Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

On the graph limit question of Vera T. Sós

Endre Csóka*

Abstract

In the dense graph limit theory, the topology of the set of graphs is defined by the distribution of the subgraphs spanned by finite number of random vertices. Vera T. Sós proposed a question that if we consider only the number of edges in the spanned subgraphs, then whether it provides an equivalent definition. We show that the answer is positive on quasirandom graphs.

1 Introduction

Graphon is a symmetric measurable function $W : [0,1]^2 \to [0,1]$, introduced in the context of graph theory by Lovász and Szegedy [2, 3]. A sampling $\mathbb{G}(n, W)$ is a distribution of n-vertex graphs constructed as follows. We choose $x_1, x_2, ..., x_n \in [0,1]^n$ uniformly at random, and we connect the ith and jth vertices with probability $W(x_i, x_j)$. Therefore, the probability of each n-vertex graph $F = ([n], E(F))$ in the sampling $\mathbb{G}(n, W)$ is

$$\mathbb{G}(n,W)(F) = \int\limits_{[0,1]^n} \prod_{\{i,j\}\in E(F)} W(x_i,x_j) \prod_{\{i,j\}\in \overline{E(F)}} \big(1 - W(x_i,x_j)\big) \ \mathrm{d}x.$$

A basic theorem about graphons tells that the sequence of samplings $\mathbb{G}(n, W)$ determine the graphon W up to weak isomorphism [2]. Vera T. Sós asked the question whether the distribution of the number of edges instead of the spanned subgraph already has this property [4]. Formally, let

$$\mathbb{N}(n,W)(k) = \sum_{F\colon |E(F)|=k} \mathbb{G}(n,W)(F).$$

Question 1. *Does the sequence $\big(\mathbb{N}(n,W)\big)_{n\in\mathbb{N}}$ determine the sequence $\big(\mathbb{G}(n,W)\big)_{n\in\mathbb{N}}$? In other words, does the sequence $\mathbb{N}(n,W)$ determine the graphon W up to weak isomorphism?*

Motivated by this question, Svante Janson (personal communication) asked the following question.

Question 2. *Does $\mathbb{N}(4,W) = \mathbb{N}(4,\frac{1}{2})$ imply $W \equiv \frac{1}{2}$, i.e. W is constant $\frac{1}{2}$ almost everywhere?*

First, Jacob Fox proved the positive answer for all $p \in [0,1]$ not only for $p = \frac{1}{2}$ in an unpublished result. At the same time, Noga Alon also proved for $p = \frac{1}{2}$. Later but independently, Jakub Sliacan proved it by flag algebra. In our paper we show a simple combinatorial proof of a stronger version of the statement. Namely, we show that 4 can be exchanged to any larger number, moreover, the sampling could be exchanged to any graph-sampling which includes a 4-cycle C_4.

*Institute of Mathematics and DIMAP, University of Warwick, Coventry, United Kingdom. Supported by ERC grant 306493

Let subgraph mean edge-subgraph, namely we keep the vertex set, but we take a subset of its edges. For a finite graph $S = ([n], E(S))$, we define the *sampling* of a graphon W by a graph S to be the following distribution $\mathbb{G}(S, W)$ of the subgraphs of S.

$$\mathbb{G}(S,W)(F) = \int\limits_{[0,1]^n} \prod_{\{i,j\}\in E(F)} W(x_i,x_j) \prod_{\{i,j\}\in E(S)\setminus E(F)} \big(1-W(x_i,x_j)\big) \;\mathrm{d}x.$$

Analogously,

$$\mathbb{N}(S,W)(k) = \sum_{F\colon |E(F)|=k} \mathbb{G}(S,W)(F).$$

Notice that $\mathbb{G}(n,W) = \mathbb{G}(K_n,W)$ and $\mathbb{N}(n,W) = \mathbb{N}(K_n,W)$ (where K_n is the complete graph on n vertices).

Theorem 3. *Let G be a graph and W a graphon. Assume that G contains a C_4 and $\mathbb{N}(G,W) = \mathbb{N}(G,p)$. Then $W \equiv p$.*

2 Proof of Theorem 3

Let S_k, P_k and C_k denote the star, the path and the cycle containing k edges, respectively. Let $|E(G)| = m$, and we define homomorphism density of F to W as

$$t(F,W) = \int\limits_{[0,1]^{V(F)}} \prod_{\{i,j\}\in E(F)} W(x_i,x_j) \;\mathrm{d}x. \tag{1}$$

Lemma 4. *For an arbitrary positive integer $k \le m$, let G_k denote the uniform random subgraph of G with k edges (namely, all $\binom{m}{k}$ subgraphs have equal probability). Then*

$$\mathsf{E}\big(t(G_k,W)\big) = p^k. \tag{2}$$

Proof. On the left hand side, there are two randomnesses: the choice of the subgraph G_k and the sampling. But if we do the sampling first and we choose the subgraph after, then we get the right hand side. Formally,

$$\mathsf{E}\big(t(G_k,W)\big) \overset{(1)}{=} \mathsf{E}\Bigg(\int\limits_{[0,1]^{V(G)}} \prod_{\{i,j\}\in E(G_k)} W(x_i,x_j) \;\mathrm{d}x\Bigg) = \int\limits_{[0,1]^{V(G)}} \mathsf{E}\Bigg(\prod_{\{i,j\}\in E(G_k)} W(x_i,x_j)\Bigg) \;\mathrm{d}x. \tag{3}$$

For a fixed $x \in [0,1]^{V(G)}$, let $X : E(G) \to \{0,1\} = \{false, true\}$ be independent events with probabilities $\mathsf{P}\big(X(\{i,j\})\big) = W(x_i,x_j)$ for all $\{i,j\} \in E(G)$. For a graph F, let us denote the number of occurring events by $\mu(F) = \sum\limits_{(a,b)\in E(F)} X(a,b)$.

$$\mathsf{E}_{G_k}\Big(\prod_{\{i,j\}\in E(G_k)} W(x_i,x_j)\Big) = \mathsf{P}_{G_k,X}\Big(\bigwedge_{\{i,j\}\in E(G_k)} X\big(\{i,j\}\big)\Big)$$

$$= \mathsf{E}_X \mathsf{P}_{G_k}\Big(\bigwedge_{\{i,j\}\in E(G_k)} X\big(\{i,j\}\big)\Big) = \mathsf{E}_X\left(\frac{\binom{\mu(G)}{k}}{\binom{m}{k}}\right) \tag{4}$$

Therefore,

$$\mathsf{E}\big(t(G_k,W)\big) \overset{(3)(4)}{=} \int\limits_{[0,1]^{V(G)}} \mathsf{E}_X\left(\frac{\binom{\mu(G)}{k}}{\binom{m}{k}}\right) \mathrm{d}x = \mathsf{E}_{x,X}\left(\frac{\binom{\mu(G)}{k}}{\binom{m}{k}}\right). \tag{5}$$

The distibution of $\mu(G)$ is $\mathbb{N}(G,W)$ by definition. We can make the same calculation with m independent edges $m \times P_1$ instead of G, which provides that

$$t(k \times P_1, W) = \mathsf{E}\Big(t\big((m \times P_1)_k, W\big)\Big) = \mathsf{E}_{x,X}\left(\frac{\binom{\mu(m\times P_1)}{k}}{\binom{m}{k}}\right). \tag{6}$$

$\mathbb{N}(G,W) = \mathbb{N}(G,p) =$ binomial distribution $B(m,p) = \mathbb{N}(m \times P_1, W)$, hence, $\mu(G) = \mu(m \times P_1)$. Therefore,

$$\mathsf{E}\big(t(G_k, W)\big) \overset{(5)}{=} \mathsf{E}_{x,X}\left(\frac{\binom{\mu(G)}{k}}{\binom{m}{k}}\right) = \mathsf{E}_{x,X}\left(\frac{\binom{\mu(m\times P_1)}{k}}{\binom{m}{k}}\right) \overset{(6)}{=} t(k \times P_1, W) = t(P_1, W)^k = p^k. \qquad \square$$

Lemma 5.

$$t(S_2, W) = p^2 \tag{7}$$

Proof. We apply Lemma 4 with $k = 2$. The support of G_2 consists of two (isomorphism classes of) graphs: two independent edges $2 \times P_1$ and S_2. Therefore, for some $\lambda > 0$,

$$p^2 = \mathsf{E}\big(t(G_2, W)\big) = \lambda \cdot t(S_2, W) + (1-\lambda) \cdot t(2 \times P_1, W) = \lambda \cdot t(S_2, W) + (1-\lambda)p^2,$$

which implies, $t(S_2, W) = p^2$. $\square$

The *degree* of a vertex $x \in [0,1]$ of a graphon W is defined as follows.

$$deg(x) = \int_0^1 W(x,y)\ \mathrm{d}y$$

Note that W is measurable, therefore, $deg(x)$ exists for almost all vertices $x \in [0,1]$.

Lemma 6. *Almost all degrees of W are p.*

Proof.

$$\mathsf{Var}\big(\deg(W)\big) = \mathsf{E}\big(\deg(W)^2\big) - \mathsf{E}\big(\deg(W)\big)^2 = t(S_2, W) - p^2 = 0. \qquad \square$$

Lemma 7. *Let F be an arbitrary graph and F' be its extension by one new vertex v and a new edge (v,w) connecting v to an arbitrary old vertex. Then*

$$t(F', W) = p \cdot t(F, W). \tag{8}$$

Proof. In short, whatever we sample by F, the probability that (v,w) maps to an edge in W is p, because all degrees are p. Formally,

$$t(F', W) \overset{(1)}{=} \int\limits_{[0,1]^{V(F')}} \prod_{\{i,j\} \in E(F')} W(x_i, x_j)\ \mathrm{d}x = \int\limits_{[0,1]^{V(F)}} \prod_{\{i,j\} \in E(F)} W(x_i, x_j) \int\limits_{[0,1]} W(x_v, x_w)\ \mathrm{d}x_v\ \mathrm{d}x_{V(F)}$$

$$= \int\limits_{[0,1]^{V(F)}} \prod_{\{i,j\} \in E(F)} W(x_i, x_j) \cdot \deg(w)\ \mathrm{d}x_{V(F)} = p \cdot \int\limits_{[0,1]^{V(F)}} \prod_{\{i,j\} \in E(F)} W(x_i, x_j)\ \mathrm{d}x \overset{(1)}{=} p \cdot t(F, W). \qquad \square$$

Lemma 8. $t(P_3, W) = t(S_3, W) = p^3$.

Proof. We apply Lemma 7 for $F = P_2 = S_2$ and $F' = P_3$ or $F' = S_3$, namely,

$$t(F', W) \overset{(8)}{=} p \cdot t(S_2, W) \overset{(7)}{=} p \cdot p^2 = p^3. \qquad \square$$

Lemma 9. *If G contains a triangle K_3, then $t(K_3, W) = p^3$.*

Proof. We apply Lemma 4 with $k = 3$. G_3 may contain only the following graphs: $3P_1$, $P_2 \sqcup P_1$, P_3, S_3 and K_3. G contains a K_3, therefore, $\mathsf{P}(G_3 = K_3) > 0$. We already know that $t(3 \times P_1, W) = t(P_2 \sqcup P_1, W) = t(P_3, W) = t(S_3, W) = p^3$, therefore, (2) implies that $t(K_3, W) = p^3$, as well. □

Lemma 10. $t(C_4, W) = p^4$.

Proof. We apply Lemma 4 with $k = 4$. Using the previous lemmas and applying Lemma 7, we see that for all subgraphs with 4 edges except C_4, the homomorphism densities are p^4. This implies $t(C_4, W) = p^4$. □

The theorem of Chung, Graham and Wilson [1], in the language of graphons [2] shows that if $t(C_4, W) = t(P_1, W)^4 = p^4$, then $W \equiv p$. □

3 Acknowledgement

I want to say thank you to Oleg Pikhurko for suggesting me this question and for useful comments on the paper.

References

[1] Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[2] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[3] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[4] Vera Sós. Some open problems on dense graphs. *Conference on Graphs and Analysis, IAS, Princeton*, 2012.