^T \left| \frac{\bar{N}_{i,k}(t)}{\bar{N}_{i,k}(T)} - \frac{S^{v_i} \hat{F}_{\hat{z}_i,k}(t)}{\hat{F}_{\hat{z}_i,k}(T)} \right|^2 dt \right\},$$

where $\hat{J}_{i,k} \equiv \{j \in \mathcal{V} : \hat{z}_j = k, j \neq i, v_j \leq v_i\}$ is the collection of nodes in the k th cluster whose time shifts are no larger than node i , and $\bar{N}_{i,k}(t) \equiv |\hat{J}_{i,k}|^{-1} \sum_{j \in \hat{J}_{i,k}} N_{i,j}(t)$ is the empirical cumulative intensity of edges between node i and nodes in $\hat{J}_{i,k}$. The optimization problem in (12) can be solved by utilizing the well-known Parseval's theorem to reframe the problem in the Frequency domain. The Frequency domain offers a suitable environment for the application of the gradient descent algorithm, which provides a solution to the problem. Details of the gradients are available in Section S3 in the Supplementary Material (Zhang and Chen (2023)).

Finally, we set the global shift as $\hat{c} \equiv \min\{t_{i,j} - \max(\hat{v}_i, \hat{v}_j) : i, j \in \mathcal{V}\}$, which is an approximation of $\inf\{t \in [0, T] : d\hat{F}_{q,k}(t) > 0 \text{ for some } q, k \in [K]\}$. Specifically, we update $\hat{\mathbf{F}}$ and $\hat{\mathbf{v}}$ as follows:

$$(13) \quad \hat{F}_{q,k} \leftarrow S^{-\hat{c}} \hat{F}_{q,k}, \quad q, k \in [K]; \quad \hat{v}_i \leftarrow \hat{v}_i + \hat{c}, \quad i \in \mathcal{V}.$$

The global shift $\hat{c}$ is the minimum of p^2 numbers. As such, for small values of p , $\hat{c}$ might be subject to a large error, which might further compromise the accuracy of $\hat{\mathbf{F}}$ and $\hat{\mathbf{v}}$.

In the clustering step (10), we update each cluster membership $\hat{z}_i$ conditioning on current $\hat{\mathbf{F}}$, $\hat{\mathbf{v}}$ and $\{\hat{z}_j : j \in \mathcal{V}, j \neq i\}$. The optimization problem in (6) reduces to, for $i \in \mathcal{V}$,

$$(14) \quad \hat{z}_i = \arg \min_{z_i \in [K]} \sum_{k \in [K]} \left\{ |\hat{C}_{k,-i}| \bar{F}_{i,k}(T) \frac{1}{T} \int_0^T \left| \frac{\bar{F}_{i,k}(t)}{\bar{F}_{i,k}(T)} - \frac{\hat{F}_{z_i,k}(t)}{\hat{F}_{z_i,k}(T)} \right|^2 dt \right. \\ \left. + \gamma |\hat{C}_{k,-i}| |\bar{F}_{i,k}(T) - \hat{F}_{z_i,k}(T)|^2 \right\},$$

where $\hat{C}_{k,-i} \equiv \hat{C}_k \setminus \{i\}$ and $\bar{F}_{i,k}(\cdot) \equiv |\hat{C}_{k,-i}|^{-1} \sum_{j \in \hat{C}_{k,-i}} S^{-\max(\hat{v}_i, \hat{v}_j)} N_{i,j}(\cdot)$. The objective function in (14) can be seen as a “distance” between node i and cluster $\hat{C}_{z_i}$ that involves differences in both shapes and scales.

The optimization problem in (6) is nonconvex; hence, the initialization scheme and the convergence criterion are crucial for the performance of the algorithm (see Remarks 2 and 3). The estimation procedure is summarized in Algorithm 1.

The algorithm to solve the optimization problem (7) closely resembles Algorithm 1. The detailed algorithm is provided in Section S2 of the Supplementary Material (Zhang and Chen (2023)). The objective function in (7) might suffer from plateaus in v , while the objective function in (6) does not encounter this issue due to the utilization of cumulative intensities (see Figure S1 in the Supplementary Material, Zhang and Chen (2023)).

REMARK 2. We initialize time shifts by the time of the first edge associated with each node, that is,

$$(15) \quad \hat{v}_i^{(0)} \equiv \min\{t_{i,j} : j \in \mathcal{V}\}, \quad i \in \mathcal{V}.$$

Letting $F_i(t|\mathbf{v}) \equiv (p-1)^{-1} \sum_{j \in \mathcal{V} \setminus \{i\}} S^{-\max(v_i, v_j)} N_{i,j}(t)$, we initialize cluster memberships by solving the following k-medoids problem:

$$(16) \quad \hat{\mathbf{z}}^{(0)}, \hat{\boldsymbol{\mu}}^{(0)} = \arg \min_{\mathbf{z} \in [K]^p, \boldsymbol{\mu} \in \mathcal{C}_1 \times \dots \times \mathcal{C}_K} \sum_{i \in \mathcal{V}} \int_0^T \left| \frac{F_i(t|\hat{\mathbf{v}}^{(0)})}{F_i(T|\hat{\mathbf{v}}^{(0)})} - \frac{F_{\mu_{z_i}}(t|\hat{\mathbf{v}}^{(0)})}{F_{\mu_{z_i}}(T|\hat{\mathbf{v}}^{(0)})} \right|^2 dt,$$

Algorithm 1: Iterative algorithm for SidSBM-C

Input: $\mathbf{N}, K, \varepsilon, \gamma$
Initialize $\hat{\mathbf{v}}^{(0)}, \hat{\mathbf{z}}^{(0)}$ via (15) (16), and $s = 0$;
while not converge **do**
 Update $\hat{\mathbf{v}}^{(s+1)}$ and $\hat{\mathbf{F}}^{(s+1)}$ via (11)–(13) with $(\mathbf{z}, \mathbf{v}) = (\hat{\mathbf{z}}^{(s)}, \hat{\mathbf{v}}^{(s)})$;
 Update $\hat{\mathbf{z}}^{(s+1)}$ via (14) with $(\mathbf{v}, \mathbf{F}, \mathbf{z}) = (\hat{\mathbf{v}}^{(s+1)}, \hat{\mathbf{F}}^{(s+1)}, \hat{\mathbf{z}}^{(s)})$;
 Evaluate the stopping criterion via (18);
 $s = s + 1$;
end
Output: $\hat{\mathbf{F}}^{(s)}, \hat{\mathbf{z}}^{(s)}, \hat{\mathbf{v}}^{(s)}$.

where $\{\mu_q : q \in [K]\}$ are the medoids of the K clusters. Optimization problem (16) is a simplification of problem (6), where $\mathbf{v}$ is given as $\hat{\mathbf{v}}^{(0)}$, $\{F_{q,k}(t) : q, k \in [K]\}$ take values in the known set $\{F_i(t|\hat{\mathbf{v}}^{(0)}) : i \in \mathcal{V}\}$, and γ is set as zero. In practice, $F_i(t|\hat{\mathbf{v}}^{(0)})$ is evaluated on an equispaced time grid as an L -vector

$$(17) \quad \frac{1}{p-1} \sum_{j \in \mathcal{V} \setminus \{i\}} \left(\mathbf{I} \left[\tilde{t}_{i,j} \leq \frac{T}{L} \right], \mathbf{I} \left[\tilde{t}_{i,j} \leq \frac{2T}{L} \right], \dots, \mathbf{I} \left[\tilde{t}_{i,j} \leq T \right] \right)^\top,$$

where $\tilde{t}_{i,j} \equiv t_{i,j} - \max(\hat{v}_i^{(0)}, \hat{v}_j^{(0)})$ is the adjusted connecting time. The tuning parameter L can be chosen such that T/L is close to temporal resolution of the data. We note that the proposed initialization scheme can not guarantee a global minimum. If feasible, we recommend to use multiple random restarts by jittering the initial time shifts and shuffling the initial cluster memberships for a small proportion of nodes.

REMARK 3. The proposed algorithm is stopped when, for a user-specified ε ,

$$(18) \quad \left[\{1 - \text{ARI}(\hat{\mathbf{z}}^{(s)}, \hat{\mathbf{z}}^{(s+1)})\} + \frac{\|\hat{\mathbf{v}}^{(s)} - \hat{\mathbf{v}}^{(s+1)}\|_2}{\|\hat{\mathbf{v}}^{(s)}\|_2} + \frac{\|\hat{\mathbf{F}}^{(s)} - \hat{\mathbf{F}}^{(s+1)}\|_2}{\|\hat{\mathbf{F}}^{(s)}\|_2} \right] \leq \varepsilon,$$

where $\|\cdot\|_2$ is the L^2 -norm and $\hat{\mathbf{z}}^{(s)}, \hat{\mathbf{v}}^{(s)}, \hat{\mathbf{F}}^{(s)}$ are estimates in s th iteration. This criterion measures the overall change of model parameters between two consecutive iterations. We use the change of model parameters rather than the change of loss function because the stabilization of model parameters can ensure the stabilization of loss function. The change of cluster memberships is measured using the Adjusted Rand Index (ARI) (Hubert and Arabie (1985)). In particular, for two partitions $\mathcal{C} \equiv \{\mathcal{C}_q : q \in [K]\}$ and $\mathcal{C}' \equiv \{\mathcal{C}'_k : k \in [K']\}$ such that $\mathcal{V} = \bigcup_{q \in [K]} \mathcal{C}_q = \bigcup_{k \in [K']} \mathcal{C}'_k$, denoting $a_q \equiv |\mathcal{C}_q|$, $b_k \equiv |\mathcal{C}'_k|$, and $d_{qk} \equiv |\mathcal{C}_q \cap \mathcal{C}'_k|$ for $q \in [K]$ and $k \in [K']$, then

$$(19) \quad \text{ARI} \equiv \frac{\sum_{q,k} \binom{d_{qk}}{2} - [\sum_q \binom{a_q}{2} \sum_k \binom{b_k}{2}] / \binom{p}{2}}{\frac{1}{2} [\sum_q \binom{a_q}{2} + \sum_k \binom{b_k}{2}] - [\sum_q \binom{a_q}{2} \sum_k \binom{b_k}{2}] / \binom{p}{2}}.$$

If the two partitions are identical up to a permutation of cluster labels, then ARI equals one. If one partition is purely random and the other one is fixed, then the expectation of ARI is zero.

5. Simulation.

5.1. *Data generation and model specification.* Random dynamic networks are generated from the proposed model. In particular, cluster memberships are generated by se-

TABLE 1
Mean (μ) and variance (σ^2) of the connecting intensities for generating synthetic data. Each column corresponds to one pair of (q, k) . The value of β is specified in Section 5.2

	(1, 1)	(1, 2)	(1, 3)	(2, 2)	(2, 3)	(3, 3)
$\mu_{q,k}$	20	20β	$20\beta^2$	$20\beta^2$	$20\beta^{1/2}$	$20\beta^{3/2}$
$\sigma_{q,k}^2$	100	$100\beta^{-2}$	$100\beta^{-2}$	100β	$100\beta^{-1}$	100β

quentially allocating p nodes into $K = 3$ clusters of equal size, that is, $\mathcal{C}_q = \{(p/K)(q - 1) + 1, \dots, (p/K)q\}$ for $q \in [K]$. For any $q, k \in [K]$, the representative connecting intensity between clusters q and k takes the form $f_{q,k} = 0.9 \times \Gamma(\mu_{q,k}^2/\sigma_{q,k}^2, \mu_{q,k}/\sigma_{q,k}^2)$, where $\Gamma(\mu_{q,k}^2/\sigma_{q,k}^2, \mu_{q,k}/\sigma_{q,k}^2)$ is the density of a Gamma distribution with mean $\mu_{q,k}$ and variance $\sigma_{q,k}^2$. Specifications of $\mu_{q,k}$'s and $\sigma_{q,k}^2$'s are provided in Table 1, where we introduce β to control the separability between representative connecting intensities. Time shifts are generated from a uniform distribution that $v_i \sim U(0, W)$, $i \in \mathcal{V}$. The connecting time between any nodes i and j is generated as $t_{i,j} = X_{i,j} + \max(v_i, v_j)$, where $X_{i,j}$ is drawn from f_{z_i, z_j} for any pair (i, j) .

By altering p , β , and W , we generate data with varying signal strengths. Intuitively, the signal strength is positively associated with p and β , since increasing p increases the sample size, and increasing β increases distinction among $(\mu_{q,k}, \sigma_{q,k}^2)$'s; the signal strength might decrease when W increases as W controls the variation in time shifts.

We compare the performance of three methods, SidSBM-C (6), SidSBM-P (7), and the Poisson process stochastic block model (PPSBM) by Matias, Rebafka and Villers (2018). We apply SidSBM-C with $K = 3$, $\epsilon = 0.01$, $\gamma = 0.01$ and $L = 200$. The same set of tuning parameters is used for SidSBM-P except for $\gamma = 0.0001$ and $\ell_0 = 4$. We apply PPSBM using the implementation in R by Giorgi et al. (2018) with the number of bins being 64.

5.2. *Simulation results.* We use two criteria to evaluate the performance of three methods described in Section 5.1. To measure the clustering performance, we calculate the ARI introduced in (19). To measure the estimation performance, we calculate the mean integrated squared error (MISE) between the estimated and the true connecting intensities

(20)
$$\text{MISE} = \frac{1}{(K^2 + K)/2} \sum_{q,k \in [K], q \leq k} \mathbb{E} \|\hat{f}_{q,k} - f_{q,k}\|_2^2,$$

where $\|\cdot\|_2$ is the L^2 -norm. Note that SidSBM-C estimates cumulative intensities rather than intensities. In order to make a fair comparison, we estimate intensities for SidSBM-C using Gaussian kernel density estimators with bandwidth selected as suggested in Silverman (2017).

Simulation results under various settings are provided in Figure 4. In all plots a smaller y-coordinate indicates a better performance. Across all settings, both proposed methods outperform PPSBM, because PPSBM suffers from the presence of unknown time shifts. In addition, SidSBM-C and SidSBM-P exhibit comparable performance, except when β reaches high values. This discrepancy may be caused by the limited ranges of supports of intensities when the value of β is high as well as the large values of W . Supporting this hypothesis, Figure S5 in the Supplementary Material (Zhang and Chen (2023)) demonstrates that a reduction in W results in a corresponding decrease of the discrepancy between the performance of SidSBM-C and SidSBM-P. Interestingly, both proposed methods still outperform the PPSBM in settings without time shifts (see Figure S8 in the Supplementary Material, Zhang and Chen (2023)).

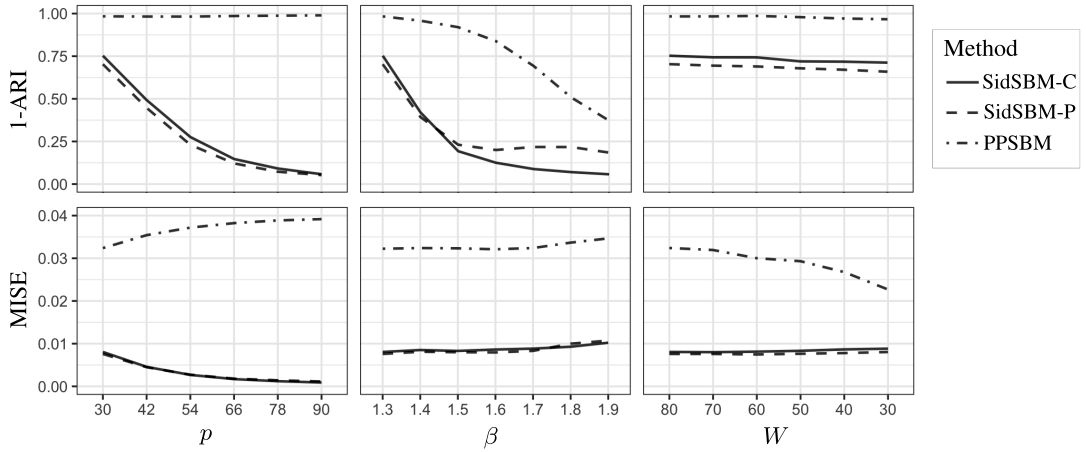


FIG. 4. Performance of SidSBM-C, SidSBM-P and PPSBM in simulation with 1500 replicates. Simulated data is generated under various p , β , W with default setting being $p = 30$, $\beta = 1.3$, and $W = 80$. The y-axis is 1-ARI in the first row and MISE in the second row, where smaller values indicate better performance. The proposals outperform PPSBM in all settings due to their effectiveness in handling time shifts. Between the two proposals, SidSBM-C and SidSBM-P have similar performance, except for when $\beta > 1.5$ due to limited intensity support ranges and large time shifts.

Furthermore, Figure 4 shows that the overall performance of the three methods improves as signal strengths increase. However, we note that there are four seemingly counterintuitive observations. First, as p increases, PPSBM has stable ARI and increasing MISE. This is because PPSBM is not able to handle time shifts, thus will converge to erroneous estimators as p increases. Second, increasing β enlarges MISE of the three methods, due to the increasing average L^2 -norm of the representative connecting intensities. Third, the proposed methods have stable performance as W decreases because estimators of cluster memberships and connecting intensities in proposed methods are invariant to time shifts. Fourth, the MISE values appear to exhibit limited sensitivity to improved clustering performance. This pattern can be attributed to several factors, including the small default values of p and β in the simulation setting as well as the wide range of the y-axis in the figure. In general cases the MISE value should exhibit sensitivity to improvements in clustering performance. Further supporting evidence is provided in Section S4.4 in the Supplementary Material (Zhang and Chen (2023)).

The computing time of the three methods under consideration is shown in Figure 5(a). The proposed methods outperform PPSBM in terms of computing time as well as scalability. In particular, the computing time of both SidSBMs seems to increase sublinearly with respect to the size of network p . To explain this phenomenon, we break down the total computing time into the number of iterations between the centering and clustering steps, and the computing time per iteration shown in Figure 5(b) and (c). We find that the computing time per iteration is roughly linear in p , whereas the number of iterations decreases as p increases for $p \geq 60$. This is a promising sign that the proposed methods can potentially handle large networks with the computing cost being linear in the size of network.

One computational concern in nonconvex optimization is the requirement of multiple initializations to achieve global minimum. During our simulation experiment, we observe that the initialization scheme in Remark 2 might lead to local minimums, for example, empty clusters. Therefore, when the size of any estimated cluster is less than $p/10$, we restart the algorithm by jittering the initial values, as discussed in Remark 2. In Figure 6(a) we see that the proposed initialization scheme offers comparable performance to random initialization with three restarts and that keep increasing the number of random restarts can only result in diminishing improvement in performance. Furthermore, the proposed initialization scheme

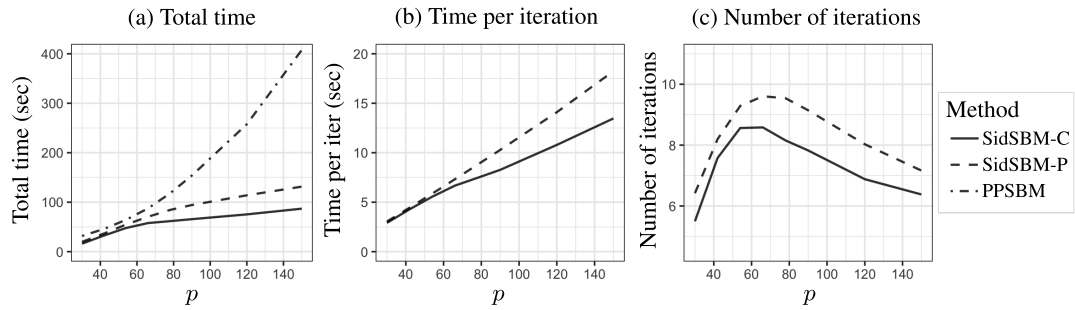


FIG. 5. Computing time of SidSBM-C, SidSBM-P and PPSBM in simulation with 1500 replicates. Simulated data is generated under $\beta = 1.3$, $W = 80$, and a range of p . The proposals are computationally more efficient than PPSBM and might be scalable to large networks.

requires only 1.07 restarts on average across the 1500 replicates, which is computationally more efficient than random initialization with three or 10 restarts.

Finally, we demonstrate the performance of integrated classification likelihood (ICL) criterion in Figure 6(b). We observe that, for both SidSBMs, ICL values reach maximum at the correct number of clusters. More specifically, for SidSBM-C, ICL correctly selects $\hat{K} = 3$ in all 1500 replicates, and for SidSBM-P, ICL selects $\hat{K} = 3$ in 84.6% of the 1500 replicates.

6. Real data application. We utilize the proposed model to analyze the neural data set from Wan et al. (2019). Wan and colleagues recorded the development of motor circuits in the spinal cord of zebrafish. The developmental process was observed from approximately 17.5 hours to 22 hours postfertilization. In the data set that we analyze, a total of 96 neurons are observed in the spinal cord. The spatial distribution of these neurons is depicted in Figure S11 in Supplementary Material (Zhang and Chen (2023)). Among the observed neurons,

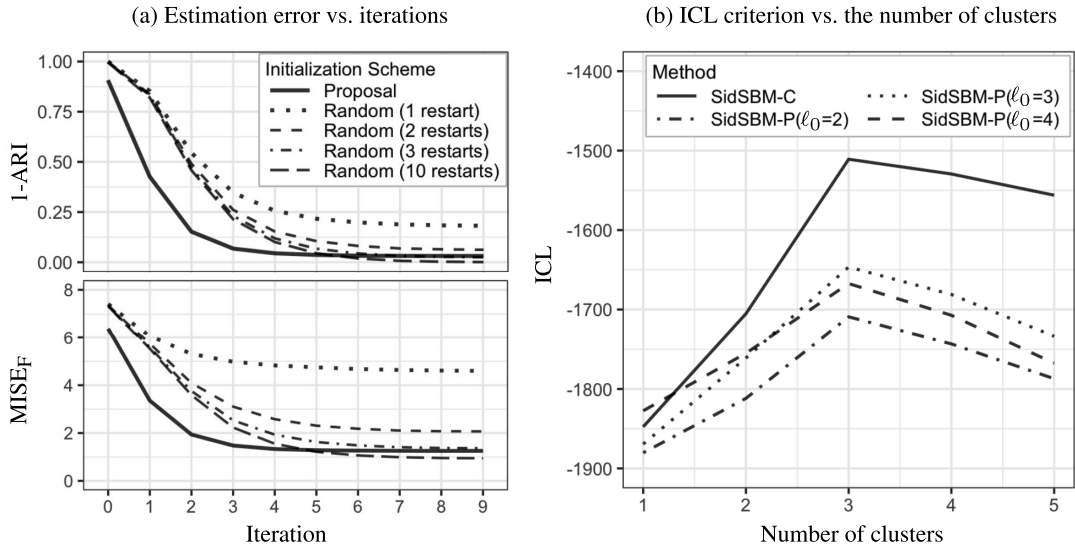


FIG. 6. Performance of proposed initialization scheme and model selection criterion (i.e., ICL) in simulation with over 1500 replicates. Simulated data is generated under the setting with $p = 30$, $\beta = 1.9$, and $V = 80$. Panel (a) displays trajectories of 1-ARI and MISE as iterations proceed, using proposed initialization or random initialization with restarts. When multiple restarts are applied, the best result is selected using the objective function in (6). The performance of proposed initialization scheme is comparable to random initialization with 10 restarts. Panel (b) displays values of the ICL criterion for various number of clusters. ICL is able to select the correct number of clusters, that is, $K = 3$.

40 neurons are located in the left spine, while the remaining 56 neurons are located in the right spine. The activities and locations of neurons were recorded with calcium imaging and processed by Wan et al. (2019). During the recorded period, neurons displayed increasing synchronization over time that indicated the functional maturation of the motor circuit. At around 22 hours postfertilization, the motor circuit was able to instruct early motor behaviors. We refer interested readers to Wan et al. (2019) for more experiment details. In this analysis we aim to learning the roles of neurons during motor circuit development.

We apply SidSBM-C to the recorded development of neural networks where nodes represent neurons and edges represent synchronization between two neurons. Roughly speaking, we define the connecting time between neurons (i.e., $t_{i,j}$'s) to be the time point when the two neurons have correlated activities. Details of preprocessing are available in Section S5 in Supplementary Material (Zhang and Chen (2023)). Neural networks from the left and right spines are treated as two separate networks in our analysis. To be specific, we apply SidSBM-C with $\varepsilon = 10^{-3}$, $\gamma = 0.1$ and $L = 336$ on the two networks separately, with estimates $(\hat{\mathbf{f}}_L, \hat{\mathbf{z}}_L, \hat{\mathbf{v}}_L)$ and $(\hat{\mathbf{f}}_R, \hat{\mathbf{z}}_R, \hat{\mathbf{v}}_R)$. Here the value of γ is selected using ICL, and it is noteworthy that the data analysis results remain unchanged while γ ranges from 0.01 to 0.3. We simultaneously select the number of clusters (i.e., K) for both spines using the sum of ICLs for the two networks. Each fit of the method takes about 30 seconds (using a machine with an AMD Epyc 7351 processor, 4 GB RAM, R 3.6.3). When accounting for the selection of both tuning parameters (i.e., $\gamma \in \{0.01, 0.03, 0.1, 0.3\}$ and $K \in \{2, 3, 4, 5\}$), the overall computational time for the obtained real data results is approximately eight minutes. Recall that Proposition 1 shows that cluster memberships and intensities are *identifiable up to a permutation* of cluster labels. Hence, we match the cluster memberships for the left spine (i.e., $\hat{\mathbf{z}}_L$) and the right spine (i.e., $\hat{\mathbf{z}}_R$) by maximizing the sum of cross-correlations between $\hat{\mathbf{f}}_L$ and $\hat{\mathbf{f}}_R$.

Figure 7 shows that estimated representative connecting intensities for left and right spines, that is, $\hat{\mathbf{f}}_L$ and $\hat{\mathbf{f}}_R$, are highly consistent, although the estimation was done separately on the two networks. In particular, intensities $(\hat{f}_{L,1,1}, \hat{f}_{L,1,2}, \hat{f}_{L,2,2})$ and $(\hat{f}_{R,1,1}, \hat{f}_{R,1,2}, \hat{f}_{R,2,2})$ are almost identical. Furthermore, intensities $\hat{f}_{L,3,k}$ and $\hat{f}_{R,3,k}$ have similar shapes for $k = 1, 2, 3$ despite having a time lag (see Figure S15 in Supplementary Material, Zhang and Chen (2023)). We hypothesize that the time lag might be due to estimation errors, given the large dispersion of $\hat{f}_{3,k}$'s and the small sizes of the third clusters. Since the two networks are from the same zebrafish, it is natural to expect consistency between them. However, it is worthwhile to emphasize that the estimation procedure does not enforce any constraints to encourage consistency between estimated representative connecting intensities of the two networks. The consistent representative connecting intensities from two sides of spine are likely reflecting the true mechanism of neural circuit development.

In addition, time shifts of neurons in the same cluster can be very different. Indeed, time shifts seem to be positively correlated with the AP-coordinates of neurons, as shown in Figure 8(a). Methods that do not account for time shifts may group neurons with similar time shifts into the same cluster, leading to less consistent estimates of the intensities between the left and right spines. For completeness we provide the results of applying PPSBM on this data set in Section S6.5 in Supplementary Material (Zhang and Chen (2023)).

The estimated representative connecting intensities and time shifts reveal different roles of the clusters during the development of neural circuits. First, neurons in Cluster 1 might be responsible for initiating neural circuit development, since they are the first to build edges both in the representative connecting intensities and have the smallest time shifts. This early-active behavior aligns with the recognized role of motoneurons (see, e.g., Blankenship and Feller (2010), Wan et al. (2019), Wenner and O'Donovan (2001)). A noteworthy association exists between neurons in Cluster 1 and motoneurons: 98% of neurons in Cluster 1 exhibit *mnx1* expression, which is commonly associated with motoneurons (Jao, Appel and Wente (2012),

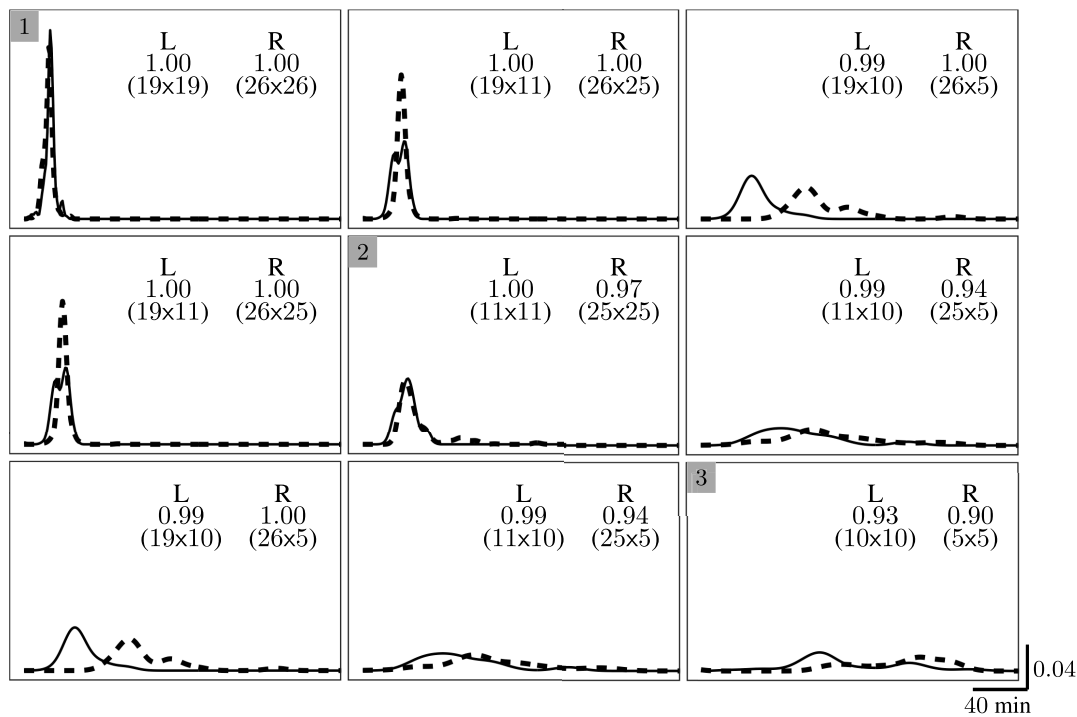


FIG. 7. Estimated connecting intensities on the neural data set from Wan et al. (2019). The rows and columns are indexed by clusters, as shown in the top-left squares on diagonal. For $q, k = 1, 2, 3$, solid and dashed curves at position (q, k) represent estimated representative connecting intensity $\hat{f}_{L,q,k}$ and $\hat{f}_{R,q,k}$ from the left and right spines, respectively. Each subfigure contains a tabular legend, where the first row shows “L” for “left” and “R” for “right,” the second row shows connecting probabilities, the third row shows the sizes of pairs of clusters. Networks in the left and right spines are fitted separately, but the estimated representative connecting intensities are highly consistent.

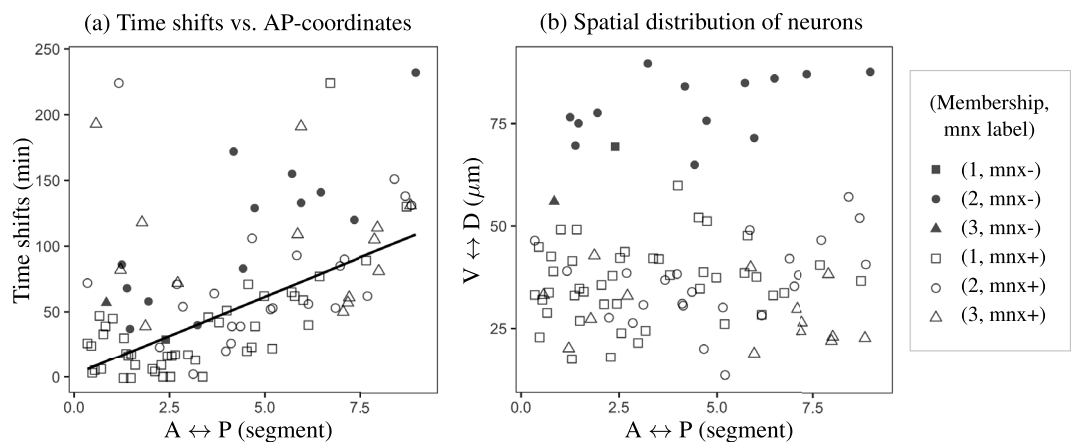


FIG. 8. Estimated time shifts and spatial distribution of neurons in both left and right spines. Each dot represents a neuron, whose shape indicates the neuron’s estimated cluster membership and its mnx label provided in the data set (see Wan et al. (2019) for details). Panel (a) shows that the time shifts might be positively correlated with the AP-coordinates of neurons, where the solid line is the regression line with least absolute deviation (i.e., median regression). Panel (b) shows that most of neurons in the dorsal side that belong to the mnx- population have the same estimated cluster membership.

Wan et al. (2019)). Second, neurons in Cluster 2 are likely to be recruited by neurons in Cluster 1, because Cluster 2 tends to establish edges with Cluster 1 faster than among themselves. The recruitment phenomenon of neurons aligns with findings in prior studies (Blankenship and Feller (2010), Nishimaru et al. (2005), Wan et al. (2019)). Lastly, neurons in Cluster 3 have lower connecting probabilities and slower edge formation speed than neurons in the other two clusters. We hypothesize that neurons in Cluster 3 may not manifest their functions during the experiment period (i.e., 17.5 hours to 22 hours postfertilization), instead, they might play essential roles in a later stage of the development, for instance, generation and coordination of rhythmic motor patterns (Menelaou and McLean (2019), Song et al. (2020)).

REMARK 4. Interestingly, Figure 8(b) shows that 86.7% of the 15 mnx^- neurons (see Wan et al. (2019) for details), which are located at the dorsal side of the spine, are grouped into the same cluster. It is worthwhile to emphasize that the proposed method does not use the spatial locations or the mnx labels of neurons. We suspect that the estimated clusters might be biologically meaningful that warrants further investigation.

REMARK 5. The data analysis results substantiate the viability of the conditions in Assumption 1. We can see from Figure 7 that the three clusters demonstrate unique intensity shapes, all intensities deviate from the constant zero function, and the values of $\hat{f}_{L,1,1}(t)$ and $\hat{f}_{R,1,1}(t)$ become nonzero as soon as t surpasses zero, which suggest the feasibility of Conditions C1, C3, and C4 in this data set. Additionally, Figure S16 in the Supplementary Material (Zhang and Chen (2023)) presents the distribution of time shifts for neurons within the three clusters, which confirms the attainability of Condition C2 in the real data context.

7. Discussion. In this article we study the problem of estimating cluster memberships and representative connecting intensities from dynamic networks with unknown time shifts. We note that there are a few directions that can be explored in future works.

First of all, the proposed method assumes that edges are persistent and the network has stabilized by the end of the experiment. These assumptions rule out networks where edges can vanish or constantly change, such as biological neural networks during learning. We can generalize the proposed model by introducing two separate intensities to characterize the stochastic formation of edges and stochastic deletion of edges as in Kreiß, Mammen and Polonik (2019), Krivitsky and Handcock (2014).

Second, we only consider one network in this analysis. A natural extension is to consider networks in a multiple-subject setting (Paul and Chen (2020), Pavlović et al. (2020), Zhang, Sun and Li (2020)). For instance, neural circuits from different zebrafish may share similar connecting intensities that can be estimated simultaneously. In this direction a key issue to address is how to match cluster labels across subjects that have different sets of nodes.

Finally, the theoretical properties of the proposed estimators have not been examined. For static stochastic block models, the minimax convergence rate of the estimated connecting probabilities has been well-studied (see, e.g., Gao, Lu and Zhou (2015)). Recently, Longepierre and Matias (2019) and Pensky (2019) establish the consistency and minimax optimality for two specific dynamic stochastic block models. Based on the established framework (Gao, Lu and Zhou (2015), Pensky (2019)), we may adapt the theory of shape invariant models (Bigot et al. (2013), Bigot and Gendre (2013)) to discuss the theoretical properties of joint estimation of time shifts and connecting intensities.

Data availability. The neural data set studied in Section 6 is available in Keller (2019).

Acknowledgments. The authors would like to express their sincere gratitude to Yinan Wan, two anonymous reviewers, the Associate Editor, and the Editor for their valuable com-

ments and suggestions. Please contact Shizhe Chen for questions or inquiries about this project.

Funding. This research was partially supported by the U.S. National Science Foundation grant DMS-1916476 and HDR:TRIPODS grant CCF-1934568.

SUPPLEMENTARY MATERIAL

Supplement to “Semiparametric estimation for dynamic networks with shifted connecting intensities” (DOI: [10.1214/23-AOAS1870SUPPA](https://doi.org/10.1214/23-AOAS1870SUPPA); .pdf). Technical derivations and proofs, details of implementation, additional results and figures are contained in this file.

Computer code for “Semiparametric estimation for dynamic networks with shifted connecting intensities” (DOI: [10.1214/23-AOAS1870SUPPB](https://doi.org/10.1214/23-AOAS1870SUPPB); .zip). R code to reproduce the simulation results and data analysis is available in this file.

REFERENCES

- BIERNACKI, C., CELEUX, G. and GOVAERT, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** 719–725. <https://doi.org/10.1109/34.865189>
- BIGOT, J. and GADAT, S. (2010). A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.* **38** 2422–2464. [MR2676894 https://doi.org/10.1214/10-AOS800](https://doi.org/10.1214/10-AOS800)
- BIGOT, J., GADAT, S., KLEIN, T. and MARTEAU, C. (2013). Intensity estimation of non-homogeneous Poisson processes from shifted trajectories. *Electron. J. Stat.* **7** 881–931. [MR3044503 https://doi.org/10.1214/13-EJS794](https://doi.org/10.1214/13-EJS794)
- BIGOT, J. and GENDRE, X. (2013). Minimax properties of Fréchet means of discretely sampled curves. *Ann. Statist.* **41** 923–956. [MR3099126 https://doi.org/10.1214/13-AOS1104](https://doi.org/10.1214/13-AOS1104)
- BLANKENSHIP, A. G. and FELLER, M. B. (2010). Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nat. Rev. Neurosci.* **11** 18–29. <https://doi.org/10.1038/nrn2759>
- BONTEMPS, D. and GADAT, S. (2014). Bayesian methods for the shape invariant model. *Electron. J. Stat.* **8** 1522–1568. [MR3263130 https://doi.org/10.1214/14-EJS933](https://doi.org/10.1214/14-EJS933)
- DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. [MR1950431 https://doi.org/10.1007/978-1-4419-6646-9](https://doi.org/10.1007/978-1-4419-6646-9)
- DAUDIN, J.-J., PICARD, F. and ROBIN, S. (2008). A mixture model for random graphs. *Stat. Comput.* **18** 173–183. [MR2390817 https://doi.org/10.1007/s11222-007-9046-7](https://doi.org/10.1007/s11222-007-9046-7)
- DIESNER, J. and CARLEY, K. M. (2005). Exploration of communication networks from the Enron email corpus. In *SIAM International Conference on Data Mining: Workshop on Link Analysis*, 3–14. Counterterrorism and Security, Newport Beach, CA.
- GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Statist.* **43** 2624–2652. [MR3405606 https://doi.org/10.1214/15-AOS1354](https://doi.org/10.1214/15-AOS1354)
- GIORGI, D., MATIAS, C., REBAFKA, T. and VILLERS, F. (2018). ppsbm: Clustering in longitudinal networks R package version 0.2.2.
- HAYTHORNTHWAITE, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Libr. Inf. Sci. Res.* **18** 323–342.
- HUBERT, L. and ARABIE, P. (1985). Comparing partitions. *J. Classification* **2** 193–218. <https://doi.org/10.1007/BF01908075>
- JAO, L.-E., APPEL, B. and WENTE, S. R. (2012). A zebrafish model of lethal congenital contracture syndrome 1 reveals G1 function in spinal neural precursor survival and motor axon arborization. *Development* **139** 1316–1326. <https://doi.org/10.1242/dev.074344>
- KELLER, P. (2019). Longitudinal functional imaging data for the zebrafish embryonic spinal cord. Janelia Research Campus. Dataset. Available at <https://doi.org/10.25378/janelia.7605824.v1>.
- KLEINBAUM, D. G. and KLEIN, M. (2012). *Survival Analysis: A Self-Learning Text*, 3rd ed. *Statistics for Biology and Health*. Springer, New York. [MR2882858 https://doi.org/10.1007/978-1-4419-6646-9](https://doi.org/10.1007/978-1-4419-6646-9)
- KNEIP, A. and ENGEL, J. (1995). Model estimation in nonlinear regression under shape invariance. *Ann. Statist.* **23** 551–570. [MR1332581 https://doi.org/10.1214/aos/1176324535](https://doi.org/10.1214/aos/1176324535)
- KREISS, A., MAMMEN, E. and POLONIK, W. (2019). Nonparametric inference for continuous-time event counting and link-based dynamic network models. *Electron. J. Stat.* **13** 2764–2829. [MR3995010 https://doi.org/10.1214/19-EJS1588](https://doi.org/10.1214/19-EJS1588)

- KRIVITSKY, P. N. and HANDCOCK, M. S. (2014). A separable model for dynamic networks. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 29–46. [MR3153932](#) <https://doi.org/10.1111/rssb.12014>
- LEI, J., CHEN, K. and LYNCH, B. (2020). Consistent community detection in multi-layer network data. *Biometrika* **107** 61–73. [MR4064140](#) <https://doi.org/10.1093/biomet/asz068>
- LIN, L., HE, Z. and PEETA, S. (2018). Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. *Transp. Res., Part C, Emerg. Technol.* **97** 258–276.
- LONGEPERRE, L. and MATIAS, C. (2019). Consistency of the maximum likelihood and variational estimators in a dynamic stochastic block model. *Electron. J. Stat.* **13** 4157–4223. [MR4021264](#) <https://doi.org/10.1214/19-EJS1624>
- LOUPOS, P., NATHAN, A. and CERF, M. (2019). Starting cold: The power of social networks in predicting non-contractual customer behavior. Available at SSRN 3001978.
- MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1119–1141. [MR3689311](#) <https://doi.org/10.1111/rssb.12200>
- MATIAS, C., REBAFKA, T. and VILLERS, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* **105** 665–680. [MR3842891](#) <https://doi.org/10.1093/biomet/asy016>
- MENELAOU, E. and MCLEAN, D. L. (2019). Hierarchical control of locomotion by distinct types of spinal V2a interneurons in zebrafish. *Nat. Commun.* **10** 4197. <https://doi.org/10.1038/s41467-019-12240-3>
- NISHIMARU, H., RESTREPO, C. E., RYGE, J., YANAGAWA, Y. and KIEHN, O. (2005). Mammalian motor neurons corelease glutamate and acetylcholine at central synapses. *Proc. Natl. Acad. Sci. USA* **102** 5245–5249.
- OSATUYI, B. (2013). Information sharing on social media sites. *Comput. Hum. Behav.* **29** 2622–2631.
- PAUL, S. and CHEN, Y. (2020). A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Ann. Appl. Stat.* **14** 993–1029. [MR4117838](#) <https://doi.org/10.1214/20-AOAS1339>
- PAVLOVIĆ, D. M., GUILLAUME, B. R. L., TOWLSON, E. K., KUEK, N. M. Y., AFYOUNI, S., VÉRTES, P. E., YEO, B. T. T., BULLMORE, E. T. and NICHOLS, T. E. (2020). Multi-subject stochastic blockmodels for adaptive analysis of individual differences in human brain network cluster structure. *NeuroImage* **220** 116611. <https://doi.org/10.1016/j.neuroimage.2020.116611>
- PENSKY, M. (2019). Dynamic network models and graphon estimation. *Ann. Statist.* **47** 2378–2403. [MR3953455](#) <https://doi.org/10.1214/18-AOS1751>
- SILVERMAN, B. W. (2017). *Density Estimation for Statistics and Data Analysis*. Routledge, New York. <https://doi.org/10.1201/9781315140919>
- SONG, J., PALLUCCHI, I., AUSBORN, J., AMPATZIS, K., BERTUZZI, M., FONTANEL, P., PICTON, L. D. and EL MANIRA, A. (2020). Multiple rhythm-generating circuits act in tandem with pacemaker properties to control the start and speed of locomotion. *Neuron* **105** 1048–1061.
- VIMOND, M. (2010). Efficient estimation for a subclass of shape invariant models. *Ann. Statist.* **38** 1885–1912. [MR2662362](#) <https://doi.org/10.1214/07-AOS566>
- WAN, Y., WEI, Z., LOOGER, L. L., KOYAMA, M., DRUCKMANN, S. and KELLER, P. J. (2019). Single-cell reconstruction of emerging population activity in an entire developing circuit. *Cell* **179** 355–372.e23. <https://doi.org/10.1016/j.cell.2019.08.039>
- WENNER, P. and O'DONOVAN, M. J. (2001). Mechanisms that initiate spontaneous network activity in the developing chick spinal cord. *J. Neurophysiol.* **86** 1481–1498. <https://doi.org/10.1152/jn.2001.86.3.1481>
- XING, E. P., FU, W. and SONG, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *Ann. Appl. Stat.* **4** 535–566. [MR2758639](#) <https://doi.org/10.1214/09-AOAS311>
- XU, K. S. (2015). Stochastic block transition models for dynamic networks. *J. Mach. Learn. Res.* **38** 1079–1087.
- XU, K. S. and HERO, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE J. Sel. Top. Signal Process.* **8** 552–562. <https://doi.org/10.1109/JSTSP.2014.2310294>
- YANG, T., CHI, Y., ZHU, S., GONG, Y. and JIN, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.* **82** 157–189. [MR3108191](#) <https://doi.org/10.1007/s10994-010-5214-7>
- ZHANG, J., SUN, W. W. and LI, L. (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *J. Amer. Statist. Assoc.* **115** 2022–2036. [MR4189774](#) <https://doi.org/10.1080/01621459.2019.1677242>
- ZHANG, Z. and CHEN, S. (2024). Supplement to “Semiparametric estimation for dynamic networks with shifted connecting intensities.” <https://doi.org/10.1214/23-AOAS1870SUPPA>, <https://doi.org/10.1214/23-AOAS1870SUPPB>