On a Question of Vera T. Sós About Size Forcing of Graphons

Oliver Cooley[1(✉)], Mihyun Kang[1], and Oleg Pikhurko[2]

[1] Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria
{cooley,kang}@math.tugraz.at

[2] Mathematics Institute and DIMAP, University of Warwick, Coventry CV4 7AL, UK
O.Pikhurko@warwick.ac.uk

Abstract. The *k-sample* $\mathbb{G}(k,W)$ from a graphon $W:[0,1]^2\to[0,1]$ is the random graph on $\{1,\dots,k\}$, where we sample $x_1,\dots,x_k\in[0,1]$ uniformly at random and make each pair $\{i,j\}\subseteq\{1,\dots,k\}$ an edge with probability $W(x_i,x_j)$, with all these choices being mutually independent. Let the random variable $X_k(W)$ be the number of edges in $\mathbb{G}(k,W)$.

Vera T. Sós asked in 2012 whether two graphons U,W are necessarily weakly isomorphic provided the random variables $X_k(U)$ and $X_k(W)$ have the same distribution for every integer $k\geqslant 2$. This question when one of the graphons W is a constant function was answered positively by Endre Csóka and independently by Jacob Fox, Tomasz Łuczak and Vera T. Sós. Here we investigate the question when W is a 2-step graphon and prove that the answer is positive for a 3-dimensional family of such graphons.

We also present some related results.

Keywords: Graphons · Weak isomorphism · Sample

1 Introduction

The *k-sample* $\mathbb{G}(k,W)$ from a graphon W (i.e. a measurable symmetric function $[0,1]^2\to[0,1]$) is the random graph on $[k]:=\{1,\dots,k\}$ obtained by sampling $x_1,\dots,x_k\in[0,1]$ uniformly at random and making each pair $\{i,j\}\subseteq[k]$ an edge with probability $W(x_i,x_j)$, with all these choices being mutually independent. The *(homomorphism) density* $t(F,W)$ of a graph F on $[k]$ in W is the probability that $E(F)\subseteq E(\mathbb{G}(k,W))$, that is, every adjacent pair in F is also adjacent in $\mathbb{G}(k,W)$. Equivalently, $t(F,W):=\int_{[0,1]^k}\prod_{\{i,j\}\in E(F)}W(x_i,x_j)\,\mathrm{d}x_1\dots\mathrm{d}x_k$. Let us call two graphons U and W *weakly isomorphic* if the random graphs

Oliver Cooley and Mihyun Kang were supported by Austrian Science Fund (FWF) Grant I3747.
Oleg Pikhurko was supported by Leverhulme Research Project Grant RPG-2018-424.

J. Nešetřil et al. (Eds.): *Extended Abstracts EuroComb 2021*, TM 14, pp. 625–630, 2021.
https://doi.org/10.1007/978-3-030-83823-2_100

$\mathbb{G}(k, U)$ and $\mathbb{G}(k, W)$ have the same distribution for every $k \in \mathbb{N}$. This is equivalent to $t(H, U) = t(H, W)$ for every connected graph H. Borgs, Chayes and Lovász [1] showed that all graphons in the weak isomorphism class of W can, roughly speaking, be obtained from W by applying measure-preserving transformations of the variables.

A *graphon parameter* f is a function that assigns to each graphon W a real number or a real vector $f(W)$ such that $f(W) = f(U)$ whenever U and W are weakly isomorphic. We say that a family $(f_i)_{i \in I}$ of graphon parameters *forces* a graphon W if every graphon U with $f_i(U) = f_i(W)$ for every $i \in I$ is weakly isomorphic to W. For example, the famous result of Chung, Graham and Wilson [2] on p-quasirandom graphs states, in this language, that the constant-p graphon is forced by $t(K_2, \cdot)$ and $t(C_4, \cdot)$, i.e. by the edge and 4-cycle densities.

Call a family $(f_i)_{i \in I}$ of graphon parameters *forcing* if it forces every graphon W. For example, the densities $t(F, \cdot)$, where F ranges over all connected graphs, form a forcing family. The authors are not aware of any results where a substantially smaller set of parameters than the densities of all connected graphs is shown to be forcing. Vera T. Sós [9] posed some questions in this direction, and in particular considered the following problem. For a graphon W and an integer $k \in \mathbb{N}$, let $X_k(W) := |E(\mathbb{G}(k, W))|$ be the size of, i.e. number of edges in, the k-sample $\mathbb{G}(k, W)$ from W. We identify the random variable $X_k(W)$ with the vector of probabilities $\mathbb{P}(X_k(W) = i)$ for $0 \leqslant i \leqslant \binom{k}{2}$, viewing it as a graphon parameter. Let $\mathcal{W}_S$ be the family of graphons W that are forced by the sequence $(X_k(W))_{k \in \mathbb{N}}$, i.e. by the distributions of sizes of samples from W.

Question 1 (Size Forcing Question (Sós [9])). *Is every graphon in $\mathcal{W}_S$?*

Alon (unpublished, see [4]) and independently Sliacan [8] proved that the constant-$\frac{1}{2}$ graphon is in the family $\mathcal{W}_S$. Then Csóka [4] and independently Fox, Łuczak and Sós [5] proved that constant-p graphon is in the family $\mathcal{W}_S$ for any $p \in (0, 1)$. A natural next step would be to try to determine whether $W \in \mathcal{W}_S$ when W is a *2-step graphon*, i.e. we have a partition of $[0, 1]$ into measurable sets A and B such that W is constant on each of the sets A^2, B^2 and $(A \times B) \cup (B \times A)$. We need four parameters to describe a 2-step graphon: the measure of A as well as the three possible values of W. We can prove that $W \in \mathcal{W}_S$ for the following 3-dimensional subset of 2-step graphons.

Theorem 1. *Let W be the 2-step graphon with parts $A := [0, a)$ and $B := [a, 1]$ such that its values on A^2, $(A \times B) \cup (B \times A)$ and B^2 are respectively 0, $p \in (0, 1]$ and $q \in (0, 1]$. If $(1 - a)q \leqslant (1 - 2a)p$, then $W \in \mathcal{W}_S$.*

We can also answer Question 1 for some other families of 2-step graphons. We present two further examples (Theorems 2 and 3) where a *finite* set of some natural real-valued parameters suffices. The first is motivated by the result of Csóka [4] who in fact proved that the constant-p graphon is forced by X_4 alone.

Theorem 2. *Let $p \in [0, 1]$ and let W be the graphon which is 0 on $[0, 1/2)^2 \cup [1/2, 1]^2$ and p everywhere else. Then W is forced by X_5 alone.*

Let the *independence ratio* $\alpha(W)$ of a graphon W be the supremum of the measure of $A \subseteq [0,1]$ such that $W(x,y) = 0$ for a.e. $(x,y) \in A^2$. As was observed by Hladký, Hu and Piguet [6, Lemma 2.4], the supremum is in fact a maximum (that is, it is attained by some A). Also, the *clique ratio* $\omega(W) := \alpha(1-W)$ is the maximum measure of $A \subseteq [0,1]$ with W being 1 a.e. on A^2.

Theorem 3. *Given $a, p \in [0,1]$, set $A := [0,a)$ and $B := [a,1]$, and let W be the graphon which is 0 on A^2, 1 on B^2, and p everywhere else. Then W is forced by (α, ω, X_4).*

By using a basic version of the container method, we show that the value of α (and thus of ω) is determined by any infinite subsequence of $(X_k)_{k\in\mathbb{N}}$. More precisely, the following holds.

Theorem 4. $\alpha(W) = \lim_{k\to\infty} \left(\mathbb{P}(X_k(W)=0)\right)^{1/k}$ *for every graphon W.*

We note that Theorem 4, by relating $\alpha(W)$ to graph densities, fills one missing entry in [7, Table 1].

By combining Theorems 3 and 4, we directly obtain the following result.

Corollary 1. *Let W be a graphon as in Theorem 3 (that is, W is 0 on $[0,a)^2$, 1 on $[a,1]^2$, and p everywhere else). Then $W \in \mathcal{W}_S$.* $\square$

Call a family $\mathcal{F}$ of graphs *forcing* if the corresponding family of parameters $(t(F,\cdot))_{F\in\mathcal{F}}$ is forcing. Sós [9] also asked if one can find substantially smaller forcing families than taking all connected graphs. We show that two natural examples, namely the family of all cycles and the family of all complete bipartite graphs, do not suffice.

Proposition 1. *(i) The family of all connected graphs with at most one cycle is not forcing. In particular, the family of all cycles is not forcing.*

(ii) For every integer d, the family of all graphs of diameter at most d is not forcing. In particular, the family of all complete bipartite graphs is not forcing.

Full proofs of all the results stated in this extended abstract can be found in [3].

2 Some Auxiliary Results

We first present some known or easy auxiliary results that we need for the proofs of the main results. For graphs H_1, H_2, let $H_1 \sqcup H_2$ denote their *disjoint union*. Let $\mathcal{G}_{k,m}$ consist of isomorphism classes of all graphs with at most k vertices and exactly m edges that do not contain any isolated vertices. For example, $\mathcal{G}_{5,3} = \{K_3, P_4, P_3 \sqcup K_2, K_{1,3}\}$.

Lemma 1. *(i) For any graphon W and for any graphs H_1 and H_2, we have $t(H_1 \sqcup H_2, W) = t(H_1, W)\, t(H_2, W)$.*

(ii) If W is a p-regular graphon and F' is obtained from a graph F by attaching a pendant edge then $t(F', W) = p\, t(F, W)$.

The following result implicitly appears in Csóka [4] (see also [3, Lemma 12]). Let $(r)_k := r(r-1)\dots(r-k+1)$ denote the falling factorial.

Lemma 2. *Let integers k and m satisfy $1 \leqslant m \leqslant \binom{k}{2}$. Then for every graphon W we have*

$$\mathbb{E}\left((X_k(W))_m\right) = \sum_{F \in \mathcal{G}_{k,m}} c_{k,F}\, t(F,W),$$

where $c_{k,F} > 0$ is the number of graphs on $[k]$ that, after discarding isolated vertices, are isomorphic to F.

A useful consequence, which we will apply frequently, is that if two graphons U, W have k-samples $X_k(U), X_k(W)$ with identical distributions, and if $t(F,U) = t(F,W)$ for all $F \in \mathcal{G}_{k,m}$ except for some F_0, then the F_0-densities are also equal.

We will also need the following bipartite analogue of the Chung-Graham-Wilson Theorem, which can be proved either by passing to finite graphs converging to U (and adapting the original proof of Chung, Graham and Wilson [2]) or, using analytic methods, by dealing directly with graphons (see [3, Lemma 14]).

Lemma 3. *Let A and B be sets of measure a and b respectively that partition $[0,1]$. (Thus $a+b=1$.) Let $p \in [0,1]$. Let U be a graphon taking value 0 on $A^2 \cup B^2$ such that $t(K_2,U) = 2abp$ and $t(C_4,U) = 2a^2b^2p^4$. Then $U(x,y) = p$ for a.e. $(x,y) \in (A \times B) \cup (B \times A)$.*

The following result can be proved using the container method (see [3, Theorem 15]). Let $\mathcal{I}(G)$ denote the family of all independent sets in a graph G and let $\mathcal{I}_k(G) := \{I \in \mathcal{I}(G) : |I| = k\}$ consist of all independent sets of size k.

Theorem 5. *For every $\delta > 0$ there exists $\varepsilon > 0$ such that for any $k \geqslant 1/\varepsilon$ there exists n_0 such that for every graph G on $n \geqslant n_0$ vertices and every real α, if $|\mathcal{I}_k(G)| \geqslant (\alpha - \varepsilon)^k \binom{n}{k}$, then there exists $A \subseteq V(G)$ with $|A| \geqslant (\alpha - \delta)n$ and $e(G[A]) \leqslant \delta n^2$.*

3 Proof Outlines of Main Results

Proof of Theorem 2. Let U be an arbitrary graphon such that the distribution of $X_5(U)$ is the same as the distribution of $X_5(W)$. Let us denote this common distribution by X_5. The aim is to successively prove the following properties of U; each step is individually relatively easy to prove given the previous properties.

- U is $(p/2)$-regular, i.e. $\deg^U(x) := \int_0^1 U(x,y)\,\mathrm{d}y = p/2$ for a.e. $x \in [0,1]$.
- If $t(H,U) = t(H,W)$ for some graph H, then $t(H',U) = t(H',W)$ for any graph H' that is obtained from H by adding a pendant edge.
- $t(H,U) = t(H,W)$, where H is any one of $K_{1,3}, P_4, P_2 \sqcup K_2, K_3, C_4$, C_4 with a pendant edge, $C_5, K_{2,3}$.
- Let the random variable Z be $\mathrm{codeg}^U(x,y) := \int_0^1 U(x,z)U(z,y)\,\mathrm{d}z$, the density of copies of P_2 which have x,y as endpoints, where x and y are chosen uniformly and independently from $[0,1]$. Then $\mathbb{P}(Z=0) = \mathbb{P}(Z = p^2/2) = \frac{1}{2}$.
- Let C consist of those $(x,y) \in [0,1]^2$ for which $\mathrm{codeg}^U(x,y) = p^2/4$ and let $\deg_C(x)$ denote the measure of $N_C(x) := \{y : (x,y) \in C\}$, for $x \in [0,1]$. Then $\deg_C(x) = 1/2$ for a.e. $x \in [0,1]$.
- For a.e. $x \in [0,1]$, the set $N_C(x)$ is independent in U.
- U is weakly isomorphic to W.

Proof of Theorem 4. For $k \in \mathbb{N}$, let $\alpha_k(W) := \mathbb{P}(X_k(W) = 0)$. It is easy to see that the limit $\alpha_\infty(W) := \lim_{k\to\infty}(\alpha_k(W))^{1/k}$ exists. Clearly, $\alpha_\infty(W)$ remains the same if we replace W by any weakly isomorphic graphon.

The inequality $\alpha_\infty(W) \geqslant \alpha(W)$ is easy to prove by picking an independent set $A \subseteq [0,1]$ in W of measure $\lambda(A) = \alpha(W)$ (which exists by [6, Lemma 2.4]) and observing that $\Pr(X_k(W) = 0) \geqslant \lambda(A)^k$.

To show the converse inequality, we pick sufficiently large $k \ll n$ (so in particular $\alpha_k(W) \approx \alpha_\infty(W)$) and let $G \sim \mathbb{G}(n, W)$ be the n-sample from W. We consider the step graphon W_G encoding the adjacency relation in G. In an appropriate sense, a typical outcome W_G is "close" to W, and therefore it suffices to show that G contains an almost independent set of size close to $\alpha_k(W) \cdot n$, which will then transfer to an independent set in W of size close to $\alpha_k(W)$. The existence of this almost independent set can be proved by applying Theorem 5.

Proof of Theorem 1. Let U be an arbitrary graphon such that for every $k \in \mathbb{N}$ the distributions of $X_k(U)$ and $X_k(W)$ are the same; let us denote this random variable by X_k.

By Theorem 4, U contains an independent set of measure a, which we may assume is $A = [0, a)$.

We next claim that for almost every $x \in B$, we have $\deg_A^U(x) \geqslant ap$. This can be proved by contradiction: if there is a set $B' \subset B$ of measure at least ε such that each point in B' has A-degree at most $ap - \varepsilon$, then some careful calculation shows that $\alpha_k(U) \geqslant \alpha_k(W)$ for sufficiently large k, which is a contradiction.

Let U' be the graphon obtained from U by averaging it over $(A\times B)\cup(B\times A)$ and over B^2. A further averaging argument considering the density of P_3 and applying the Cauchy-Schwarz inequality shows that in fact $U' = W$.

Next, we show that for almost every $(x, y) \in B^2$, we have that $\mathrm{codeg}_A^U(x, y) = ap^2$ where we denote $\mathrm{codeg}_A^U(x, y) := \int_A U(x, z)U(y, z)\,\mathrm{d}z$. If this were not true, then some careful calculation shows that $\mathbb{P}(X_k(U) \leqslant 1) > \mathbb{P}(X_k(W) \leqslant 1)$ for sufficiently large k, which is a contradiction.

We next deduce that the triangle densities in U, W are identical even if we specify which vertices are in A and B, and that the same is true for triangles with a pendant edge. It follows from Lemma 2 that U and W have the same density of 4-cycles. Since due to codegree considerations U and W have the same densities of $ABAB$-cycles, Lemma 3 implies that U is constant p on $A \times B$. Finally, since we know the densities of all types of 4-cycles except those lying inside B, we also know the density of these 4-cycles, and the Chung-Graham-Wilson Theorem implies that U is also constant on B, so $U = W$.

Proof of Theorem 3. There are subsets $C, D \subseteq [0,1]$ of measures a and $1-a$ respectively such that U is 0 on C^2 a.e. and U is 1 on D^2 a.e., and we may assume that $C = A$ and $D = B$.

We first show that $\deg_B^U(x) = (1-a)p$ for a.e. $x \in A$ by showing that, viewed as a random variable when x is chosen uniformly at random from A, the first and third moments are $(1-a)p$ and $((1-a)p)^3$ respectively, which is only possible if the random variable is constant $(1-a)p$ a.e. on A. Similarly we can show that $\deg_A^U(x) = ap$ for a.e. $x \in B$.

Let K_4^- be the 4-clique minus an edge, the unique graph on 4 vertices with 5 edges. A 4-sample from U can only form a K_4^- if it has either two or three vertices in B. Since we know the density of the second type, and the densities of K_4^- in U and W are identical, we can also deduce the density of the second type, which exactly matches the density of $ABAB$-cycles in both U and W. It follows that U is constant a.e. on $A \times B$, and therefore weakly isomorphic to W.

Proof of Proposition 1. To prove (i), take the unit vectors

$$\boldsymbol{x}_1 := \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}, \quad \boldsymbol{x}_2 := \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\0\end{pmatrix}, \quad \text{and} \quad \boldsymbol{x}_3 := \frac{1}{\sqrt{6}}\begin{pmatrix}2\\-1\\-1\end{pmatrix},$$

let $\varepsilon := 1/4$ and set

$$A := \boldsymbol{x}_1\boldsymbol{x}_1^T + \varepsilon\,\boldsymbol{x}_2\boldsymbol{x}_2^T \quad \text{and} \quad A' := \boldsymbol{x}_1\boldsymbol{x}_1^T + \varepsilon\,\boldsymbol{x}_3\boldsymbol{x}_3^T.$$

Let W and W' be the 3-step graphons, with steps of measure $1/3$, whose values are given by the symmetric matrices $A, A' \in [0,1]^{3\times 3}$. It is simple to calculate that W, W' have the same densities of k-cycles for every k, and indeed the same densities of all unicyclic graphs, but are not weakly isomorphic since their limiting density of K_k as $k \to \infty$ is different.

To prove (ii), let $G := P_{d+2} \sqcup P_{d+2}$ and $G' := P_{d+3} \sqcup P_{d+1}$, and let W, W' be the step graphons with $2d+4$ steps of equal measure encoding their adjacency relations. They are not weakly isomorphic because the induced density of P_{d+3} is zero in W but not in W', but their densities of any graph of diameter at most d are identical, so this family is not forcing.

References

1. Borgs, C., Chayes, J., Lovász, L.: Moments of two-variable functions and the uniqueness of graph limits. Geom. Func. Anal. **19**, 1597–1619 (2010)
2. Chung, F.R.K., Graham, R.L., Wilson, R.M.: Quasi-random graphs. Combinatorica **9**, 345–362 (1989)
3. Cooley, O., Kang, M., Pikhurko, O.: On a question of Vera T. Sós about size forcing of graphons. arXiv:2103.09114
4. Csóka, E.: On the graph limit question of Vera T. Sós. J. Combin. Theory (B) **116**, 306–311 (2016)
5. Fox, J.: Personal communication
6. Hladký, J., Hu, P., Piguet, D.: Komlós's tiling theorem via graphon covers. J. Graph Theory **90**, 24–45 (2019)
7. Hladký, J., Rocha, I.: Independent sets, cliques, and colorings in graphons. Eur. J. Comb. **88**, 103108 (18 p.) (2020)
8. Sliacăn, J.: Flag algebras and flagmatic. M.Phil. thesis, University of Warwick (2015)
9. Sós, V.T.: Some open problems on dense graphs. In: Conference on Graphs and Analysis, IAS, Princeton (2012)