ESTIMATION OF EXCHANGEABLE GRAPH MODELS BY STOCHASTIC BLOCKMODEL APPROXIMATION

Stanley H. Chan [1,2] *, Thiago B. Costa* [1,2] *and Edoardo M. Airoldi* [2]

[1] School of Engineering and Applied Sciences, and [2] Department of Statistics,
Harvard University, Cambridge, MA 02138, USA.

ABSTRACT

We consider a non-parametric perspective of analyzing network data. Our goal is to seek a limiting object of a sequence of exchangeable random arrays called the *graphon*. We propose a numerically efficient algorithm for estimating graphons and we show that the proposed algorithm yields a consistent estimate as the size of the graph grows. Preliminary experiments show that the algorithm is effective in estimating stochastic block-models and continuous graphons.

***Index Terms*—** Network analysis, exchangeable random graph model, stochastic blockmodel, non-parametric estimation, graphon, graphlet.

1. INTRODUCTION

1.1. Models of Network Data

Network analysis is a rapidly evolving research topic. From the pioneer work of Erdős to the recent surge of mining data from large-scale online social networks, the statistical tools for analyzing networks have been continuously advancing [1]. Neverthelss, these methods share the same common goal - how do we extract the meaningful but hidden structure from the data we observed?

Among many statistical methods and stylized models proposed in the literature [2], perhaps the most commonly used strategy is to seek a parametric model that best describes the data so that subsequent inferences can be made. Some well-known examples include, for example, exponential random graph models [3, 4], stochastic block models [5, 6], mixed membership stochastic block models [7], latent space models [8], graphlets [9] and many others [10].

In this paper, we attempt to answer the same question, but from a non-parametric and frequentists' perspective. This non-parametric model is originated from the theory of graph limits [11, 12] for studying exchangeable random arrays [13]. The theory predicts that every convergent sequence of graphs has a limit object, which we call it *graphon*, that preserves local and global properties of the graphs in the sequence. The goal of this paper is to develop new statistical tools to analyze network data from the graphon model.

This work was supported, in part, by a Croucher Foundation Postdoctoral Research Fellowship to SHC, and by an Alfred P. Sloan Research Fellowship and an NSF CAREER award to EMA.

Emails: schan@seas.harvard.edu,{tcosta,airoldi}@fas.harvard.edu.

1.2. Generating Random Graphs from a Graphon

To have a more precise understanding of a graphon, it is helpful to first understand how it generates a random graph.

First, we define a graphon as a measurable function $w : [0,1]^2 \to [0,1]$. The input to the graphon is a pair of random variables $(u_i, u_j) \in [0,1]^2$ drawn from a uniform distribution, *i.e.*, $u_i \sim \text{Uniform}[0,1]$, which can be interpreted as random labels of the nodes in a network. Thus, for a network of n nodes, there will be a set of n labels $\{u_i\}_{i=1}^n$.

Conditioned on (u_i, u_j), the output of the graphon is a scalar $w(u_i, u_j) \in [0,1]$, which specifies the (conditional) probability that the ith node and the jth node are linked given the labels (u_i, u_j), *i.e.*,

$$\Pr\Big[G[i,j] = 1 \,\Big|\, u_i, u_j\Big] = w(u_i, u_j) \tag{1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, n$, where $G[i,j]$ denotes the (i,j)th entry of the graph adjacency matrix. A pictorial illustration of the graph generating process is shown in Figure 1. To elaborate the ideas further, we consider the following two examples.

Example 1 (Erdős-Rényi Graph)**.** *The Erdős-Rényi graph specifies that an edge linking any two nodes of the graph has a probability p, identical for all labels (u_i, u_j). Therefore, $\Pr[G[i,j] = 1 \,|\, u_i, u_j] = p$ for all i, j. Hence, the graphon is $w(u,v) = p$, for all $(u,v) \in [0,1]^2$.*

Example 2 (Stochastic Block Models)**.** *A finite-class stochastic block model is specified by the inter- and intra- class probabilities of the nodes. Consider a network of K membership classes, we partition the unit interval $[0,1]$ equally into K sub-intervals $I_1, \ldots, I_K$, and assign the probability of linking nodes i and j given the membership classes be $\Pr[G[i,j] = 1 \,|\, u_i \in I_k, u_j \in I_l] = p_{kl}$, for some constant p_{kl}, $k \in \{1, \ldots, K\}$ and $l \in \{1, \ldots, K\}$. Therefore, the graphon is a piecewise constant function: $w(u_i, u_j) = p_{kl}$.*

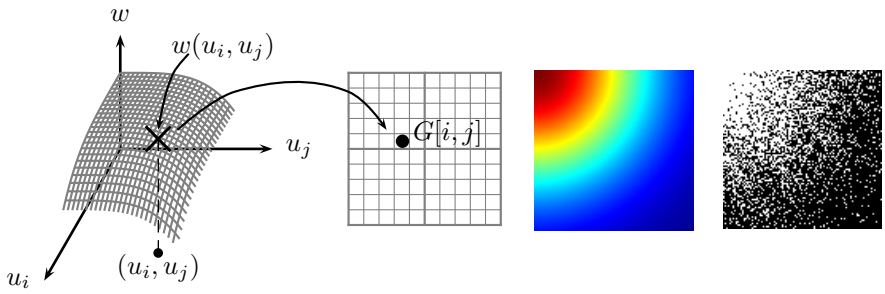


Fig. 1: [Left] Given a graphon $w : [0,1]^2 \to [0,1]$, we draw i.i.d. samples u_i, u_j from Uniform[0,1] and assign $G[i,j] = 1$ with probability $w(u_i, u_j)$. [Middle] Heat map of an example graphon w. [Right] A random graph generated by the graphon shown in the middle. The rows and columns of the graph are ordered by increasing u_i (instead of i) for better visualization.

1.3. Problem Statement: How to Estimate a Graphon?

The problem of interest is the following fundamental question: Given some observed graph adjacency matrices generated from a graphon w, can we make an estimate $\widehat{w}$ of w such that $\widehat{w} \to w$ with high probability as $n \to \infty$?

The proposed method in this paper is called the Stochastic Block-model Approximation (SBA) algorithm. The SBA algorithm, to our best knowledge, is the first frequentist's approach in the literature to estimate the graphon from the observed data. As the name suggested, the main idea of the SBA algorithm is to approximate w by a two-dimensional step function $\widehat{w}$ (the stochastic block model). The intuition is that as n grows, the density of the spatially normalized graph on $[0,1]^2$ should also grow. Consequently, the approximation error should converge to 0 as $n \to \infty$.

The rest of the paper is organized as follow. In Section 2 we present the SBA algorithm and discuss its properties. Some preliminary experimental results are shown in Section 3, and a concluding remark is given in Section 4. Due to limited space, theories and proofs mentioned in this paper are left to a follow up paper.

2. STOCHASTIC BLOCK APPROXIMATION

2.1. Overview of SBA Algorithm

The idea of the proposed SBA algorithm is to approximate a continuous graphon w by a piecewise constant function $\widehat{w}$. In order to do that, we should first cluster the nodes $\{1, \ldots, n\}$ into K membership classes $\widehat{B}_1, \ldots, \widehat{B}_K$ (called the *blocks*). Once the blocks are identified, the probability of linking two nodes (*i.e.*, the value on the graphon) can be estimated from the empirical frequency of the number of edges in the blocks:

$$\widehat{w}(u_i, u_j) = \frac{1}{|\widehat{B}_k||\widehat{B}_l|} \sum_{x \in \widehat{B}_k} \sum_{y \in \widehat{B}_l} G[x,y] \stackrel{\text{def}}{=} p_{kl},$$

where $i \in \widehat{B}_k$ and $j \in \widehat{B}_l$. Therefore, the key question is how to perform the clustering using only the observed graphs.

Example 3. *Consider the graphon shown in Figure 2 where there are $K = 3$ blocks. Assume that the n nodes have already been perfectly partitioned into these 3 blocks. Then, the probability p_{kl} $(k,l \in \{1,2,3\})$ can be estimated from the empirical frequency of the edges in the (k,l)th sub-matrix of the graph.*

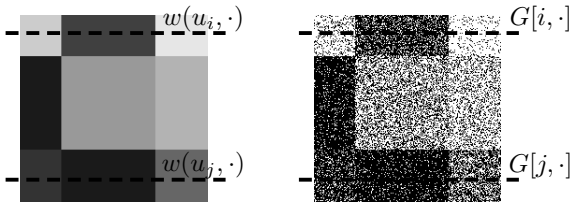


Fig. 2: An example graphon of $K = 3$ blocks (See Example 3). Dotted lines in the left denote the graphon slices: $w(u_i, \cdot)$ and $w(u_j, \cdot)$; Dotted lines in the right denote the corresponding slices of $G[i, \cdot]$ and $G[j, \cdot]$ (See Example 4).

2.2. Similarity of Graphon Slices

Our proposed clustering algorithm is based on the observation that if two nodes i and j belongs to the same block, then the interaction from *all* other nodes to i should also be the same as to j. Correspondingly, the horizontal cross-sections of the graphon $w(u_i, \cdot)$ and $w(u_j, \cdot)$ must be equal, so does the vertical cross-sections $w(\cdot, u_i)$ and $w(\cdot, u_j)$. Therefore, if we ought to group $\{1, \ldots, n\}$ into K blocks, a natural way is to compute the distance

$$d_{ij} = \frac{1}{2}\Big(\int_0^1 \big(w(u_i, y) - w(u_j, y)\big)^2 \, dy + \ldots + \int_0^1 \big(w(x, u_i) - w(x, u_j)\big)^2 \, dx \Big), \qquad (2)$$

and group nodes according to d_{ij}, because (2) is the L_2 distances between the cross-sections.

In practice, d_{ij} is never known and has to be estimated from the observed graphs. Consider a set of $2T$ (an even number of) observed graphs $G_1, \ldots, G_{2T}$, our estimator $\widehat{d}_{ij}$ for d_{ij} can be derived by first expressing d_{ij} as

$$d_{ij} = \frac{1}{2}\Big[(r_{ii} - r_{ij} - r_{ji} + r_{jj}) + (c_{ii} - c_{ij} - c_{ji} + c_{jj})\Big],$$

where each $c_{ij} = \int_0^1 w(x, u_i)w(x, u_j)dx$ is one of the terms that can be found by expanding the squares in (2), so does $r_{ij} = \int_0^1 w(u_i, y)w(u_j, y)dy$. Inspecting this expression, we propose to use the following estimator $\widehat{d}_{ij}$ for d_{ij}:

$$\widehat{d}_{ij} = \frac{1}{2}\Big[\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}}\Big\{(\widehat{r}^k_{ii} - \widehat{r}^k_{ij} - \widehat{r}^k_{ji} + \widehat{r}^k_{jj}) + \ldots + (\widehat{c}^k_{ii} - \widehat{c}^k_{ij} - \widehat{c}^k_{ji} + \widehat{c}^k_{jj})\Big\}\Big], \quad (3)$$

where $\mathcal{S} = \{1, \ldots, n\}\backslash\{i, j\}$ is the set of summation indices, and so the summation $\sum_{k\in\mathcal{S}}\{\cdots\}$ is analogous to the integration in (2). The individual terms in (3) are

$$\widehat{c}^k_{ij} = \frac{1}{T^2}\left(\sum_{1\le t_1\le T} G_{t_1}[k, i]\right)\left(\sum_{T<t_2\le 2T} G_{t_2}[k, j]\right), \quad (4)$$

$$\widehat{r}^k_{ij} = \frac{1}{T^2}\left(\sum_{1\le t_1\le T} G_{t_1}[i, k]\right)\left(\sum_{T<t_2\le 2T} G_{t_2}[j, k]\right), \quad (5)$$

and index k is equivalent to dummy variables x and y in (2).

Example 4. *To interpret* (4) *and* (5)*, it is helpful to refer to Figure 2. If we want to determine d_{ij}, we need to compute terms like $\int_0^1 w(u_i, y)w(u_j, y)dy$ in* (2)*. Since $w(u_i, \cdot)$ and $w(u_j, \cdot)$ are not known, we approximate $w(u_i, \cdot)$ by $G[i, \cdot]$ and $w(u_j, \cdot)$ by $G[j, \cdot]$.*

The following theorem is the basis for a series of consistency theories, of which details are left a follow up paper.

Theorem 1. *The estimator $\widehat{d}_{ij}$ for d_{ij} is unbiased,* i.e.*, $\mathbb{E}[\widehat{d}_{ij}] = d_{ij}$. Further, for any $\epsilon > 0$,*

$$\Pr\left[\left|\widehat{d}_{ij} - d_{ij}\right| > \epsilon\right] \le 8e^{-\frac{S\epsilon^2}{4(1/2T + 2\epsilon/3)}}, \quad (6)$$

where S is the size of the neighborhood $\mathcal{S}$, and $2T$ is the number of observations.

2.3. Clustering

The similarity metric $\widehat{d}_{ij}$ discussed above suggests one simple method to cluster the unknown labels $\{u_1, \ldots, u_n\}$ using a greedy approach as shown in Algorithm 1. Starting with $\Omega = \{u_1, \ldots, u_n\}$, we randomly pick a node i_p and call it the *pivot*. Then for all other vertices $i_v \in \Omega\backslash\{i_p\}$, we compute the distance $\widehat{d}_{i_p,i_v}$ and check whether $\widehat{d}_{i_p,i_v} < \Delta^2$ for some precision parameter $\Delta > 0$. If $\widehat{d}_{i_p,i_v} < \Delta^2$, then we assign i_v to the same block as i_p. Ffter scanning through Ω once, a block $\widehat{B}_1 = \{i_p, i_{v_1}, i_{v_2}, \ldots\}$ will be defined. By updating Ω as $\Omega \leftarrow \Omega\backslash\widehat{B}_1$, the process repeats until $\Omega = \emptyset$.

Algorithm 1 Stochastic Block-model Approximation

Input: A set of observed graphs $G_1, \ldots, G_{2T}$ and the precision parameter Δ.
Output: Estimated stochastic blocks $\widehat{B}_1, \ldots, \widehat{B}_K$.
Initialize: $\Omega = \{1, \ldots, n\}$, and $k = 1$.
while $\Omega \neq \emptyset$ **do**
  Randomly choose a vertex i_p from Ω and assign it as the pivot for $\widehat{B}_k$: $\widehat{B}_k \leftarrow i_p$.
  for Every other vertices $i_v \in \Omega\backslash\{i_p\}$ **do**
    Compute the distance estimate $\widehat{d}_{i_p,i_v}$.
    If $\widehat{d}_{i_p,i_v} \le \Delta^2$, then assign i_v as a member of $\widehat{B}_k$: $\widehat{B}_k \leftarrow i_v$.
  end for
  Update Ω: $\Omega \leftarrow \Omega\backslash\widehat{B}_k$.
  Update counter: $k \leftarrow k + 1$.
end while

The complexity of this algorithm is $\mathcal{O}(TSKn)$, where T is half the number of observations, S is the size of the neighborhood $\mathcal{S}$, K is the number of blocks and n is number of vertices of the graph.

3. PRELIMINARY EXPERIMENTAL RESULTS

In this section we provide preliminary evaluation results of the proposed SBA algorithm. For the purpose of comparison, we implemented two of the most recent algorithms that could be potentially used for estimating graphons. The first algorithm is the universal singular value thresholding (USVT)[14]. The second algorithm is the Largest Gap Algorithm (LG) [15], a stochastic block estimation algorithm to detect block structures if there are large gaps in the degree distribution.

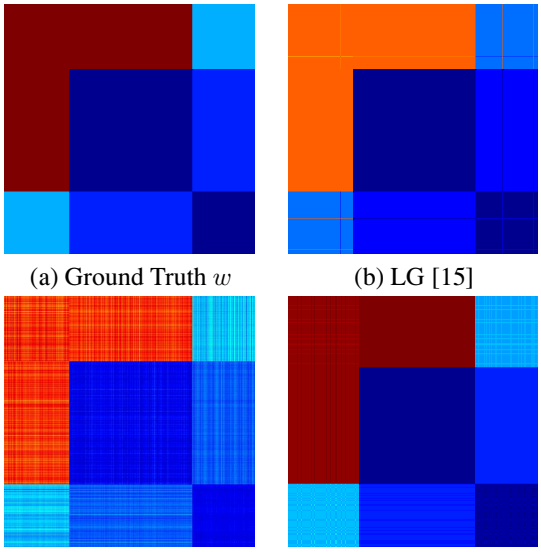
The quality of the estimation is determined by the mean squared error (MSE), defined as $\text{MSE} = \|w - \widehat{w}\|_2^2/n$.

3.1. Estimating Stochastic Blocks

Our first experiment is to evaluate the proposed SBA algorithm for estimating stochastic blocks. For this purpose, we generate a $K = 3$ graphon

$$w = \begin{bmatrix} 0.8 & 0.8 & 0.3 \\ 0.8 & 0.1 & 0.2 \\ 0.3 & 0.2 & 0.1 \end{bmatrix},$$

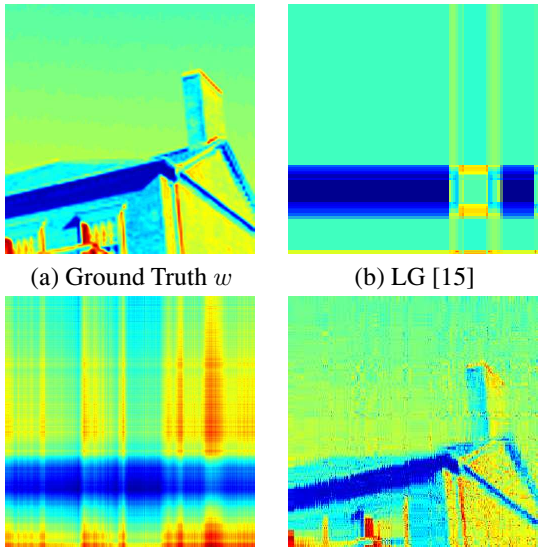
and constructed a random graph of $n = 500$ nodes with 2 observations ($T = 1$). The result is shown in Figure 3.



(a) Ground Truth w (b) LG [15]

(c) USVT [14] (d) SBA (Proposed)

Fig. 3: Estimating stochastic block models of $K = 3$ classes. The MSE values are: LG: $\text{MSE} = 5.9 \times 10^{-4}$, USVT: $\text{MSE} = 1.2 \times 10^{-3}$, SBA (Proposed) $\text{MSE} = 3.8 \times 10^{-5}$. Here, no. observations is 2 and $n = 500$.



(a) Ground Truth w (b) LG [15]

(c) USVT [14] (d) SBA (Proposed)

Fig. 4: Estimating an arbitrary graphon w. The MSE values are: LG: $\text{MSE} = 1.089 \times 10^{-2}$, USVT: $\text{MSE} = 1.116 \times 10^{-2}$, SBA (Proposed) $\text{MSE} = 2.969 \times 10^{-3}$. Here, no. observations is 20 and $n = 500$.

3.2. Estimating Continuous Graphon

Our next experiment to consider a continuous graphon. In order to show the ability of the proposed SBA algorithm, we define w as a gray-scaled image on $[0, 1]^2$, shown in Figure 4(a), and generate 20 observed graphs ($T = 10$), each with $n = 500$ nodes. The results are shown in Figure 4.

4. CONCLUSION

Graphons are powerful non-parametric models for network analysis, subsuming a wide range of existing parametric models and providing great flexibilities to model relational data. The proposed stochastic block-model approximation (SBA) algorithm is a numerically efficient algorithm to estimate the graphon from a small collection of graphs that it generates. From the preliminary experimental results, we found that SBA yields the good estimation results on both stochastic block models and continuous graphons. Future work will be focused on theoretical aspects of the SBA algorithm and applications of the graphon.

5. REFERENCES

[1] E. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*, Springer, 2009.

[2] E. Airoldi, X. Bai, and K. Carley, "Network sampling and classification: An investigation of network model representations," *Decision Support Systems*, vol. 51, pp. 506–518, Jun. 2011.

[3] S. Wasserman, *Social Network Analysis: Methods and Applications*, Cambridge University Press, 1994.

[4] D. R. Hunter and M. S. Handcock, "Inference in curved exponential family models for networks," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 565–583, 2006.

[5] K. Nowicki and T. Snijders, "Estimation and prediction of stochastic blokstructures," *Journal of American Statistical Association*, vol. 96, pp. 1077–1087, 2001.

[6] D. Choi, P. Wolfe, and E. Airoldi, "Stochastic blockmodels with a growing number of classes," *Biometrika*, vol. 99, pp. 273–284, 2012.

[7] E. Airoldi, D. Blei, S. Fienberg, and E. Xing, "Mixed-membership stochastic blockmodels," *Journal of Machine Learning Research*, vol. 9,, pp. 1981–2014, 2008.

[8] P. Hoff, A. Raftery, and M. Handcock, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1090–1098, 2002.

[9] H. Soufiani and E. M. Airoldi, "Graphlet decomposition of a weighted network," *Journal of Machine Learning Research, W&CP*, vol. 22, pp. 54–63, 2012.

[10] A. Goldenberg, A. Zheng, S. Fienberg, and E. Airoldi, "A survey of statistical network models," *Foundations and Trends in Machine Learning*, vol. 2, pp. 129–233, 2009.

[11] P. Diaconis and S. Janson, "Graph limits and exchangeable random graphs," arXiv:0712.2749, Dec 2007.

[12] L. Lovasz and B. Szegedy, "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, vol. 96, pp. 933–957, 2006.

[13] O. Kallenberg, *Probabilistic Symmetries and Invariance Principles*, Springer, 2005.

[14] S. Chatterjee, "Matrix estimation by universal singular value thresholding," arXiv:1212.1247, Dec 2012.

[15] A. Channarond, J. Daudin, and S. Robin, "Classification and estimation in the Stochastic Blockmodel based on the empirical degrees," *Electronic Journal of Statistics*, vol. 6, pp. 2574–2601, 2012.