On the Typical Structure of Graphs in a Monotone Property

Svante Janson* Andrew J. Uzzell†

Department of Mathematics
Uppsala University
P.O. Box 480
SE-751 06 Uppsala, Sweden
{svante.janson,andrew.uzzell}@math.uu.se

Submitted: Apr 9, 2014; Accepted: Aug 20, 2014; Published: Aug 28, 2014
Mathematics Subject Classifications: 05C75, 05C80, 05C30

Abstract

Given a graph property $\mathcal{P}$, it is interesting to determine the typical structure of graphs that satisfy $\mathcal{P}$. In this paper, we consider monotone properties, that is, properties that are closed under taking subgraphs. Using results from the theory of graph limits, we show that if $\mathcal{P}$ is a monotone property and r is the largest integer for which every r-colorable graph satisfies $\mathcal{P}$, then almost every graph with $\mathcal{P}$ is close to being a balanced r-partite graph.

Keywords: Graph limits; Monotone properties; Structure of graphs

1 Introduction and main results

Given a graph property $\mathcal{P}$, it is natural to study the structure of a typical graph that satisfies $\mathcal{P}$. A graph property is *monotone* if it is closed under taking subgraphs and *hereditary* if it is closed under taking induced subgraphs. Thus, every monotone property is also hereditary. Many authors have studied the structure of typical graphs in various hereditary properties—see, e.g., [1, 3, 7, 8, 9, 11], as well as the survey [4]. In this note, we use results from graph limit theory to study the structure of a typical graph in a general monotone property.

Before stating our main result, let us recall certain basic notions of graph limit theory. For more details, see, e.g., [5, 6, 13], as well as the monograph [12]. Here, we simply recall that certain sequences of graphs are defined to be *convergent*. A convergent sequence has a limit, called a *graph limit*, which is unique if it exists.

*Partly supported by the Knut and Alice Wallenberg Foundation.
†Supported by the Knut and Alice Wallenberg Foundation.

Lovász and Szegedy [13] showed that a graph limit Γ may be represented by a *graphon*, a symmetric, measurable function $W : [0,1]^2 \to [0,1]$. (So, abusing notation slightly, we will sometimes write $G_n \to W$ if the sequence $\{G_n\}_{n=1}^{\infty}$ converges to the graph limit Γ represented by W.) More than one graphon may represent the same graph limit; we say that the graphons W_1 and W_2 are *equivalent*, and write $W_1 \cong W_2$, if they represent the same graph limit.

Let $X_1,\ X_2, \ldots$ be i.i.d. uniformly distributed random variables in $[0,1]$. Given a graphon W, the *W-random graph* $G(n,W)$ is a graph with vertex set $[n]$ in which vertices i and j are adjacent with probability $W(X_i,X_j)$, independently of all other edges.

Let $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ denote the binary entropy function. The *entropy* of a graphon W is

$$\mathrm{Ent}(W) = \int_0^1 \int_0^1 h\big(W(x,y)\big)\,\mathrm{d}\mu(x)\mathrm{d}\mu(y),$$

where μ denotes the Lebesgue measure. As noted in [9], if $W_1 \cong W_2$, then $\mathrm{Ent}(W_1) = \mathrm{Ent}(W_2)$. In other words, entropy is a property of a graph limit, rather than of the graphon that represents it. Thus, we may define the entropy $\mathrm{Ent}(\Gamma)$ of a graph limit Γ to be the entropy of any graphon that represents it.

Hatami, Janson, and Szegedy [9] posed the question of which graphons may arise as limits of sequences of graphs with a given property $\mathcal{P}$. In addition to the intrinsic interest of this question, it turns out that if $\mathcal{P}$ is hereditary, then certain limits of sequences of graphs in $\mathcal{P}$ (namely, those with maximum entropy) give a great deal of information about the number and typical structure of graphs in $\mathcal{P}$. (We do not distinguish between a graph property and the class of graphs with that property.) In order to state these results, we need to introduce more notation.

Let $\mathcal{U}_n$ denote the set of unlabeled graphs on n vertices and let $\mathcal{L}_n$ denote the set of labeled graphs with vertex set $[n]$. Given a graph property $\mathcal{P}$, we let $\mathcal{P}_n = \mathcal{P} \cap \mathcal{U}_n$ denote the set of unlabeled elements of $\mathcal{P}$ with n vertices and let $\mathcal{P}_n^L$ denote the set of labeled elements of $\mathcal{P}$ with vertex set $[n]$. The function $n \mapsto |\mathcal{P}_n|$ is called the (unlabeled) *speed* of $\mathcal{P}$; the labeled speed is defined similarly. Observe that

$$\big|\mathcal{P}_n\big| \leqslant \big|\mathcal{P}_n^L\big| \leqslant n!\big|\mathcal{P}_n\big|. \tag{1}$$

Given a graph property $\mathcal{P}$, we let $\widehat{\mathcal{P}}$ denote the set of graph limits of sequences in $\mathcal{P}$. We furthermore let $\widehat{\mathcal{P}}^*$ denote the set of elements of $\widehat{\mathcal{P}}$ of maximum entropy, i.e.,

$$\widehat{\mathcal{P}}^* = \bigg\{\Gamma \in \widehat{\mathcal{P}}\ :\ \mathrm{Ent}(\Gamma) = \max_{\Gamma' \in \widehat{\mathcal{P}}} \mathrm{Ent}(\Gamma')\bigg\}.$$

We will also use these symbols to refer to the set of graphons (respectively, the set of maximum-entropy graphons) representing limits of sequences in $\mathcal{P}$. It is shown in [9] that if $\mathcal{P}$ is hereditary (and not finite), then $\max_{\Gamma\in\widehat{\mathcal{P}}} \mathrm{Ent}(\Gamma)$ is achieved—in other words, $\widehat{\mathcal{P}}^*$ is nonempty.

In [9, Theorem 1.6], Hatami, Janson, and Szegedy showed that if a hereditary property $\mathcal{P}$ has a single graph limit Γ of maximum entropy, then a typical element of $\mathcal{P}$ is close to Γ (in terms of the standard cut metric on the space of graph limits).

Theorem 1. *Suppose that* $\mathcal{P}$ *is a hereditary property and that* $\max_{\Gamma\in\widehat{\mathcal{P}}} \operatorname{Ent}(\Gamma)$ *is attained by a unique graph limit* $\Gamma_{\mathcal{P}}$*. Then*

(i) if $G_n \in \mathcal{U}_n$ *is a uniformly random unlabeled element of* $\mathcal{P}_n$*, then* G_n *converges in probability to* $\Gamma_{\mathcal{P}}$ *as* $n \to \infty$*;*

(ii) if $G_n \in \mathcal{L}_n$ *is a uniformly random labeled element of* $\mathcal{P}_n^L$*, then* G_n *converges in probability to* $\Gamma_{\mathcal{P}}$ *as* $n \to \infty$*.*

Now we define a special class of graphons. All of these graphons will be defined on $[0,1)^2$, rather than on $[0,1]^2$; it is easy to see that this change is immaterial. Given $r \in \mathbb{N}$ and $i \in [r]$, let $I_i = [(i-1)/r, i/r)$ and let $E_r = \cup_{i\neq j} I_i \times I_j$. We also let $E_\infty = [0,1)^2$. Given $r \in \mathbb{N} \cup \{\infty\}$, we let R_r denote the set of graphons W such that $W(x,y) = 1/2$ if $(x,y) \in E_r$ and $W(x,y) \in \{0,1\}$ otherwise. It is easy to see that if $W \in R_r$, then

$$\operatorname{Ent}(W) = \iint_{E_r} h(1/2)\,\mathrm{d}\mu(x)\mathrm{d}\mu(y) = \mu(E_r) = 1 - \frac{1}{r}.$$

For $r \in \mathbb{N}$ and $0 \leqslant s \leqslant r$, we let $W^*_{r,s}$ denote the graphon in R_r that equals 1 on $I_i \times I_i$ for $i \leqslant s$ and equals 0 on $I_i \times I_i$ for $s+1 \leqslant i \leqslant r$. Observe that R_∞ consists only of the graphon that equals $1/2$ everywhere on $[0,1)^2$; for notational convenience, we denote this graphon by $W^*_{\infty,0}$.

Given $r \in \mathbb{N}$ and $0 \leqslant s \leqslant r$, we let $\mathcal{C}(r,s)$ denote the class of graphs whose vertex sets can be partitioned into s (possibly empty) cliques and $r-s$ (possibly empty) independent sets. In particular, $\mathcal{C}(r,0)$ is the class of r-colorable graphs. Observe that for each r and s, the class $\mathcal{C}(r,s)$ is hereditary, and that $\mathcal{C}(r,0)$ is monotone.

It is shown in [9, Theorem 1.9] that if $\mathcal{P}$ is a hereditary property, then the maximum entropy of an element of $\widehat{\mathcal{P}}$ takes one of countably many values, and furthermore that this value determines the asymptotic speed of $\mathcal{P}_n$.

Theorem 2. *If* $\mathcal{P}$ *is a hereditary property, then there exists* $r \in \mathbb{N} \cup \{\infty\}$ *such that* $\max_{\Gamma\in\widehat{\mathcal{P}}} \operatorname{Ent}(\Gamma) = 1 - 1/r$ *and such that every graph limit* $\Gamma \in \widehat{\mathcal{P}}^*$ *can be represented by a graphon* $W \in R_r$*. Moreover,*

$$|\mathcal{P}_n| = 2^{\left(1-\frac{1}{r}+o(1)\right)\binom{n}{2}}.$$

Given a graph F, we say that a graph G is *F-free* if no subgraph of G is isomorphic to F. Given a (possibly infinite) family of graphs $\mathcal{F}$, we say that G is *$\mathcal{F}$-free* if it is F-free for every $F \in \mathcal{F}$. Observe that for any family $\mathcal{F}$, the class of $\mathcal{F}$-free graphs is monotone. (Conversely, every monotone class $\mathcal{P}$ equals the class of $\mathcal{F}$-free graphs for some family $\mathcal{F}$—for example, $\mathcal{F} = \mathcal{U} \setminus \mathcal{P}$.) We write $\operatorname{Forb}(\mathcal{F})$ for the class of $\mathcal{F}$-free graphs and write $\operatorname{Forb}(F)$ when $\mathcal{F} = \{F\}$. Note in particular that $\operatorname{Forb}(\emptyset)$ equals the class of all unlabeled finite graphs, which we denote by $\mathcal{U}$.

The *coloring number* of a family of graphs $\mathcal{F}$ is

$$\operatorname{col}(\mathcal{F}) = \min_{F \in \mathcal{F}} \chi(F).$$

In particular, we define

$$\operatorname{col}(\emptyset) = \infty. \tag{2}$$

Our main result says that if $\operatorname{col}(\mathcal{F}) = r + 1$, then a typical element of $\operatorname{Forb}(\mathcal{F})$ resembles a balanced r-partite graph in which cross-edges are present independently with probability 1/2.

Theorem 3. *Let $\mathcal{F}$ be a family of graphs and let $r = \operatorname{col}(\mathcal{F}) - 1$. If $\mathcal{P} = \operatorname{Forb}(\mathcal{F})$, then as n tends to ∞, a sequence of uniformly random unlabeled (respectively, labeled) elements of $\mathcal{P}_n$ (respectively, elements of $\mathcal{P}_n^L$) converges in probability to the graph limit Γ_r represented by $W^*_{r,0}$.*

Note that the quantity r in the statement of the theorem also equals the largest integer t for which every t-colorable graph is in $\operatorname{Forb}(\mathcal{F})$.

It follows from Theorems 2 and 3 that if $\operatorname{col}(\mathcal{F}) = r + 1$ then

$$\left|\operatorname{Forb}(\mathcal{F})_n\right| = 2^{\left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}}, \tag{3}$$

which was first shown in [7]. Let us also note that Balogh, Bollobás, and Simonovits [2] obtained a fairly sharp bound on the error term in (3).

Remark 4. The proof of Theorem 3 shows that if $r = \operatorname{col}(\mathcal{F}) - 1$ and $\mathcal{P} = \operatorname{Forb}(\mathcal{F})$, then $W^*_{r,0}$ is the unique maximum-entropy element of $\widehat{\mathcal{P}}$. For certain families $\mathcal{F}$, it is also possible to describe the set of all $\mathcal{F}$-free graph limits. For example, the set of limits of bipartite graphs is determined in [9, Example 2.1], and a very similar argument holds for r-partite graphs when $r \geqslant 3$. However, we know of no representation of the set of all $\mathcal{F}$-free graph limits for arbitrary $\mathcal{F}$.

Remark 5. Erdős, Frankl, and Rödl [7] showed that if $\chi(F) = r + 1$, then every F-free graph G may be made K_{r+1}-free by removing $o(n^2)$ edges from G. This result is similar in spirit to Theorem 3, but we see no direct implication: if $\{G_n\}_{n=1}^{\infty}$ is a sequence of uniformly random F-free graphs and $\{G'_n\}_{n=1}^{\infty}$ is the sequence of resulting K_{r+1}-free graphs, then the distribution of G'_n need not be uniform in $\operatorname{Forb}(K_{r+1})_n$.

Remark 6. Theorem 3 says that if $\operatorname{col}(\mathcal{F}) = r+1$ then almost every (labeled or unlabeled) $\mathcal{F}$-free graph is close to a balanced r-partite graph. (Conversely, every r-partite graph is trivially $\mathcal{F}$-free.) In the case of labeled graphs, Prömel and Steger [14] proved a much stronger result for a specific class of monotone properties: they characterized the graphs F for which almost every labeled F-free graph *is* $(\chi(F) - 1)$-partite. Given a graph F, we say that $e \in E(F)$ is *critical* if $\chi(F - e) < \chi(F)$. Prömel and Steger showed that if $\chi(F) = r + 1$ then

$$\left|\operatorname{Forb}(F)_n^L\right| = \big(1 + o(1)\big)\left|\mathcal{C}(r, 0)_n^L\right|$$

if and only if F contains a critical edge. They also showed that if F does not contain a critical edge, then there exists a constant $c_r > 0$ not depending on F such that

$$\left|\mathrm{Forb}(F)_n^L\right| \geqslant c_r n \left|\mathcal{C}(r,0)_n^L\right| \tag{4}$$

for all n sufficiently large. Theorem 3 shows that if $\mathcal{F}$ is any family of graphs with $\mathrm{col}(\mathcal{F}) = r+1$, then $\mathrm{Forb}(\mathcal{F})^L$ and $\mathcal{C}(r,0)^L$ have roughly the same asymptotic speed. Note that this result does not contradict (4) when $\mathcal{F} = \{F\}$ and F does not contain a critical edge: if $\chi(F) = r+1$, then, in view of (1) and (3), Theorem 3 implies the weaker statement that $|\mathrm{Forb}(F)_n^L|$ and $|\mathcal{C}(r,0)_n^L|$ differ by a factor of $2^{o(n^2)}$.

2 Proof of Theorem 3

Lemma 7. *Let $\mathcal{P}$ be a monotone property and let $W \in \widehat{\mathcal{P}}$. If W' is a graphon such that $W' \leqslant W$ pointwise, then $W' \in \widehat{\mathcal{P}}$.*

Proof. Consider the sequences of random graphs $\{G(n,W)\}_{n=1}^\infty$ and $\{G(n,W')\}_{n=1}^\infty$. Since $W' \leqslant W$ pointwise, a standard argument shows that the two sequences can be coupled so that for each n, $G(n,W') \subseteq G(n,W)$ almost surely. It is shown in [10, Theorem 3.1] that if $W \in \widehat{\mathcal{P}}$ then, for each n, $G(n,W) \in \mathcal{P}$ almost surely. It follows that for each n, we almost surely have $G(n,W') \in \mathcal{P}$, as well. Finally, it is shown in [5, Theorem 4.5] that $G(n,W') \to W'$ almost surely as $n \to \infty$, which implies that $W' \in \widehat{\mathcal{P}}$, as claimed. □

Now we prove our main result, Theorem 3.

Proof of Theorem 3. We begin by showing that, up to equivalence of graphons, $\widehat{\mathcal{P}}$ contains a unique element of maximum entropy. By Theorem 2, there exists $t \in \mathbb{N} \cup \{\infty\}$ such that $\widehat{\mathcal{P}}^* \subseteq R_t$ up to equivalence of graphons.

First, suppose that $t < \infty$. Observe that if $W \in \widehat{\mathcal{P}}^* \cap R_t$, then $W \geqslant W^*_{t,0}$ pointwise, which by Lemma 7 implies that $W^*_{t,0} \in \widehat{\mathcal{P}}^*$. We claim that, up to equivalence of graphons, $W^*_{t,0}$ is in fact the only maximum-entropy element of $\widehat{\mathcal{P}}$. Indeed, let $W' \in \widehat{\mathcal{P}}^* \cap R_t$ and suppose that $\mu(W' = 1) > 0$. But then Lemma 7 implies that $\min\{W', 1/2\}$ is a graphon in $\widehat{\mathcal{P}}$ with entropy strictly larger than $1 - 1/t$, which contradicts the definition of t.

Now we show that $t = \mathrm{col}(\mathcal{F}) - 1 = r$. Suppose that $t < \infty$. It is observed in [9, Remark 1.10] that if $\mathcal{P}$ is hereditary and $0 \leqslant s \leqslant r < \infty$, then $W^*_{r,s} \in \widehat{\mathcal{P}}$ if and only if $\mathcal{C}(r,s) \subseteq \mathcal{P}$. By the definition of $\mathrm{col}(\mathcal{F})$, it is easy to see that if $u \leqslant r$, then $\mathcal{C}(u,0) \subseteq \mathcal{P}$ and hence $W^*_{u,0} \in \widehat{\mathcal{P}}$. On the other hand, $\mathcal{F}$ contains some element of $\mathcal{C}(r+1,0)$, which implies that $\mathcal{C}(r+1,0) \nsubseteq \mathcal{P}$. This implies that $W^*_{u,0} \notin \widehat{\mathcal{P}}$ when $u \geqslant r+1$, and hence that $t = r$.

If $t = \infty$, then we claim that $\mathcal{P} = \mathrm{Forb}(\emptyset) = \mathcal{U}$; the conclusion then follows from (2). Suppose to the contrary that $\mathcal{P}$ does not contain some graph F. Then $\mathcal{C}(\chi(F),0) \nsubseteq \mathcal{P}$, which implies that $W^*_{\chi(F),0} \notin \widehat{\mathcal{P}}$. However, $W^*_{\chi(F),0} \leqslant W^*_{\infty,0}$ pointwise, so Lemma 7 implies that $W^*_{\infty,0} \notin \widehat{\mathcal{P}}$, which is a contradiction.

Finally, since $\mathcal{P}$ is a hereditary property, it follows from Theorem 1 that a uniformly random (labeled or unlabeled) element of $\mathcal{P}$ converges in probability to Γ_r, as claimed. □

References

[1] N. Alon, J. Balogh, B. Bollobás, and R. Morris. The structure of almost all graphs in a hereditary property. *J. Combin. Theory Ser. B*, 101(2):85–110, 2011.

[2] J. Balogh, B. Bollobás, and M. Simonovits. The number of graphs without forbidden subgraphs. *J. Combin. Theory Ser. B*, 91(1):1–24, 2004.

[3] J. Balogh, B. Bollobás, and M. Simonovits. The typical structure of graphs without given excluded subgraphs. *Random Structures Algorithms*, 34(3):305–318, 2009.

[4] B. Bollobás. Hereditary and monotone properties of combinatorial structures. In *Surveys in Combinatorics 2007*, volume 346 of *London Math. Soc. Lecture Note Ser.*, pages 1–39. Cambridge Univ. Press, Cambridge, 2007.

[5] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

[6] P. Diaconis and S. Janson. Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)*, 28(1):33–61, 2008.

[7] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.*, 2(1):113–121, 1986.

[8] P. Erdős, D.J. Kleitman, and B.L. Rothschild. Asymptotic enumeration of K_n-free graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, pages 19–27. Atti dei Convegni Lincei, No. 17. Accad. Naz. Lincei, Rome, 1976.

[9] H. Hatami, S. Janson, and B. Szegedy. Graph properties, graph limits, and entropy. Preprint, arXiv:1312.5626, 2013.

[10] S. Janson. Graph limits and hereditary properties. *European J. Combin.*, to appear.

[11] S. Janson and A.J. Uzzell. On string graph limits and the structure of a typical string graph. Preprint, arXiv:1403.2911, 2014.

[12] L. Lovász. *Large networks and graph limits*, volume 60 of *Amer. Math. Soc. Colloq. Publ.* Amer. Math. Soc., Providence, RI, 2012.

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

[14] H.J. Prömel and A. Steger. The asymptotic number of graphs not containing a fixed color-critical subgraph. *Combinatorica*, 12(4):463–473, 1992.