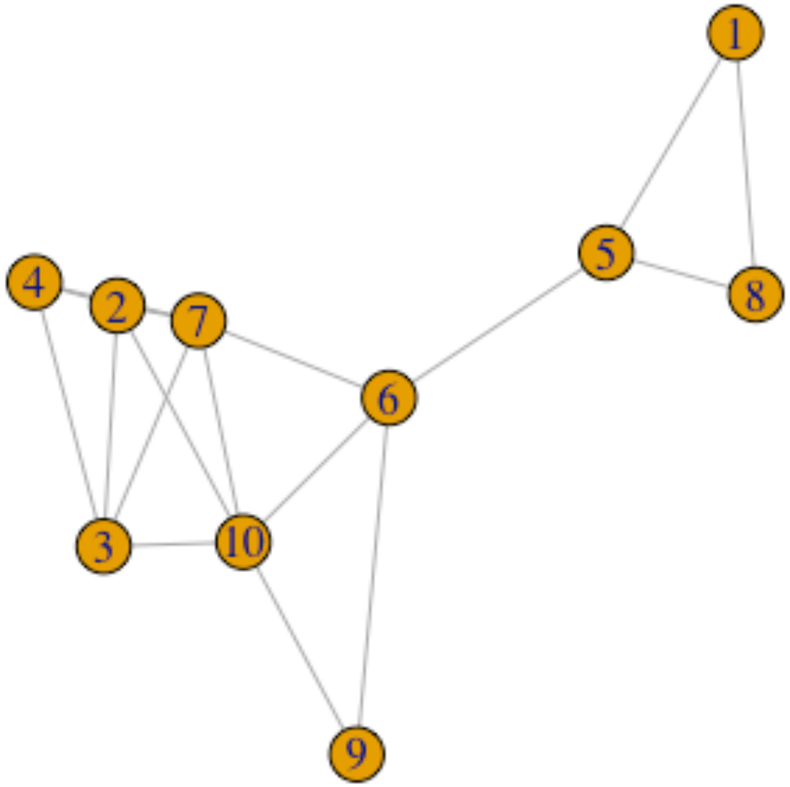
Introduction to graphons

Sevvandi Kandanaarachchi, Cheng Soon Ong
CSIRO

NeurIPS Education Track 2026

A graph is ...

- Undirected
- Unweighted
- No loops
- No node or edge attributes

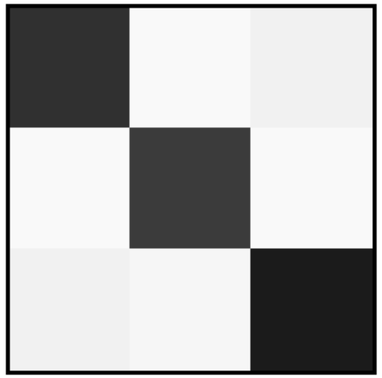


Graphons as graph generators

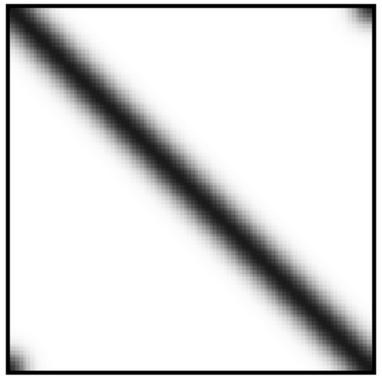


What are graphons?

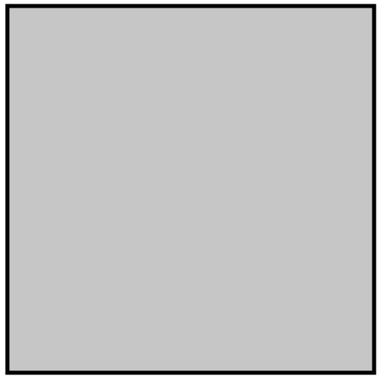
- A symmetric, measurable function $W : [0,1]^2 \to [0,1]$



Stochastic Block Model graphon



Ring graphon

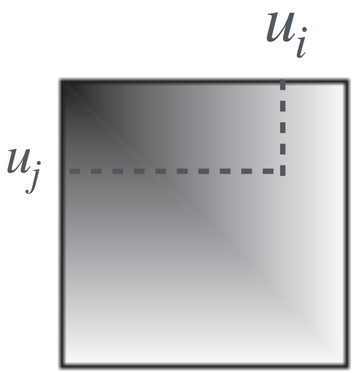


Erdos-Renyi graphon

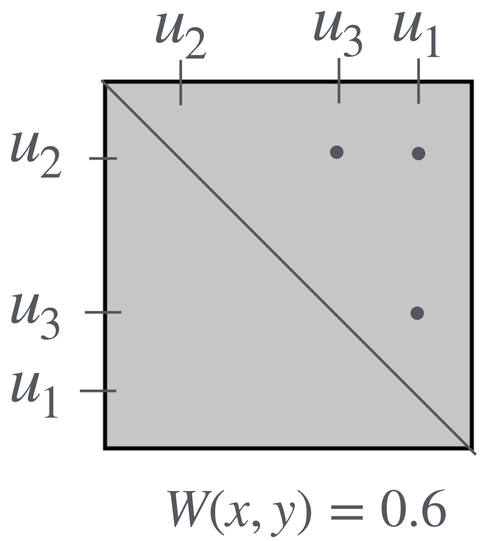
- Compact domain, bounded function, no need to be continuous

Sampling graphs from a graphon

- Say your graph has n nodes labeled $\{1,2,\ldots,n\}$
- Uniformly sample n points from the interval $[0,1]$ - $\{u_1, u_2, \ldots, u_n\}$
- For every (u_i, u_j) the value $W(u_i, u_j)$ is the edge probability
- Edge $e_{ij} \sim$ Bernoulli$(W(u_i, u_j))$
- Toss a coin with probability $W(u_i, u_j)$ and connect the nodes i and j if you get heads
- Each edge is connected independently - we'll get back to this later on!

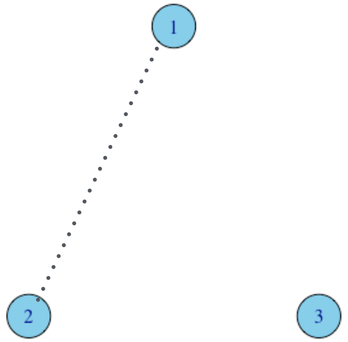


Example



$W(x, y) = 0.6$

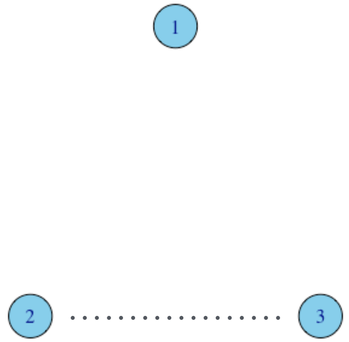
Consider edge e_{21}



Bernoulli$(W(u_2, u_1)) = 0$

=> No edge

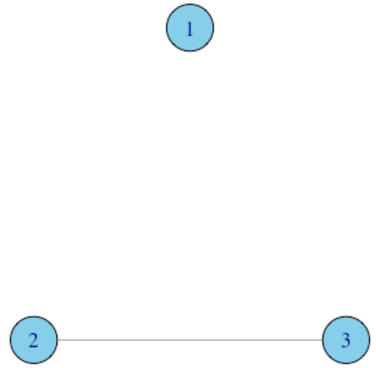
Consider edge e_{23}



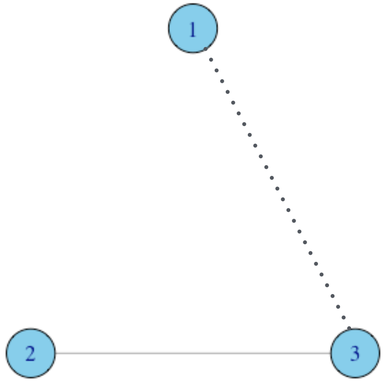
Bernoulli$(W(u_2, u_3)) = 1$

=> Add edge

Resulting graph



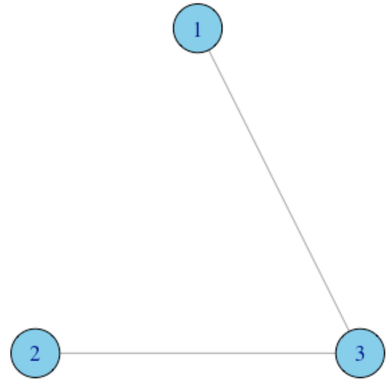
Consider edge e_{31}



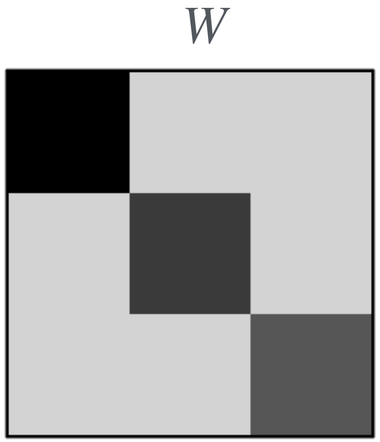
Bernoulli$(W(u_3, u_1)) = 1$

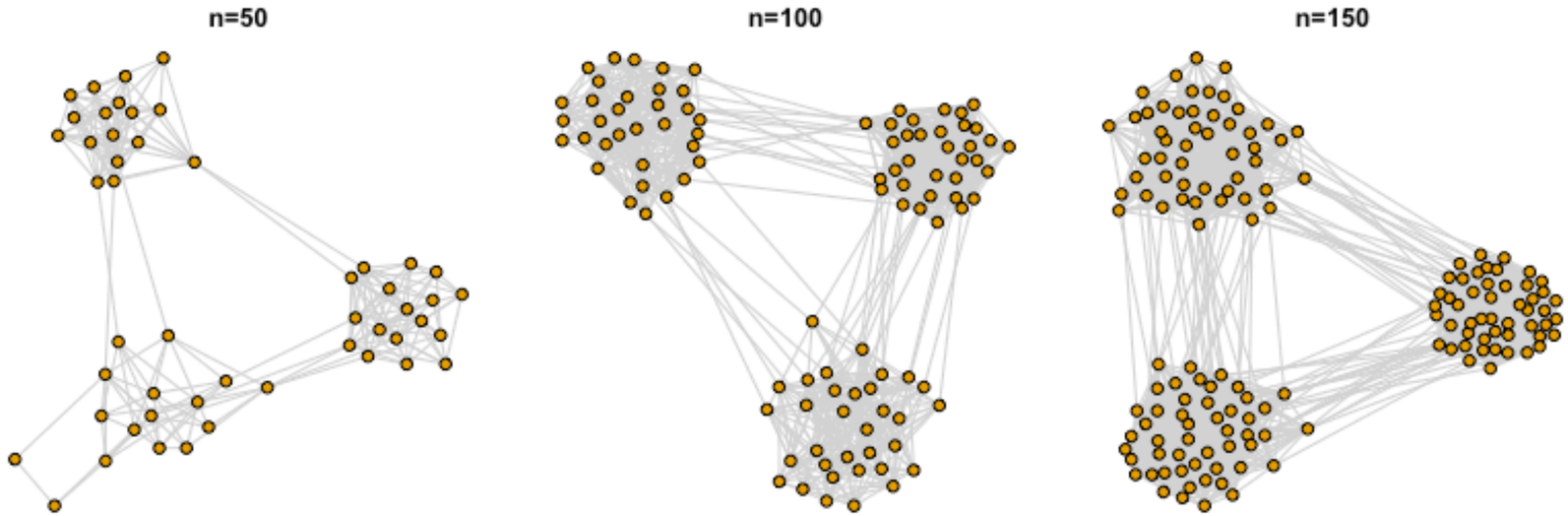
=> Add edge

Resulting graph



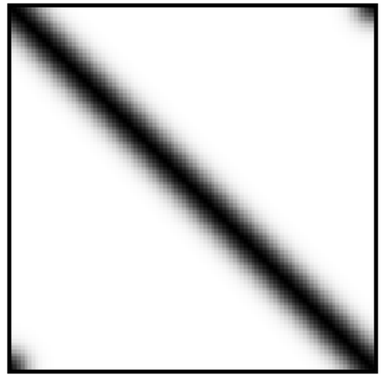
Stochastic block model graphon

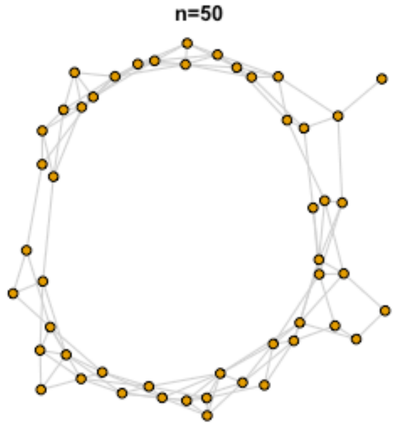




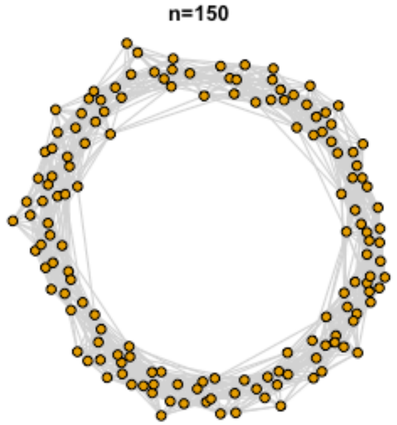
Ring graphon

W

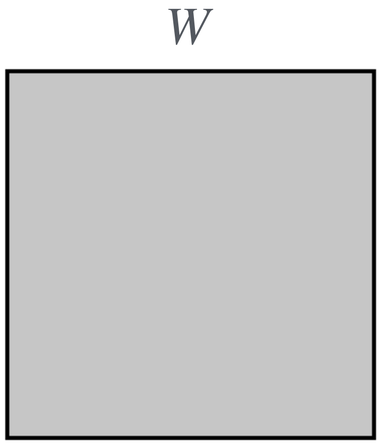


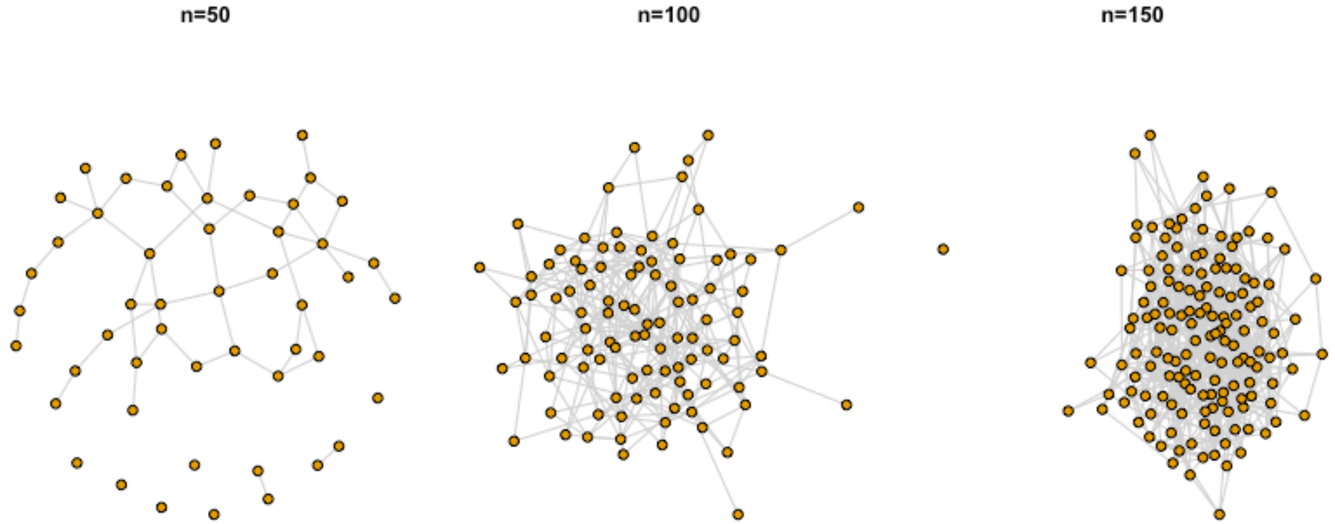






Constant graphon



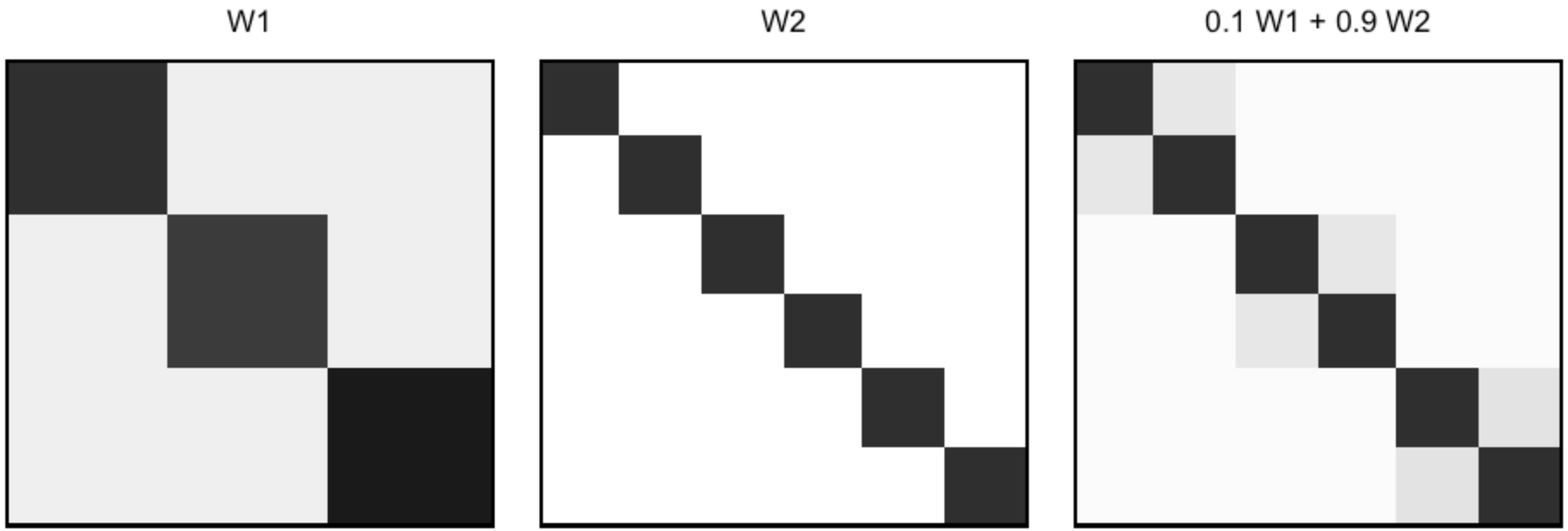


Combining graphons



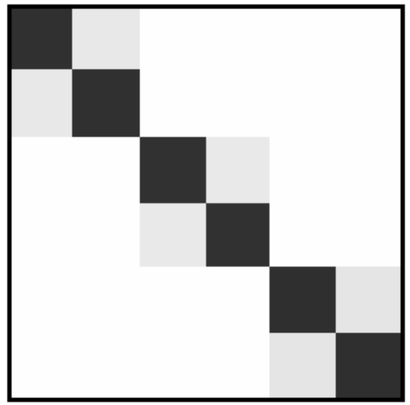
Convex combination of graphons

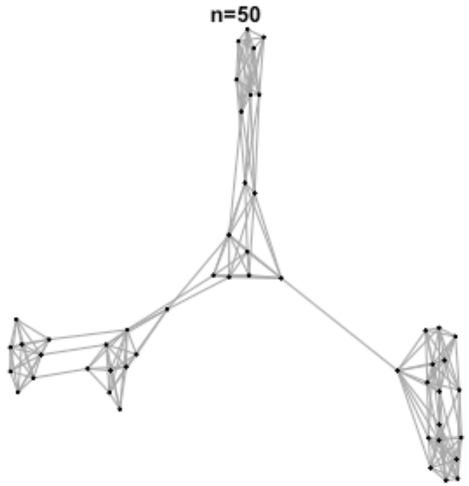
- Let W_1 and W_2 be graphons
- Consider $W = \alpha W_1 + (1 - \alpha) W_2$ for $\alpha \in [0,1]$

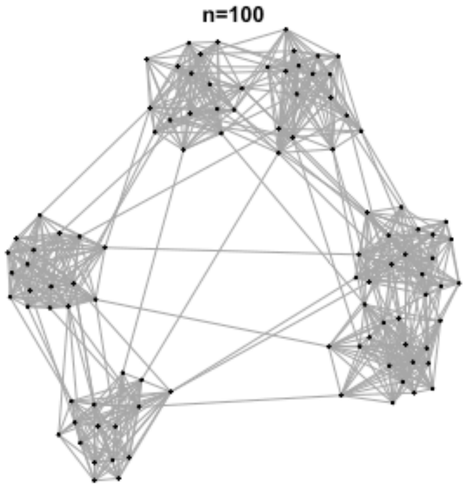


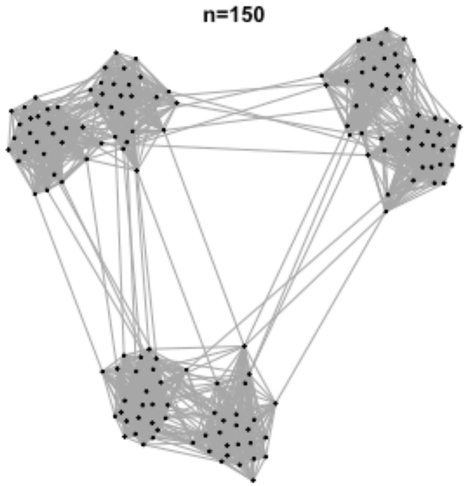
Convex combination

W



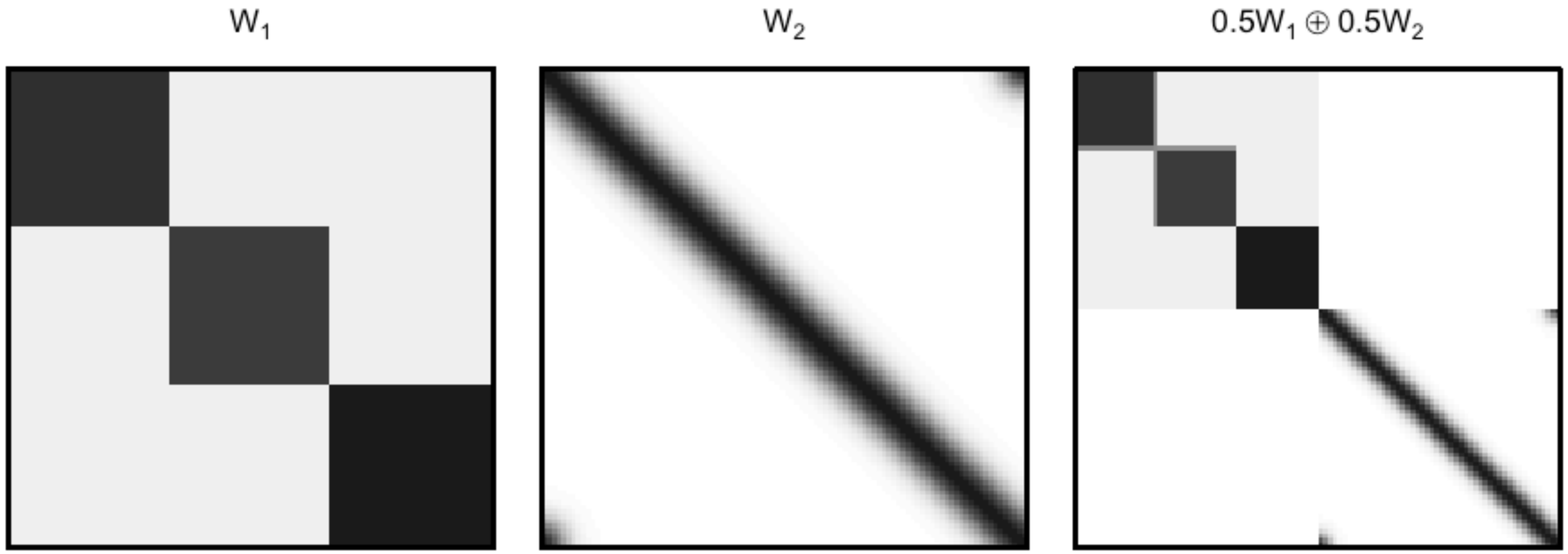






Disjoint union of graphs

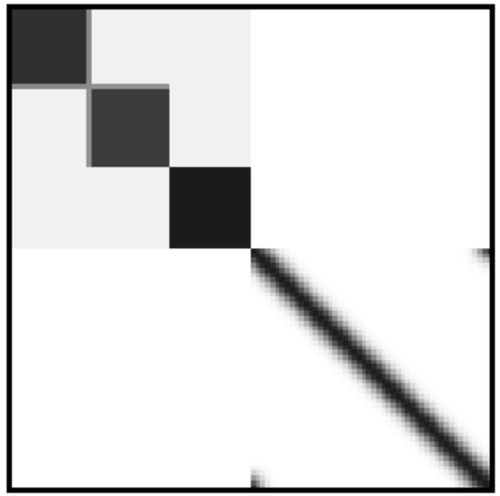
- Let W_1 and W_2 be graphons
- Let $W = \alpha W_1 \oplus (1-\alpha) W_2$ for $\alpha \in [0,1]$ visualized by placing a scaled graphon W_1 on $[0,\alpha]^2$ and a scaled graphon W_2 on $[\alpha,1]^2$



Janson, S. (2008). Connectedness in graph limits. *arXiv preprint arXiv:0802.3795*.

Disjoint union - example 1

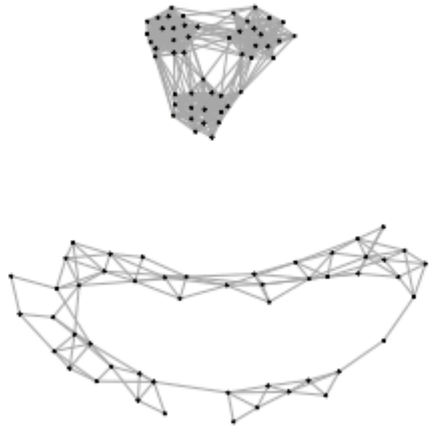
W



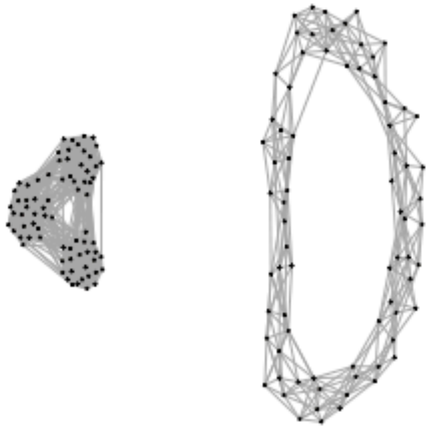
n=70



n=100

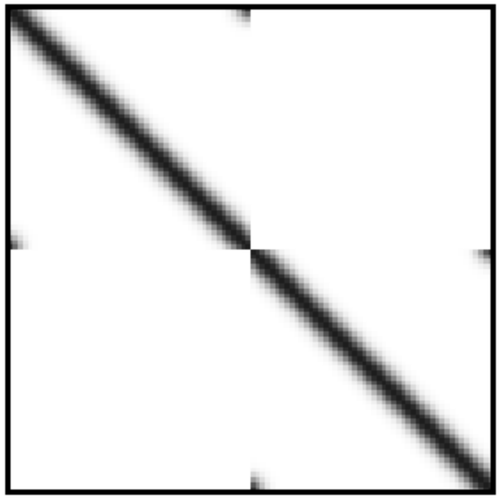


n=150



Disjoint union - example 2

W



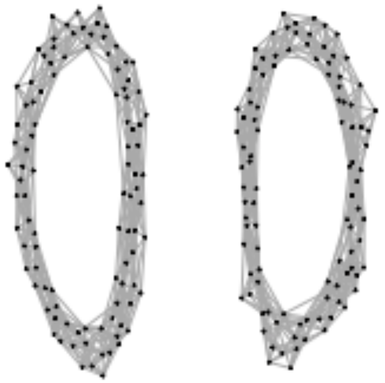
n=100



n=150

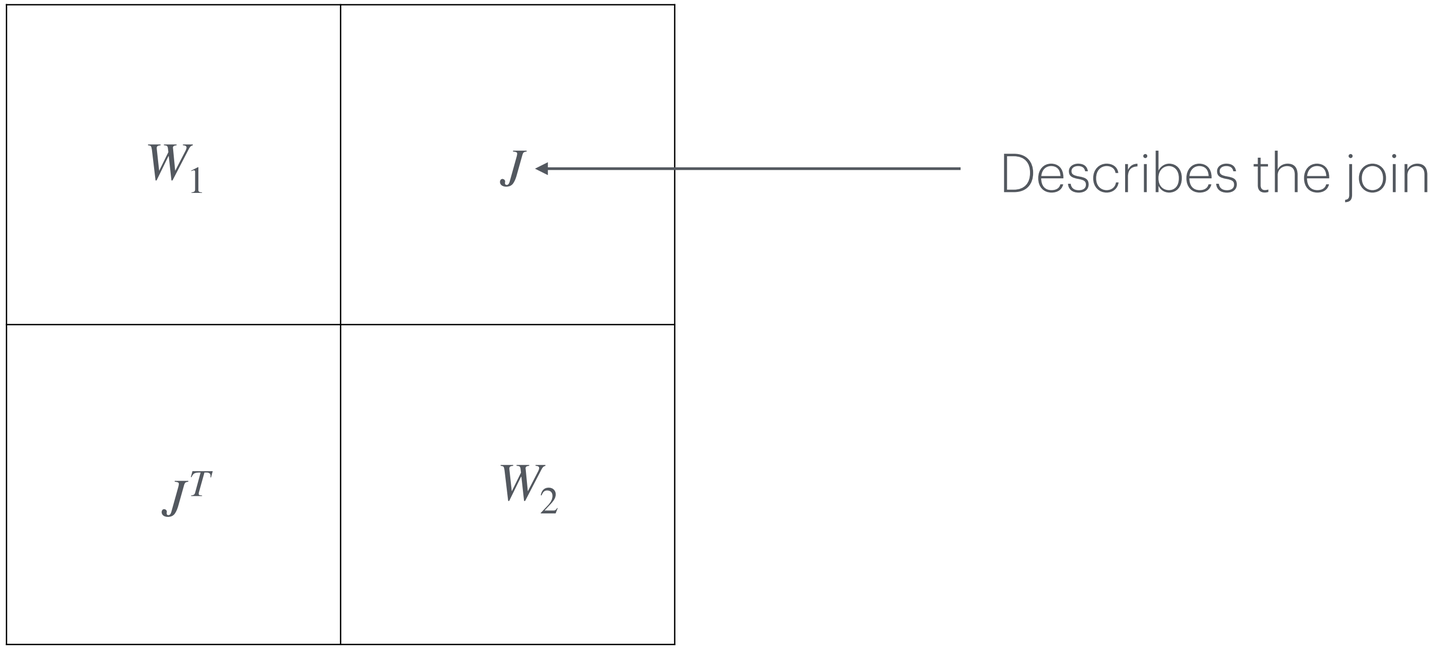


n=200



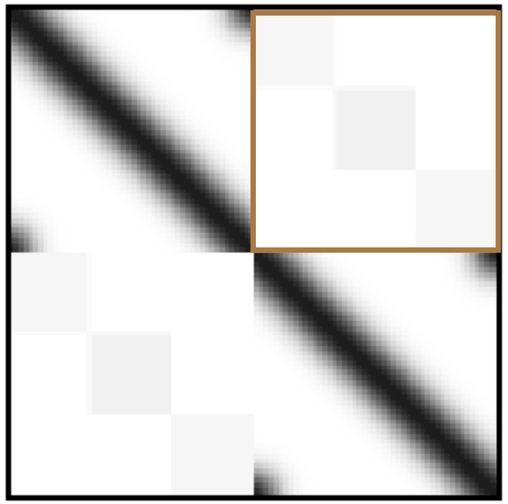
Joining disjoint unions

- Let W_1 and W_2 be graphons, and $J : [0,1]^2 \to [0,1]$, but J does not need to be symmetric
- Scale W_1, W_2 and J as necessary.

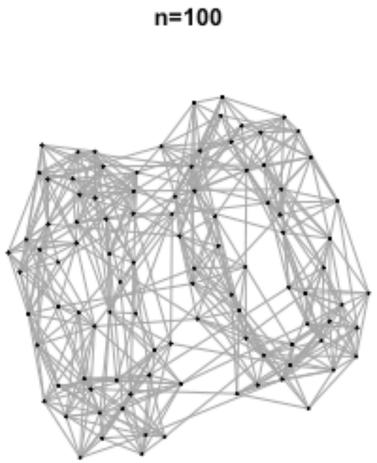


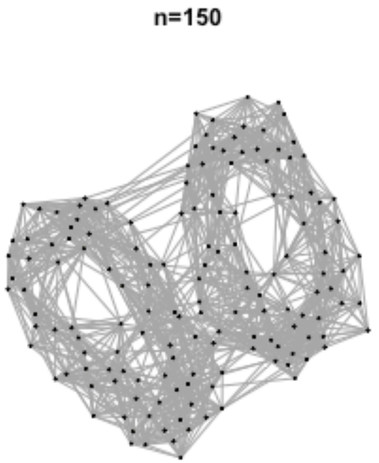
Joined graphs

Describes the join









Mathematical Properties of Graphons

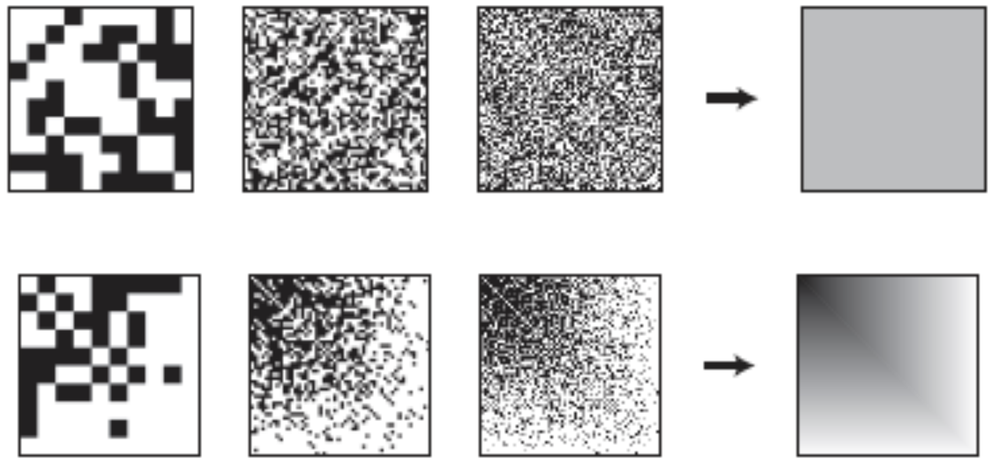
Sevvandi Kandanaarachchi, Cheng Soon Ong
CSIRO

NeurIPS Education Track 2026

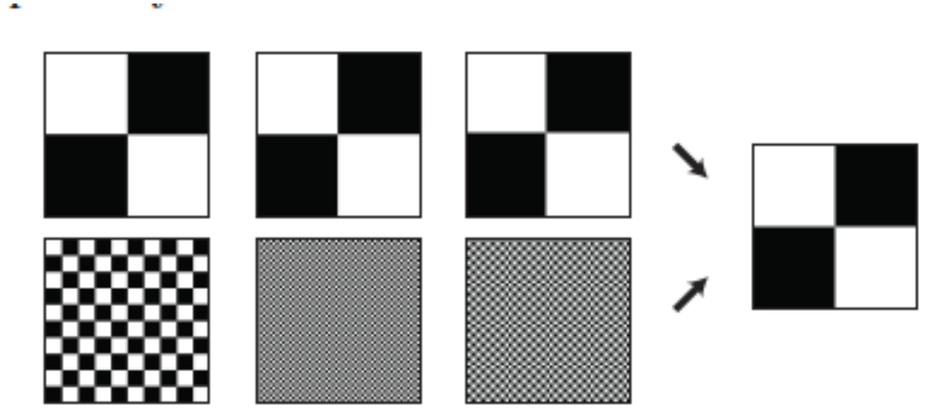
Graphons are graph limits

- Map the adjacency matrix to the unit square (empirical graphons) and take a limit

Erdos-Renyi graphs



Growing uniform attachment graphs

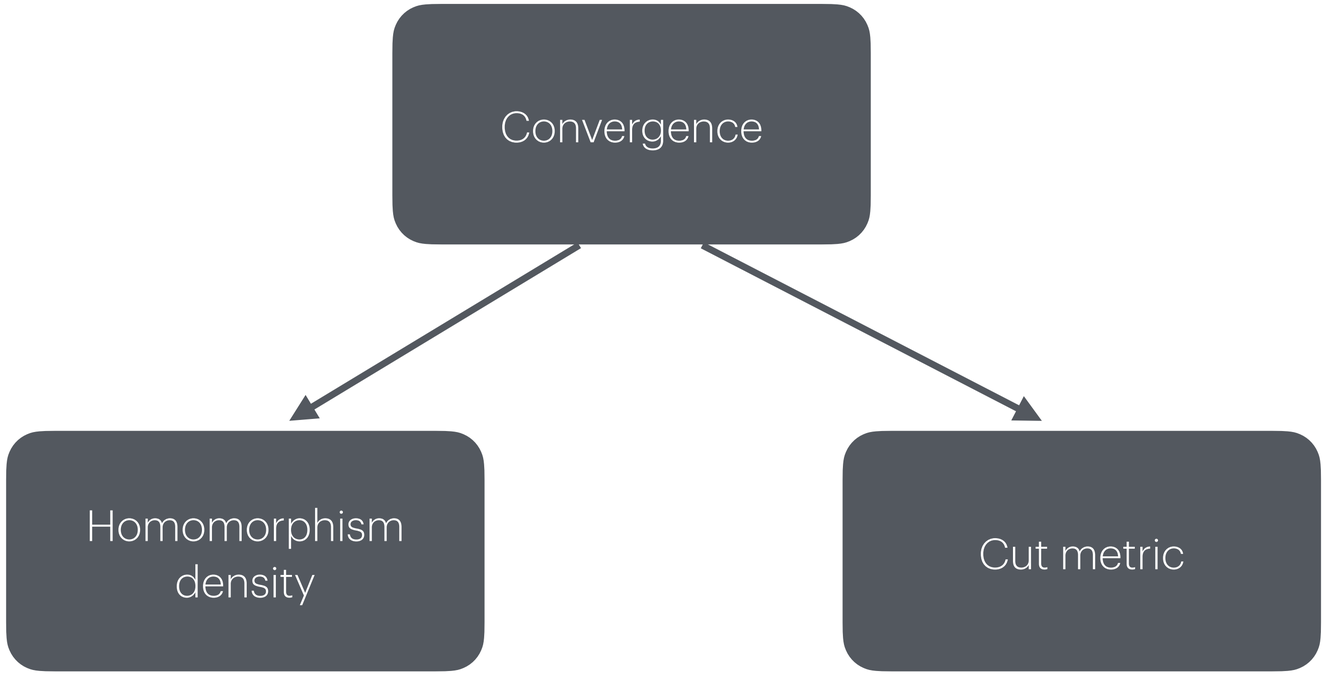


Bi-partite graphs

Glasscock, D. (2015). What is… a graphon. *Notices of the AMS*, *62*(1), 46-48.

How is convergence defined?

- Sequence of graphs $(G_1, G_2, \ldots, G_n, \ldots)$



Borgs, C., Chayes, J., Lovász, L., Sós, V., & Vesztergombi, K. (2011). Limits of randomly grown graph sequences. *European Journal of Combinatorics*, *32*(7), 985-999

Homomorphism density



Graph homomorphism convergence

- Subgraph density convergence

- A graph homomorphism from F to G is a map $f : V(F) \to V(G)$ such that if $uv \in E(F)$ then $f(u)f(v) \in E(G)$ (Maps edges to edges). Let $Hom(F, G)$ be the set of all such homomorphisms and let $\hom(F, G) = |Hom(F, G)|$. Then **homomorphism density** is defined as

- $$t(F, G) = \frac{\hom(F, G)}{|V(G)|^{|V(F)|}}, \quad \text{note that} \quad t(F, G) \in \mathbb{R}$$

- A graph sequence $\{G_n\}_n$ is said to be convergent if $t(F, G_n)$ converges as n goes to infinity for any simple graph F.

Graph homomorphisms

- $t(\bullet\!\!-\!\!\bullet, G) = 2\dfrac{|E(G)|}{n^2}$, where $|E(G)|$ is the number of edges and n is the number of vertices
- This is similar to edge density $2\dfrac{|E(G)|}{n(n-1)}$
- Consider the graph sequence $(G_1, G_2, \ldots)$
- Does the sequence $\left(2\dfrac{|E(G_1)|}{n_1^2}, 2\dfrac{|E(G_2)|}{n_2^2}, \ldots\right)$ converge?
- Does $(t(F, G_1), t(F, G_2), \ldots)$ converge for all simple graphs F?

Graph homomorphisms

- If $(t(F, G_1), t(F, G_2), \ldots)$ converges for all simple graphs F, then the graphs converge to a graphon W
- $\{G_n\}_n$ converges iff there is a W such that $t(F, G_n) \to t(F, W)$ for all simple graphs F.
- $$t(F, W) = \int_{[0,1]^{|V(F)|}} \prod_{ij \in E(F)} W(x_i, x_j)\, dx$$
- And for every W there is a convergent sequence of graphs like above.

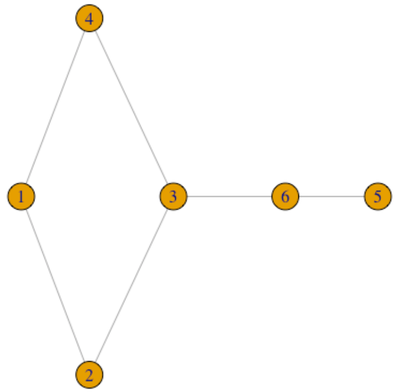
Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., & Vesztergombi, K. (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. Advances in Mathematics, 219(6), 1801–1851. https://doi.org/10.1016/j.aim.2008.07.008

The catch: convergence must hold for every simple graph F

Cut metric



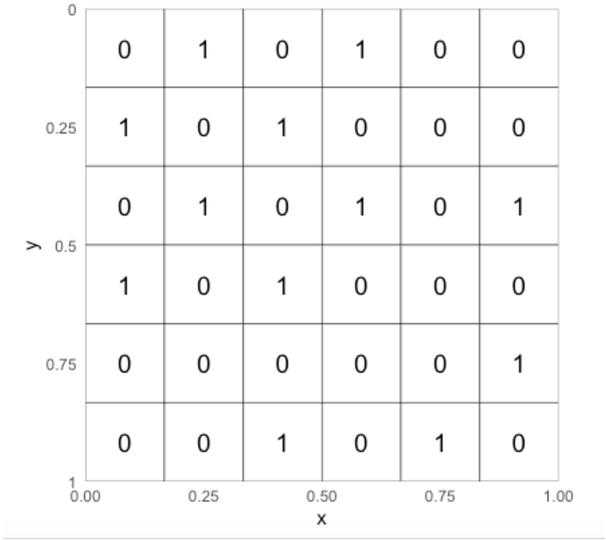
Three related objects



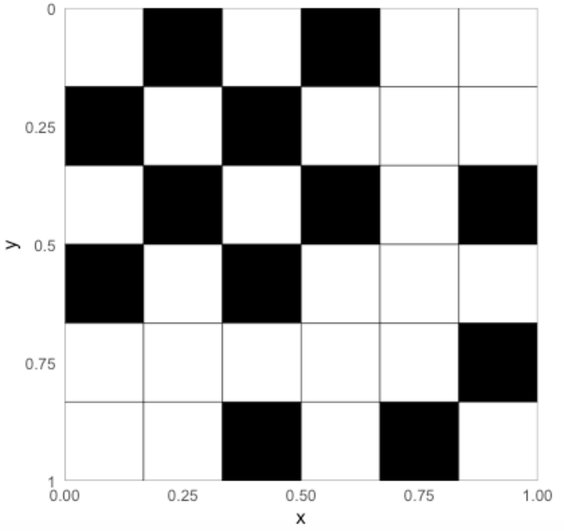
- Adjacency matrix - Empirical graphon - pixel pictures

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]
[1,]	0	1	0	1	0	0
[2,]	1	0	1	0	0	0
[3,]	0	1	0	1	0	1
[4,]	1	0	1	0	0	0
[5,]	0	0	0	0	0	1
[6,]	0	0	1	0	1	0

Adjacency matrix



Empirical graphon



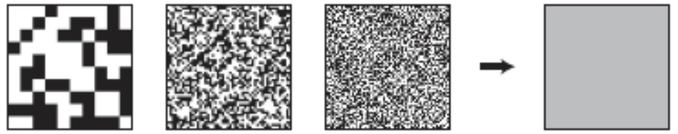
Pixel picture

Matrix gets converted to function on a unit square

1 => black
0 => white

Convergence - cut metric

- Pixel picture/empirical graphon convergence

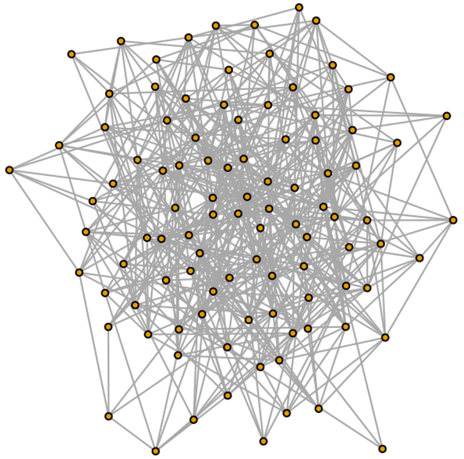


- The cut norm of a graphon W is defined as $\|W\|_{\square} = \sup_{S,T} \left| \int_{S \times T} W(x,y)\, dxdy \right|$ where supremum is taken over all measurable sets S and T of $[0,1]$
- Given two graphons W_1 and W_2, the **cut metric** is defined as $\delta_{\square}(W_1, W_2) = \inf_{\varphi} \left\| W_1 - W_2^{\varphi} \right\|_{\square}$ where φ is a measure preserving bijection from $[0,1]$ to $[0,1]$.
- Why φ? Because nodes can be re-named or permuted.
- Convergence in homomorphism density is equivalent to convergence in cut-metric. (Borgs et al 2011)

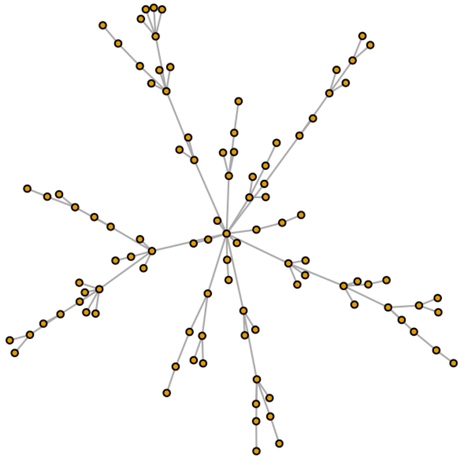
Dense and sparse graphs

- Consider a graph sequence $\{G_n\}_{n'}$ where G_n has n nodes and m edge
- Dense graphs: $\liminf_{n\to\infty} \frac{m}{n^2} = c > 0$
  - **Edges grow quadratically with nodes**
- Sparse graphs: $\lim_{n\to\infty} \frac{m}{n^2} = 0$
  - **Edges grow subquadratically with nodes**

Dense



Sparse



Exchangeability



A bit on exchangeability

Exchangeability: order of the variables does not change the joint distribution.

	Discrete	**Continuous**
1D (Sequences/ Point processes)	de Finetti (1931)	Bühlmann (1960)
2D (Arrays Joint/ separate)	Aldous (1981) and Hoover (1979)	Kallenberg (1990)

de Finetti, B. (1931) Funzione caratteristica di un fenomeno aleatorio. Atti R. Acad. Nazn. Lino, Ser. 6, 4, 251- 299.

Bühlmann, H. (1960). *Austauschbare stochastische Variabeln und ihre Grenzwertsätze* (Doctoral dissertation, ETH Zurich)

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

Hoover, D. N. (1979). Relations on Probability Spaces and Arrays of. *Preprint, Institute for Advanced Study,* Princeton

Kallenberg, O. (1990). Exchangeable random measures in the plane. *Journal of Theoretical Probability*, *3*(1), 81-136.

Exchangeable binary matrices

- Aldous-Hoover representation theorem

$$A_{ij} \mid U_i, U_j, W \sim \text{Ber}(W(U_i, U_j))$$

with $U_i \overset{\text{iid}}{\sim} \text{Unif}(0,1)$, and $W : [0,1]^2 \to [0,1]$ a random function

- Infinite exchangeable binary matrices imply a mixture model representation exist
- For undirected graphs, this mixture is specified by the **graphon** (Borgs et al 2008, Lovász et al 2006)
- For graphs the connection to Aldous-Hoover made by Diaconis & Janson

Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T., & Vesztergombi, K. (2008). Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, *219*(6), 1801-1851.

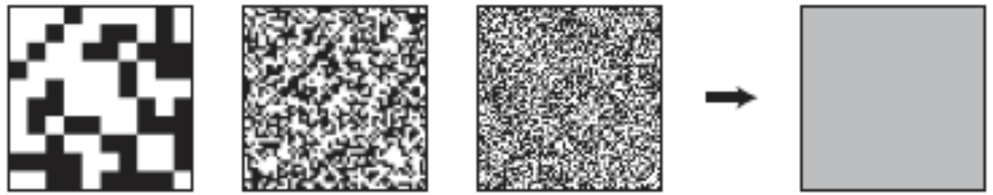
Janson, S., & Diaconis, P. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII*, 33-61.

Lovász, L., & Szegedy, B. (2006). Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, *96*(6), 933-957.

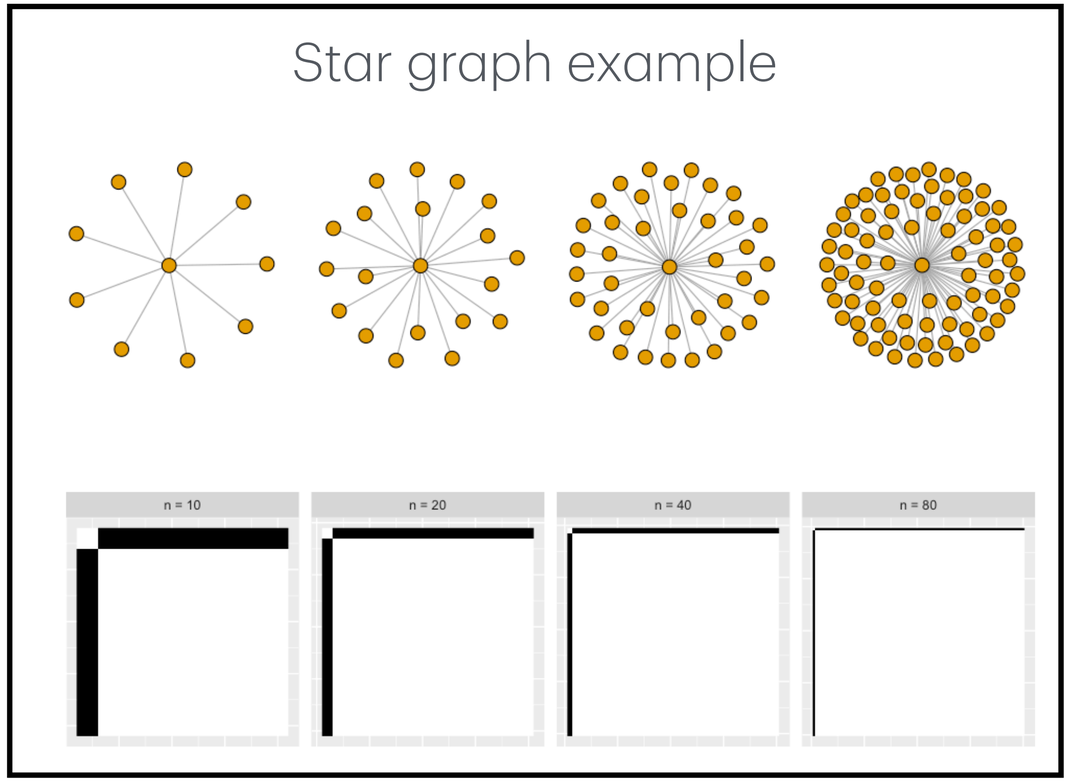
Lovász, L. (2012). *Large networks and graph limits* (Vol. 60). American Mathematical Soc..

Limitation of traditional graphons

- ***Graphs represented by an exchangeable matrix are either trivially empty or dense***
- These graphons can model dense graphs but not sparse graphs



- Why?
  - The volume from $(x, y, W(x, y))$ converges to $\lim_{n\to\infty} \frac{2m}{n^2}$
  - For sparse graphs $\lim_{n\to\infty} \frac{m}{n^2} = 0 \iff W = 0$



What happens when $W = 0$?

$W = 0$

- Sample a graph from $W = 0$
- You have a graph with n nodes, but no edges — not representative of sparse graphs

This is the fundamental problem of modeling sparse graphs using graphons!

Some Approaches for Sparse Graphs

Sevvandi Kandanaarachchi, Cheng Soon Ong
CSIRO

Kallenberg Exchangeable Graphs

- Introduced by Caron and Fox (2017) and later used by many others
- Nodes $\theta_i \in \mathbb{R}^+$
- $W : \mathbb{R}_+^2 \to [0,1]$
  defined on the positive quadrant

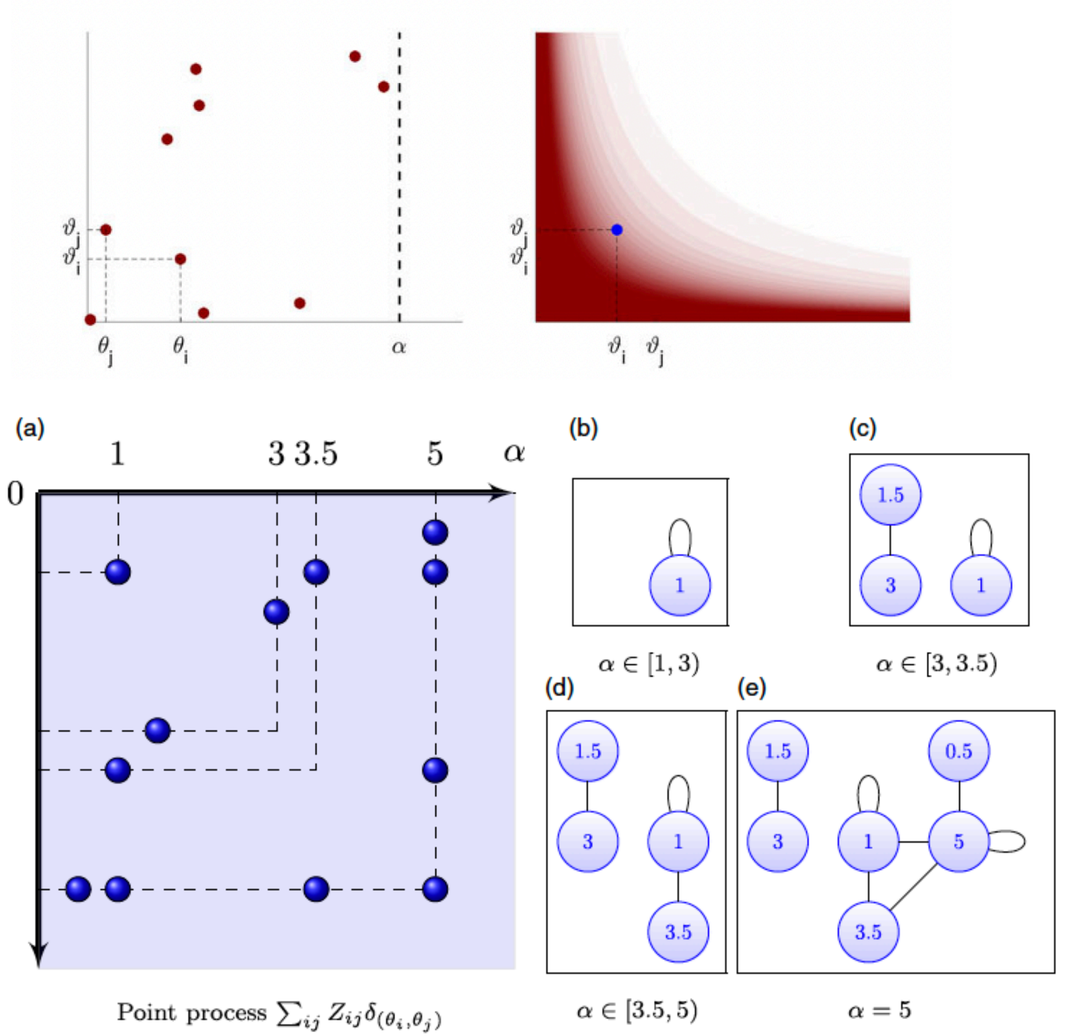


FIGURE 3. Illustration of the connection between the point process on the plane and the graphex process. (a) Point process $\sum_{ij} Z_{ij}\delta_{(\theta_i,\theta_j)}$ on the plane. (b–e) Associated graphs $\mathcal{G}_\alpha$ for (b) $\alpha \in [1, 3)$, (c) $\alpha \in [3, 3.5)$, (d) $\alpha \in [3.5, 5)$, and (e) $\alpha = 5$. Note that the graph is empty for $\alpha < 1$.

Caron, F., & Fox, E. B. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(5), 1295-1366.

Caron, F., Panero, F., & Rousseau, J. (2023). On sparsity, power-law, and clustering properties of graphex processes. *Advances in Applied Probability*, *55*(4), 1211-1253.

Rescaled and stretched graphons

- Many methods proposed by Borgs & Chayes research group
- Rescaled $\widetilde{W} = \frac{W}{\|W\|_1}$
- Stretched graphon defined on $\mathbb{R}^2_+$.
  $\widetilde{W}(x,y) = W(\|W\|_1^{0.5}x, \|W\|_1^{0.5}y)$

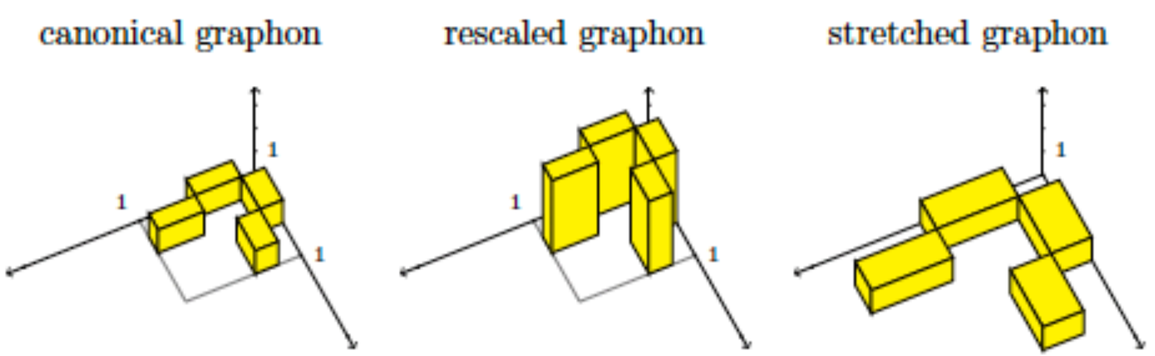


Figure 3: The figure shows three graphons associated with the same simple graph G on five vertices. In the classical theory of graphons all simple sparse graphs converge to the zero graphon. We may prevent this by renormalizing the graphons, either by rescaling the height of the graphon (middle) or by stretching the domain on which it is defined (right). The first approach was chosen by Bollobás and Riordan (2009) and Borgs, Chayes, Cohn, and Zhao (2014a,b), and the second approach is chosen in this paper. In our forthcoming paper (Borgs, Chayes, Cohn, and Holden, 2017) we choose a combined approach, where the renormalization depends on the observed graph.

Borgs, C., Chayes, J. T., Cohn, H., & Holden, N. (2018). Sparse exchangeable graphs and their limits via graphon processes. *Journal of Machine Learning Research*, *18*(210), 1-71.

Edge exchangeability

- Crane and Dempsey, Cai et al, Janson and others

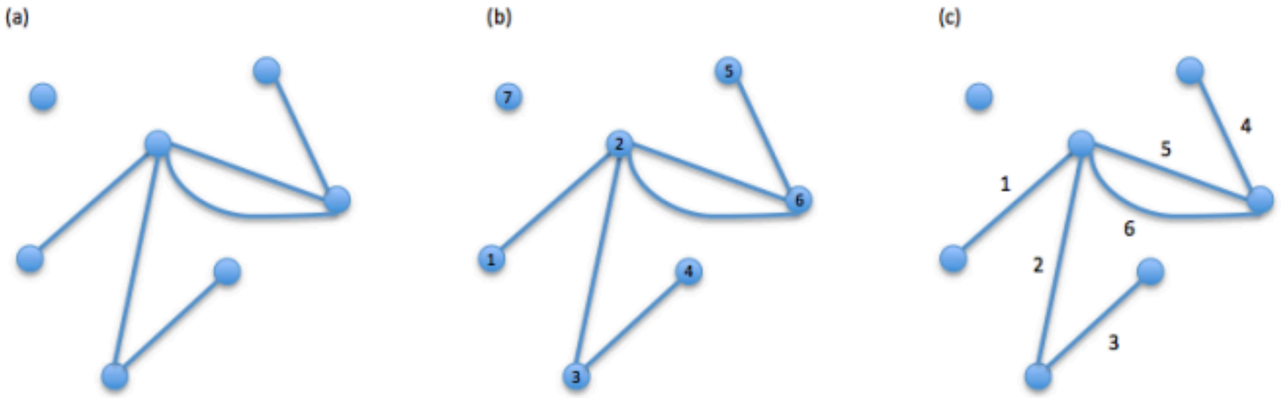


FIGURE 1. (a) A network structure derived from some process of interactions. Neither vertices nor edges come equipped with labels. (b) Network obtained by labeling vertices of network in Panel (a). (c) Network obtained by labeling edges of network in Panel (a).

Crane, H., & Dempsey, W. (2015). A framework for statistical network modeling. *arXiv preprint arXiv:1509.08185*.

Crane, H., & Dempsey, W. (2016). Edge exchangeable models for network data. *arXiv preprint arXiv:1603.04571*.

Cai, D., Campbell, T., & Broderick, T. (2016). Edge-exchangeable graphs and sparsity. *Advances in Neural Information Processing Systems*, *29*.

Janson, S. (2018). On edge exchangeable random graphs. *Journal of statistical physics*, *173*(3), 448-484.

Graphexes

- Originally introduced by Veitch & Roy
- A triple (I, S, W) where $I \in \mathbb{R}$ models isolated edges, $S : \mathbb{R}_+ \to \mathbb{R}_+$ integrable, models infinite stars and $W : \mathbb{R}_+^2 \to [0,1]$ is the standard graphon. But they consider $I = 0$ and $S = 0$
- Borgs and Chayes group - high degree hubs give the skeleton with low degree vertices forming the body

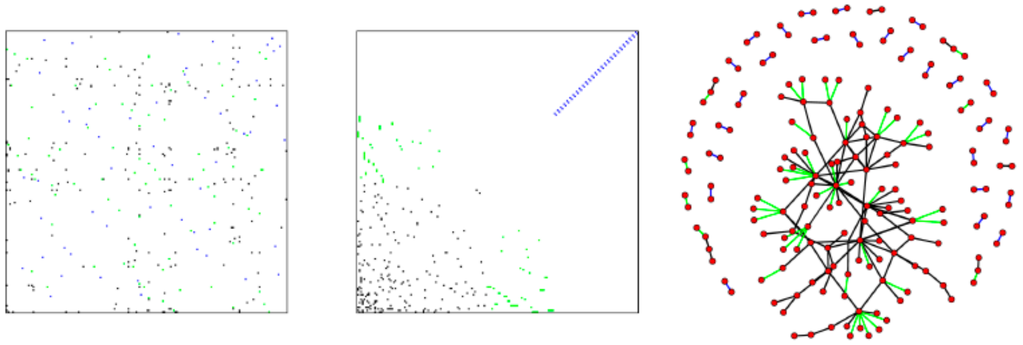


FIGURE 3. Realization of unlabeled graphex process generated by $\mathcal{W} = (I, S, W)$ at size $s = 15$ (right panel), and associated dilated empirical graphon (left and center panels). The generating graphex is $W = (x+1)^{-2}(y+1)^{-2}$, $S = 1/2\exp(-(x+1))$, and $I = 0.1$. The observation size is $s = 15$. The dilated empirical graphex is

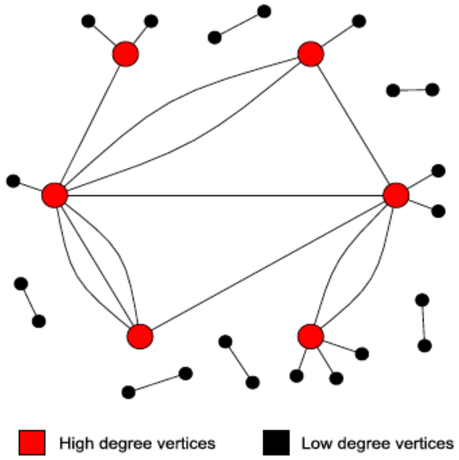


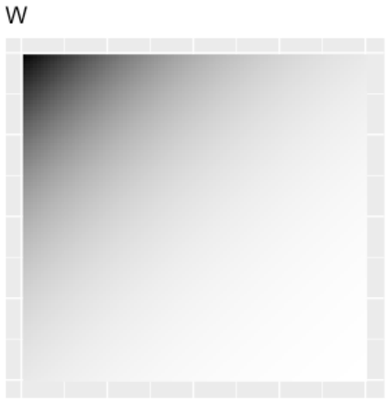
FIG 1. *Typical structure of* $\mathrm{Smpl}(\mathrm{CM}_n(\boldsymbol{d}), t/\sqrt{\ell_n})$.

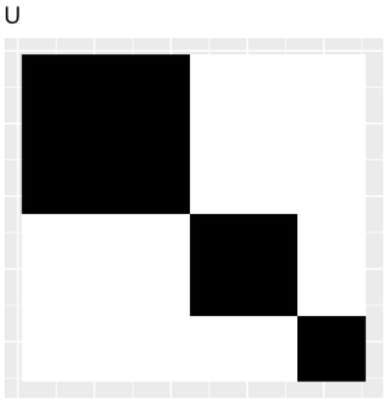
Veitch, V., & Roy, D. M. (2015). The class of random graphs arising from exchangeable random measures. *arXiv preprint arXiv:1512.03099*.

Veitch, V., & Roy, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs.

Borgs, C., Chayes, J. T., Dhara, S., & Sen, S. (2021). Limits of sparse configuration models and beyond. *The Annals of Probability*, *49*(6), 2830-2873.

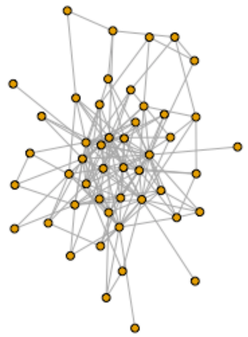
Graphon mixtures

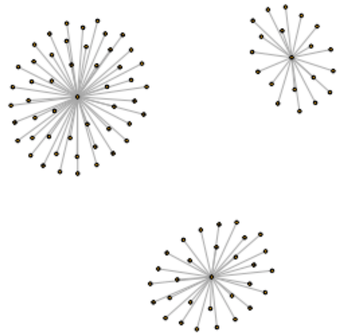




Sample dense graph

G_{n_d}





Mixture graph after joined



Sample sparse graph

$G_{n_s} = L^{-1}(H_{m_s})$

First sample disjoint cliques

Then compute the inverse line graph

- Parameters: dense nodes n_d, sparse edges m_s, the number of edges when joining m_{new}

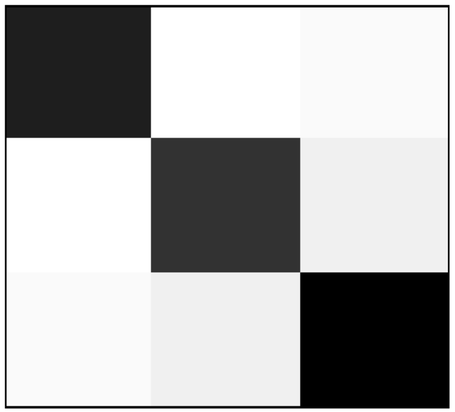
Kandanaarachchi, S., & Ong, C. S. (2026). Graphon Mixtures. *AISTATS, 2026*

Graphettes

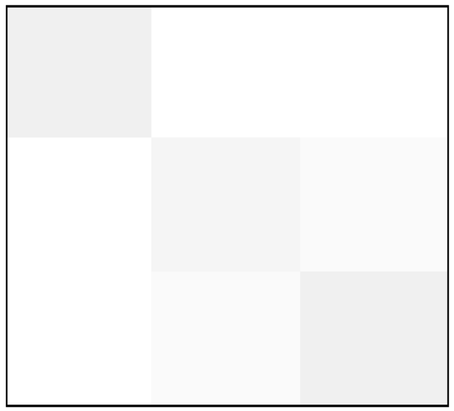
- Graphette $\mathscr{W} = (W, \rho_n, f)$
- W is a graphon on $[0,1]^2$
- $\{\rho_n\}_n = (\rho_1, \ldots, \rho_i, \ldots)$, a sequence, can converge to zero
- f is a graph edit function - eg. ring addition, star addition, cycle deletion

Wijesinghe, A., Kandanaarachchi, S., Steinberg, D. M., & Ong, C. S. (2026). Flowette: Flow Matching with Graphette Priors for Graph Generation. *arXiv preprint arXiv:2602.23566*.

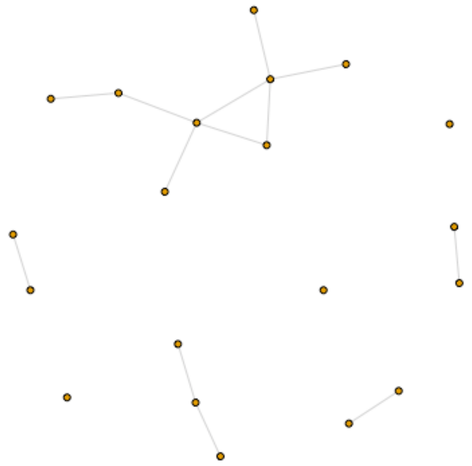
Example $\mathscr{W} = (W, \rho_n, f)$ with $\rho_n = c/n$ and $f = R$



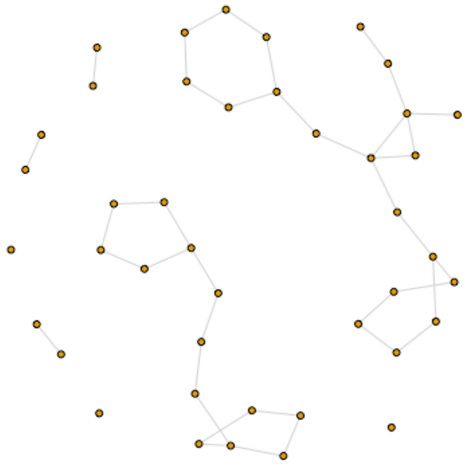
W



$W_n = \rho_n W$



$\tilde{G}_n$



$G_n = R(\tilde{G}_n)$

- R denotes ring addition - Bernoulli r.v.
- First $\tilde{G}_n \sim W_n = \rho_n W$ and then $G_n = R(\tilde{G}_n)$

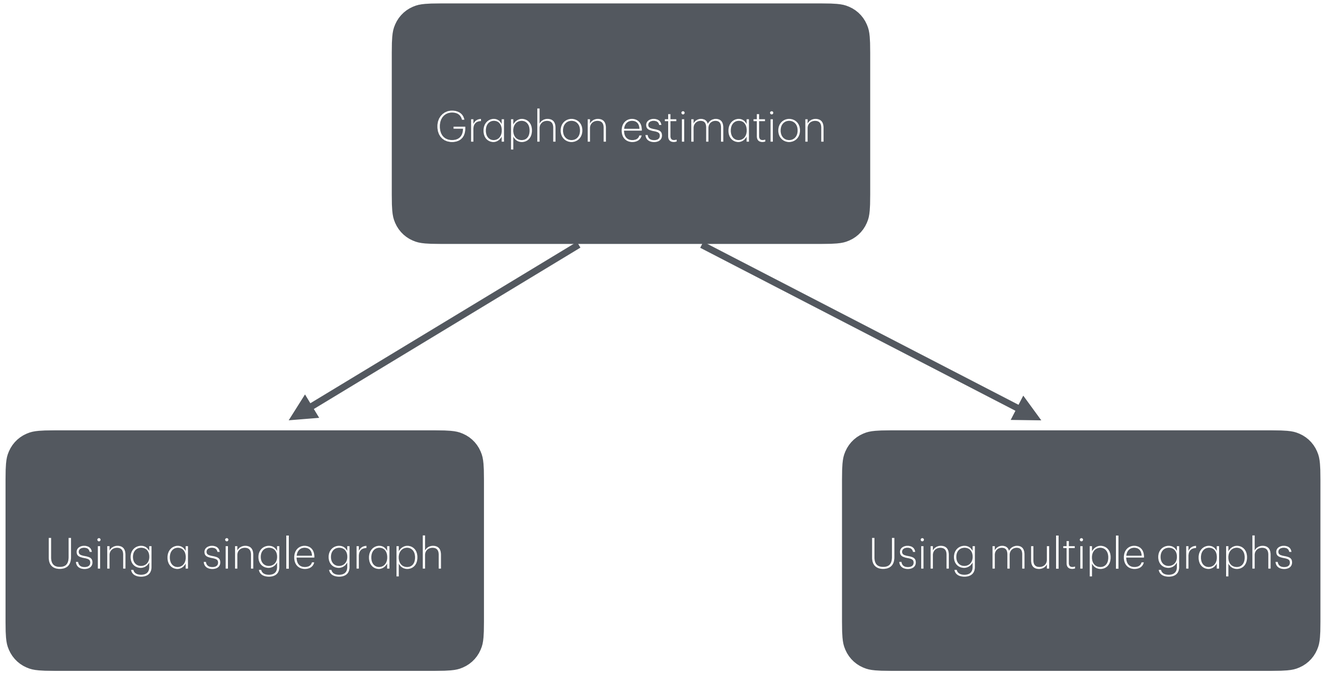
Wijesinghe, A., Kandanaarachchi, S., Steinberg, D. M., & Ong, C. S. (2026). Flowette: Flow Matching with Graphette Priors for Graph Generation. *arXiv preprint arXiv:2602.23566*.

Graphon Estimation

Sevvandi Kandanaarachchi, Cheng Soon Ong
CSIRO

Why estimate?

- Recover the latent structure
- Given a graph, simulate growth and generate a larger graph
- Prediction purposes



- Easier estimation problem
- Only 1 sample from graphon

- Difficult estimation problem (nodes not aligned)
- Multiple samples from graphon

Using a single graph

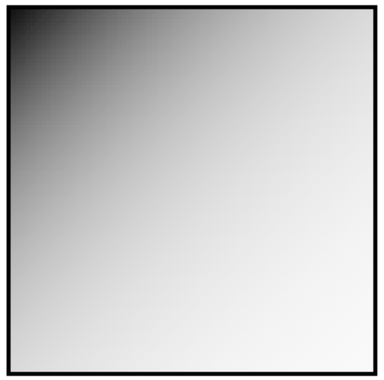


Monotonic degree condition

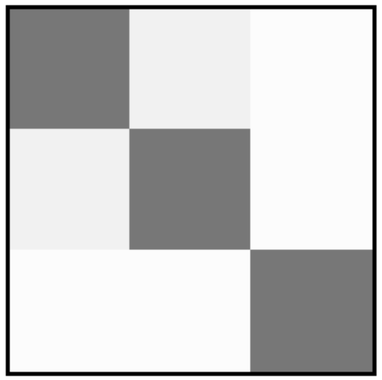
- Degree function $g(x) = \int_0^1 W(x, y)\, dy$

- Condition/Assumption: There exists a measure-preserving bijection ϕ such that the degree function g_ϕ associated with $W^\phi = W(\phi(x), \phi(y))$ such that W^ϕ is **strictly** increasing.

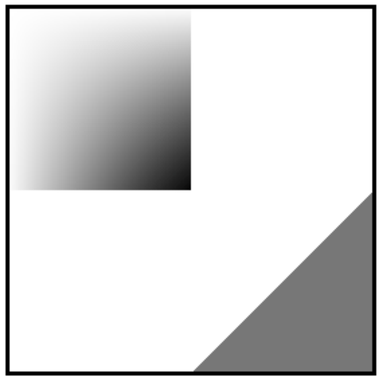
Monotonic



Non-monotonic



Non-monotonic



$W(x, y) = \exp(-2(x + y))$

Janson, S. (2020). A graphon counter example. *Discrete Mathematics*, *343*(6), 111836.

Monotonic degree condition - Why?

- Makes it a lot easier to estimate the graphon
- Sort the degrees => Then estimate!
- Chan, S., & Airoldi, E. (2014). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* (pp. 208-216). PMLR.

The sorting and smoothing algorithm

- Graphon estimation using a single graph
- Sort the graph by degree
- Smoothing
  - Network histogram estimation
  - Total variation minimization

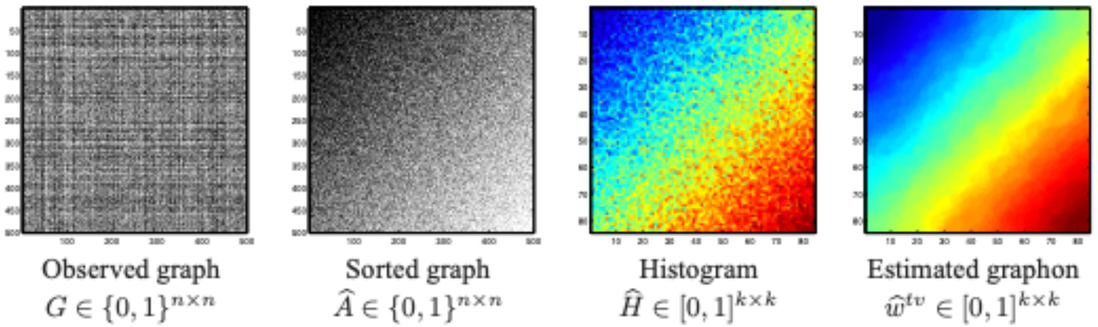


Figure 2. Illustration of the SAS algorithm. Given an observed graph G, we first sort G using the empirical degrees to get $\widehat{A}$. Then, a local histogram $\widehat{H}$ is computed and a total variation minimization is used to determine an estimate $\widehat{w}^{tv}$.

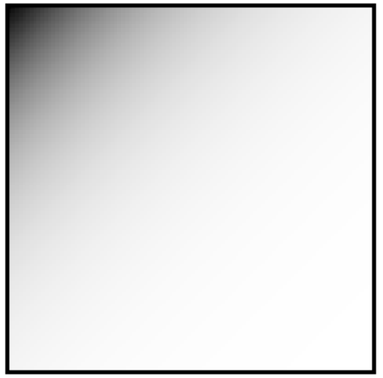
Chan, S., & Airoldi, E. (2014, January). A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning* (pp. 208-216). PMLR.

Using multiple graphs



Multiple graphs: node alignment between graphs

W



- Surprisingly, the estimation problem is harder
- $G_1 \sim W$, suppose node labels are ordered from $[0,1]$ on graphon axis
  - Degrees decrease with the node labels
- $G_2 \sim W$, suppose node labels are all random from $[0,1]$ on graphon axis
  - Degrees are all over the place
- To estimate the graphon, you need to relabel nodes so that labeling is consistent with the graphon

Popular fix: Assume graphon has monotonic degree property

Two selected methods

- Joint Graph Sorting (JGS)
  - Extends Sorting and Smoothing Algorithm by Chan and Airoldi
- Implicit Graphon Neural Representation (IGNR)
  - SIREN + GW distance

Sogan, R. B., & Rebafka, T. (2026). Low-Complexity and Consistent Graphon Estimation from Multiple Networks. In *International Conference on Artificial Intelligence and Statistics*

Xia, X., Mishne, G., & Wang, Y. (2023). Implicit graphon neural representation. In *International Conference on Artificial Intelligence and Statistics* (pp. 10619-10634). PMLR.

Joint Graph Sorting (JGS)

- Degree Monotonic assumption applies
- Take union of nodes and sort by normalized degree
- Merged adjacency matrix
- Histogram estimator
- Optional smoothing step

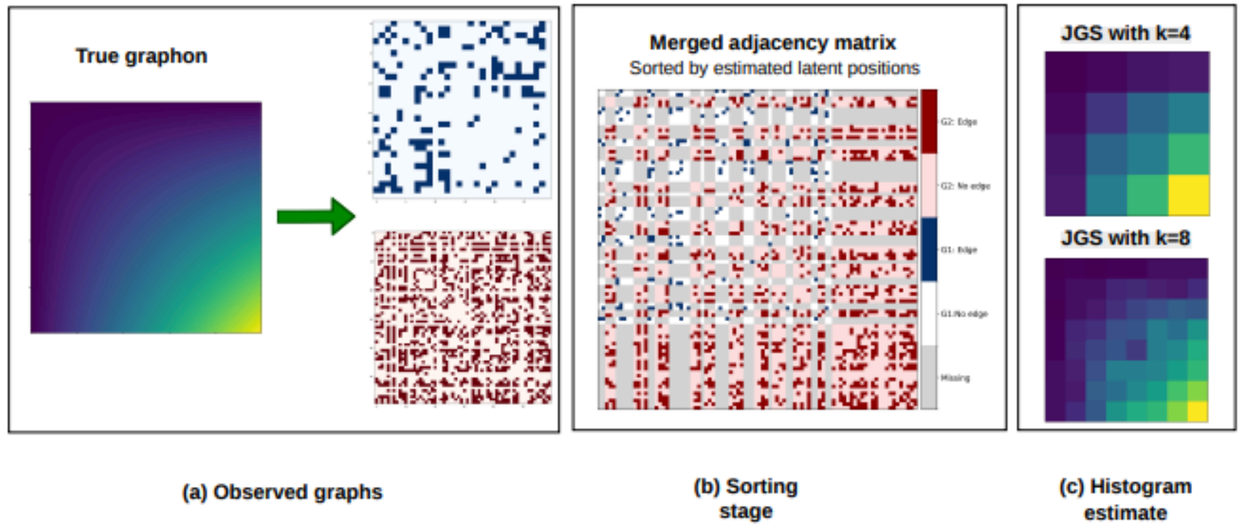
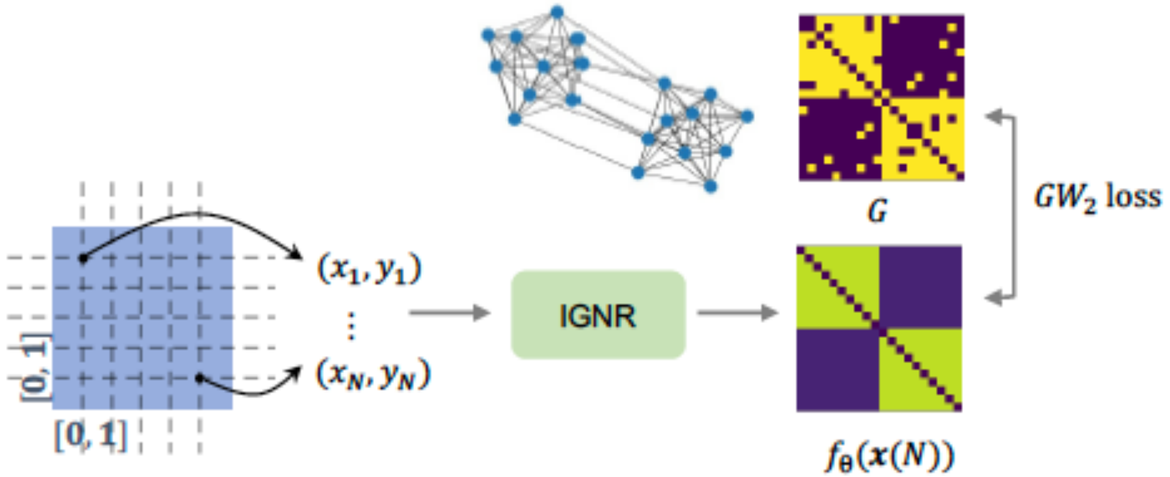


Figure 1: Illustration of the Joint Graph Sorting (JGS) estimator. (a) Networks are generated from a graphon. (b) Nodes are jointly sorted across networks using normalized degrees, yielding a merged adjacency matrix with missing entries (grey). Dark blue and white denote edges and non-edges in the first graph, respectively, while dark red and misty rose denote edges and non-edges in the second graph, respectively. (c) JGS graphon estimates with different resolutions k, where colors represent estimated edge probabilities.

Sogan, R. B., & Rebafka, T. (2026). Low-Complexity and Consistent Graphon Estimation from Multiple Networks. In *International Conference on Artificial Intelligence and Statistics*

IGNR

- No degree monotonicity condition
- SIREN implementation (sine activation functions)
- $$\min_{\theta} \frac{1}{M} \sum_{i=1}^{M} GW_2(f_\theta(\mathbf{x}(N_i)), G_i)$$
- $f_\theta(\mathbf{x}(N_i))$ is a weighted adjacency matrix sampled from IGNR
  - Use optimal transport



Xia, X., Mishne, G., & Wang, Y. (2023). Implicit graphon neural representation. In *International Conference on Artificial Intelligence and Statistics* (pp. 10619-10634). PMLR.

Graphons in ML

Sevvandi Kandanaarachchi, Cheng Soon Ong
CSIRO

Applications

- Transferability of GNNs
  - Graphon neural networks and the transferability of graph neural networks
- Neural Network Pruning
  - The Graphon Limit Hypothesis: Understanding Neural Network Pruning via Infinite Width Analysis

Ruiz, L., Chamon, L., & Ribeiro, A. (2020). Graphon neural networks and the transferability of graph neural networks. *Advances in Neural Information Processing Systems*, *33*, 1702-1712.

Pham, H., Jacobs, T., Burkholz, R., & Tran-Thanh, L. (2026). The graphon limit hypothesis: Understanding neural network pruning via infinite width analysis. *Advances in Neural Information Processing Systems*, *38*, 53283-53319.

Transferability of GNNs



Graphon NNs and Transferability

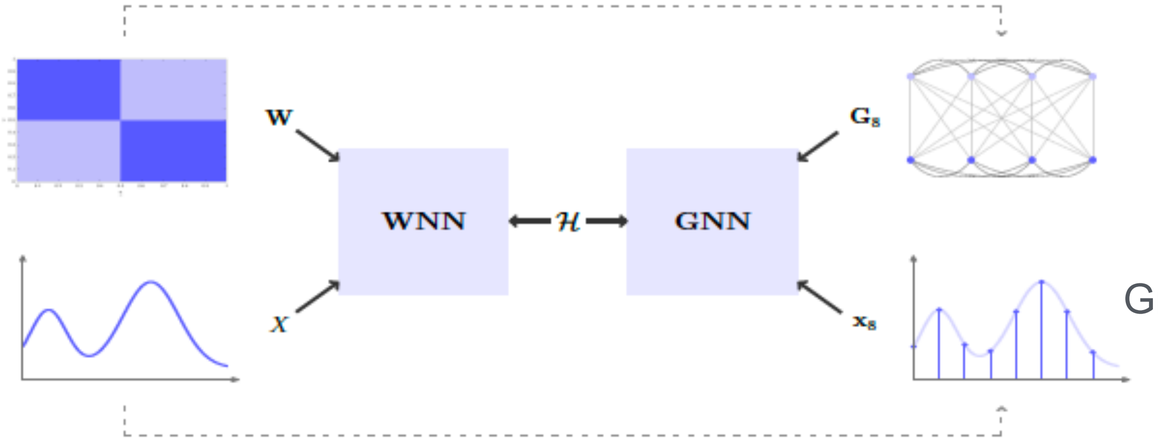
Theorem (GNN transferability, informal) Let $\mathbf{\Phi}(\mathbf{G})$ be a GNN with fixed parameters. Let $\mathbf{G}_{n_1}$ and $\mathbf{G}_{n_2}$ be deterministic graphs with n_1 and n_2 nodes obtained from a graphon $\mathbf{W}$. Under mild conditions, $\|\mathbf{\Phi}(\mathbf{G}_{n_1}) - \mathbf{\Phi}(\mathbf{G}_{n_2})\| = \mathcal{O}(n_1^{-0.5} + n_2^{-0.5})$.

Number of nodes	$n = 271$	$n = 542$	$n = 1354$
Accuracy on subnetwork $\mathbf{G}_n$	25.3%	29.5%	36.6%
Accuracy on full network	27.9%	35.4%	41.8%
Relative accuracy difference	23.1%	22.5%	15.4%

Accuracy on node classification using Cora citation network

Ruiz, L., Chamon, L., & Ribeiro, A. (2020). Graphon neural networks and the transferability of graph neural networks. *Advances in Neural Information Processing Systems*, *33*, 1702-1712.

Graphon NNs



Graph signal = data at node i

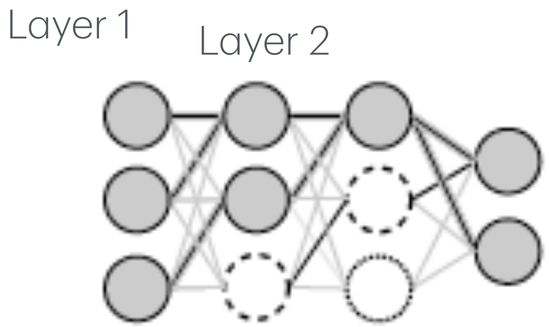
Figure 1: Example of a graphon neural network (WNN) given by $\mathbf{\Phi}(\mathcal{H}; \mathbf{W}; X)$, and of a graph neural network (GNN) instantiated from it as $\mathbf{\Phi}(\mathcal{H}; \mathbf{S}_8; \mathbf{x}_8)$. The graphon $\mathbf{W}$, shown on the top left corner, is a stochastic block model with intra-community probability $p_{c_i c_i} = 0.8$ and inter-community probability $p_{c_i c_j} = 0.2$, and the graphon signal X is plotted on the bottom left corner. The graph $\mathbf{G}_8$ (top right corner) and the graph signal $\mathbf{x}_8$ (bottom right corner) are obtained from $\mathbf{W}$ and X according to (11). Note that the parameter set $\mathcal{H}$ is shared between the WNN and the GNN.

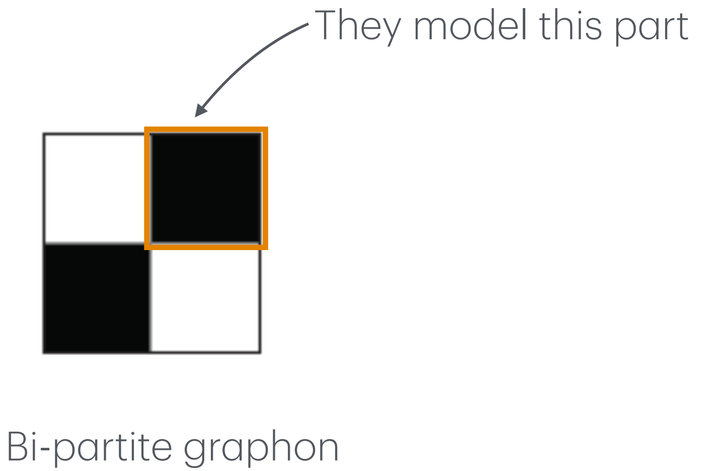
Ruiz, L., Chamon, L., & Ribeiro, A. (2020). Graphon neural networks and the transferability of graph neural networks. *Advances in Neural Information Processing Systems*, *33*, 1702-1712.

NN Pruning



NN Pruning at Initialization





- Bi-partite graph from layer 1 to layer 2

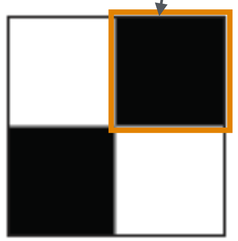
Given a sparsity level, each network pruning method defines sequences of binary masks $(M_n^{(l)})$ ***that converges layer-wisely to graphons*** $W^{(l)}$ ***in cut distance as the network width*** $n \to \infty$.

Pham, H., Liu, S., Xiang, L., Le, D., Wen, H., & Tran-Thanh, L. (2023). Towards data-agnostic pruning at initialization: What makes a good sparse mask?. *Advances in neural information processing systems*, *36*, 80044-80065.

Pham, H., Jacobs, T., Burkholz, R., & Tran-Thanh, L. (2026). The graphon limit hypothesis: Understanding neural network pruning via infinite width analysis. *Advances in Neural Information Processing Systems*, *38*, 53283-53319.

NN Pruning at Initialization

They model this part



Bi-partite graphon

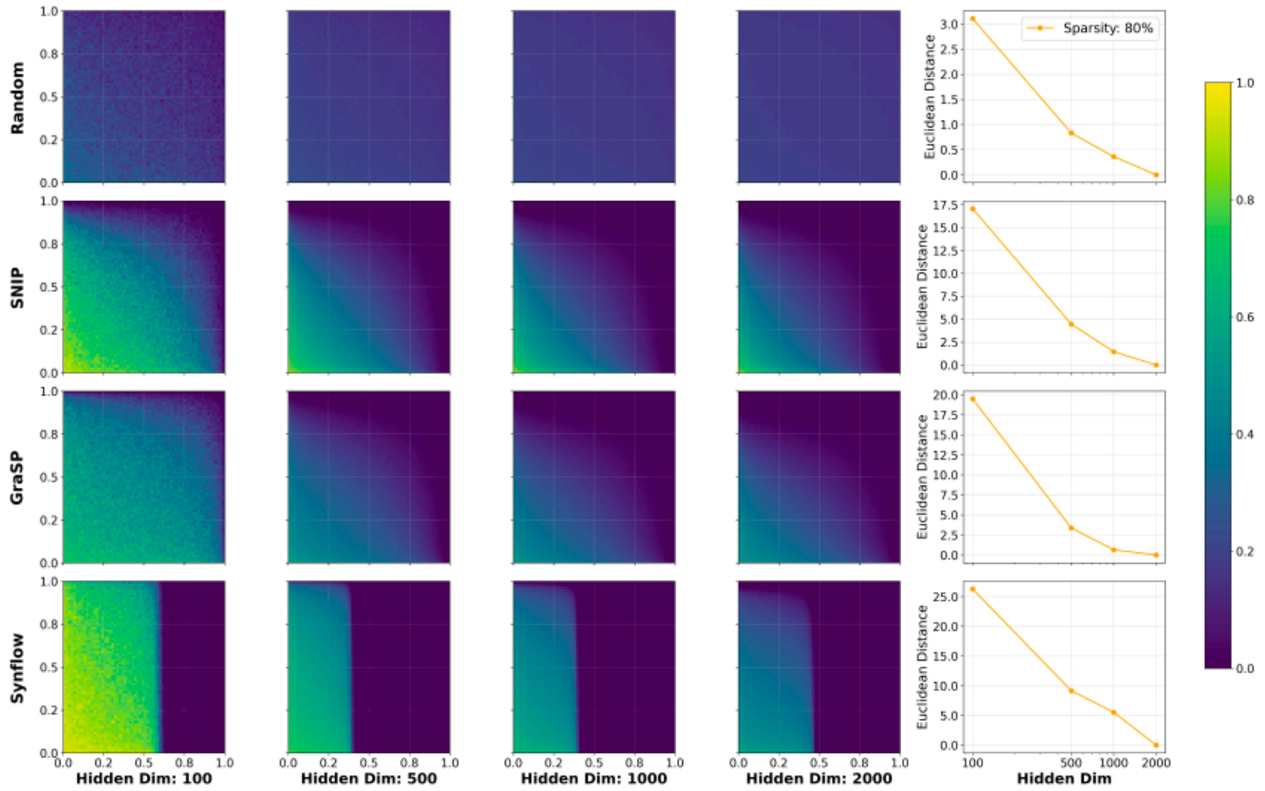


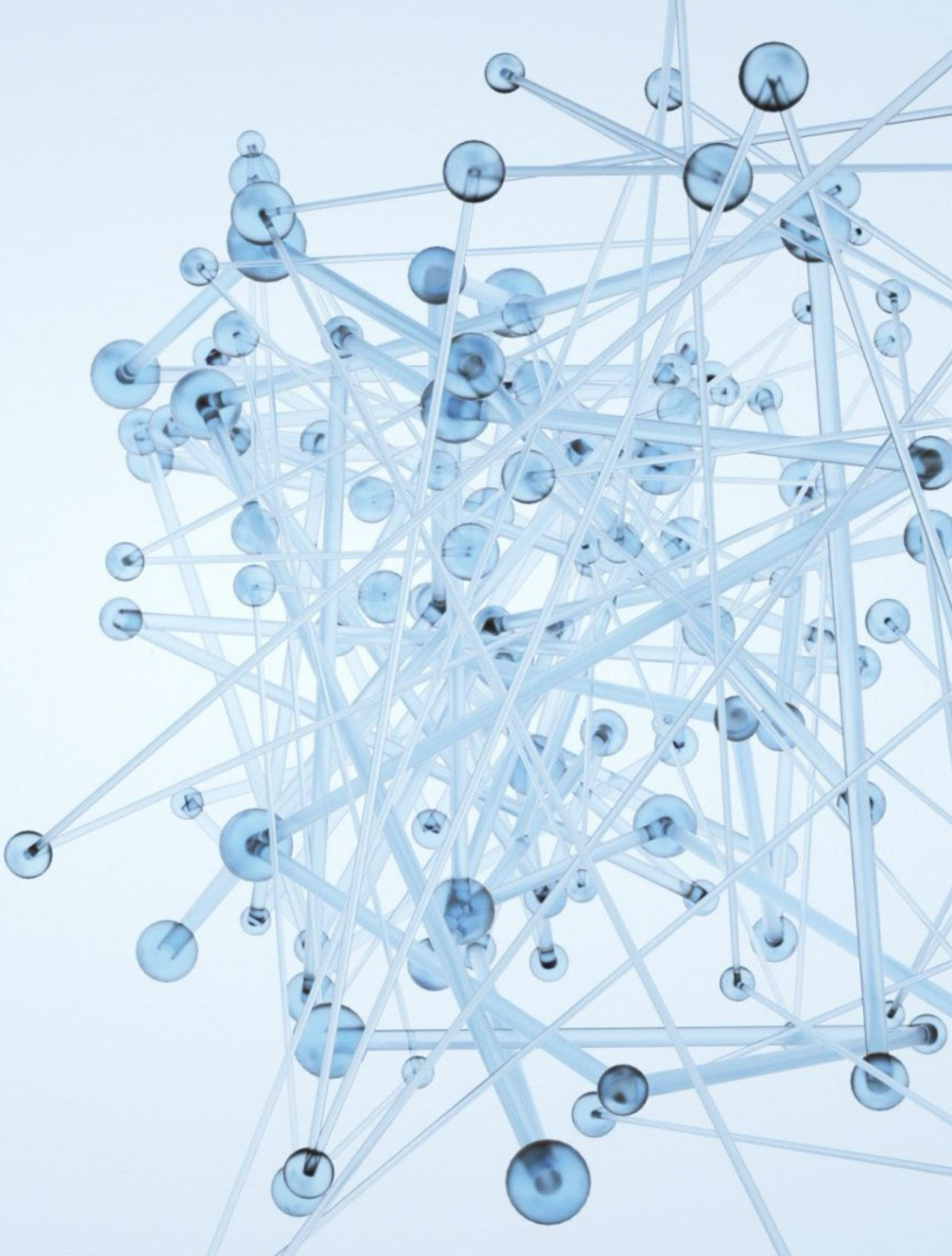
Figure 1: Graph limit of subnetworks' mask produced by PaI methods at 80% sparsity and the corresponding convergence of graphons via Euclidean distances.

Pham, H., Liu, S., Xiang, L., Le, D., Wen, H., & Tran-Thanh, L. (2023). Towards data-agnostic pruning at initialization: What makes a good sparse mask?. *Advances in neural information processing systems*, *36*, 80044-80065.

Pham, H., Jacobs, T., Burkholz, R., & Tran-Thanh, L. (2026). The graphon limit hypothesis: Understanding neural network pruning via infinite width analysis. *Advances in Neural Information Processing Systems*, *38*, 53283-53319.

CSIRO

Thank you!



NEURAL INFORMATION PROCESSING SYSTEMS