

# Understanding developing networks with stochastic block models\*

Zitong Zhang<sup>†,‡,??</sup>, Shizhe Chen<sup>§,??</sup>

*Address of the First and Second authors  
Usually a few lines long  
e-mail: [zztzhang@ucdavis.edu](mailto:zztzhang@ucdavis.edu)*

**Abstract:** In neuroscience, it is not well understood how newborn neurons form a mature nervous system, due to the lack of observations. A recent study (Wan et al. 2019) made available a dataset of this functional maturation process on zebrafish. This novel data, however, introduce inherent challenges for the analysis of the formation of the nervous system. First, the formation process is transient by nature. The non-stationarity of the process makes the amount of data pale in comparison to the size of the neural network. Moreover, combining observations on multiple subjects are not straightforward since the neural circuits are not identical across subjects. In this talk, we propose a model for describing the emergence of a coordinated network from isolated nodes. To this end, we adapt and generalize the stochastic block model for random graphs. The proposed method learns the transferable features across subjects, while allowing for individual variabilities. Briefly, the proposed method classifies nodes into different functional groups by identifying typical connecting behavior. We further employ the shape invariant models to handle the nodal delays due to neuron-specific delays. We establish the consistency and minimax optimality of the proposed estimator. We demonstrate the performance of our algorithm on simulation experiments, and on the real zebrafish dataset.

**MSC 2010 subject classifications:** Primary 60K35, 60K35; secondary 60K35.

**Keywords and phrases:** sample, L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>.

## 1. Introduction

Dynamic networks emerge in many area, such as the neuronal network in the brain during disease[] or learning tasks[], the social network in a time period[], etc. A lot of work has been done in order to study the dynamic networks, but most of these research focus on analyzing the dynamics in a well-developed network. In this paper, we concentrate on the growing networks where isolated nodes develop into a mature functioning network. To the best of our knowledge, there is no existing study about the growing networks, partly because of the lack of data. Fortunately, the data collected by Wan et al. [12] provides us the possibility to pursue this study.

---

\*Footnote to the title with the “thankstext” command.

†Some comment

‡First supporter of the project

§Second supporter of the project

In this paper, we focus on the neural data provided by [12]. Since the neuronal network is complicated, we try to break it down and propose to model it by identifying the typical roles of individual neurons. As supported by [12], neurons in a growing network play different roles in terms of active time and connecting patterns. The connecting pattern of a neuron can be described as the occurrence time of edges that include this neuron. Neurons with the same roles perform similar activities and thus have similar connecting patterns. Learning these roles can help us understand the development procedure of growing networks.

However, being able to identify the typical roles in a single growing network is infeasible because the growing network is transient — there is only one measurement of a growing network due to its non-stationarity. Due to this limit, we combine networks of other subjects as additional samples. This is not trivial, because the neurons in different networks are not one-to-one mapped, and the roles identified from different subjects are not transferable. This constrains our ability to study the common features across subjects. For this reason, we propose to use the stochastic block model as it allows to combine multiple networks and hence resolve the above problems. [SBM is.... It has been applied to ...]

There are, nevertheless, some unique properties of the data that are beyond the scope of the stochastic block model. First, the connection between two neurons is measured over time. Second, the connecting pattern between two neurons are determined not only by their roles but also by their active time, which varies from node to node. Third, the connection is also effected by the spatial distance between neurons — connection cannot occur if two neurons are too away from each other. To incorporate such uniqueness, we propose a generalized stochastic block model in this paper. [Note that these properties also appear in other problems, e.g. venmo...]

### ***Related work***

The stochastic block model is first proposed by Holland, Blackmond and Leinhardt [5]. It has many dynamic extensions, Yang et al. [15], Xu and Hero [14], Matias and Miele [9], Xu [13] use the Markov chain to model the time-varying connecting probabilities and/or the clustering matrix. (Variational) EM algorithm is commonly used for inference. Matias, Rebafka and Villers [10] adapt the stochastic block model to the context of recurrent interaction events in continuous time, where the recurrent events are modeled by Poisson processes with intensities determined by the nodes' group memberships. A semiparametric version of the variational EM algorithm is proposed, where the maximization over intensities is obtained by a histogram approach and/or a kernel estimator.

There are many other fitting methods of the stochastic block model that have been proposed, including spectral clustering Liu et al. [8], convex relaxation Li, Chen and Xu [7], Chen, Sanghavi and Xu [3], semidefinite programming relaxation Amini and Levina [1], low-rank approximation Le, Levina and Vershynin [6] etc.

Optimal rate of convergence is also studied. Gao, Lu and Zhou [4] provides an optimal rate under the mean squared error for the stochastic block model.

Pensky [11] derives a penalized least square estimator in a dynamic network setting, and shows that the estimator satisfies an oracle inequality and attains the minimax lower bound for the risk.

### **Contribution**

In this paper, we propose a method for analyzing the growing networks. Our method is able to identify the roles of individual nodes and the connecting patterns. We adapt the stochastic block model to the growing networks context by generalizing the connecting probabilities to intensities of point processes. In addition, we incorporate the time delay of each node so that our model is able to handle the network where nodes become active over time. We derive a least square estimator and show that the estimator converges [in a certain rate]. Finally, an algorithm combining the k-means method and the shape invariant method is proposed for estimation.

### **Future work**

Future working directions include (but not limited to) (i) identifying clusters with similar connecting pattern but different active time phase or different vertex degree, (ii) incorporating the movement of nodes, (iii) seeking for a convex relaxation method that convexify over both clustering matrix and time lags (convex relaxation can also be adapted to solve the penalized least square problem in Pensky [11]), (iv) try other clustering methods.

### **Organization**

The rest of this paper is organized as follows. In Section 2, we review the stochastic block model and introduce the proposed dynamic generalization of the stochastic block model. We introduce the least square estimator and the estimation algorithm in Section 3. Theoretical results are provided in Section 4. Section 5 shows the numerical experiments.

## **2. Model**

### **2.1. Stochastic block model**

A set of  $n$  nodes  $\Gamma = \{1, \dots, n\}$  is partitioned into  $k$  clusters  $\Gamma_1, \dots, \Gamma_k$ . The cluster of node  $i$  is represented by  $z_i \in \{1, \dots, k\}$ , and the vector of clusters is  $\mathbf{z} = (z_i)_{i=1}^n$ . Define the adjacency matrix  $\mathbf{A} \in \{0, 1\}^{n \times n}$  where  $A_{i,j} = 1$  if an edge is observed between node  $i$  and node  $j$  and  $A_{i,j} = 0$  otherwise. We set  $A_{i,i} \equiv 0$  for any  $i = 1, \dots, n$ , and assume that  $A_{i,j}$ 's are conditionally independent given the cluster vector  $\mathbf{z}$ :

$$A_{i,j}|z_i = q, z_j = l \stackrel{ind}{\sim} \text{Bernoulli}(C_{q,l}), \quad i \neq j,$$

where  $\mathbf{C} \in [0, 1]^{k \times k}$  denotes the connecting probability matrix.

## 2.2. Dynamic stochastic block model for point processes

We consider a developing network where pairwise connections between  $n$  nodes are observed during time interval  $[0, T]$ . Each node belongs to one of  $k$  groups denoted by  $\Gamma_1, \dots, \Gamma_k$ , and the group membership for node  $i$  is denoted by  $z_i$ . The connection between node  $i$  and  $j$  is modeled by a point process  $N_{i,j}(\cdot)$  with conditional intensity depending only on their group memberships  $z_i$  and  $z_j$ . Given  $z_i = q, z_j = l$ , the conditional intensify function of  $N_{i,j}$  is

$$\lambda_{i,j}^*(t) = \lambda_{q,l}^*(t), \quad t \in [0, T], \quad i, j = 1, \dots, n.$$

Here  $*$  is used to denote that the density is conditional on the past. Similar to the stochastic block model, we set  $\lambda_{i,i}^*(\cdot) \equiv 0$  for  $i = 1, \dots, n$ .

Note that edges in the developing networks are considered as physical connections — they do not disappear after constructed, so one can observe at most one event (appearance of an edge) in each point process. This allows us to define the event time  $t_{i,j}$  in  $N_{i,j}$ , where  $t_{i,j} = \infty$  if there is no event during  $[0, T]$ . The conditional distribution of  $t_{i,j}$  given that there is an event during  $[0, T]$  is then expressible in the form

$$p_{q,l}(t_{i,j}) \equiv p(t_{i,j} | N_{i,j}[0, T] = 1) = \frac{L(t_{i,j}; \lambda_{q,l}^*)}{\int_0^T L(u; \lambda_{q,l}^*) du},$$

where  $L(t; \lambda^*) = \lambda^*(t) \exp\left(-\int_0^T \lambda^*(s) ds\right)$  is the likelihood of the realization  $t_1 = t$  of  $N$  over  $[0, T]$ . This simplifies our model to

$$t_{i,j} | t_{i,j} < \infty, z_i = q, z_j = l \sim p_{q,l}(\cdot). \quad (2.1)$$

Here we rule out the probability of having a connection between node  $i$  and  $j$ , because it may vary drastically from node to node and therefore does not contribute to identifying clusters.

Let us now take into account the edge-specific time lags caused by the different active time of neurons. Denoting the time lag for  $t_{i,j}$  by  $\tau_{i,j}$ , our observation becomes  $\tilde{t}_{i,j} = t_{i,j} + \tau_{i,j}$ . It is believed (?) in neuroscience that every connection is led by one of the two neurons, so the time lags are actually controlled by the active time of the leader neurons. We incorporate the underlying structure of  $\tau_{i,j}$  by making the following assumption

$$\tau_{i,j} = v_i \mathbf{1}\{L_{q,l} = 0\} + v_j \mathbf{1}\{L_{q,l} = 1\}, \text{ given } i \in \Gamma_q, j \in \Gamma_l, \quad (2.2)$$

where  $\mathbf{v}_{n \times 1}$  is a vector of relative active time satisfying  $\min_{i \in \Gamma_q} v_i = 0$ ,  $L_{k \times k}$  is a matrix of relative leadership — the edge time delay  $\tau_{i,j}$  is governed by node  $i$  if  $L_{q,l} = 0$ , otherwise by node  $j$ .

### Objective function

With this model, we propose the following least-square loss function that measures the overall distance between each neuron and its group

$$\min_{\mathbf{z}, \mathbf{f}, \tau} \sum_{i \in [n]} \left[ \sum_{l \in [k]} w_{i,l} \cdot \|f_{i,l} - p_{z_i,l}\|_2^2 \right], \quad (2.3)$$

where  $f_{i,l}(\cdot)$  is the kernel density estimation of sequence  $\{\tilde{t}_{i,j} - \tau_{i,j} : \tilde{t}_{i,j} < \infty\}_{j \in \Gamma_l}$  with some proper bandwidth,  $w_{i,l}$  is the weight measuring the contribution of cluster  $\Gamma_l$  to the distance between node  $i$  and cluster  $z_i$ .

## 3. Method

### 3.1. Evaluating the distance function

In practice, the distance between a point process  $N$  and a probability distribution function  $f$  is evaluated by computing distance between their (smoothed) probability distribution functions, because the distance between cumulative distribution functions can be misleading in some certain cases. We illustrate this through figure 1. Although the black dash pdf is closer to the green pdf than the red pdf, it is closer to the red curve in terms of cdf.

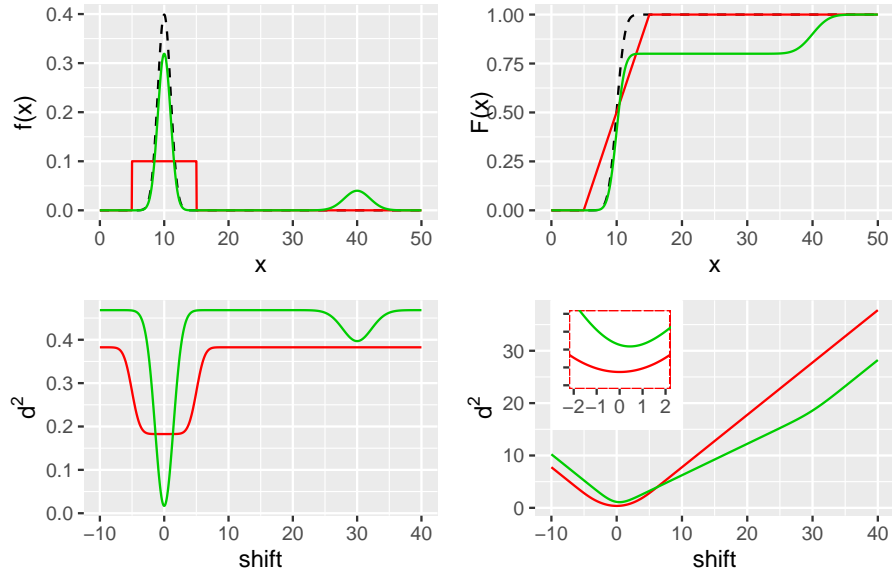


Fig 1: Distance between cdfs is misleading. Top left: pdf to be shifted (black dash line) and target pdf (green and red solid line). Top right: corresponding cdf. Bottom left: distance between pdfs as a function of shift. Bottom right: distance between cdfs as a function of shift.

However, as shown in figure 2, aligning cdf can help us to locate the minimizer of distance function, especially when there is a flat area between two pdfs.

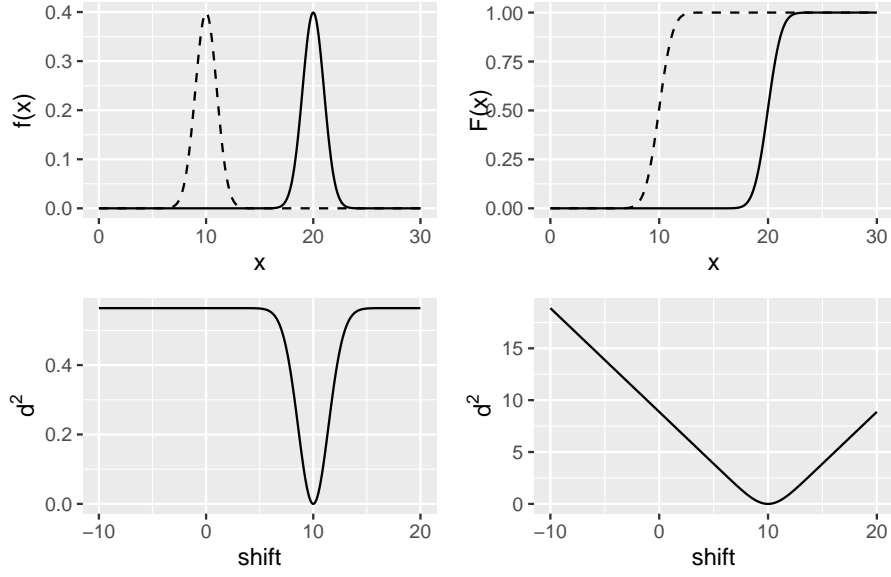


Fig 2: cdf can help to locate the minimizer. Top left: pdf to be shifted (dash line) and target pdf (solid line). Top right: corresponding cdfs. Bottom left: distance between pdfs as a function of shift. Bottom right: distance between cdfs as a function of shift.

Based on these properties of cdf and pdf, we align pdfs using gradient descent algorithm with initialization obtained by aligning their corresponding cdfs. We use the shape-invariant method proposed by Bigot and Gendre [2]. The idea is to find the optimal shift that aligns  $f$  and  $g$  best in the Fourier domain. Let  $\theta_j$  and  $\gamma_j, j = -(N-1)/2, \dots, (N-1)/2$  be the discrete Fourier coefficients of  $f$  and  $g$ , where  $N$  is the length of discretized  $f$  and  $g$ . The time shift parameter  $\tau$  can be estimated by solving the following problem by gradient descent

$$\hat{n} = \arg \min_{|n| \leq (N-1)/2} \sum_{|j| \leq (N-1)/2} \left| \theta_j e^{i2\pi j n/N} - \gamma_j \right|^2, \quad (3.1)$$

where  $n = N \cdot \tau / 2T$ . Initialization is given by the alignment result of their cdfs. For details see [A](#).  
[add details for obtaining smooth pdf.]

### 3.2. Main steps

We iterate between three main steps: re-center step, re-cluster step, and re-align step. In the re-center step, we update connecting patterns based on current clustering and time lags; in the re-cluster step, we update the clusters based

on updated connecting patterns and current time lags; in the re-align step, we update time lags based on updated clusters.

#### Re-center step

Based on current estimation of clusters  $\{\Gamma_q^{(c)}\}_{q \in [k]}$  and time shifts  $\{\tau_{i,j}^{(c)}\}_{i,j \in [n]}$ , we can obtain shifted event times  $t_{i,j}^{(c)} = \tilde{t}_{i,j} - \tau_{i,j}^{(c)}$ . Let  $N_{q,l}$  be a point process with events at time  $T_{q,l} \equiv \{t_{i,j}^{(c)} : t_{i,j}^{(c)} < \infty, i \in \Gamma_q^{(c)}, j \in \Gamma_l^{(c)}\}$ , the connecting patterns are then estimated by kernel density estimation

$$p_{q,l}^{(u)}(t) = \frac{1}{h|T_{q,l}|} \sum_{t' \in T_{q,l}} K\left(\frac{t-t'}{h}\right), \quad (3.2)$$

where  $K$  is a kernel function and  $h$  is the bandwidth.

#### Re-cluster step

Given  $\{p_{q,l}^{(u)}\}_{q,l \in [k]}$  and  $\{\tau_{i,j}^{(c)}\}_{i,j \in [n]}$ , we need to solve

$$z_i^{(u)} = \arg \min_{q \in [k]} \sum_l w_{i,l} \cdot \|f_{i,l} - p_{q,l}^{(u)}\|_2^2, \quad i = 1, \dots, n.$$

Recall that  $f_{i,l}$  is the density of  $\{t_{i,j}^{(c)} = \tilde{t}_{i,j} - \tau_{i,j}^{(c)} : t_{i,j}^{(c)} < \infty, j \in \Gamma_l\}$  depending on the cluster  $\Gamma_l$  that is to be estimated, so the estimation of  $z_i^{(u)}$  is not straightforward. To handle this problem, we propose to evaluate  $f_{i,l}$  by computing the density of  $\{t_{i,j}^{(c)} : t_{i,j}^{(c)} < \infty, j \in \Gamma_l^{(c)}\}$ . The optimization problem then becomes

$$z_i^{(u)} = \arg \min_{q \in [k]} \sum_l w_{i,l} \cdot \|f_{i,l^{(c)}} - p_{q,l}^{(u)}\|_2^2, \quad i = 1, \dots, n.$$

Furthermore, to rule out the effect of time lag between  $f_{i,l^{(c)}}$  and  $p_{q,l}^{(u)}$  and only consider the distance between shapes, we substitute the  $\ell_2$ -distance by its shift-invariant version  $d(f, g) = \inf_{\tau} \|f(\cdot + \tau) - g\|_2$  and write down the final optimization problem

$$z_i^{(u)} = \arg \min_{q \in [k]} \sum_l w_{i,l} \cdot d(f_{i,l^{(c)}}, p_{q,l}^{(u)})^2, \quad i = 1, \dots, n. \quad (3.3)$$

The evaluation of this shift-invariant distance will be discussed later. Computing and comparing all the desired distances will solve this problem.

#### Re-align step

Recall the structure of  $\tau$  in (2.2). Based on current clusters, we will estimate  $\mathbf{v}^{(u)}$  and  $\mathbf{L}^{(u)}$  which determines  $\tau^{(u)}$ .



For each cluster  $\Gamma_q^{(u)}$ , randomly select a node  $i^* \in \Gamma_q^{(u)}$  as a point of reference. By aligning node  $i \in \Gamma_q^{(u)}$  and  $i^*$  with respect to each cluster, their time lag  $v_i^{(u)} - v_{i^*}^{(u)}$  has  $k$  possible values

$$\Delta v_{(i,i^*),l} = \arg \min_{\tau} \|\tilde{f}_{i,l^{(u)}}(\cdot + \tau) - \tilde{f}_{i^*,l^{(u)}}(\cdot)\|_2, \quad l \in [k],$$

where  $\tilde{f}_{i,l^{(u)}}$  is the density of  $\{\tilde{t}_{i,j} : \tilde{t}_{i,j} < \infty, j \in \Gamma_l^{(u)}\}$ . We will use the one with the largest magnitude as the estimated time lag:

$$v_i^{(u)} - v_{i^*}^{(u)} = \Delta v_{(i,i^*),l_0}, \quad \text{where } l_0 = \arg \max_{l \in [k]} |\Delta v_{(i,i^*),l}|. \quad (3.4)$$

The reason behind this is the following. If  $L_{q,l} = 0$ , the connection between node  $i/i^*$  and cluster  $\Gamma_l$  is led by nodes in  $\Gamma_l$ , and  $\tilde{t}_{i,j} = t_{i,j} + v_i$ ,  $\tilde{t}_{i^*,j} = t_{i^*,j} + v_{i^*} \stackrel{d}{=} t_{i,j} + v_{i^*}$  for  $j \in \Gamma_l$ , as a result  $\Delta v_{(i,i^*),l}$  will be close to  $v_i - v_{i^*}$ ; otherwise  $\Delta v_{(i,i^*),l}$  will be close to zero. Thus non-zero values are considered as valid estimation of time lags.

Setting  $\min\{v_i^{(u)}\}_{i \in \Gamma_q^{(u)}}$  as zero yields estimated  $\{v_i^{(u)}\}_{i \in \Gamma_q^{(u)}}$ .

$\mathbf{L}$  is then updated by comparing within group variances. For cluster pair  $(q, l)$ , let

$$\begin{aligned} \text{var}_1 &= \text{var} \left[ \{\tilde{t}_{i,j} - v_i^{(u)}\}_{i \in \Gamma_q^{(u)}, j \in \Gamma_l^{(u)}} \right], \\ \text{var}_2 &= \text{var} \left[ \{\tilde{t}_{i,j} - v_j^{(u)}\}_{i \in \Gamma_q^{(u)}, j \in \Gamma_l^{(u)}} \right], \end{aligned}$$

then

$$L_{q,l}^{(u)} = \mathbf{1}\{\text{var}_1 < \text{var}_2\}. \quad (3.5)$$

Finally, the time shifts are updated as

$$\tau_{i,j}^{(u)} = v_i^{(u)} \mathbf{1}\{L_{q,l}^{(u)} = 0\} + v_j^{(u)} \mathbf{1}\{L_{q,l}^{(u)} = 1\}. \quad (3.6)$$

### 3.3. Initialization

Estimate  $\{v_i^{(0)}\}_{i \in [n]}$  based on  $\Gamma_1^{(0)} = [n]$ . Initialize time lags by  $\tau_{i,j}^{(c)} = v_i^{(0)}$ . Clustering  $\{\Gamma_q^{(c)}\}_{q \in [k]}$  is then initialized by applying k-means++ algorithm to

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix},$$

where  $f_i$  is the density of  $\{\tilde{t}_{i,j} - v_i^{(0)} : \tilde{t}_{i,j} < \infty, j \in [n]\}$ .

### 3.4. Summary of algorithm

The algorithm is summarized as following.

---

**Algorithm 1:** [name of algorithm]

---

Initialize  $\{\Gamma_q^{(c)}\}_{q \in [k]}$  and  $\{\tau_{i,j}^{(c)}\}_{i,j \in [n]}$ ;  
**while**  $\{\Gamma_q^{(c)}\}_{q \in [k]} \neq \{\Gamma_q^{(u)}\}_{q \in [k]}$  **do**  
    Update  $\{p_{q,l}^{(u)}\}_{q,l \in [k]}$  via (3.2) with  $\{\Gamma_q^{(c)}\}_{q \in [k]}$  and  $\{\tau_{i,j}^{(c)}\}_{i \in [n]}$  ;  
    Update  $\{\Gamma_q^{(u)}\}_{q \in [k]}$  via (3.3) with  $\{p_{q,l}^{(u)}\}_{q,l \in [k]}$ ,  $\{\Gamma_q^{(c)}\}_{q \in [k]}$  and  
         $\{\tau_{i,j}^{(c)}\}_{i \in [n]}$  ;  
    Update  $\{\tau_{i,j}^{(u)}\}_{i \in [n]}$  via (3.6) with  $\{\Gamma_q^{(u)}\}_{q \in [k]}$  ;  
    Evaluate the stopping criterion ;  
     $\{p_{q,l}^{(c)}\}_{q,l \in [k]} \leftarrow \{p_{q,l}^{(u)}\}_{q,l \in [k]}$  ;  
     $\{\Gamma_q^{(c)}\}_{q \in [k]} \leftarrow \{\Gamma_q^{(u)}\}_{q \in [k]}$  ;  
     $\{\tau_{i,j}^{(c)}\}_{i \in [n]} \leftarrow \{\tau_{i,j}^{(u)}\}_{i \in [n]}$  ;  
**end**  
**Output:**  $\{p_{q,l}^{(c)}\}_{q,l \in [k]}$ ,  $\{\Gamma_q^{(c)}\}_{q \in [k]}$ ,  $\{\tau_{i,j}^{(c)}\}_{i \in [n]}$ .

---

## 4. Theory

Assume  $\mathbf{F} = (F_i)_{i=1}^n$  is from the parameter space

$$\mathcal{F}_k = XXX.$$

**Theorem 4.1.** *For any constant  $C' > 0$ , there is a constant  $C > 0$  only depending on  $C'$ , such that*

$$\frac{1}{n} \sum_{i=1}^n \left\| \hat{F}_i - F_i \right\|^2 \leq C(XXX),$$

with probability at least  $1 - \exp(-C'XXX)$ , uniformly over  $\mathbf{F} \in \mathcal{F}_k$ .

*Proof.* This is a sketch of proof and is based on the proof in [4].

We denote the true value by  $\theta_i^* = F_{z_i^*}^*(\cdot - \tau_i^*)$ . For the estimated  $\hat{z}$ , define  $\tilde{\theta} = \arg \min_{\theta \in \Theta_k(\hat{z})} \|\theta^* - \theta\|^2$ . By the definition of the estimator, we have

$$L(\hat{F}, \hat{Z}, \hat{\tau}) \leq L(F^*, Z^*, \tau^*),$$

which can be rewritten as

$$\|\hat{\theta} - F^{obs}\|^2 \leq \|\theta^* - F^{obs}\|^2. \quad (4.1)$$

The left-hand side of (4.1) can be decomposed as

$$\|\hat{\theta} - \theta^*\|^2 + 2\langle \hat{\theta} - \theta^*, \theta^* - F^{obs} \rangle + \|\theta^* - F^{obs}\|^2. \quad (4.2)$$

Combining (4.1) and (4.2), we have

$$\|\hat{\theta} - \theta^*\|^2 \leq 2\langle \hat{\theta} - \theta^*, F^{obs} - \theta^* \rangle. \quad (4.3)$$

The right-hand side of (4.3) can be bounded as

$$\begin{aligned} \langle \hat{\theta} - \theta^*, F^{obs} - \theta^* \rangle &= \langle \hat{\theta} - \tilde{\theta}, F^{obs} - \theta^* \rangle + \langle \tilde{\theta} - \theta^*, F^{obs} - \theta^* \rangle \\ &\leq \|\hat{\theta} - \tilde{\theta}\| \left| \left\langle \frac{\hat{\theta} - \tilde{\theta}}{\|\hat{\theta} - \tilde{\theta}\|}, F^{obs} - \theta^* \right\rangle \right| \end{aligned} \quad (4.4)$$

$$+ \left( \|\tilde{\theta} - \hat{\theta}\| + \|\hat{\theta} - \theta^*\| \right) \left| \left\langle \frac{\tilde{\theta} - \theta^*}{\|\tilde{\theta} - \theta^*\|}, F^{obs} - \theta^* \right\rangle \right|. \quad (4.5)$$

Using Lemmas XXX, the following three terms:

$$\|\hat{\theta} - \tilde{\theta}\|, \quad \left| \left\langle \frac{\hat{\theta} - \tilde{\theta}}{\|\hat{\theta} - \tilde{\theta}\|}, F^{obs} - \theta^* \right\rangle \right|, \quad \left| \left\langle \frac{\tilde{\theta} - \theta^*}{\|\tilde{\theta} - \theta^*\|}, F^{obs} - \theta^* \right\rangle \right| \quad (4.6)$$

can all be bounded by XXX with probability at least

$$XXX.$$

Combining these bounds with (4.4), (4.5) and (4.3), we get

$$\|\hat{\theta} - \theta^*\|^2 \leq XXX$$

with probability at least XXX.  $\square$

Now we present the lemmas, which bound the three terms in (4.6), respectively.

**Lemma 1.** *For any constant  $C' > 0$ , there exists a constant  $C > 0$  only depending on  $C'$ , such that*

$$\|\hat{\theta} - \tilde{\theta}\| \leq CXXX,$$

with probability at least XXX.

*Proof of Lemma 1.*

*Step 1: Control  $\mathbb{E}\|\hat{\tau} - \tilde{\tau}\|^2$ .*

Fix a group  $\hat{\Gamma}_k$ , denote the group size by  $J = |\Gamma_k|$ . For notation simplicity, relabel the nodes in group  $\hat{\Gamma}_k$  with  $\{1, \dots, J\}$ . Let

$$M(\tau_1, \dots, \tau_J) = \frac{1}{J} \sum_{j=1}^J \sum_{|k| \leq k_0} \left| c_{j,k} e^{i2\pi k \tau_j} - \frac{1}{J} \sum_{j'=1}^J c_{j',k} e^{i2\pi k \tau_{j'}} \right|^2,$$

where  $c_{j,k}, k \in \mathbb{Z}$ , are the Fourier coefficients of the distribution function  $F_{z_j^*}^*(\cdot - \tau_j^*)$ . Let

$$\hat{M}(\tau_1, \dots, \tau_J) = \frac{1}{J} \sum_{j=1}^J \sum_{|k| \leq k_0} \left| \hat{c}_{j,k} e^{i2\pi k \tau_j} - \frac{1}{J} \sum_{j'=1}^J \hat{c}_{j',k} e^{i2\pi k \tau_{j'}} \right|^2,$$

where  $\hat{c}_{j,k}, k \in \mathbb{Z}$ , are the Fourier coefficients of the empirical distribution function  $F_j^{obs}$ . Then  $\hat{\tau}$  is the minimizer of  $\hat{M}$ , and  $\tilde{\tau}$  is the minimizer of  $M$ .

By Proposition 3.1 in Bigot and Gendre [2], under proper assumptions of  $F$  and distribution of  $\tau$ ,

$$\frac{1}{J} \|\hat{\tau} - \tilde{\tau}\|^2 \leq C^{-1} \cdot (M(\hat{\tau}_1, \dots, \hat{\tau}_J) - M(\tilde{\tau}_1, \dots, \tilde{\tau}_J)).$$

Note that  $M(\hat{\tau}) - M(\tilde{\tau}) = M(\hat{\tau}) - \hat{M}(\hat{\tau}) + \hat{M}(\hat{\tau}) - M(\tilde{\tau}) \leq 2 \sup_{\tau} |M(\tau) - \hat{M}(\tau)|$ , so it suffices to control  $\mathbb{E} \sup_{\tau} |M(\tau) - \hat{M}(\tau)|$ .

$\mathbb{E} \sup_{\tau} |M(\tau) - \hat{M}(\tau)|$  is controlled by  $F_j^{obs} - F_j^*$  or  $\hat{c}_{j,k} - c_{j,k}$ ?

*Step 2: Bound  $\|\hat{\theta} - \tilde{\theta}\|$ .*

□

**Lemma 2.** *For any constant  $C' > 0$ , there exists a constant  $C > 0$  only depending on  $C'$ , such that*

$$\left| \left\langle \frac{\tilde{\theta} - \theta^*}{\|\tilde{\theta} - \theta^*\|}, F^{obs} - \theta^* \right\rangle \right| \leq CXXX,$$

with probability at least  $XXX$ .

*Proof.* Note that

$$\tilde{\theta}_i - \theta_i^* = \tilde{F}_{\tilde{z}_i}(\cdot - \tilde{\tau}_i) - F_{z_i^*}^*(\cdot - \tau_i^*)$$

is a function of the partition  $\hat{\mathbf{z}}$ , then we have

$$\left| \sum_i \left\langle \frac{\tilde{\theta}_i - \theta_i^*}{\sqrt{\sum_i \|\tilde{\theta}_i - \theta_i^*\|^2}}, F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right| \leq \max_{\mathbf{z} \in \mathcal{Z}_{n,k}} \left| \sum_i \left\langle \gamma_i(\mathbf{z}), F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right|$$

where

$$\gamma_i(\mathbf{z}) \propto \tilde{F}_{z_i}(\cdot - \tilde{\tau}_i) - F_{z_i^*}^*(\cdot - \tau_i^*)$$

satisfies  $\sum_i \|\gamma_i(\mathbf{z})\|^2 = 1$ . By [some inequality similar to Hoeffding's inequality] and union bound, we have

$$\begin{aligned} & \mathbb{P} \left( \max_{\mathbf{z} \in \mathcal{Z}_{n,k}} \left| \sum_i \left\langle \gamma_i(\mathbf{z}), F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right| > t \right) \\ & \leq \sum_{\mathbf{z} \in \mathcal{Z}_{n,k}} \mathbb{P} \left( \left| \sum_i \left\langle \gamma_i(\mathbf{z}), F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right| > t \right) \\ & \leq |\mathcal{Z}_{n,k}| \exp(-C_1 t^2), \end{aligned}$$

for some universal constant  $C_1 > 0$ . Choosing  $t \propto \sqrt{n \log k}$ , the proof is complete.  $\square$

**Lemma 3.** *For any constant  $C' > 0$ , there exists a constant  $C > 0$  only depending on  $C'$ , such that*

$$\left| \left\langle \frac{\hat{\theta} - \tilde{\theta}}{\|\hat{\theta} - \tilde{\theta}\|}, F^{obs} - \theta^* \right\rangle \right| \leq CXXX,$$

with probability at least XXX.

*Proof.* Note that

$$\hat{\theta}_i - \tilde{\theta}_i = \hat{F}_{\hat{z}_i}(\cdot - \hat{\tau}_i) - \tilde{F}_{\tilde{z}_i}(\cdot - \tilde{\tau}_i)$$

is a function of both the partition  $\hat{\mathbf{z}}$  and the observations  $F^{obs}$ . For each  $\mathbf{z} \in \mathcal{Z}_{n,k}$ , define the set  $\mathcal{B}_{\mathbf{z}}$  by

$$XXX$$

Thus, we have the bound

$$\left| \sum_i \left\langle \frac{\tilde{\theta}_i - \theta_i^*}{\sqrt{\sum_i \|\tilde{\theta}_i - \theta_i^*\|^2}}, F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right| \leq \max_{\mathbf{z} \in \mathcal{Z}_{n,k}} \sup_{c \in \mathcal{B}_{\mathbf{z}}} \left| \sum_i \left\langle c_i, F_i^{obs} - F_{z_i^*}^*(\cdot - \tau_i^*) \right\rangle \right|$$

If set  $\mathcal{B}_{\mathbf{z}}$  is not too large, applying union bound (and Hoeffding-like inequality) completes the proof.  $\square$

## 5. Simulation

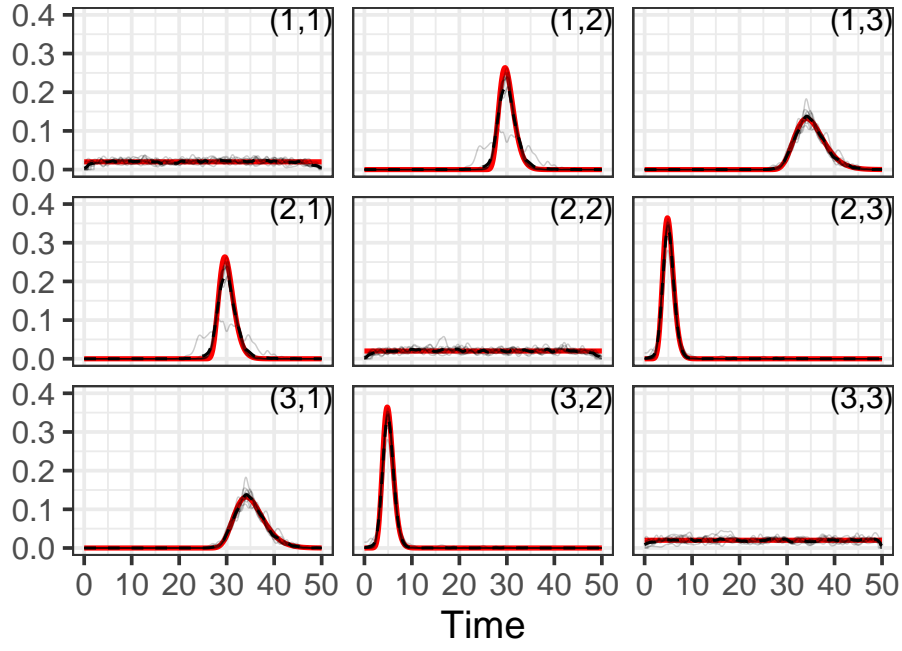


Fig 3: Estimated connecting patterns (ten trials). Dash line: average curves of estimated connecting patterns. Red line: true connecting patterns.

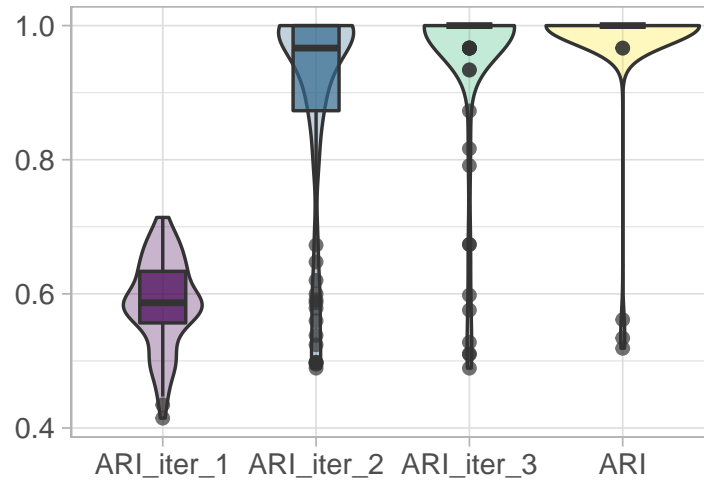


Fig 4: Clustering result.

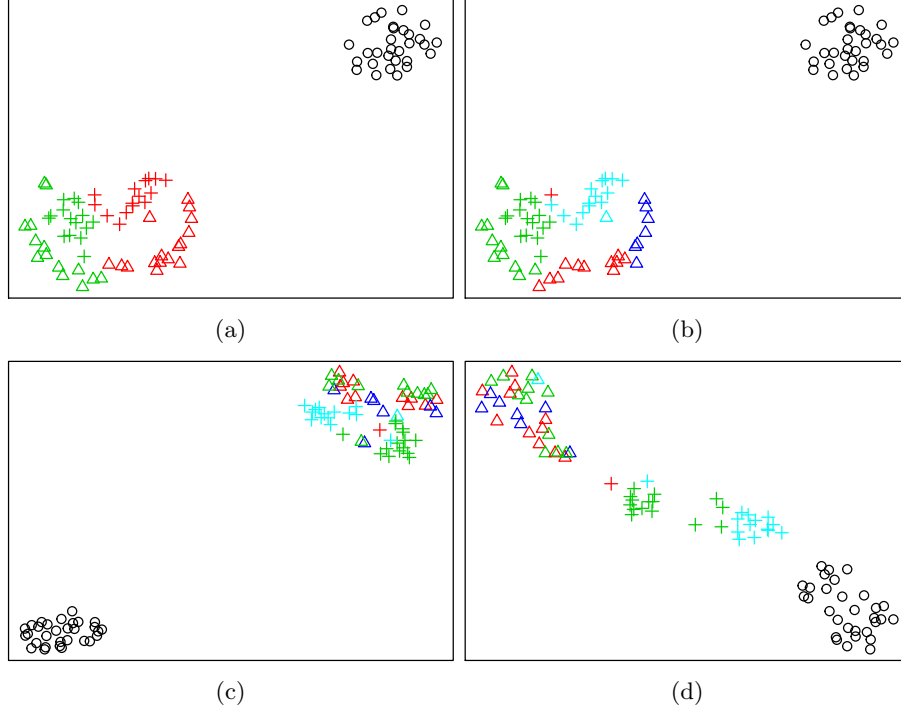


Fig 5: 2D representation by t-SNE. (a) Aggregated point processes with estimated three clusters. (b) Aggregated point processes with estimated five clusters. (c) Point processes incorporated clustering results from (a). (d) Point processes incorporated clustering results from (b).

## 6. Conclusion

### Appendix A: Appendix section

#### A.1. Derivation of distance between pdfs and cdfs

Suppose a pdf  $f$  is defined on  $[0, T]$ . Extend  $f$  by zero to  $[-T, 0]$ , and define the shifted pdf  $S_\tau \circ f(\cdot)$  by

$$S_\tau \circ f(t) = \begin{cases} 0, & t \in [-T, -\tau) \\ f(t + \tau), & t \in [-\tau, T - \tau) \\ 0, & t \in [T - \tau, T] \end{cases} \quad (\text{A.1})$$



Suppose the extended  $f$  (defined on  $[-T, T]$ ) is represented by  $X_{-(N-1)/2}, \dots, X_{(N-1)/2}$ , the Fourier coefficients are

$$\theta_j = \sum_{k=-\frac{(N-1)}{2}}^{\frac{N-1}{2}} X_k \cdot \exp\{-i2\pi(k/N)j\}, \quad j = -\frac{(N-1)}{2}, \dots, \frac{(N-1)}{2}.$$

Inverse transformation

$$X_k = \frac{1}{N} \sum_{j=-\frac{(N-1)}{2}}^{\frac{N-1}{2}} \theta_j \cdot \exp\{i2\pi(j/N)k\}, \quad k = -\frac{(N-1)}{2}, \dots, \frac{(N-1)}{2}.$$

The Fourier coefficients of  $S_\tau \circ f$  is

$$\begin{aligned} \theta'_j &= \sum_{k=-\frac{(N-1)}{2}}^{\frac{N-1}{2}-n_0} X_{k+n_0} \cdot \exp\{-i2\pi(k/N)j\} + \sum_{k=\frac{(N-1)}{2}-n_0+1}^{\frac{N-1}{2}} 0 \\ &= \sum_{k=-\frac{(N-1)}{2}+n_0}^{\frac{N-1}{2}} X_k \cdot \exp\{-i2\pi((k-n_0)/N)j\} + \sum_{k=\frac{-(N-1)}{2}+n_0-1}^{\frac{-(N-1)}{2}+n_0-1} 0 \\ &= \sum_{k=-\frac{(N-1)}{2}}^{\frac{N-1}{2}} X_k \cdot \exp\{-i2\pi((k-n_0)/N)j\} \\ &= \theta_j \cdot \exp\{i2\pi(n_0/N)j\}. \end{aligned}$$

Suppose another pdf  $g$  is also defined on  $[0, T]$  and extended to  $[-T, T]$  with Fourier coefficients  $\gamma_j$ ,  $j = -\frac{(N-1)}{2}, \dots, \frac{(N-1)}{2}$ . The shift-invariant distance be-

tween  $f$  and  $g$  is then

$$\begin{aligned}
d(f, g) &= \min \left\{ \inf_{\tau \in [0, T]} \left( \int_{-T}^T |S_\tau \circ f(t) - g(t)|^2 dt \right)^{1/2}, \right. \\
&\quad \left. \inf_{\tau \in [0, T]} \left( \int_{-T}^T |f(t) - S_\tau \circ g(t)|^2 dt \right)^{1/2} \right\} \\
&= \min \left\{ \inf_{\tau \in [0, T]} \left( \frac{2T}{N} \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |X'_k - Y_k|^2 \right)^{1/2}, \inf_{\tau \in [0, T]} \left( \frac{2T}{N} \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |X_k - Y'_k|^2 \right)^{1/2} \right\} \\
&= \min \left\{ \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N} \frac{N}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta'_j - \gamma_j|^2 \right)^{1/2}, \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N} \frac{N}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta_j - \gamma'_j|^2 \right)^{1/2} \right\} \\
&= \min \left\{ \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta_j \cdot \exp\{i2\pi(n_0/N)j\} - \gamma_j|^2 \right)^{1/2}, \right. \\
&\quad \left. \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta_j - \gamma_j \cdot \exp\{i2\pi(n_0/N)j\}|^2 \right)^{1/2} \right\}.
\end{aligned}$$

Here  $n_0 = N \cdot \tau / 2T$ .

For loss function

$$L(n_0) = \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta_j \cdot \exp\{i2\pi(n_0/N)j\} - \gamma_j|^2,$$

the gradient is

$$\nabla_{n_0} = \frac{4\pi}{N} \cdot \sum_{|j| \leq (N-1)/2} j \cdot \text{Im} \left( \theta_j \overline{\gamma_j} e^{i2\pi n_0 j / N} \right).$$

For cdf  $F$  and  $G$ , define the shifted (towards left) cdf as

$$S_\tau \circ F(t) = \begin{cases} 0, & t \in [-T, -\tau) \\ F(t + \tau), & t \in [-\tau, T - \tau) \\ 1, & t \in [T - \tau, T] \end{cases} \quad (\text{A.2})$$

The Fourier coefficients of  $S_\tau \circ F$  is

$$\begin{aligned}
\theta'_j &= \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}-n_0} X_{k+n_0} \cdot \exp\{-i2\pi(k/N)j\} + \sum_{k=\frac{(N-1)}{2}-n_0+1}^{\frac{N-1}{2}} 1 \cdot \exp\{-i2\pi(k/N)j\} \\
&= \sum_{k=\frac{-(N-1)}{2}+n_0}^{\frac{N-1}{2}} X_k \cdot \exp\{-i2\pi((k-n_0)/N)j\} + \sum_{k=\frac{-(N-1)}{2}}^{\frac{-(N-1)}{2}+n_0-1} 0 + \sum_{k=\frac{(N-1)}{2}-n_0+1}^{\frac{N-1}{2}} 1 \cdot \exp\{-i2\pi(k/N)j\} \\
&= \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} X_k \cdot \exp\{-i2\pi((k-n_0)/N)j\} + \sum_{k=\frac{(N-1)}{2}-n_0+1}^{\frac{N-1}{2}} 1 \cdot \exp\{-i2\pi(k/N)j\} \\
&= \theta_j \cdot \exp\{i2\pi(n_0/N)j\} + \frac{e^{i2\pi j(n_0/N)} - 1}{1 - e^{i2\pi j/N}} \cdot \mathbf{1}_{\{j \neq 0\}} + n_0 \cdot \mathbf{1}_{\{j=0\}}.
\end{aligned}$$

The shift-invariant distance between  $F$  and  $G$  is then

$$\begin{aligned}
d(F, G) &= \min \left\{ \inf_{\tau \in [0, T]} \left( \int_{-T}^T |S_\tau \circ F(t) - G(t)|^2 dt \right)^{1/2}, \right. \\
&\quad \left. \inf_{\tau \in [0, T]} \left( \int_{-T}^T |F(t) - S_\tau \circ G(t)|^2 dt \right)^{1/2} \right\} \\
&= \min \left\{ \inf_{\tau \in [0, T]} \left( \frac{2T}{N} \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |X'_k - Y_k|^2 \right)^{1/2}, \inf_{\tau \in [0, T]} \left( \frac{2T}{N} \sum_{k=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |X_k - Y'_k|^2 \right)^{1/2} \right\} \\
&= \min \left\{ \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N} \frac{N}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta'_j - \gamma_j|^2 \right)^{1/2}, \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N} \frac{N}{N^2} \sum_{j=\frac{-(N-1)}{2}}^{\frac{N-1}{2}} |\theta_j - \gamma'_j|^2 \right)^{1/2} \right\} \\
&= \min \left\{ \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N^2} \left[ \sum_{0 < |j| \leq (N-1)/2} \left| \theta_{i,j} e^{i2\pi j n_0/N} + \frac{e^{i2\pi j n_0/N} - 1}{1 - e^{i2\pi j/N}} - \gamma_{l,j} \right|^2 + |\theta_0 + n - \gamma_0|^2 \right] \right)^{1/2}, \right. \\
&\quad \left. \min_{0 \leq n_0 \leq (N-1)/2} \left( \frac{2T}{N^2} [\dots] \right)^{1/2} \right\}.
\end{aligned}$$

Loss function

$$\begin{aligned}
L(n_0) &= \sum_{0 < |j| \leq (N-1)/2} \left| \theta_j e^{i2\pi j(n_0/N)} + \frac{e^{i2\pi j(n_0/N)} - 1}{1 - e^{i2\pi j/N}} - \gamma_j \right|^2 + |\theta_0 + n_0 - \gamma_0|^2 \\
&= \sum_{0 < |j| \leq (N-1)/2} \left| \left( \theta_j + \frac{1}{1 - e^{i2\pi j/N}} \right) e^{i2\pi j(n_0/N)} - \left( \gamma_j + \frac{1}{1 - e^{i2\pi j/N}} \right) \right|^2 + \\
&\quad |\theta_0 + n_0 - \gamma_0|^2 \\
&\triangleq \sum_{0 < |j| \leq (N-1)/2} \left| \theta_j'' e^{i2\pi j(n_0/N)} - \gamma_j'' \right|^2 + |\theta_0 + n_0 - \gamma_0|^2.
\end{aligned}$$

Gradient:

$$\nabla_{n_0} = \frac{4\pi}{N} \cdot \sum_{0 < |j| \leq (N-1)/2} j \cdot \text{Im} \left( \theta_j'' \overline{\gamma_j''} e^{i2\pi n_0 j/N} \right) + 2n_0 + 2\text{Re}(\theta_0 - \gamma_0). \quad (\text{A.3})$$

TO DO:

1. Finish proposal [by Tuesday]
2. Modify slides, practice, think about real data[by Monday]
3. Literature review: optimization method, Committee members' paper.
4. fMRI data.
5. EM algorithm vs least squared loss
6. Theory: mimic others' work.

## Acknowledgements

See [Supplement A](#) for the supplementary material example.

## Supplementary Material

### Supplement A: Title of the Supplement A

(<http://www.e-publications.org/ims/support/download/imsart-ims.zip>). Dum es-set rex in accubitu suo, nardus mea dedit odorem suavitatis. Quoniam confortavit seras portarum tuarum, benedixit filiis tuis in te. Qui posuit fines tuos

## References

- [1] AMINI, A. A. and LEVINA, E. (2018). On semidefinite relaxations for the block model. *Ann. Stat.* **46** 149–179.
- [2] BIGOT, J. and GENDRE, X. (2013). Minimax properties of Fréchet means of discretely sampled curves. *Ann. Stat.* **41** 923–956.
- [3] CHEN, Y., SANGHAVI, S. and XU, H. (2014). Improved graph clustering. *IEEE Trans. Inf. Theory* **60** 6440–6455.

- [4] GAO, C., LU, Y. and ZHOU, H. H. (2015). Rate-optimal graphon estimation. *Ann. Stat.* **43** 2624–2652.
- [5] HOLLAND, P. W., BLACKMOND, K. and LEINHARDT, S. (1983). Stochastic blockmodels: first steps Technical Report.
- [6] LE, C. M., LEVINA, E. and VERSHYNIN, R. (2016). Optimization via low-rank approximation for community detection in networks. *Ann. Stat.* **44** 373–400.
- [7] LI, X., CHEN, Y. and XU, J. (2018). Convex Relaxation Methods for Community Detection.
- [8] LIU, F., CHOI, D., XIE, L. and ROEDER, K. (2018). Global spectral clustering in dynamic networks. *Proc. Natl. Acad. Sci. U. S. A.* **115** 927–932.
- [9] MATIAS, C. and MIELE, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **79** 1119–1141.
- [10] MATIAS, C., REBAFKA, T. and VILLERS, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* **105** 665–680.
- [11] PENSKEY, M. (2019). Dynamic network models and graphon estimation. *Ann. Stat.* **47** 2378–2403.
- [12] WAN, Y., WEI, Z., LOOGER, L. L., KOYAMA, M., DRUCKMANN, S., KELLER CORRESPONDENCE, P. J. and KELLER, P. J. (2019). Single-Cell Reconstruction of Emerging Population Activity in an Entire Developing Circuit. *Cell* **179**.
- [13] XU, K. S. (2015). Stochastic block transition models for dynamic networks. *J. Mach. Learn. Res.* **38** 1079–1087.
- [14] 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.
- [15] 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.