

## SBM

Denote the  $p$  nodes by  $n_1, n_2, \dots, n_p$ . Let  $Z_i \in \{1, 2, \dots, K\}$  be the cluster that node  $n_i$  belongs to, where  $K$  is the number of clusters. Denote by  $C_{k \times k}$  the connecting probability matrix, where  $C_{kl} := P(n_i, n_j \text{ are connected} | Z_i = k, Z_j = l)$ . The observed adjacency matrix  $A_{p \times p}$  is defined as

$$A_{ij} = \begin{cases} 1, & n_i \text{ and } n_j \text{ are connected;} \\ 0, & \text{otherwise.} \end{cases}$$

SBM models this matrix by Bernoulli distribution, that is  $A_{ij} \sim \text{Bernoulli}(C_{Z_i Z_j})$ .

## Dynamic Generalization of SBM

Under the assumption that the cluster is static and the edges do not disappear once constructed, the sequence of adjacency matrices  $A(t)$  can be uniquely determined by  $T_{p \times p}$  with entries

$$T_{ij} = \min \{t : A_{ij}(t) = 1\}.$$

On the other hand, one can model the observed adjacency matrices using point processes with intensity function

$$C_{kl}(t) = \frac{P(dA_{ij}(t) = 1 | A_{ij}(t^-) = 0, Z_i = k, Z_j = l)}{dt}$$

where  $dA_{ij}(t) = A_{ij}(t + dt) - A_{ij}(t)$ .

If we estimate the clustering vector  $Z$  based on the observed connecting time matrix  $T$ , local ensembles will be recognized. More specifically, the peripheral neurons that are connected to the same MN might be grouped together, and the neuron they are connecting with can then be clustered as a MN.

One could also estimate  $Z$  based on the intensity  $\Lambda_i(t)$  of  $n_i$ .  $\Lambda_i(t)$  has an explicit expression

$$\Lambda_i(t) = P\left(\sum_{j=1}^p dA_{ij}(t) \geq 1\right) / dt$$

Should we allow more than one  
new edges at a single timepoint?  
If so, is this a proper definition?

$$\begin{aligned} &\leq \sum_{j=1}^p P(dA_{ij}(t) = 1) / dt \\ &= \sum_{j=1}^p P(A_{ij}(t^-) = 0) \cdot C_{Z_i Z_j}(t) \\ &= \sum_{k=1}^K w_{Z_i k} \cdot C_{Z_i k}(t), \end{aligned}$$

where  $w_{Z_ik} = \sum_{j:Z_j=k} P(A_{ij}(t^-) = 0)$ .

Clustering based on  $\Lambda_i(t)$  yields clusters corresponding to cell types.

How to choose a proper time period for clustering?

If the time window is too large, similarity may be weakened due to containing various phases. For example, when clusters correspond to ensembles, several neurons in the same ensemble may connect to different neurons in the phase where ensembles merge with each other; when clusters correspond to cell types, neurons of the same cell type may grow to different subtypes later on.

On the other hand, the time window cannot be too small. It should cover all the activities of the same type (e.g. recruiting).

So how to detect the phase change? And what if the neurons get into a new phase at various time points?