Graphon estimation review 2

Chanwoo Lee
February 22, 2021

1 Literature of graphon estimation

Aldous [1981] and Hoover [1979] independently proved the representation theorem for exchangeable arrays. In the jointly exchangeable array $\boldsymbol{A}_{ij}$ case, one gets a representation for the array as

$$\boldsymbol{A}_{ij} = f(\xi_i, \xi_j, \xi_{ij}), \text{ for } i, j \in \mathbb{N},$$

for some function $f\colon [0,1]^3 \to \{0,1\}$ such that $f(u,v,w) = f(v,u,w)$ for all u, v, w and some independent Unif$(0,1)$ random variables $\xi_i, \xi_{ij} = \xi_{ji}$, $1 \le i \le j$. This distribution are naturally parrametrized by function

$$w(u,v) = \mathbb{P}[\boldsymbol{A}_{ij} = 1 | \xi_i = u, \xi_j = v].$$

Diaconis and Janson [2007] showed that such $w(\cdot,\cdot)$ is not uniquely determine $\mathbb{P}$ but if w_1 and w_2 define the same $\mathbb{P}$, then there exists measure preserving transformation ϕ such that $w_1(u,v) = w_2(\phi(u), \phi(v))$. In addition, they viewed a finite adjacency matrix $\boldsymbol{A}_{ij} \in \mathbb{R}^{n\times n}$ as a partial observation of the limiting object under Bernoulli sampling. Based on a measure preserving transformation that makes $\int_0^1 w(u,v)dv$ monotone nondecreasing, one can have canonical representation of $w_{CAN}(u,v)$ [Bickel and Chen, 2009].

Early development of graphon estimation focused on community detection for the fixed number of group (k) or determined k by algorithms. Newman and Girvan [2004] first suggested modularity that evaluate community structure in networks. Newman-Girvan modularity measures the fraction of the edges on the graph within the same group minus the expected value of the same quantity on a graph with the same community but random connection. Later Newman [2006] developed spectral method to maximize the modularity and extended to multiple groups. Many modifications of Newman-Grivan modularity has been studied [Fortunato, 2010]. Bickel and Chen [2009] compared the Nerman-Girvan algorithm to likelihood-based modularity and constructed theoretical consistency when $\mathbb{E}(\text{Degree})/\log n \to \infty$.

Paradigm of graphon estimation has changed from nonparametric histogram approaches [Wolfe and Olhede, 2013, Olhede and Wolfe, 2014, Chan and Airoldi, 2014]. First, it is not necessary to assume the data to have been generated by SBM. Second, instead of SBM assumption, the graph $w(\cdot,\cdot)$ is assumed to be Holder-continuous [Olhede and Wolfe, 2014] or satisfy strict monotonicity of degree condition [Chan and Airoldi, 2014]. Third, the number of community $k = \lfloor n/h \rfloor$ is now considered as a bandwidth. This concept is shifted away from the standard use of the SBM. Instead of representing community structure, a block model is used as a universal mechanism to represent an arbitrary unlabeled network. The goal in the setting of exchangeable networks is therefore, no longer to discover latent community structure, but rather simply to group together nodes whose patterns of interaction are similar. Like kernel smoothing approaches, optimal bandwidth h is set to be a function of the number of vertices n, smoothness parameter M, and sparsity parameter ρ_n balancing different types of errors. Lastly, the goal of graphon estimation becomes minimizing MSE type of errors not clustering misclassification errors by the above reasons.

References

David J Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598, 1981.

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216. PMLR, 2014.

Persi Diaconis and Svante Janson. Graph limits and exchangeable random graphs. *arXiv preprint arXiv:0712.2749*, 2007.

Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

Douglas N Hoover. Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ*, 2, 1979.

Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

Sofia C Olhede and Patrick J Wolfe. Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41):14722–14727, 2014.

Patrick J Wolfe and Sofia C Olhede. Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*, 2013.