Survey on sparse graph limits + A toy example

Mei Yin

Department of Mathematics, University of Denver

September 30, 2022

Dense graph limits

Graph limits for dense graphs and the associated classical graphon representation emerged as a subfield of graph theory about 15 years ago. The theory immediately spurred wide interest due to its wide interdisciplinary connections. See the book by Lovász (2012) and the book by van der Hofstad (2017, 2023) and the many references therein.

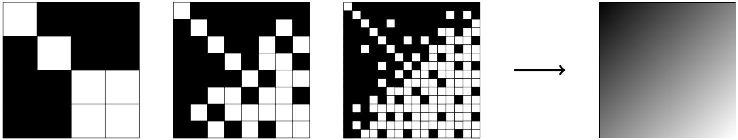
Graphon space $\mathcal{W}$ is the space of all symmetric measurable functions $h(x, y)$ from $[0, 1]^2$ into $[0, 1]$. The interval $[0, 1]$ represents a 'continuum' of vertices, and $h(x, y)$ denotes the probability of putting an edge between x and y.

For classical graphons, there are several quite different ways to define topology and convergence: left convergence (convergence of local properties), right convergence (convergence of global properties), metric convergence, convergence of quotients, etc.

The space of classical graphons (modulo equivalence) is compact, and the different notions of dense graph convergence are all equivalent under some technical conditions. See Borgs, Chayes, Lovász, Sós, and Vesztergombi (2008, 2012) and Chatterjee and Diaconis (2013).

*Many people made contributions from the statistical physics perspective. See the papers by den Hollander, Radin, ...

The following figure gives a visual realization of the convergence of a sequence of dense graphs and the limiting graphon. We can think of these pixel pictures as representations of graphs where each pixel indicates the presence, black, or absence, white, of an edge. As the number of vertices increases without bound, the limiting pixel picture is a representation of the measurable function.



Sparse graph limits

The original graph limit theory however did not apply to many real-world networks, since the theory dealt with dense graphs while networks in the real world tend to be sparser.

We survey the limiting constructions introduced in Veitch and Roy (2015, 2019), Herlau, Schmidt, and Mørup (2016), Borgs, Chayes, Cohn, and Holden (2017), Janson (2017), and Borgs, Chayes, Dhara, and Sen (2019).

Alternative limiting constructions were introduced in Bollobás and Riordan (2009) and Borgs, Chayes, Cohn, and Zhao (2019).

*Consistent estimation using graphons were carried out in Wolfe and Olhede (2013), Mukherjee (2020), and Bhattacharya and Ramanan (2021), and the work of many others from the statistics community. Quite a number of the experts are in attendance at this workshop!

From array to measure:

Sequences of dense graphs sampled from a possibly random graphon are characterized by a natural notion of exchangeability via the Aldous-Hoover exchangeable arrays (1979, 1981).

Analogously, the sparse graph limit theory may be built on exchangeable random measures. See Kallenberg's representation theorem (2002, 2005) and a first step in this direction in Caron and Fox (2014).

*Not all random graphs may be addressed by this "exchangeable" theory, for example the Barabási-Albert preferential attachment model or the model discussed by Lutz Warnke on Wednesday.

Reformulating the Kallenberg representation theorem for graphs:

A random graph is characterized by three (potentially random) components: a non-negative real I in $\mathbb{R}_+$, an integrable function $S : \mathbb{R}_+ \to \mathbb{R}_+$, and a symmetric measurable function $W : \mathbb{R}_+^2 \to [0,1]$ that satisfies several weak integrability conditions. The triple (I, S, W) is called a graphex (or simply, a (generalized) graphon). W is the main component of the graphex, and often $I = S = 0$.

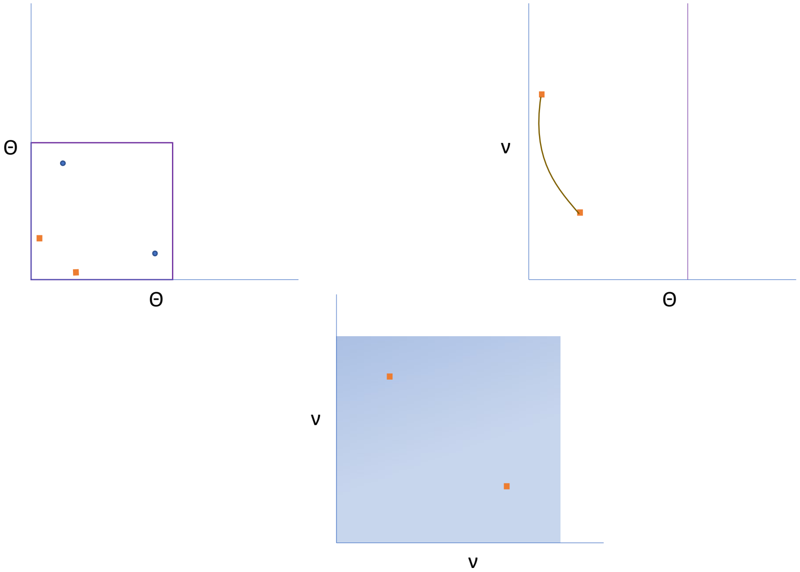
This construction can be extended further to σ-finite measure space.

Generating a random graph from the graphex W:

Take realizations of independent unit-rate Poisson processes $\Xi = \{(\theta_i, \upsilon_i)\}_i$ on $\mathbb{R}_+^2$. θ_i's are potential vertex labels. Independently an edge is placed in between (θ_i, θ_j) with probability $W(\upsilon_i, \upsilon_j)$. Typically we only keep non-isolated vertices born before a certain time (whose vertex labels are below some threshold).

S and I, should they be non-zero, would respectively contribute the stars and isolated edges of the random graph and are of minor interest relatively.

The generated graphs are projective and grow over time.



Kallenberg exchangeable graph (simple illustration).

Main takeaway:

The generated random graph is dense if and only if the corresponding integrable graphex has compact support (i.e. the graphex is equal to the dilation of some classical graphon). The classical graphon model is a special case of the graphex model.

The distinction that allows for more general graphs in the Kallenberg exchangeable graphs setting is that the latent variables υ_i associated with each vertex θ_i are not independent, and the sizes of the graphs are random.

The space of probability distributions on dense graphs can be parameterized by the space of graphons. The space of probability distributions on sparse graphs can be parameterized by the space of graphexes. Graphons and graphexes are respectively the ergodic measures in the family of distributions.

*Please recall Peter Orbanz's very informative talk on Tuesday for background on ergodic measures!

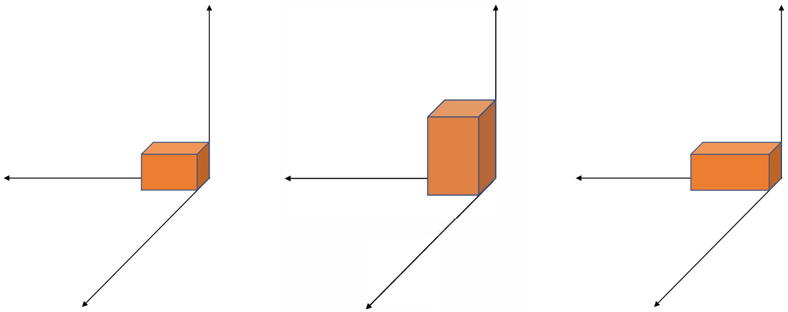
Modes of metric convergence for exchangeable random graphs:

- Cut metric (dense graphs): $W(x, y)$.

$$\delta_{\Box}(W_1, W_2) = \inf_{\sigma_1,\sigma_2} \sup_{S,T\subseteq\mathbb{R}_+} \left| \int_{S\times T} (W_1^{\sigma_1}(x,y) - W_2^{\sigma_2}(x,y))\, dxdy \right|.$$

- Rescaled cut metric (sparse graphs): $\|W\|_1^{-1} W(x, y)$.
- Stretched cut metric (sparse graphs): $W\left(\|W\|_1^{1/2}x, \|W\|_1^{1/2}y\right)$.
- And more...

We are particularly interested in the stretched cut metric as it strips away the size information of the observed graph (alternatively, graph size is unobserved).



Classical, rescaled, and stretched graphons.

A toy example

We now investigate a power-law random graph model and cast it in the sparse graph limit theory framework. As discussed in Yin (2022), intriguing phenomena arise even in this naïve-looking model, as we will soon see.

Our model is closely connected to a motivating example in Borgs, Chayes, Cohn, and Zhao (2019). We introduce their model first and our model next.

The original model in Borgs, Chayes, Cohn, and Zhao (2019):

Consider a discrete graph of n vertices numbered 1 through n. Connect vertices i, j with probability

$$p_n(i,j) = \min\left\{1, n^{\beta-2/\alpha}(i/n)^{-1/\alpha}(j/n)^{-1/\alpha}\right\},$$

where $\alpha > 1$ and $\beta \in (0, 2/\alpha)$ are parameters.

Intuitively, the edge connection probability between vertices i, j behaves like $(ij)^{-1/\alpha}$, but boosted by a factor of n^β in case it becomes too small.

The graph is sparse with expected edge density $n^{\beta-2/\alpha}$.

The limiting graphon in the rescaled cut metric is

$$W^r(x,y) = (1-1/\alpha)^2 (xy)^{-1/\alpha},$$

which lies in $L^p([0,1])^2$ for any $p < \alpha$.

The adapted model in Yin (2022):

Consider a discrete graph of n vertices numbered 1 through n. Connect vertices i, j with probability

$$p_n(i,j) = \mathbf{1}_{\{X_i X_j / a_n > 1\}},$$

where $a_n = n^{-\beta+2/\alpha}$ is a parameter, $X_i \overset{d}{=} U_i^{-1/\alpha}$, and U_i are i.i.d. $(0,1)$-uniform random variables.

Graphs may be equivalently formulated as adjacency measures, and there are standard Poisson convergence results at the critical regime ($a_n \sim n^{2/\alpha}$): A typical realization of this adapted model exhibits a small clique and large numbers of follower vertices asymptotically. See Dabrowski, Dehling, Mikosch, and Sharipov (2002). The limit structure of the model away from criticality on the other hand is less understood, and will be the central focus of this talk.

Original model: a power-law random graph with Bernoulli edges

Adapted model: a power-law random graph without Bernoulli edges

Difference between the two models lies in the edge connection probability:

$$\min\left\{1, n^{\beta-2/\alpha}(i/n)^{-1/\alpha}(j/n)^{-1/\alpha}\right\} \to \min\left\{1, \frac{X_i X_j}{a_n}\right\} \to \mathbf{1}_{\left\{\frac{X_i X_j}{a_n}>1\right\}},$$

where $a_n = n^{-\beta+2/\alpha}$.

The parameter range $\alpha > 1$ and $\beta > 0$ in the original model translates to $\alpha > 1$ and $a_n \ll n^{2/\alpha}$ in the adapted model, and will be referred to as the super-critical regime in a moment.

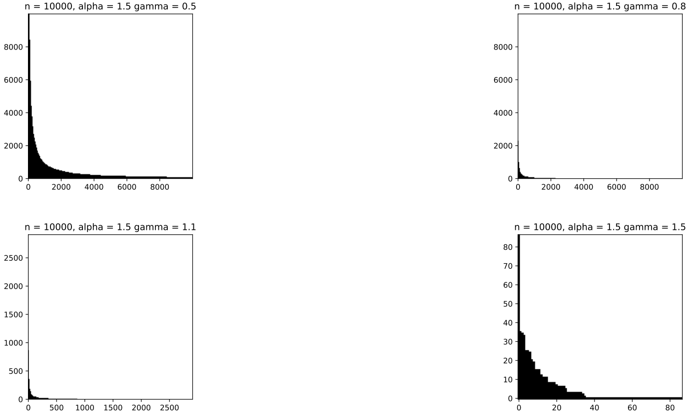
The first step in the adaptation continualizes the discrete normalized vertex labels into a uniform measure, and implicitly relabels the vertices $1, \ldots, n$ using the order statistics of their associated random variables $X_1, \ldots, X_n$, the latter not having a real impact on the structure of the graph.

The second step in the adaptation is more significant. For edges that our adapted model connects, the original model connects them too. Call these "hard edges". However, the original construction is not that strict with those edges that we drop. Instead they choose whether to connect them or not depending on a Bernoulli sampling probability between $[0, 1]$. Call these "Bernoulli edges".

We introduce an auxiliary parameter $\gamma > 0$ and set $a_n^\alpha = n^\gamma \log n$. The sub-critical regime $a_n \gg n^{2/\alpha}$ translates to $\gamma \geq 2$ and the super-critical regime $a_n \ll n^{2/\alpha}$ translates to $\gamma < 2$. (The $\log n$ factor is not essential and more for technical convenience.)

Recall that the critical regime corresponds to $a_n \sim n^{2/\alpha}$.

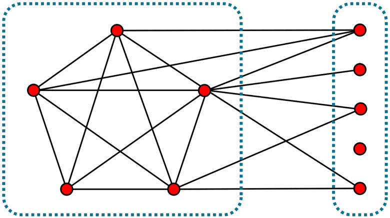
In the sub-critical regime, we will show that although there is no graph in the limit, in the rare event that we do see a non-empty graph, typically it contains exactly one edge. Contrarily, in the super-critical regime, we will show that a limit random graph exists in the stretched cut metric, and universality emerges in the limiting graphon.



Some simulations of the empirical graphons of our adapted model.
Vertices are labeled according to decreasing vertex values.

First estimates:

Given a realization $X_1, \ldots, X_n$ of vertex values and a chosen normalization a_n, we group the non-isolated vertices of the graph into two parts depending on whether $X_i > \sqrt{a_n}$ or $X_i \leq \sqrt{a_n}$, respectively referred to as "clique" and "followers". Since two vertices only get connected when the product of their vertex values exceeds a_n, a split graph is produced, as vertices are all connected within the clique and form a complete subgraph, while follower vertices can only be connected to clique vertices but not to themselves.



A split graph: clique vs. followers.

Straightforward calculations:

We compute the expected number of edges of the random graph:

$$\mathbb{E}|E_n| \sim \frac{\gamma}{2} n^{2-\gamma},$$

and the expected number of non-isolated vertices of the random graph:

$$\mathbb{E}|V_n| \sim \begin{cases} 0 & \gamma > 2, \\ (\gamma - 1) n^{2-\gamma} & \gamma \in (1, 2], \\ n^{\frac{\log \log n}{\log n}} & \gamma = 1, \\ n & \gamma \in (0, 1). \end{cases}$$

The sub-critical regime ($\gamma \geq 2$):

Let $K_{n,0}$ denote the number of vertices with large weight (vertex value $> \sqrt{a_n}$). It is Binomial distributed with parameter $\left(n, a_n^{-\alpha/2}\right)$. These vertices are clique vertices if they are in addition non-isolated.

The conditional law of $\mathcal{L}(K_{n,0} \mid K_{n,0} \geq 1)$ is:

$$\mathbb{P}\left(K_{n,0} \geq 1\right) \sim \mathbb{P}\left(K_{n,0} = 1\right) \sim \frac{n^{1-\gamma/2}}{\log^{1/2} n},$$

Therefore, given the appearance of a non-trivial random graph, the clique part (conditioning on non-empty) typically only contains one vertex.

Fine point:

$\mathbb{P}(K_{n,0} = 1)$ does not necessarily imply the appearance of a star graph as the edge number may still be zero. Let $K_{n,1}$ denote the number of follower vertices. Then

$$\begin{aligned}\mathbb{P}(\text{one clique vertex}) &\sim \mathbb{P}(\text{one clique vertex}, K_{n,1} = 1)\\ &\sim \left(\frac{\gamma}{2} - 1\right) n^{2-\gamma}.\end{aligned}$$

Conclusion: Given that the graph is non-empty, in the limit predominantly it has exactly two vertices, one clique vertex and one follower vertex.

Physical interpretation:

For a typical behavior, with probability going to one we would not see any graph eventually. In the rare event that we do see one, we would need certain "extra energy" (than typical) to push some of the X_i values up, and the most "economical" way to do so is to push one up to the clique and another up as a follower. Pushing up two to the clique or pushing up more than one follower or any other construction, by comparison, might be too costly.

*To infuse physical intuition into math models, Frank den Hollander and Charles Radin are good sources!

The super-critical regime ($\gamma < 2$):

Let $K_{n,0}$ denote the number of vertices with large weight (vertex value $> \sqrt{a_n}$).

Height function: For $0 \leq x \leq 1$, let $H_n(x)$ denote the number of not-in-clique vertices that are connected to the top $\lceil xK_{n,0} \rceil$ clique vertices, where clique vertices are ordered according to increasing vertex values. (More technical details later...)

- $H_n(1)$ is the number of followers of the leader from the clique.
- $H_n(0) \equiv 0$ by convention.

Main Theorem:

$$\frac{1}{\sqrt{\mathbb{E}K_{n,0}}}\left\{H_n(x)-\mathbb{E}K_{n,0}\cdot\frac{x}{1-x}\right\}_{x\in[0,1)} \overset{f.d.d.}{\Rightarrow} \left\{\mathbb{B}_{x/(1-x)}+\mathbb{G}_x\right\}_{x\in[0,1)},$$

where $f.d.d.$ indicates convergence of finite-dimensional distributions, $\{\mathbb{B}_t\}_{t\in[0,\infty)}$ is a standard Brownian motion, $\{\mathbb{G}_x\}_{x\in[0,1)}$ is a generalized Brownian bridge with covariance function

$$\mathrm{Cov}\left(\mathbb{G}_x,\mathbb{G}_y\right)=\frac{\min(x,y)(1-\max(x,y))}{(1-x)^2(1-y)^2},\ \ x,y\in[0,1),$$

and $\mathbb{B}$ and $\mathbb{G}$ are independent.

*Svante Janson strengthened $f.d.d.$ to process convergence in the Skorokhod topology in the space $D[0,1)$ (personal communications).

Implications of Main Theorem:

$$\mathbb{E}H_n(x) \sim \mathbb{E}K_{n,0}\frac{x}{1-x},$$
$$\mathrm{Var}H_n(x) \sim \mathbb{E}K_{n,0}\frac{x}{1-x}\left(1+\frac{1}{(1-x)^2}\right).$$

Switching from increasing to decreasing order statistics introduces a simple transformation $x \mapsto 1-x$, so the boundary line

$$h(x) = 1 + \lim_{n\to\infty}\frac{\mathbb{E}H_n(x)}{\mathbb{E}K_{n,0}} = \frac{1}{x}.$$

Having the same asymptotic order for the expected value and the variance of the height function $H_n(x)$ also explains why the simulations look so regular.

Further implications of Main Theorem:

Let W_n denote the graphon of our model with n vertices without scaling, i.e. a $\{0,1\}$-valued function on $[0,n]^2$. We have

$$W_n(\mathbb{E}K_{n,0} \cdot x,\ \mathbb{E}K_{n,0} \cdot y) \to W^s(x,y) = \mathbf{1}_{\{xy \le 1\}}, \qquad x, y \in (0,\infty).$$

This is a universal result independent of the parameters.

Recall that $\|W_n\|_1 = \mathbb{E}|E_n| \sim (\gamma/2)n^{2-\gamma}$ while $\mathbb{E}K_{n,0} \sim n^{1-\gamma/2}/(\log n)^{1/2}$. In the super-critical regime where there is a non-trivial random graph in the limit, our adapted model may be viewed as an example that lies at the boundary of stretched convergence.

Idea of proof:

Recall that $X_i \stackrel{d}{=} U_i^{-1/\alpha}$ and U_i are i.i.d. $(0,1)$-uniform random variables. Introduce two i.i.d. sequences of random variables: $\{Y_{n,i}\}_{i\in\mathbb{N}}$ are i.i.d. with law as $\mathcal{L}(X_1 \mid X_1 > \sqrt{a_n})$ and $\{Z_{n,i}\}_{i\in\mathbb{N}}$ are i.i.d. with law as $\mathcal{L}(X_1 \mid X_1 \le \sqrt{a_n})$ (with scaling adjustment).

We order $\{Y_{n,i}\}_{i=1,\ldots,K_n}$ in *increasing* order statistics

$$Y_{n,K_n:K_n} > \cdots > Y_{n,1:K_n} > \sqrt{a_n} > \frac{a_n}{Y_{n,1:K_n}} > \cdots > \frac{a_n}{Y_{n,K_n:K_n}},$$

where listed on the right hand side of $\sqrt{a_n}$ are the thresholds for different groups of followers.

For $x \in (0,1)$, define

$$\tau_n(x) = \frac{a_n}{Y_{n,\lceil xK_n \rceil:K_n}}, \qquad B_{n,i}(x) = \mathbf{1}_{\{Z_{n,i} > \tau_n(x)\}}.$$

The height function may be written as

$$H_n(x) = \sum_{i=1}^{n-K_n} B_{n,i}(x).$$

We then study the order statistics of $Y_{n,i}^{-1}$ and nested Bernoulli random variables $B_{n,i}(x)$, utilizing convergence of quantile processes in Shorack (1972, 1973).

More details:

Introduce $\mathcal{K}_n := \sigma(K_n, Y_{n,1}, \ldots, Y_{n,K_n})$. Then

$$\sum_{i=1}^{n-K_n} (B_{n,i}(x) - \mathbb{E}(B_{n,i}(x)|\mathcal{K}_n))$$

contributes to the $\mathbb{B}$ part (Brownian motion), and the difference between the above expression and $H_n(x)$ contributes to the $\mathbb{G}$ part (Brownian bridge).

*Recall Bhaswar Bhattacharya's talk on Monday and Siva Athreya's talk on Thursday for more background on these methods. They are experts!

Some remarks: Let $W(x,y) = \mathbf{1}_{\{xy \le 1\}}$.

- $W(x,y)$ is not L^1-integrable and so does not exactly fit in the stretched convergence framework as discussed in Borgs, Chayes, Cohn, and Holden (2017).
- $W(x,y)$ is however locally finite and satisfies weak integrability conditions as discussed in Veitch and Roy (2015, 2019), with graphex marginal $\mu_W(x) = 1/x$.
- After scaling with $\mathbb{E}K_{n,0}$, the explicit asymptotics for this toy model agree well with the generic formulas in Veitch and Roy (2015, 2019), including the number of edges, non-isolated vertices, and degree distribution.
- Certain technical conditions of the toy model may be relaxed, for example, distributions with regularly varying tail.
- The Brownian phenomenon is intriguing. Is this purely conincidental? Intimately tied to "hard edges"? Some intuitive arguments are suggested by Svante Janson.

Main References:

- Borgs, C., Chayes, J.T., Cohn, H., Holden, N.: Sparse exchangeable graphs and their limits via graphon processes. J. Mach. Learn. Res. 18: 7740-7810 (2017)
- Borgs, C., Chayes, J.T., Cohn, H., Zhao, Y.: An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. Trans. Amer. Math. Soc. 372: 3019-3062 (2019)
- Janson, S.: On convergence for graphexes. arXiv: 1702.06389 (2017)
- Lovász, L.: Large Networks and Graph Limits. American Mathematical Society, Providence. (2012)
- Veitch, V., Roy, D.M.: The class of random graphs arising from exchangeable random measures. arXiv: 1512.03099 (2015)
- Veitch, V., Roy, D.M.: Sampling and estimation for (sparse) exchangeable graphs. Ann. Statist. 47: 3274–3299 (2019)
- Yin, M.: Remarks on power-law random graphs. Stochastic Process. Appl. 153: 183-197 (2022)

Thank You! Questions?