031421 `tensorsparse` package simulation

Chanwoo Lee, March 15, 2021

1 Simulation model

I found that function `est.SBA` (which is the main estimation function) in `graphon` package needs to have at least 4 adjacency matrices drawn from the same graphon. If there is only one adjacency matrix, it gives us a matrix whose entries are all averaged value across all entries of the adjacency matrix. Therefore, I focus on simulations of hypergraphon based on the package `tensorsparse`.

I have checked if `tensorsparse` successfully finds the true probability tensor and the estimation error decays similar to the theoretical error bound. I simulated two different settings: stochatic block model and smooth-hypergrahpon model.

1.1 Stochastic block model

I constructed stochastic block model as follows. First, for a given group size k, I generated $\mathcal{W} \in [0,1]^{k\times k\times k}$ core group tensor whose entries are drawn from i.i.d. Unif$[0,1]$. From the number of vertices n and generated core group tensor $\mathcal{W}$, I constructed an adjacency tensor of 3-uniform hyper graph $\mathcal{A} \in \{0,1\}^{n\times n\times n}$ based on the following rule.

$$\mathcal{A}_{\omega_1,\omega_2,\omega_3} \sim \text{Bernoulli}(\mathcal{W}_{z^{-1}(\omega_1),z^{-1}(\omega_2),z^{-1}(\omega_3)}), \text{ for } \omega = (\omega_1,\omega_2,\omega_3) \in E$$

where $z\colon [n] \to [k]$ is a membership function. z is chosen to have the balanced group size. `hgmodel.block` is a R function for such data generation.

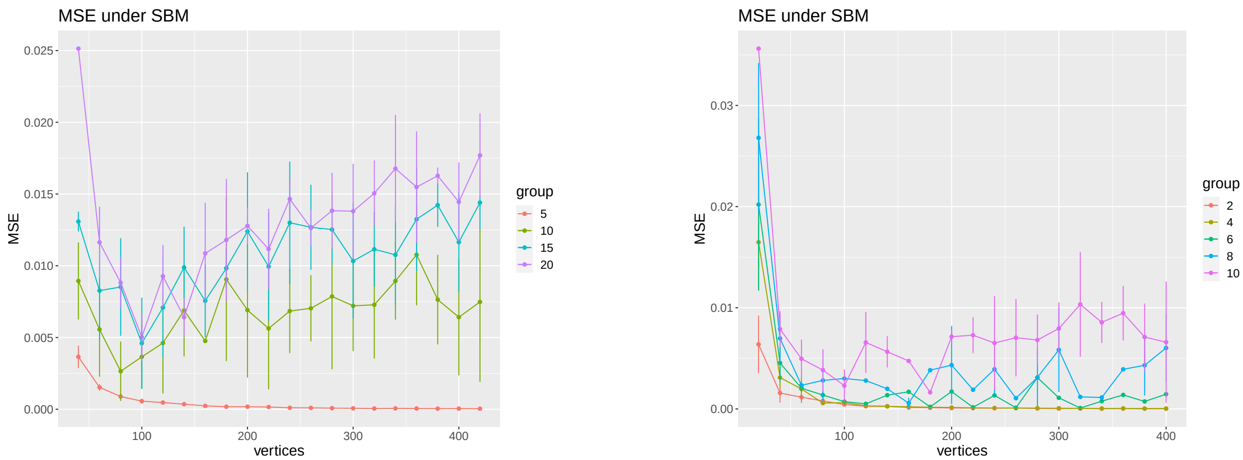
Based on the this model, I run `tbmClustering` and calculated MSE for the different group number $k \in \{2,4,5,6,8,10,15,20\}$ and the number of vertices $n \in \{20,40,\ldots,420\}$. Figure 1 shows the simulation results. MSE decays fast and converges to as n increases when the group number is small. However, I find out that the algorithm performance becomes unstable when the number of group increases. Possible explanation is since the current algorithm does not consider symmetricity of hyper graph, the error from false cluster is getting larger as the group number increases. I have checked that group clusters among modes are the same when the group size is small while the algorithm cannot capture symmetricity when the group size is large. Therefore, I am updating the algorithm to add symmetric option.

1.2 Smooth-hypergrahpon model

I generated a probability tensor from smooth hypergraphon such that

$$\Theta_{\omega_1,\omega_2,\omega_3} = f(\xi_{\omega_1},\xi_{\omega_2},\xi_{\omega_3}) = \frac{1}{1+\exp(-(\xi_{\omega_1}^2+\xi_{\omega_2}^2+\xi_{\omega_3}^2))},$$

where $\xi_1,\ldots,\xi_n$ are drawn from i.i.d Unif$[0,1]$. From the given probability tensor $\Theta \in [0,1]^{n\times n\times n}$, I generated an adjacency tensor $\mathcal{A} \in \{0,1\}^{n\times n\times n}$ by Bernoulli trial. Figure 2 shows the MSE across $k \in \{2,4,8,16\}$ and $n \in \{20,40,\ldots,400\}$. In this setting, the algorithm shows more stable performance estimating the true probability tensor compared to the stochastic block model and MSE decays smoothly as the number of observation increases.



(a) When the group number is quite large (5,10,15,20).

(b) When the group number is quite small (2,4,6,8,10).

Figure 1: MSE depending on the group number k and the number of vertices n under stochastic block model.

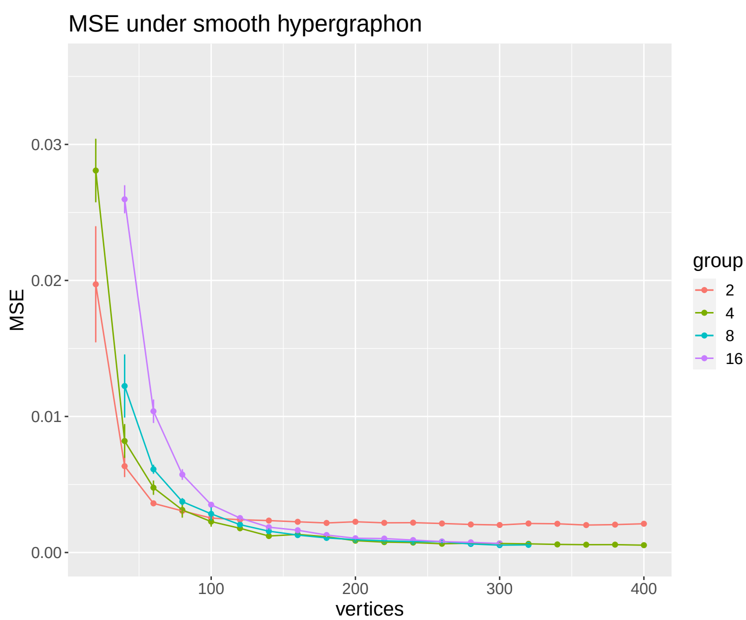


Figure 2: MSE depending on the group number k and the number of vertices n under smooth hypergraphon model.