Supplemental Material for Sampling and Estimation for (Sparse) Exchangeable Graphs

1. Additional Figures.

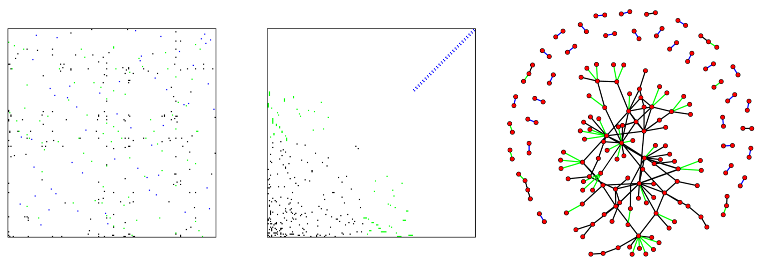

Fig 1: Realization of unlabeled KEG generated by $\mathcal{W} = (I, S, W)$ at size $s = 15$ (right panel), and associated dilated empirical graphon (left and center panels). The generating graphex is $W = (x+1)^{-2}(y+1)^{-2}$, $S = 1/2\exp(-(x+1))$, and $I = 0.1$. The observation size is $s = 15$. The dilated empirical graphon is pictured as two equivalent representations $\hat{W}_{(G,15)}$ and $\hat{W}'_{(G,15)}$, each with support $[0,12)^2$ (180 vertices at size 15). Edges from the W component are shown in black, edges from the S component are shown in green, and edges from the I component are shown in blue. Recall that the ordering of the dilated empirical graphon is arbitrary, so the left and center panels depict different representations of the same estimator. The leftmost panel shows the dilated empirical graphon with a random ordering. The middle panel shows the dilated empirical graphon sorted to group the I, S, and W edges, with the W edges sorted as in Fig. 2. The middle panel gives some intuition for why the dilated empirical graphon is able to estimate the entire graphex triple: When a KEG is generated according to $\hat{W}_{(G,15)}$ with latent Poisson process Π, the disjoint structure of the dilated graphon regions due to the I, S, and W components induces a natural partitioning of Π into independent Poisson processes that reproduce the independence structure used in the full generative model Eq. (2.1).

2. Proofs of Convergence in Distribution of Random Labelings.

Here we give proofs of the results of Section 4.1 establishing that that, almost surely, $\mathrm{P}(\mathsf{Lbl}_{s_k}(G_k) \in \cdot \mid G_k) \to \mathrm{uKEG}(\mathcal{W}, \infty)$ weakly as $k \to \infty$.

We adapt techniques from the distributional convergence of point processes. We will need the following definition and technical lemma: A separating class for a locally compact second countable Hausdorff space S is a class $\mathcal{U} \subset S$ such that for any compact open sets with $K \subset G$ there is some $U \in \mathcal{U}$ with $K \subset U \subset G$.

LEMMA 2.1. *Let $\phi, \phi_1, \phi_2, \ldots$ be simple point processes on a locally compact second countable Hausdorff space S. If*

$$\phi_n(U) \xrightarrow{d} \phi(U), \ n \to \infty$$

weakly for all U in some separating class for S then

$$\phi_n \xrightarrow{d} \phi, \ n \to \infty$$

weakly.

PROOF. By [22, Thms. 16.28 and 16.29], it suffices to check that $\mathrm{P}(\phi_n(U) = 0) \to \mathrm{P}(\phi(U) = 0)$ and that $\limsup_n \mathrm{P}(\phi_n(U) > 1) \leq \mathrm{P}(\phi(U) > 1)$. Because $\phi_n(U)$ is a non-negative integer a.s., both conditions are implied by $\phi_n(U) \xrightarrow{d} \phi(U)$. □

THEOREM 4.6. *Let Γ be a KEG generated by a non-trivial graphex $\mathcal{W}$, let $s_1, s_2, \ldots$ be a sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ as $k \to \infty$ and let $G_k = \mathcal{G}(\Gamma_{s_k})$ for all k. Then $\mathrm{P}(\mathsf{Lbl}_{s_k}(G_k) \in \cdot \mid G_k) \to \mathrm{KEG}(\mathcal{W})$ weakly almost surely.*

PROOF. For each $k \in \mathbb{N}$, conditional on G_k, let ξ^k be a point process with law $\mathrm{P}(\mathsf{Lbl}_{s_k}(G_k) \in \cdot \mid G_k)$. Note that $\Gamma \sim \mathrm{KEG}(\mathcal{W})$. Observe that the collection $\mathcal{U}$ of finite unions of rectangles with rational end points is a separating class for $\mathbb{R}_+^2$. Further, ξ^k is simple for all $k \in \mathbb{N}$, as is Γ. Thus by Lemma 2.1, to show the claimed result it will suffice to show that, for all $U \in \mathcal{U}$, $\mathrm{P}(\xi^k(U) \in \cdot \mid G_k) \to \mathrm{P}(\Gamma(U) \in \cdot)$ weakly as $k \to \infty$.

Fix U. To establish this condition we first show that for all bounded continuous functions f, it holds that $\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k] = \mathbb{E}[f(\Gamma(U))]$ a.s. Let $\mathcal{F}_{-s}$ be the σ-algebra generated by partially labelled graph derived from Γ by forgetting the labels of all nodes with label $\theta_i < s$. Formally, $\mathcal{F}_{-s}$ is the sigma algebra generated by $\mathcal{G}(\Gamma_s)$ and $\Gamma(\cdot \cap \mathbb{R}_+^2 \backslash [0, s)^2)$. Take $r \in \mathbb{R}_+$ large enough so that $U \subset [0, r]^2$. Then for $s_k > r$,

$$\mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s_k}] = \mathbb{E}[f(\xi^k(U)) \mid G_k].$$

Define $U_t = U + (t, t)$ for $t \in \mathbb{R}_+$ and let

$$X_s^{(r)} = \frac{1}{s - r} \int_0^{s-r} f(\Gamma(U_t)) \mathrm{d}t.$$

Observe that for s such that $t \leq s - r$, the joint exchangeability of Γ implies

$$\mathbb{E}[f(\Gamma(U_t)) \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}].$$

Moreover, by the linearity of conditional expectation, for $s > r$, it holds that $\mathbb{E}[X_s^{(r)} \mid \mathcal{F}_{-s}] = \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-s}]$.

A standard result [2, Ex. 5.6.2] shows that $\lim_{k\to\infty} \mathbb{E}[X_{s_k}^{(r)} \mid \mathcal{F}_{-s_k}] = \mathbb{E}[X_\infty^{(r)} \mid \mathcal{F}_{-\infty}]$ a.s. if $X_{s_k}^{(r)} \to X_\infty^{(r)}$ a.s. and there is some integrable random variable that dominates $X_{s_k}^{(r)}$ for all k; the second condition holds because f is bounded. Notice that $Y_t = f(\Gamma(B_t))$ is a stationary stochastic process. Moreover, it's easy to see from the Kallenberg exchangeable graph construction that Y_t and $Y_{t'}$ are independent whenever $|t - t'| > r$, so (Y_t) is mixing. The ergodic theorem then gives $\lim_{k\to\infty} X_{s_k}^{(r)} = \mathbb{E}[f(\Gamma(U))]$ a.s. This means

$$\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k] \to \mathbb{E}[f(\Gamma(U))] \text{ a.s.,}$$

as promised.

For $l \in \mathbb{Z}_+$, let $f_l(\cdot) = 1[\cdot \le l]$, let $A_l^{(U)}$, for each $U \in \mathcal{U}$, be the set on which

$$\lim_{k\to\infty} \mathbb{E}[f_l(\xi^k(U)) \mid G_k] = \mathbb{E}[f_l(\Gamma(U))]$$

and let $A_U = \bigcap_l A_l^{(U)}$. We have shown that $\mathrm{P}(A_l^{(U)}) = 1$, and so $\mathrm{P}(A_U) = 1$ and on A_U it holds that $\lim_{k\to\infty} \mathrm{P}(\xi^k(U) \in \cdot \mid G_k) = \mathrm{P}(\Gamma(U) \in \cdot)$ weakly. Let $A = \bigcap_{U\in\mathcal{U}} A_U$, then $\mathrm{P}(A) = 1$ and on A it holds that

$$\lim_{k\to\infty} \mathrm{P}(\xi^k(U) \in \cdot \mid G_k) = \mathrm{P}(\Gamma(U) \in \cdot)$$

weakly for all $U \in \mathcal{U}$, completing the proof. □

Next, we extend this result to allow the observation sizes to be the jump sizes of the Kallenberg exchangeable graph.

THEOREM 4.8. *Let Γ be a Kallenberg exchangeable graph generated by a non-trivial graphex $\mathcal{W}$, and let $\tau_1, \tau_2, \ldots$ be the jump sizes of Γ. Let $G_k = \mathcal{G}(\Gamma_{\tau_k})$ for each $k \in \mathbb{N}$. Then $\mathrm{P}(\mathsf{Lbl}_{\tau_k}(G_k) \in \cdot \mid G_k, \tau_k) \to \mathrm{KEG}(\mathcal{W})$ weakly almost surely as $k \to \infty$.*

PROOF. For each $k \in \mathbb{N}$, let ξ^k be a point process with law $\mathrm{P}(\mathsf{Lbl}_{\tau_k}(G_k) \in \cdot \mid G_k, \tau_k)$.

As in the proof of Theorem 4.6, to establish the claim it suffices to show that, for all bounded continuous functions f and all rectangles U, it holds that

$$\lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k, \tau_k] = \mathbb{E}[f(\Gamma(U))] \text{ a.s.}$$

Let $\mathcal{F}_{-s}$ be as in proof of Theorem 4.6. Because $\tau_k \uparrow \infty$ a.s. as $k \to \infty$, by the same argument used in Theorem 4.6 it holds that

$$\lim_{k\to\infty} \mathbb{E}[f(\Gamma(U)) \mid \mathcal{F}_{-\tau_k}] = \lim_{k\to\infty} \mathbb{E}[f(\xi^k(U)) \mid G_k, \tau_k] \text{ a.s.}$$

It is clear that $\mathcal{F}_{-\tau_k} \subset \mathcal{F}_{-\tau_{k-1}}$ for all k. The l.h.s. thus converges by reverse martingale convergence, so the r.h.s. also exists a.s.

It remains to identify the limit. To that end, we will define a coupling between the counts on test set U at a subsequence of the jump sizes and the counts on U at some deterministic sequence, which is known to converge to the desired limit. Let $s_k = \sum_{n=1}^{k} \frac{1}{n}$, let $\{\tau_{k_j}\}$ be a subsequence of the jump sizes defined such that at most one point in $\{\tau_{k_j}\}$ lies in $[s_l, s_{l+1})$ for all l and define s_{k_j} to be the subsequence of $\{s_k\}$ such that s_{k_j} is the smallest value in $\{s_k\}$ that is larger than τ_{k_j}. This gives a subsequence of the jump sizes and a subsequence of $\{s_k\}$ such that the points s_{k_j} and τ_{k_j} become arbitrarily close as $j \to \infty$.

Let $\xi^{(s,j)} = \mathsf{Lbl}_{s_{k_j}}(G_{k_j})$, and let $\xi^{(\tau,j)} = \mathsf{Lbl}_{\tau_{k_j}}(G_{k_j})$. Our goal is to couple $\xi^{(\tau,j)}(U)$ and $\xi^{(s,j)}(U)$.

Consider the case $U = [0,r]^2$ for some $r \in \mathbb{R}_+$. Each vertex of $\xi^{(s,j)}$ is included in $[0,r]$ independently with probability r/s_{k_j}, and each vertex of $\xi^{(\tau,j)}$ independently with probability r/τ_{k_j}. It thus suffices to couple r/s_{k_j}-sampling from G_{k_j} with r/τ_{k_j}-sampling from G_{k_j}. Because $s_{k_j} > \tau_{k_j}$, we may produce the r/s_{k_j}-sample by τ_{k_j}/s_{k_j}-sampling from the r/τ_{k_j}-sample. The two graphs will be equal if every edge of the r/τ_{k_j}-sample is selected by τ_{k_j}/s_{k_j}-sampling. This is indeed the case, because: (i) by construction, $\tau_{k_j}/s_{k_j} \to 1$ a.s. as $j \to \infty$. And, (ii) the number of edges of the r/τ_{k_j}-sample is $\xi^{(\tau,j)}(U)$, which is dominated by $\mathsf{Lbl}_{s_{k_{j-1}}}(G_{k_j})(U)$, which converges in expectation by Theorem 4.6. That is, the expected number of edges *not* selected by the τ_{k_j}/s_{k_j}-sampling goes to 0 by dominated convergence. Thus,

$$\begin{aligned}\lim_{j\to\infty} \mathbb{E}[f(\xi^{k_j}(U)) \mid G_{k_j}, \tau_{k_j}] = \lim_{k_j\to\infty} \mathbb{E}[f(\xi^{(\tau,j)}(U)) \mid G_{k_j}, \tau_{k_j}] \\ \mathbb{E}[f(\xi^{(s,j)}(U)) \mid G_{k_j}].\end{aligned}$$

Theorem 4.6 then establishes that the limit is $\mathbb{E}[f(\Gamma(U))]$, as required.

It is straightforward to extend this argument to general rectangles U by choosing $r \in \mathbb{R}_+$ such that $U \subset [0,r]^2$. This establishes the result. □

3. Proof of Consistent Estimation for Known Sizes. This section contains a proof of the main consistent estimation result for the known sizes setting.

Theorem 4.12 establishes convergence in distribution of the adjacency measures generated by the dilated empirical graphon. Consistent estimation requires us to establish converegence in distribution of unlabeled graphs generated by the dilated empirical graphon. Note that the map taking an adjacency measure to its associated graph is measurable, but not continuous, and so this result does not follow from a naive application of the continuous mapping theorem. The remaining technical development is mainly concerned with overcoming this difficulty.

LEMMA 3.1. *Let S be a discrete space, T a metric space, $Q_1, Q_2, \dots$ a tight sequence of probability measures on S, and K a probability kernel from S to T, such that K is injective when considered as a map from probability measures on S to probability measures on T. If $Q_1 K, Q_2 K, \dots$ converge weakly to QK then $Q_1, Q_2, \dots$ converges weakly to Q.*

PROOF. Assume otherwise. Case 1: $Q_n \to Q' \neq Q$ weakly. By [22, Lem. 16.24] and the discreteness of S, $Q_n K \to Q'K$ weakly. Since K is injective $Q'K \neq QK$, a contradiction.

Case 2: Q_n does not converge weakly. Since the sequence Q_n is tight it does converge subsequentially. Choose two infinite subsequences $Q_{i_1}, Q_{i_2}, \dots$ and $Q_{j_1}, Q_{j_2}, \dots$ with respective limits Q', Q'' with $Q' \neq Q''$. But then, by [22, Lem. 16.24] and the discreteness of S, $Q'_{i_k} K \to Q'K$ and $Q''_{j_k} K \to Q''K$, hence $Q'K = QK = Q''K$, but K is injective, hence $Q' = Q = Q''$, a contradiction. □

The motivating application of this last lemma is showing that a sequence of graphs $G_1, G_2, \dots$ converge in distribution if and only if their random labelings into $[0, s]$ for some s also converge in distribution. To parse the following theorem, note that when G is a finite random graph, and $s \in \mathbb{R}_+$, then $\mathrm{P}(G \in \cdot)\mathrm{P}(\mathsf{Lbl}_s(G) \in \cdot \mid G) = \mathrm{P}(\mathsf{Lbl}_s(G) \in \cdot)$.

LEMMA 3.2. *For $s \in \mathbb{R}_+$, Let K_s be the kernel defined by $K_s(G) = \mathrm{P}(\mathsf{Lbl}_s(G) \in \cdot)$ for all finite graphs G, let $Q, Q_1, Q_2, \dots$ be probability measures on the space of almost surely finite random graphs, let $\zeta_k = Q_k K_s$ and let $\zeta = QK_s$. Then, $Q_k \to Q$ weakly as $k \to \infty$ if and only if $\zeta_k \to \zeta$ weakly as $k \to \infty$.*

PROOF. The forward direction (convergence in distribution of the random graphs implies convergence in distribution of the random adjacency measures) follows immediately from the discreteness of the space of finite graphs and [22, Lem. 16.24].

Conversely, suppose that $\zeta_k \to \zeta$ weakly as $k \to \infty$, and, for every $n \in \mathbb{N}$, let E_n be the set of adjacency measures ξ such that $\xi([0,s]^2) \leq n$, i.e., E_n is the event that the graph has fewer than n edges. Note that E_n is a ζ-continuity set by the definition of K_s, and therefore, by weak convergence, $\zeta_k(E_n) \to \zeta(E_n)$ as $k \to \infty$ for every $n \in \mathbb{N}$. Let E'_n be the set of graphs with fewer than n edges. By definition, $Q_k(E'_n) = \zeta_k(E_n)$ and $Q(E'_n) = \zeta(E_n)$, hence $Q_k(E'_n) \to Q(E'_n)$. But E'_n is a finite (hence, compact) set, hence $\{Q_k\}_{k\in\mathbb{N}}$ is tight. Noting in addition that K_s is injective, the result follows from Lemma 3.1. □

We now have the necessary tools to prove the main estimation result:

THEOREM 4.13. *Let Γ be a Kallenberg exchangeable graph generated by non-trivial graphex $\mathcal{W}$ and let $s_1, s_2, \ldots$ be a (possibly random) sequence in $\mathbb{R}_+$ such that $s_k \uparrow \infty$ almost surely as $k \to \infty$. Let $G_k = \mathcal{G}(\Gamma_{s_k})$ for $k \in \mathbb{N}$. Suppose that either*

1. *(s_k) is independent of Γ_k, or*
2. *$s_k = \tau_k$ for all $k \in \mathbb{N}$, where $\tau_1, \tau_2, \ldots$ are the jump sizes of Γ.*

Then, for every infinite sequence $N \subseteq \mathbb{N}$, there exists an infinite subsequence $N' \subseteq N$, such that

$$\hat{W}_{(G_k, s_k)} \to_{\mathrm{GP}} \mathcal{W} \text{ a.s.}$$

along N'.

PROOF. We first treat the case (1) where the times (s_k) are independent of Γ.

Let $N \subseteq \mathbb{N}$ be an infinite sequence. Theorem 4.12 implies that there is some infinite subsequence $N' \subseteq N$ such that, for all $r \in \mathbb{R}_+$, $\mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r) \to \mathrm{KEG}(\mathcal{W}, r)$ weakly almost surely along N'.

Let $r \in \mathbb{R}_+$ and $K_r(\cdot)$ as in Lemma 3.2. For all $k \in N'$,

$$\mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r)K_r = \mathrm{KEG}(\hat{W}_{(G_k,s_k)}, r) \text{ a.s.},$$

and $\mathrm{uKEG}(\mathcal{W}, r)K_r = \mathrm{KEG}(\mathcal{W}, r)$. Moreover, the graph corresponding to a size-r Kallenberg exchangeable graph is almost surely finite. Thus Lemma 3.2 applies and we have that $\mathrm{uKEG}(\hat{W}_{(G_k,s_k)}, r) \to \mathrm{uKEG}(\mathcal{W}, r)$ weakly a.s. along N'. This holds for all $r \in \mathbb{R}_+$, so we have even that $\hat{W}_{(G_k,s_k)} \to_{\mathrm{GP}} \mathcal{W}$ a.s. along N'.

The same proof mutatis mutandis applies for convergence along the jump sizes. □

Bibliography.

[1] DALEY, D. J. and VERE-JONES, D. (2003). *An introduction to the theory of point processes: volume I: elementary theory and methods*, Second ed. Springer Science & Business Media.

[2] DURRETT, R. (2010). *Probability: Theory and Examples. 4th Edition*. Cambridge U Press.

[22] KALLENBERG, O. (2001). *Foundations of Modern Probability*, 2nd ed. Springer.

V. VEITCH
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVE
NEW YORK, NY 10027, USA
E-MAIL: vcv2109@columbia.edu

D. M. ROY
DEPARTMENT OF STATISTICAL SCIENCES
SIDNEY SMITH HALL
100 ST GEORGE STREET
TORONTO, ONTARIO, M5S 3G3, CANADA
E-MAIL: droy@utstat.toronto.edu