STOCHASTICITY FOR RANDOM GRAPHS

Jan Reimann, Penn State

June 18, 2015

RANDOM GRAPHS

- Fix $0 < p < 1$. $\mathbb{G}(n, p)$ is a simple, undirected graph with n vertices where each edge is present (indepedently) with probability p.
- A natural "limit object" for $n \to \infty$ is $\mathbb{G}(\mathbb{N}, p)$, a *countable* p*-random graph*.
- This is known as the *Erdös-Renyi model*.

CONUNDRUMS OF RANDOM GRAPHS

- For any $0 < p < q < 1$, two graphs $\mathbb{G}(\mathbb{N}, p)$ and $\mathbb{G}(\mathbb{N}, q)$ are almost surely equivalent.
- There exists a computable graph $\mathcal{G}$ on $\mathbb{N}$ such that for every p, almost surely $\mathbb{G}(\mathbb{N}, p)$ is isomorphic to $\mathcal{G}$. [Rado]

CONSEQUENCES

What does this imply for algorithmic randomness?

- We can fix a probability distribution and develop randomness for labeled graphs and try to keep it "as invariant as possible".
  - Since there is a recursive copy, no approach with even modest computational power will include **all** copies of the random graph.
  - Will the randomness be "in the isomorphism"?
    [Fouché]

CONSEQUENCES

- We can accept the fact that a random graph is so highly symmetric (the automorphism group is extremely rich) that we have a recursive copy.
  - The situation then seems similar to *normal numbers*.
  - They satisfy many randomness properties (particularly from a dynamical point of view).
  - This suggests to look at random graphs from a stochasticity point of view (but what is a *normal* graph?).

CONSEQUENCES

As we will see, both aspects are closely related.

Does algorithmic randomness (in the "classical" sense) have anything significant to add to the picture?

RANDOM GRAPHS AS HOMOGENEOUS STRUCTURES

- The reason for the rich symmetry of the random graph can be seen in its homogeneity.
- A countable (relational) structure $\mathcal{M}$ is *homogeneous* if every isomorphism between finite substructures of $\mathcal{M}$ extends to an automorphism of $\mathcal{M}$.
- The Rado graph $\mathcal{G}$ is homogeneous by virtue of the *I-property*:

For any $x_1, \ldots, x_n, y_1, \ldots, y_m$ *there exists* z

$$z \sim x_i, z \nsim y_j$$

for all $1 \le i \le n, 1 \le j \le m.$

HOMOGENEOUS STRUCTURES

- **Fraissé**: Any homogeneous structure arises as a *amalgamation process* of finite structures over the same language (Fraissé limits).
- Examples:
  - $(\mathbb{Q}, <)$,
  - the Rado (random) graph
  - the universal K_n-free graphs, $n \geq 3$ (Henson)
- Homogeneous structures (over finite languages) are $\aleph_0$-*categorical*, i.e. their theory has only one model up to isomorphism.

RANDOMNIZED CONSTRUCTIONS

- Many homogeneous structures can obtained (almost surely) by adding new points according to a randomized process.
  - $(\mathbb{Q}, <)$: add the n-th point between (or at the ends) of any existing point with uniform probability $1/n$.
  - Rado graph: add the n-th vertex and connect to every previous vertex with probability p (uniformly and independently).
  - Vershik: Urysohn space, Droste and Kuske: universal poset
  - Henson graph: ???

CONSTRUCTIONS "FROM BELOW"

- A naive approach to "randomize" the construction of the Henson graph would be as follows:
  - In the n-th step of the construction, pick a one-vertex extension uniformly among all possible extensions that preserve K_n-freeness.
- However: **Erdös, Kleitman, and Rothschild** showed that (as n goes to ∞) almost all graphs missing a K_n are bipartite.
  - The Henson graph(s), in contrast, has to contain every finite K_n-free graph as an induced subgraph, in particular, C_5 and hence cannot be bipartite.

CONSTRUCTIONS "FROM ABOVE"

- **Petrov and Vershik** (2010) showed how to construct universal K_n-free graphs probabilistically by *sampling them from a continuous graph*.
- These continuous graphs, known as **graphons**, have been studied extensively over the past decade.
  - See, for example the recent book by Lovasz, *Large networks and graph limits* (2012).

GRAPHONS

- One basic motivation behind graphons is to capture the asymtotic behavior of growing sequences of dense graphs, e.g. with respect to subgraph densities.
- While the Rado graph can be seen as the limit object of a sequence (G_n) of finite random graphs, it does not distinguish between the distributions with which the edges are produced.
- For any $0 < p < 1$, $\mathbb{G}(n, p)$ "converges" almost surely to (an isomorphic copy of) the Rado graph.
  - However, $p_1 \ll p_2$, $\mathbb{G}(n, p_1)$ will exhibit very different subgraph densities than $\mathbb{G}(n, p_2)$

CONVERGENCE

- Let (G_n) be a graph sequence with $|V(G_n)| \to \infty$.
- We say (G_n) **converges** if

> *for every finite graph F, the relative number $t_i(F, G_n)$ of embeddings of F into G_n converges.*

QUASIRANDOM GRAPHS

- A sequence of graphs (G_n), $|G_n| = n$ is *quasirandom* if for every graph F on k vertices,
$$t_i(F, G_n) \approx 2^{-\binom{k}{2}} \text{asymptotically.}$$
- That means every fixed finite graph occurs with the "right" frequency.
- Hence quasirandom sequences converge in the above sense.

QUASIRANDOM GRAPHS

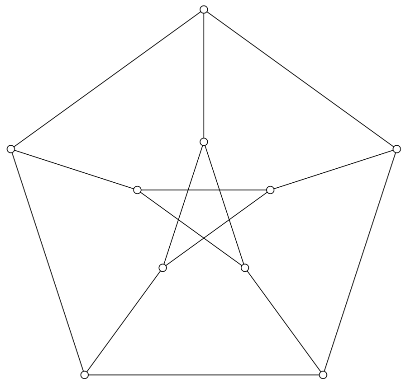
- Quasirandom graph sequences form a natural analog to normal sequences.
- However, the additonal structure of graphs makes them more robust. Chung, Graham and Wilson (1989) showed that it suffices to satisfy the asymptotic frequency condition for K_2 (one edge) and C_4 (squares) only.
- One can take quasirandom graphs as a basis for "classical" *stochasticity* for graphs.
- *How robust are they under various kinds of selection rules?*
  - This is an ongoing project of Penn State graduate student Jake Pardo.

GRAPHONS

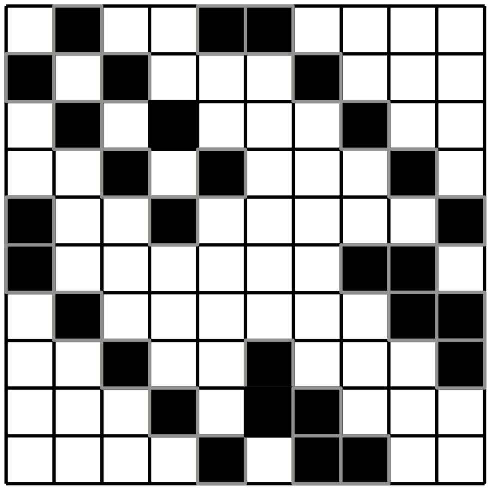
- $W : [0,1]^2 \to [0,1]$ measurable, and for all x, y,
$$W(x,x) = 0 \text{ and } W(x,y) = W(y,x).$$
- Think: $W(x,y)$ is the probability there is an edge between x and y.
- Subgraph densities:
  - edges: $\int W(x,y)\,dx\,dy$
  - triangles: $\int W(x,y)W(y,z)W(z,x)\,dx\,dy\,dz$
  - this can be generalized to define $t_i(F, W)$.

GRAPHONS AND GRAPH LIMITS

Basic idea: **"pixel pictures"**



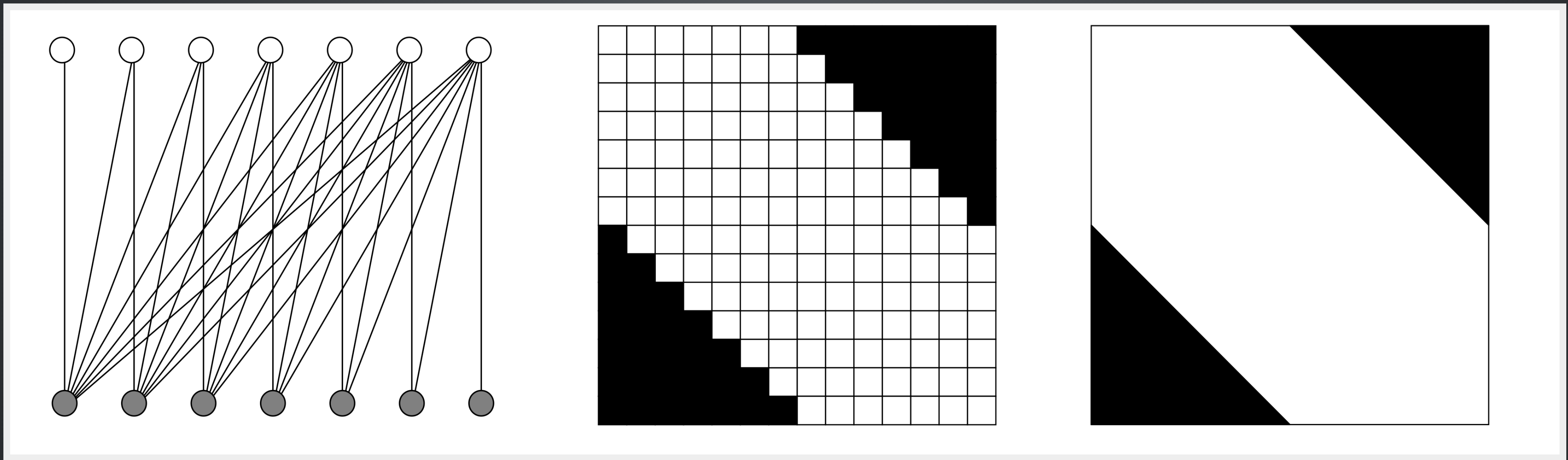
$$\begin{matrix}
0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0
\end{matrix}$$



from Lovasz (2012), Large networks and graph limits

GRAPHONS AND GRAPH LIMITS

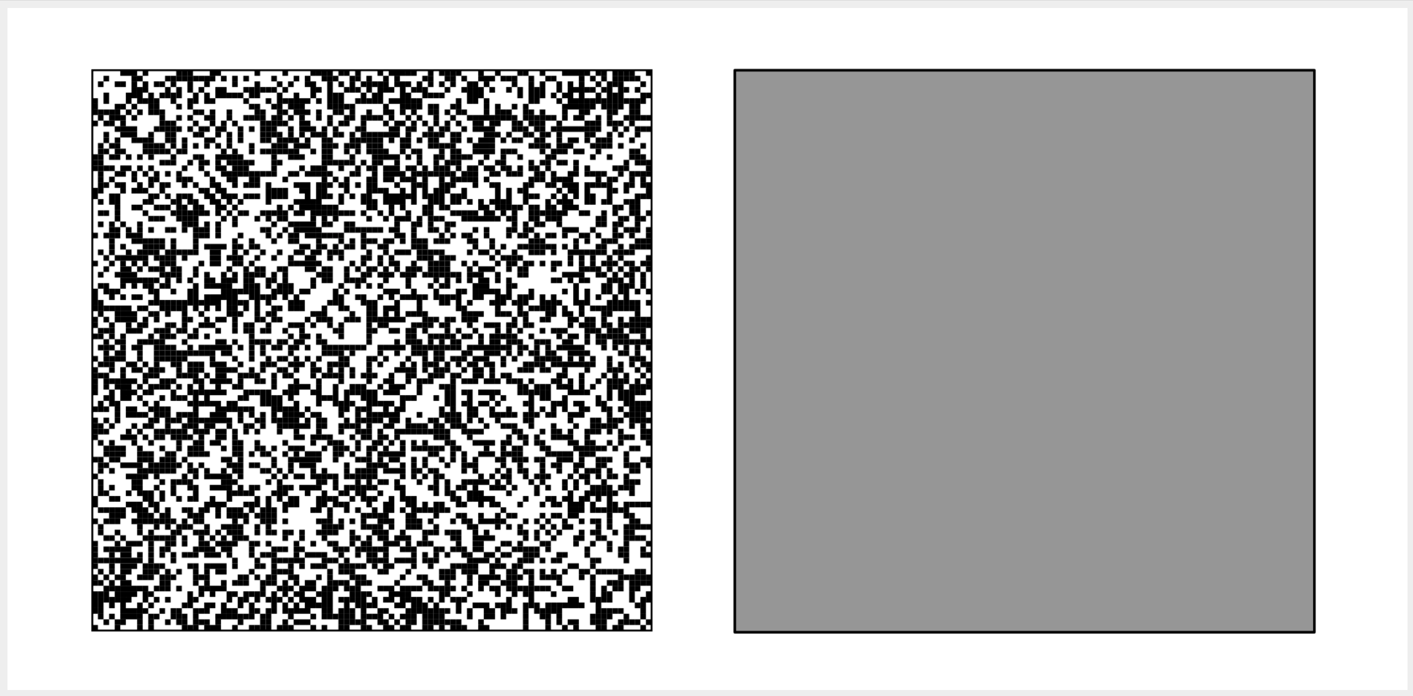
Convergence of pixel pictures



from Lovasz (2012), Large networks and graph limits

GRAPHONS AND GRAPH LIMITS

Convergence of pixel pictures



from Lovasz (2012), Large networks and graph limits

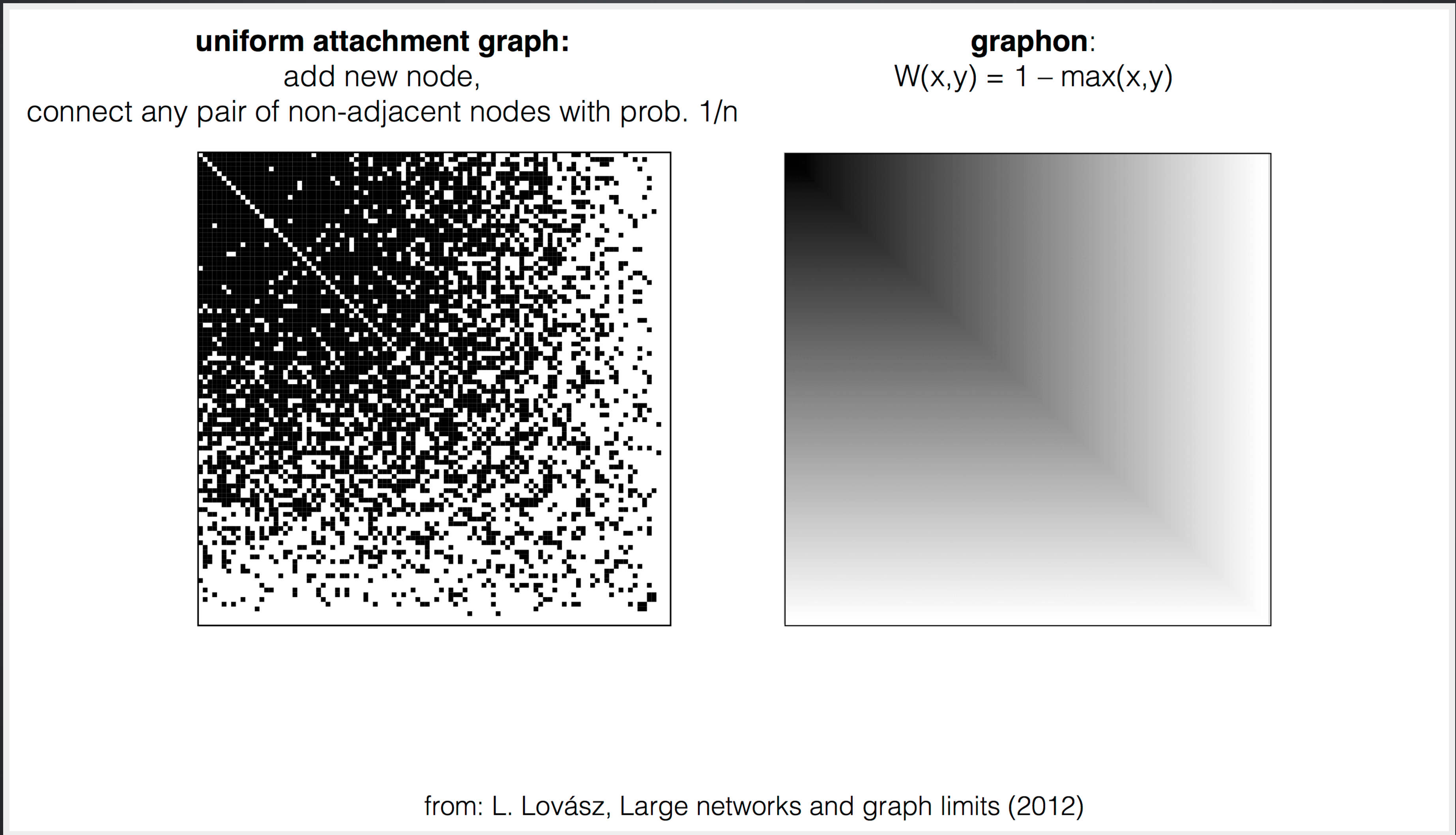
THE LIMIT GRAPHON

THM: For every convergent graph sequence (G_n) there exists (up to weak isomorphim) exactly one graphon W such that for all finite F:

$$t_i(F, G_n) \longrightarrow t_i(F, W).$$

THE LIMIT GRAPHON

Example: Uniform attachment graphs



from Lovasz (2012), Large networks and graph limits

from Lovász (2012), Large networks and graph limits

A COMPATIBLE METRIC

- *Edit distance:* $d_1(F,G) = \| A_F - A_G \|_1$.
- *Cut distance:* $d_\square(F,G) = \| A_F - A_G \|_\square$, where $\|.\|_\square$ is the **cut norm**

$$\|A\|_\square = \frac{1}{n^2} \max_{S,T \subseteq [n]} \Big| \sum_{i \in S, j \in T} A_{ij} \Big|.$$

- $d_\square$ can be extended to graphs of different order...
- ... and to graphons:

$$\|W\|_\square = \sup_{S,T \subseteq [0,1]} \int_{S \times T} W(x,y)\, dx\, dy.$$

A COMPATIBLE METRIC

- A sequence (G_n) converges iff it is a Cauchy sequence with respect to $d_\square$.
- $G_n \to W \quad \text{iff} \quad d_\square(G_n, W) \to 0$

SAMPLING FROM GRAPHONS

- We can obtain a finite graph $\mathbb{G}(n, W)$ from W by (independently) sampling n points $x_1, \ldots, x_n$ from W and filling edges according to probabilities $W(x_i, x_j)$.
  - almost surely, we get a sequence with $\mathbb{G}(n, W) \to W$.
- If we sample ω-many points from $W(x, y) \equiv 1/2$, we almost surely get the random graph.

THE PETROV-VERSHIK GRAPHON

- **Petrov and Vershik** (2010) constructed, for each $n \geq 3$, a graphon W such that we almost surely sample a Henson graph for n.
  - The graphons are (necessarily) {0,1}-valued.
  - Such graphons are called **random-free**.
  - The constructions resembles a finite extension construction with simple geometric forms, where each step satisfies a new type requiring attention.
  - The method can also be used to construct random-free graphons from which we sample the Rado graph.

INVARIANT MEASURES

- The Petrov-Vershik graphon also yields a measure on the set of countable infinite graphs concentrating on the set of universal, homogeneous K_n-free graphs.
- This measure will be invariant under the "logic action", the natural action of S_∞ on the space of countable (relational) structures with universe $\mathbb{N}$.
- This method was generalized by *Ackerman*, *Freer*, *and Patel* (2014) to other homogeneous structures.
- It can be used to define algorithmic randomness for such structures (as suggested by Nies and Fouché).

UNIVERSAL GRAPHONS

- A random-free graphon is *countably universal* if for every set of distinct points from $[0, 1], x_1, x_2, \ldots, x_n, y_1, \ldots, y_m$, the intersection
$$\bigcap_{i,j} E_{x_i} \cap E_{y_j}^C$$
has non-empty interior.
  - Here $E_x = \{y: W(x, y) = 1\}$ is the neighborhood of x.

- For *countably K_n-free universal* graphs, we require this to hold only for such tuples where the induced subgraph by the x_i has no induced K_{n-1}-subgraph,
  - additionally, require that there are no n-tuples in X which induce a K_n.

THE TOPOLOGY OF GRAPHONS

- Neighborhood distance:

  $$r_W(x, y) = \parallel W(x, .\,) - W(y, .\,) \parallel_1 = \int |W(x, z) - W(y, z)|$$

  and mod out by $r_W(x, y) = 0$.
- Example: $W(x, y) \equiv p$ is a singleton space.
- **THM: (Freer & R.)** (informal) If W is a random-free universal graphon obtained via a "tame" extension method, then W is no compact in the r_W topology.

"TAME" EXTENSIONS

- **DEF:** A random-free graphon W has *continuous realization of extensions* if there exists a function
$$f : (x_1, \ldots, x_n), (y_1, \ldots, y_m) \mapsto (l, r)$$
that is continuous a.e. such that for all $\vec{x}, \vec{y}$,
$$[l, r] \subseteq \bigcap_{i,j} E_{x_i} \cap E^C_{y_j}.$$
- Here $E_x = \{y \colon W(x, y) = 1\}$ is the neighborhood of x.
- The Petrov-Vershik graphons have uniformly continuous realization of extensions.

NON-COMPACTNESS

THM: *If a countably (K_n-free) universal graphon has uniformly continuous realization of extensions, then it is not compact in the r_W-topology.*

FEATURES OF THE PROOF

- Building a "Cantor sequence" in W.
- Apply the Szemeredi regularity lemma to pass to a sequence of stepfunctions that approximate the graphon *uniformly*.
- Use universality to find the next splitting.
- Uniform continuity guarantees that the Szemeredi "squares" are filled with the right measure.

REGULARITY LEMMA

- For every $\epsilon > 0$ there is an $S(\epsilon) \in \mathbb{N}$ such that every graph G with at least $S(\epsilon)$ vertices has an equitable partition of V into k pieces (1/ε ≤ k ≤ S(ε)) such that for all but ϵk^2 pairs of indices i, j, the bipartite graph $G[V_i, V_j]$ is ϵ-regular.
- For every graphon W and $k \geq 1$ there is stepfunction U with k steps such that

$$d_\square(W, U) < \frac{2}{\sqrt{\log k}} \parallel W\|_2$$

COMPLEXITY OF UNIVERSAL GRAPHONS

construction:	fully random	tame deterministic	general deterministic
complexity of graphon	low (singleton)	high (non-compact, infinite Minkowski dimension)	?

Also: Is there a robust notion of a stochastic graphon?