ARTICLE

Universal and unavoidable graphs

Matija Bucić, Nemanja Draganić and Benny Sudakov[*][†]

Department of Mathematics, ETH, Zürich, Switzerland
*Corresponding author. Email: benjamin.sudakov@math.ethz.ch

Abstract

The Turán number $ex(n, H)$ of a graph H is the maximal number of edges in an H-free graph on n vertices. In 1983, Chung and Erdős asked which graphs H with e edges minimise $ex(n, H)$. They resolved this question asymptotically for most of the range of e and asked to complete the picture. In this paper, we answer their question by resolving all remaining cases. Our result translates directly to the setting of universality, a well-studied notion of finding graphs which contain every graph belonging to a certain family. In this setting, we extend previous work done by Babai, Chung, Erdős, Graham and Spencer, and by Alon and Asodi.

1. Introduction

The following question of Turán dating back to 1941 [27] is one of the most classical problems of graph theory. Given a fixed graph H, what is the maximal number of edges one can have in an n vertex graph which does not contain a copy of H as a subgraph? The answer to this question, denoted $ex(n, H)$, is called the Turán number of H. Turán numbers have been extensively studied in the last 70 years, see, for example, the surveys [25, 23, 20, 26, 21].

Turán's problem leads to another very natural extremal question—what is the largest size of a graph which we cannot avoid as a subgraph in any graph on n vertices and e edges? In other words, what kind of a graph H with a fixed number of edges (possibly depending on n) has minimal Turán number?

This question was first asked by Chung and Erdős [10] in 1983 and many questions of a similar flavour were later considered in [11, 8, 12, 17]. Some of these questions have also featured in an Erdős open problem collection [9].

In this paper, we revisit the original question of Chung and Erdős. They say a graph H is (n, e)-*unavoidable* if every graph on n vertices and e edges contains a copy of H as a subgraph. Let $f(n, e)$ be the maximum number of edges in an (n, e)-unavoidable graph. Chung and Erdős obtain the following bounds on $f(n, e)$:

i. $f(n, e) = \Theta\left(\lceil e/n \rceil^2\right)$ if $e \leq n^{4/3}$.

ii. $f(n, e) = \Theta\left(\frac{\sqrt{e}\log n}{\log(\binom{n}{2}/e)}\right)$ if $cn^{4/3} < e < \binom{n}{2} - n^{1+c}$ for any $0 < c < 1$.

iii. $\Omega\left(\frac{m^2}{\log^2 m}\right) \leq \binom{n}{2} - f(n, e) \leq \mathcal{O}\left(\frac{m^2 \log\log m}{\log m}\right)$ where $m = \binom{n}{2} - e$, and $\binom{n}{2} - cn < e$, for some $c > 0$.

In fact for certain regimes from parts (i) and (ii) they obtain even more precise bounds, some of which appear in a later paper [11]. However, there is a gap between regimes of part (ii) and (iii) for which their methods fail to give an answer. Given this, they naturally ask what is the correct behaviour of $f(n, e)$ towards the end of the range, so when $e > \binom{n}{2} - n^{1+o(1)}$. In this paper, we resolve this question and determine $f(n, e)$ for the remaining values of e.

Theorem 1.1. *For any $\varepsilon > 0$, if we let $m = \binom{n}{2} - e$*

$$f(n, e) = \begin{cases} \binom{n}{2} - \Theta\left(\frac{m^2}{\log^2 m}\right) & \text{if } m \leq n \log n \\ \Theta\left(\frac{n^3 \log n}{m}\right) & \text{if } n \log n < m < n^{3/2-\varepsilon} \end{cases}.$$

It is worth noting that the unavoidable graph we use in order to obtain the lower bound in the second part of the above theorem is the random Erdős–Renyi graph. This is in stark contrast to the very structured graph used by Chung and Erdős in regimes of (ii)[a] and (iii) above. In particular, they use a disjoint union of complete bipartite graphs. Furthermore, for part of the regime both of these very different examples are extremal, up to a constant factor.

1.1 Universality

An (n, e)-unavoidable graph H on n vertices is contained in every graph G on n vertices and e edges. Another way of saying this is that the complement of H contains the complement of G. Since G was arbitrary, this means that H is (n, e)-unavoidable if and only if its complement contains every graph on n vertices and $m = \binom{n}{2} - e$ edges as a subgraph. This observation, made by Chung and Erdős, links unavoidability to perhaps an even more natural and well-studied notion, namely that of universality.

For a given family of graphs $\mathcal{H}$, we call a graph $\mathcal{H}$-universal if it contains a copy of each graph in $\mathcal{H}$. Given $\mathcal{H}$ one usually wants to find the smallest $\mathcal{H}$-universal graph with respect to the number of edges or both vertices and edges. Chung and Graham [13] were the first to use this general notion of universality in 1979. In [13], they survey a number of results in this setting. Universality problems have been extensively studied ever since, for some examples, see [14, 16, 24, 4, 6] and references therein.

Following the above observation by Chung and Erdős, the relevant family in our case is that of all graphs on n vertices and m edges, which we denote as $\mathcal{H}(n, m)$. We denote by $g(n, m)$ the minimum number of edges in an $\mathcal{H}(n, m)$-universal graph on n vertices. The above observation of Chung and Erdős boils down to the following relation between the functions f and g:

$$g(n, m) = \binom{n}{2} - f(n, e), \tag{1}$$

where $m = \binom{n}{2} - e$. This relation allows us to easily translate results between unavoidability and universality. For example, Theorem 1.1 is equivalent to the following statement.

---

[a]Note that the bound of Chung and Erdős agrees, after some simplification, with our bound in Theorem 1.1, for the part of the range where they overlap.

Theorem 1.2. *For any $\varepsilon > 0$ we have*

$$g(n, m) = \begin{cases} \Theta\left(\frac{m^2}{\log^2 m}\right) & \text{if } m \leq n \log n \\ \binom{n}{2} - \Theta\left(\frac{n^3 \log n}{m}\right) & \text{if } n \log n < m < n^{3/2-\varepsilon} \end{cases}.$$

While in the unavoidability case, one might arguably think that the very end of the regime is not that interesting since it only involves determining the behaviour of the second-order term, we see here that this second-order term becomes the main and only term for the case of universality. This regime, in particular is related to previous work on universality for the family $\mathcal{E}(m)$ of graphs with exactly m edges (and no isolated vertices). This was first considered by Babai, Chung, Erdős, Graham and Spencer [5] in 1982. They show that there exist $\mathcal{E}(m)$-universal graphs with at most $\mathcal{O}\left(\frac{m^2 \log \log m}{\log m}\right)$ edges. This was later improved upon by Alon and Asodi [1] who show the upper bound of $\mathcal{O}\left(\frac{m^2}{\log^2 m}\right)$, which matches the lower bound. Our bound on $g(2m, m)$ recovers the result of Alon and Asodi as any graph on m edges can use at most $2m$ vertices, so any graph which is $\mathcal{H}(2m, m)$-universal is also $\mathcal{E}(m)$-universal. On the other hand when $m > n/2$ their results cannot be used to obtain a bound on $g(n, m)$ for the following reason. In order to bound $g(n, m)$ one needs to find an $\mathcal{H}(n, m)$-universal graph on n vertices. This can never come from an $\mathcal{E}(m)$-universal graph as any such graph needs to have at least $2m > n$ vertices, in order to contain a matching of m edges. Furthermore, $\mathcal{E}(m)$-universal construction of Alon and Asodi uses more than $4m$ vertices. Since any graph in $\mathcal{E}(m)$ has at most $2m$ vertices, this affords them a lot of leeway during the embedding process. In contrast, in our problem we need to find spanning universal graphs, which requires a different embedding technique. Proving such spanning results is often a harder problem. While neither our nor the universal graphs used by Alon and Asodi are explicit, our embedding technique allows us to work with a weaker and simpler property of random graphs, which might be helpful in answering their question about finding explicit universal graphs.

Another related notion of universality deals with the family $\mathcal{E}(n, d)$ of graphs on n vertices with degree bounded by d. Alon and Capalbo [3] show that $\Theta\left(n^{2-2/d}\right)$ is the least possible number of edges in an $\mathcal{E}(n, d)$-universal graph. The $\mathcal{E}(n, d)$-universal graph they construct has more than $(1+\varepsilon)n$ vertices and it is a seemingly hard open problem in the area to obtain such a graph on exactly n vertices. The best result in this direction is due to the same authors [2] where they exhibit such a graph with $\mathcal{O}\left(n^{2-2/d} \log^{4/d} n\right)$ edges. As already mentioned, we need to overcome a similar difficulty in our setting since we require our universal graphs to be of the same order as the graphs we want to embed.

Notation. We will abbreviate $\mathcal{H}(n, m)$-universal by (n, m)-universal throughout the paper. Let $G = (V, E)$ be a graph and $U \subseteq V$. We denote with $G[U]$ the subgraph of G induced by U. We denote with $\mathcal{G}(n, p)$ the standard Erdős–Renyi random graph, i.e. the probability distribution on the set of all graphs on n vertices where each graph H has probability measure $p^{e(H)}(1-p)^{\binom{n}{2}-e(H)}$. We say that $\mathcal{G}(n, p)$ satisfies a property with high probability (whp) if a sample from $\mathcal{G}(n, p)$ satisfies this property with probability tending to 1 as n tends to infinity. Let f, g be functions from $\mathbb{N}$ to $\mathbb{R}^+$. Then $f = \mathcal{O}(g)$ if there exists a constant C such that $f(n) \leq Cg(n)$ for all $n \in \mathbb{N}$. Also, $f = \Omega(g)$ if $g = \mathcal{O}(f)$. Furthermore, $f = \Theta(g)$ if $g = \mathcal{O}(f)$ and $g = \Omega(f)$. If $\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$ then we write $f = o(g)$ and $g = \omega(f)$. With $\log n$ we denote the natural logarithm of n. We omit floors and ceilings whenever they are not essential.

2. Lower bounds for $g(n, m)$

We start our proof of Theorem 1.2 with the lower bounds.

Lemma 2.1. *Let G be an (n, m)-universal graph with t edges. Then the following holds:*

- *If $m > n \log n$ then $t \geq \binom{n}{2} \left(1 - \frac{n \log n}{m} \right)$.*
- *If $m \leq n \log n$ then $t \geq \frac{1}{4} \binom{m/\log m}{2}$.*

Proof. The idea behind the proof is quite simple, we will count how many non-isomorphic graphs there are on n vertices and m edges and we will count how many subgraphs with m edges we can find in our a graph with t edges. The existence of an (n, m)-universal graph G on t edges implies that the former count must be smaller than the latter (since all n-vertex m-edge graphs must appear as subgraphs of G); the desired bounds then follow after some simplifications, which will differ depending on the regime.

The number of non-isomorphic graphs on n vertices and m edges is at least

$$\binom{\binom{n}{2}}{m} \frac{1}{n!},$$

here the first term counts the number of labelled graphs on n vertices and m edges, and we counted each graph at most $n!$ many times.

Since G contains all these graphs, and there are $\binom{t}{m}$ subgraphs of G with m edges, we get

$$\binom{\binom{n}{2}}{m} \frac{1}{n!} \leq \binom{t}{m}.$$

By using the inequality $\binom{b}{c} \leq \binom{a}{c}(\frac{b}{a})^c$ that holds for all $a > b > c \geq 0$ we conclude

$$\frac{1}{n!} \leq \left(\frac{t}{\binom{n}{2}} \right)^m,$$

which gives

$$t > \binom{n}{2}(n!)^{-1/m} \geq \binom{n}{2} n^{-n/m}. \tag{2}$$

Since $n^{-n/m} \geq 1 - \frac{n \log n}{m}$ the first claim follows. For the case when $m < n \log n$ let $k = m/\log m < n$. So G is also (k, m)-universal and the bound (2) implies

$$t > \binom{k}{2} k^{-k/m} = \binom{k}{2} \left(\frac{m}{\log m} \right)^{-1/\log m} > \binom{k}{2} m^{-1/\log m} = \binom{k}{2} e^{-1}. \qquad \square$$

3. Upper bounds for $g(n, m)$

In order to give an upper bound on $g(n, m)$ one needs to exhibit an example of an (n, m)-universal graph. We distinguish two regimes, corresponding to the relations between n and m which give us different behaviours in Theorem 1.2. Namely, in the first regime, we will work with $n \log n < m < n^{3/2-\varepsilon}$ and in the second with $m \leq n \log n$. In both regimes, we have different examples but they share some common traits. We begin with a few results which will be useful for both regimes.

Let us state an immediate observation which we will use frequently, without further mention.

Observation 3.1. *Let G be a graph with m edges. If G' is the subgraph of G obtained by deleting k vertices of highest degree, then $\Delta(G') \leq 2m/k$. In fact, any vertex of G' has degree at most $2m/k$ in G.*

The following lemma will allow us to pass from almost spanning universal graphs to spanning ones. Here, and in the rest of the paper when we say that we add full-degree vertices to a graph, we mean that we create a new graph, which consists of an induced copy of the original graph and a number of new 'full-degree' vertices that are adjacent to every other vertex of the new graph.

Lemma 3.2. *Let $n > k$ and suppose there is a k-vertex $\left(k - \left\lfloor \frac{k(n-k)}{2m+k} \right\rfloor, m\right)$ universal graph U. By adding $n - k$ new full-degree vertices to U we obtain an n-vertex (n, m)-universal graph.*

Proof. Let U be the universal graph given by the assumption. We construct G by adding a set F of $n - k$ new vertices to U and joining each vertex of F to all other vertices in $U \cup F$. We claim that G is (n, m)-universal. To show this we consider an arbitrary n-vertex graph H with m edges. We prove that H can be embedded into G.

Before turning to the details let us explain the idea how we will achieve this. Let $\ell := \lfloor k(n - k)/(2m + k) \rfloor$ so that U is $(k - \ell, m)$-universal. We will first find an independent set of S of ℓ vertices in H with the property that there are at least k vertices of H having no edges to S (including vertices of S themselves). We embed the remaining $n - k$ vertices into vertices of G of full degree, we then embed the remaining $k - \ell$ vertices not in S inside U, using the universality of U, finally we may embed the vertices in S as we please to the remainder of U since they send no edges to the part of H embedded in U and will be joined to every other vertex in G.

Let us now fill in the details. We pick a vertex v_1 in $H_0 := H$ of minimum degree and remove v_1 and all its neighbours from H to obtain H_1. We repeat ℓ times, in step i we pick a vertex v_i in H_{i-1} of minimum degree and create H_i by removing v_i and all its neighbours in H_{i-1}. By induction it is easy to see that after each step, we are left with at least k vertices in H_i. Indeed, if H_{i-1} has at least k vertices, then for $j \leq i$ vertex v_j has a degree of at most $2|E(H_{j-1})|/|H_{j-1}| \leq 2m/k$ in H_{j-1}, since v_j is a vertex of minimum degree in H_{j-1} and $|H_{j-1}| \geq |H_{i-1}| \geq k$. So after deleting all such v_j's and their neighbours, we have in total deleted at most $i \cdot (2m/k + 1) \leq \ell(2m/k + 1) \leq n - k$ vertices, so indeed we do have at least k vertices in H_i for each $i \in [\ell]$.

In particular, we obtain a subgraph H_ℓ of H of size at least k and an independent set $S = \{v_1 \ldots, v_\ell\}$ of ℓ vertices each with no neighbours in H_ℓ. We embed all vertices of $H \setminus (H_\ell \cup S)$ into F (which we can since $|H \setminus H_\ell| \leq n - k$) as well as as many vertices of H_ℓ as we can. If we are only left with vertices of S we can embed them into U (since they make an independent set). Otherwise we are left with a graph H' such that $S \subseteq H' \subseteq H_\ell \cup S$. Note that H' has $|G \setminus F| = k$ vertices, ℓ of which are independent (depicted by the first and last part of H in Figure 1). Since U was $(k - \ell, m)$-universal we can embed $H' \setminus S$ into U and place S into unused vertices. This is permissible since vertices in S were isolated in H'. Furthermore, since all other vertices got embedded into F (whose vertices have full degree in G), we have found our embedding of H. □

This lemma shows that if we can establish a sufficiently strong version of almost spanning universality it only costs us a very few extra vertices of full degree in order to obtain spanning universality. On the other hand since we want universality for n-vertex graphs with m edges, we know that provided $m \geq n - 1$ the host graph must contain some vertices of full degree, as the graphs we want to embed might also have some vertices of full degree. We note that this simple lemma is one of the main ingredients that allow us to find spanning universal graphs, which is usually a very hard problem. In addition, it allows us to greatly simplify certain technical parts of our arguments.

We will construct a part of our universal graph G to satisfy a property of the following type, which will allow us to embed any graph on n vertices and m edges into G.
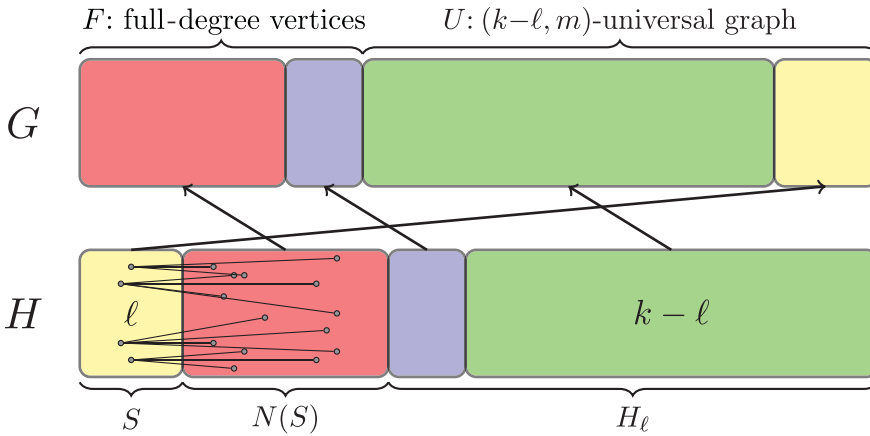
Figure 1. Illustration of the final state of the embedding process when passing from almost spanning universality to spanning universality.

Definition. A graph is said to have the (r, s, t)-*domination property* if for any set R of size r and any t disjoint sets $S_1, \ldots, S_t$ of size s which are disjoint from R, there is a vertex v in R and a set S_i such that v is a common neighbour of all vertices in S_i.

A part of our universal graphs which will satisfy the above property is going to be provided by a random graph. Towards this end, the following lemma establishes a condition on the parameters r, s, t, which implies that the random graph $\mathcal{G}(n, p)$ has the (r, s, t)-domination property.

Lemma 3.3. *The random graph $\mathcal{G}(n, p)$ has the (r, s, t)-domination property whp provided*

$$3 \log n \leq p^s \cdot \min\left(\frac{r}{s}, t\right).$$

Proof. The probability that a fixed vertex is a common neighbour of every vertex in a fixed set S of size s is p^s. By independence, the probability that a fixed set R of size r and t fixed disjoint sets $S_1, \ldots, S_t$ of size s are such that no vertex in R is a common neighbour of every vertex in some S_i is $(1 - p^s)^{tr}$. So, by a union bound, the probability that some choice of sets $R, S_1, \ldots, S_t$ fails the desired condition is at most

$$\binom{n}{r}\binom{n}{s}^t (1 - p^s)^{tr} \leq n^{r+ts} e^{-p^s tr} = n^{r+ts - p^s tr/\log n} \leq n^{-r} = o(1). \qquad \square$$

3.1 The first regime

Let us now turn to the first regime, so when $n \log n < m < n^{3/2-\varepsilon}$. As part of our construction of an (n, m)-universal graph with n vertices we first find a smaller almost spanning universal graph, in particular a graph on $k < n$ vertices, which is $(k - r, m)$-universal for appropriately chosen parameters k and r. We then finish the construction by applying Lemma 3.2.

Let $\varepsilon > 0$ and k, m be integers such that $k \log k < m < 3k^{3/2-\varepsilon}$ and k is sufficiently large depending only on ε. We let G be a k-vertex graph consisting of

- a set F of $k/2$ vertices joined to every vertex and
- a set V of $k/2$ vertices inducing a graph G' with the (r, s, t)-domination property missing at least $\frac{\varepsilon k^3 \log k}{2^8 m}$ edges
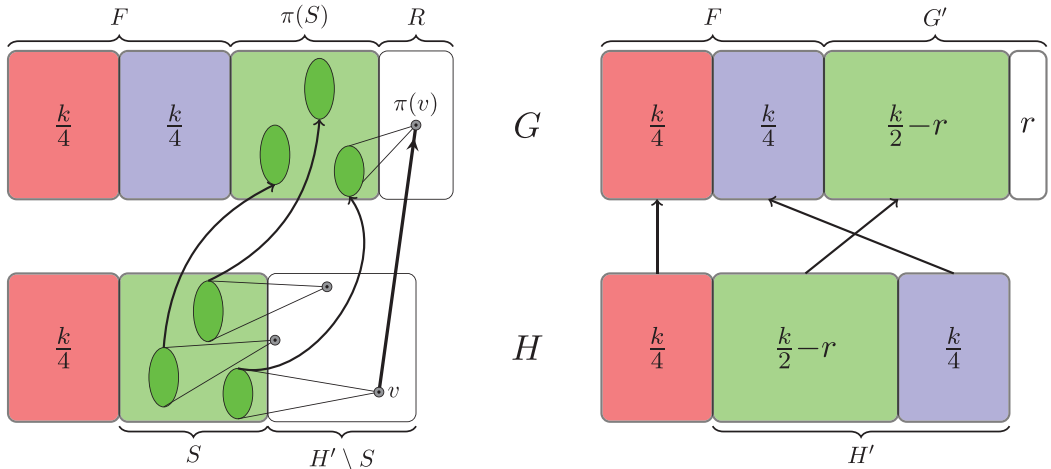
Figure 2. The picture on the left illustrates the embedding process while the one on the right illustrates the final state of the embedding.

where $t = \frac{k^3}{2^9 m^2}$, $s = \frac{8m}{k}$ and $r = \frac{k^2}{4m}$. In order to ensure we can find such a graph $G' = G[V]$ we let $p = 1 - \frac{\varepsilon k \log k}{16m} > \frac{1}{2}$. Then we have

$$p^s = (1 - (1 - p))^s \geq e^{-2(1-p)s} = k^{-\varepsilon},$$

where we used that $1 - x \geq e^{-2x}$, provided $0 \leq x < 1/2$. In addition, since $m < 3k^{3/2-\varepsilon}$, we also have

$$\min\left(\frac{r}{s}, t\right) = t = k^3/(2^9 m^2) \geq k^{2\varepsilon}/(3^2 2^9) > 3k^\varepsilon \log k,$$

as k is large enough compared to ε. So, by Lemma 3.3 $\mathcal{G}(k/2, p)$ has the (r, s, t)-domination property whp. Notice further that $\mathcal{G}(k/2, p)$ is also missing at least $(1 - p)k^2/16 = \frac{\varepsilon k^3 \log k}{2^8 m}$ edges whp. So, in particular, it provides us with our G'. We now show G is universal for almost spanning graphs with m edges.

Claim. G is $(k - r, m)$-universal.

Proof. Let H be a graph on $k - r$ vertices and m edges. Our goal is to embed H into G. Let us first delete $k/4$ vertices of largest degrees from H to obtain $H' \subseteq H$ with $\Delta(H') \leq 8m/k$ and $|H'| = 3k/4 - r$. The deleted $k/4$ vertices we embed into F, leaving us with $k/4$ remaining vertices in F.

We continue by embedding vertices from H' one by one into G', making sure that whenever two embedded vertices make an edge in H', their images also make an edge in G'. Let S be the set of already embedded vertices from H', i.e. assume that we have already found a distinct vertex $\pi(v) \in V(G')$ for every $v \in S \subseteq V(H')$. If $|S| \geq k/2 - r$ we stop, so let us assume that $|S| < k/2 - r$, which implies that we have at least $k/4$ vertices in $H' \setminus S$. For each $v \in H' \setminus S$, let $S_v = \pi(S \cap N_{H'}(v))$, so the set of vertices of G' which were already assigned to a neighbour of v (see Figure 2). We know that $|S_v| \leq 8m/k = s$ since $\Delta(H') \leq 8m/k$. Furthermore, the same max degree condition tells us that each S_v can intersect at most $(8m/k)^2$ other such sets (since S_v has size at most $8m/k$ and every vertex of S_v can belong to at most $8m/k$ other S_u's). This implies the existence of a family of at least $\frac{|H' \setminus S|}{(8m/k)^2 + 1} \geq \frac{k^3}{2^9 m^2} = t$ disjoint sets S_v. We can find such a family by greedily choosing the S_v's.

On the other hand we have a set R of at least $|G'| - |S| \geq r$ yet unassigned vertices of G'. Since G' is (r, s, t)-dominating, we know that there is a vertex u in R and one of the sets S_v such that u is a common neighbour of every vertex in S_v. So we may embed $\pi(v) = u$ and repeat.

When we stop (i.e. when $|S| \geq k/2 - r$), we will have embedded at least $k/2 - r$ vertices of H'. We now embed the remaining $k/4$ vertices of H' to the remaining $k/4$ vertices of F to obtain the desired copy of H' in G'. $\qquad\square$

Theorem 3.4. *Let $\varepsilon > 0$ and $n \log n < m < n^{3/2-\varepsilon}$. There is an n-vertex graph missing at least $\frac{\varepsilon n^3 \log n}{2^{12} m}$ edges which is (n, m)-universal, provided n is large enough depending only on ε.*

Proof. Let $k = n/2$, so $k \log k < m < 3k^{3/2-\varepsilon}$. By the previous claim G is a $(k - r, m)$-universal graph that misses at least $\frac{\varepsilon k^3 \log k}{2^8 m} \geq \frac{\varepsilon n^3 \log n}{2^{12} m}$ edges. Since $r = \frac{k^2}{4m} \leq \left\lfloor \frac{k(n-k)}{2m+k} \right\rfloor$ Lemma 3.2 implies that adding $n - k = n/2$ vertices of full degree to G gives us the desired graph. $\qquad\square$

Note that our final example for above theorem consisted of $3n/4$ vertices of full degree and $n/4$ vertices spanning a graph satisfying certain properties, which hold for the random graph $\mathcal{G}(n/4, p)$ whp. Upon taking complements, what we have shown is that whp the random graph $\mathcal{G}(n/4, q)$, where $q = 1 - p$, has Turán number on n vertices of at most $\binom{n}{2} - m = \binom{n}{2} - \frac{\varepsilon n \log \frac{n}{2}}{32 q} = (1 - \Theta(\frac{\log n}{nq}))\binom{n}{2}$, where we may choose an arbitrary q such that $1/2 > q > n^{-1/2+\varepsilon}$. Since in this regime $\chi(\mathcal{G}(n, q)) = \Theta(nq/\log n)$ this is approximately what we would obtain by applying Erdős–Stone–Simonovits theorem, despite the fact that our graph is of linear order while the usual Erdős–Stone–Simonovits applies only to small forbidden graphs. We postpone further discussion to concluding remarks.

3.2 The second regime

In this subsection, we deal with the case when $n/2 < m < n \log n$. In this regime, our goal is to construct an (n, m)-universal graph on $\Theta\left(\frac{m^2}{\log^2 m}\right)$ edges to show the desired upper bound.

A part of our construction will again use a graph satisfying an appropriate domination property with few edges, similarly as in the first regime. We, however, require slightly different relation between parameters. The next lemma shows that the random graph still provides us with such a graph.

Lemma 3.5. *$\mathcal{G}(n, p)$ whp has the $(n^{\frac{3}{4}}, \frac{\log n}{2 \log \frac{1}{p}}, n^{\frac{3}{4}})$-domination property for $p < \frac{1}{2}$.*

Proof. Using Lemma 3.3 and setting $r = n^{\frac{3}{4}}$, $s = \frac{\log n}{2 \log \frac{1}{p}}$ and $t = n^{\frac{3}{4}}$, it is enough to show that $3 \log n \leq p^s \cdot \min\left(\frac{r}{s}, t\right)$, which is equivalent to $3 \log n \leq \frac{1}{\sqrt{n}} \cdot \frac{n^{\frac{3}{4}}}{s}$ and since $s \leq \log n$, we are done. $\qquad\square$

To upper bound $g(n, m)$ we will give a recursive construction of an (n, m)-universal graph. It will not be hard to 'extract' the constructed graph later, but we use the recursive definition as it provides us with a convenient way of controlling the bounds.

Lemma 3.6. *For $n/2 \leq m \leq \frac{n \log n}{2^{10}}$ we have*

$$g(n, m) \leq \frac{32 m^3}{n \log^3 n} + g\left(n', m\right),$$

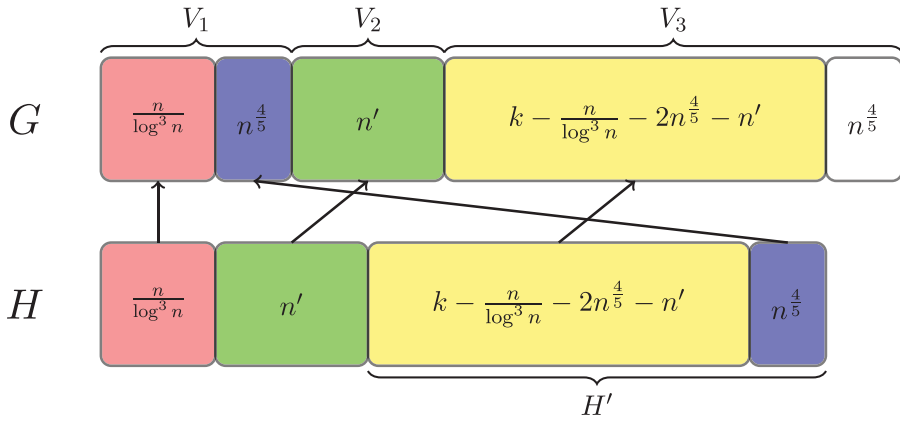*where $n' = 32 m \cdot \log\left(\frac{n \log n}{m}\right) / \log n$.*

Figure 3. Final state of the embedding process in the second regime.

Proof. Note that n' is increasing in m for $m \le \frac{n \log n}{2^{10}}$ so

$$n' \le 32 \frac{n \log n}{2^{10}} \cdot \frac{\log (2^{10})}{\log n} \le \frac{2}{3} n. \tag{3}$$

Let $k = n - \frac{n}{\log^3 n}$. Our initial goal is to find a k-vertex, $(k - n^{\frac{4}{5}}, m)$-universal graph G with at most $g(n', m) + \frac{17 m^3}{n \log^3 n}$ edges. After this is done, we will finish the proof by applying Lemma 3.2.

We construct G as follows. Let $p = \frac{m^3}{n^3 \log^3 n} < \frac{1}{2}$. Let the vertex set V of G be the union of three disjoint sets $V = V_1 \cup V_2 \cup V_3$, of sizes $|V_1| = \frac{n}{\log^3 n} + n^{\frac{4}{5}}$, $|V_2| = n'$ and $|V_3| = k - |V_1| - |V_2|$.

We construct the edge set in three steps as follows:

- Let V induce a graph with the $(n^{\frac{3}{4}}, \frac{\log n}{2 \log \frac{1}{p}}, n^{\frac{3}{4}})$-domination property that has at most $n^2 p$ edges.
- Make each vertex in V_1 adjacent to all other vertices.
- Finally, add at most $g(n', m)$ edges within V_2 to make $G[V_2]$ an (n', m)-universal graph.

The graph from the first step exists due to Lemma 3.5, as $\mathcal{G}(n, p)$ has both the $(n^{\frac{3}{4}}, \frac{\log n}{2 \log \frac{1}{p}}, n^{\frac{3}{4}})$-domination property and at most $n^2 p$ edges whp, and since both these properties are hereditary we may take any subgraph on $k < n$ vertices for our graph. Observe that G has at most

$$n^2 p + n |V_1| + g(n', m) \le n^2 p + \frac{2n^2}{\log^3 n} + g(n', m) \le \frac{17 m^3}{n \log^3 n} + g(n', m)$$

edges, as $n/2 < m$.

We now show that G is indeed $(k - n^{\frac{4}{5}}, m)$-universal. Let H be an arbitrary graph on $k - n^{\frac{4}{5}}$ vertices and m edges. Our task is to find an embedding of H into G (see Figure 3 for an illustration of the final state of the embedding process).

First, embed the $\frac{n}{\log^3 n}$ vertices of highest degree from H arbitrarily into V_1, and note that $n^{\frac{4}{5}}$ vertices in V_1 remain free. Second, embed the next n' vertices of highest degree into V_2. As the subgraph of H induced by those vertices has less than m edges, it can be embedded into V_2, as $G[V_2]$ is (n', m)-universal.

Let H' be the subgraph consisting of the remaining vertices of H. Note that $|H'| = k - n^{\frac{4}{5}} - \frac{n}{\log^3 n} - n' = |V_3|$. We embed vertices of H' one by one into V_3 and delete them from H' until only $n^{\frac{4}{5}}$ are left. For each $v \in H'$ let S_v be the image set of all v's neighbours in H, which are already embedded into $V_2 \cup V_3$. At each step, we argue that we can find a vertex $v \in H'$ and a free vertex $u \in V_3$ such that u is a common neighbour of S_v and we embed v into u.

Note that the size of each S_v at each step is at most $\Delta_1 = 2m/n' = \frac{2\log n}{32\log(\frac{n\log n}{m})} \leq \frac{\log n}{2\log \frac{1}{p}} \leq \log n$, because Δ_1 is an upper bound on the degree of vertices in H'. Note also that each vertex in $V_2 \cup V_3$ can only be in at most $\Delta_2 = 2m/(\frac{n}{\log^3 n}) \leq \log^5 n$ sets S_v, as vertices in $V_2 \cup V_3$ cannot be images of one of the $\frac{n}{\log^3 n}$ vertices of highest degree in H, so their degree in H is at most $2m/(\frac{n}{\log^3 n})$. Now we find a disjoint collection of S_v's, by choosing them one by one and each time deleting all other sets which intersect the chosen one. Thus we get a collection of at least $\frac{|H'|}{\Delta_1\Delta_2+1} \geq n^{\frac{3}{4}}$ disjoint sets S_i, since $|H'| \geq n^{\frac{4}{5}}$.

To summarise we have found $n^{\frac{3}{4}}$ disjoint sets S_v each of size at most $\frac{\log n}{2\log(\frac{1}{p})}$. The set of remaining free vertices of V_3 has size $|H'| \geq n^{\frac{4}{5}}$ and is disjoint from these S_v's. Therefore, the domination property of G implies that there is a free vertex in V_3, which is a common neighbour of all vertices in some set S_v. We embed v into this vertex. We update each S_v after every single embedding. When fewer than $n^{\frac{4}{5}}$ vertices are left in H' we embed them into the remaining free part of V_1 and are done.

We have shown the existence of a k-vertex graph G which is $(k - n^{\frac{4}{5}}, m)$-universal. By making use of Lemma 3.2 and noting that

$$k - \left\lfloor \frac{k(n-k)}{2m+k} \right\rfloor \leq k - \frac{n(n-k)}{9m} \leq k - \frac{n-k}{\log n} \leq k - \frac{n}{\log^4 n} \leq k - n^{\frac{4}{5}},$$

we conclude that by adding $n - k = \frac{n}{\log^3 n}$ full-degree vertices to G we get an n-vertex, (n, m)-universal graph. The number of edges of this graph is at most

$$g(n', m) + \frac{17m^3}{n\log^3 n} + \frac{n^2}{\log^3 n} \leq g(n', m) + \frac{25m^3}{n\log^3 n}$$

as $n/2 \leq m$, which finishes the proof. □

Corollary 3.7. *For $n/2 \leq m$ we have*

$$g(n, m) \leq \mathcal{O}\left(\frac{m^2}{\log^2 m}\right).$$

Proof. If $n < \frac{16m}{\log m}$ then the statement holds trivially since a graph on n vertices has at most $\binom{n}{2}$ edges.

For any fixed m we will use the previous lemma and induction on n to show

$$g(n, m) \leq \frac{2^{22}m^2}{\log^2 m}\left(1 - \frac{2m}{n\log n}\right) \tag{4}$$

holds for all n such that $\frac{16m}{\log m} \leq n \leq 2m$. The lower bound on n implies that $n \geq \sqrt{m}$ and therefore $\log m \leq 2\log n$. For the base of the induction consider all n in our range such that $n \leq \frac{2^{11}m}{\log m}$. In

this case the inequality (4) holds, since $n \geq \frac{16m}{\log m}$ implies $\frac{2m}{n \log n} \leq \frac{1}{2}$ and $n^2 \leq \frac{2^{22}m^2}{\log^2 m}$ so the RHS is at least $n^2/2$ and as before any graph on n vertices has at most $\binom{n}{2}$ edges.

We now proceed to the induction step. Let $2m \geq n \geq \frac{2^{11}m}{\log m}$ and assume that the statement holds for all smaller n (but still larger than $\frac{16m}{\log m}$). Since our bounds on n imply $\frac{n \log n}{2^{10}} \geq m \geq n/2$ we have that

$$g(n, m) \leq \frac{32m^3}{n \log^3 n} + g(n', m),$$

where $n' = \frac{32m}{\log n} \cdot \log \left(\frac{n \log n}{m} \right)$ is given by the previous lemma. By inequality (3) we have $n' < n$ and from the definition $n' \geq \frac{32m}{\log n} \geq \frac{16m}{\log m}$ so we can apply the induction hypothesis to n' to obtain

$$g(n, m) \leq \frac{32m^3}{n \log^3 n} + \frac{2^{22}m^2}{\log^2 m} \left(1 - \frac{2m}{n' \log n'} \right)$$

$$= \frac{2^{22}m^2}{\log^2 m} \left(1 + \frac{m \log^2 m}{2^{17}n \log^3 n} - \frac{2m}{n' \log n'} \right).$$

In order to finish the proof it is enough to show that:

$$\frac{2m}{n' \log n'} - \frac{m \log^2 m}{2^{17}n \log^3 n} \geq \frac{2m}{n \log n}$$

or equivalently

$$\frac{2 \log n}{n' \log n'} - \frac{\log^2 m}{2^{17}n \log^2 n} \geq \frac{2}{n}.$$

Recall that $\log m \leq 2 \log n$ and $\log n' < \log n$ (since $n' < n$). Thus it is enough to show

$$\frac{2}{n'} \geq \frac{3}{n},$$

which is true by inequality (3), so we are done. □

If one looks at what kind of graph this recursive argument builds, in each step it will add a few more vertices of full degree, in total at most $\mathcal{O}(\frac{m^2}{n \log^2 m})$ since otherwise we would have used too many edges. The rest of the graph consists of several blocks which are initially small and have a large number (of randomly chosen) neighbours and progressively the blocks become bigger and bigger but they have less (randomly chosen) neighbours. The number of blocks we see is the number of times we needed to call upon the recursion and from definition of n' we roughly have $\frac{n' \log n'}{m} \approx \log \left(\frac{n \log n}{m} \right)$ where we used that $\log n' \approx \log n$ as throughout the argument we stay in the same range depending on m. What this means is that in each step, we take a logarithm of the current value of $\frac{n \log n}{m}$ up until it reaches a constant. That is, we need $\mathcal{O} \left(\log^* \left(\frac{n \log n}{m} \right) \right)$[b] many steps. See Figure 4 for an illustration of the constructed graph.

---

[b]$\log^* x$ is the function defined as the number of times we need to apply the logarithm function to x in order to get to 1, applied to the number of atoms in the universe it evaluates to about 5.
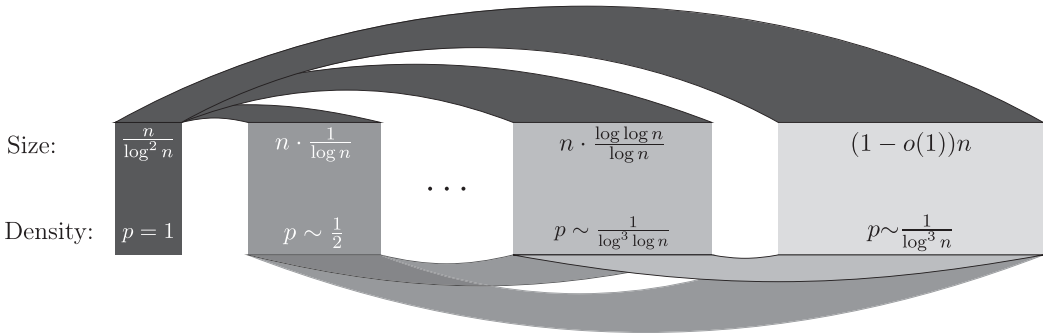
Figure 4. The (n, n)-universal graph constructed by the recursion. Each vertex in a block has an edge towards another vertex in its own block or a vertex in a subsequent block with probability p. The density of the random edges, depicted by different shades of grey, decreases as the size of the block increases.

3.3 Completing the picture

We are now ready to combine our results to show our main theorem.

Proof of Theorem 1.2. The lower bounds follow directly from Lemma 2.1. In Corollary 3.7 we have proven the desired upper bound for $g(n, m)$ when $n/2 \leq m \leq n \log n$. Theorem 3.4 gives us the bound when $n \log n < m < n^{3/2-\varepsilon}$.

The remaining case is when $m < n/2$. Notice that in this case $g(n, m) \leq g(2m, m) = \Theta\left(\frac{m^2}{\log^2 m}\right)$ since any graph with e edges has at most $2m$ non-isolated vertices. $\qquad\square$

Remark. We have shown that the function $g(n, m)$ exhibits different behaviour in the two regimes. Namely, when $m = o(n \log n)$ then $g(n, m) = \Theta\left(\frac{m^2}{\log^2 m}\right) = o(n^2)$ while if $m = \omega(n \log n)$ we have that $g(n, m) = \binom{n}{2} - \Theta\left(\frac{n^3 \log n}{m}\right) = (1 - o(1))\binom{n}{2}$. From our arguments one can see that as we transition from one regime to another both bounds become quadratic in n. The following corollary shows that $m = \Theta(n \log n)$ is precisely the transitioning point between the two behaviours.

Corollary 3.8. *Let $\mu > 0$ be a positive constant. Then there exist positive constants $0 < c_1, c_2 < 1$ such that $c_1\binom{n}{2} \leq g(n, \mu n \log n) \leq c_2\binom{n}{2}$ for all positive integers n.*

Proof. The lower bound follows from Lemma 2.1. For the upper bound we have the following cases:

- When $m > n \log n$ the claim follows from Theorem 3.4.
- When $m < \frac{n \log n}{2^{12}}$ the claim follows from (4).
- If $\frac{n \log n}{2^{12}} \leq e \leq n \log n$ then we use the inequality $g(n, x) \leq 2g(n/2, x) + (n/2)^2$ iterating it a constant number of times until we are able to use Theorem 3.4. This inequality holds as one can construct an (n, x)-universal graph by taking two disjoint copies of a $(n/2, x)$-universal graph and making every two vertices from different copies adjacent. $\qquad\square$

4. Concluding remarks

In this paper, we complete the study of the behaviour of $f(n, e)$ defined as the maximum size of an (n, e)-unavoidable graph.

As already mentioned in the introduction, in order to get a lower bound on $f(n, e)$, we want to find a graph H with $f(n, e)$ edges and Turán number $ex(n, H) \leq e$. Inverting this statement, we

want to find graphs with fixed number of edges which minimise the Turán number. The universal graph we used in the proof of Theorem 3.4 consisted of a collection of vertices joined to every other vertex and a random graph on the remainder. By transferring this result into unavoidability language, as discussed at the end of Section 3.1, we obtain that whp the appropriate random graph $\mathcal{G}(n/4, q)$ is the desired minimiser of the Turán number, provided $1/2 > q > n^{-1/2+\varepsilon}$. Being a bit more careful with our estimates one can even obtain an almost spanning version of this result. Namely we can get:

Theorem 4.1. *Let $\varepsilon > 0$. There is a $\delta > 0$ such that for $G \sim \mathcal{G}\big((1-\varepsilon)n, q\big)$ we have whp that*
$$ex(n, G) = \left(1 - \Theta\left(\frac{\log n}{nq}\right)\right)\binom{n}{2}, \text{ provided } n^{-1/2+\varepsilon} < q < \delta.$$

Note that for this range of q the chromatic number of the above random graph G whp satisfies $\chi(G) = \Theta\left(\frac{nq}{\log n}\right)$. Interestingly, this shows that $ex(n, G)$ behaves essentially the same as we would expect from the Erdős–Stone–Simonovits theorem [19, 18], i.e. $ex(n, G) = \left(1 - \frac{\Theta(1)}{\chi(G)}\right)\binom{n}{2}$, despite the fact that G is almost spanning! In contrast, in order for the Erdős–Stone–Simonovits theorem to apply, the size of the host graph is required to be significantly larger than the graph being embedded. The question of the exact requirement on the parameters was considered by Bollobás [7] and Chvátal and Szemerédi [15] who showed that the best one can hope for in general is that the Erdős–Stone–Simonovits theorem holds for graphs of order $O(\log n)$, even in our approximate sense.

It could be interesting to determine whether the above theorem extends for values of q smaller than $n^{-1/2}$. The main obstacle for our argument is that our current embedding strategy requires too strong a domination property, which in particular is no longer satisfied by $\mathcal{G}(n, q)$. In light of this, it might be interesting to try to find a weaker version of our property, which would suffice for our embedding argument but is still satisfied by the sparser random graphs. Another possible benefit of such a weaker property is that it could possibly allow one to construct explicit universal graphs which one could use for our argument and answer a question of Alon and Asodi [1] and later Hetterich, Parczyk and Person[22].

Several bounded-degree analogues of our universality problem arise quite naturally. Alon and Capalbo [3] show that the minimal number of edges in a graph which is universal for the family $\mathcal{E}(n, d)$ of n-vertex graphs of maximum degree at most a constant d is $\Theta\left(n^{2-2/d}\right)$. This is asymptotically very different from $g(n, dn) = \Theta\left(\frac{n^2}{\log^2 n}\right)$ which we get from Theorem 1.2. We determine $g(n, dn)$ up to a constant factor for any values of n and d, even if we allow d to depend on n. However, very little seems to be known about the above bounded-degree problem if one allows d to grow with n. In particular, what is the minimal number of edges in an $\mathcal{E}(n, d)$-universal graph when d is allowed to depend on n. Another, possibly an even closer analogue, is what happens if we consider the spanning variant of this problem. So, what is the smallest number of edges in a graph on exactly n vertices which is $\mathcal{E}(n, d)$-universal, where d is allowed to depend on n?

Acknowledgements

References

[1] Alon, N. and Asodi, V. (2002) Sparse universal graphs. *J. Comput. Appl. Math.* **142**(1) 1–11.
[2] Alon, N. and Capalbo, M. (2007) Sparse universal graphs for bounded-degree graphs. *Random Struct. Algorithms* **31**(2) 123–133.

[3]   Alon, N. and Capalbo, M. (2008) Optimal universal graphs with deterministic embedding. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, ACM, New York, pp. 373–378.

[4]   Alon, N., Krivelevich, M. and Sudakov, B. (2007) Embedding nearly-spanning bounded degree trees. *Combinatorica* **27**(6) 629–644.

[5]   Babai, L., Chung, F. R. K., Erdős, P., Graham, R. L. and Spencer, J. H. (1982) On graphs which contain all sparse graphs. In *Theory and Practice of Combinatorics*, Vol. 60 of *North-Holland Mathematics Studies*, pp. 21–26.

[6]   Bhatt, S. N., Chung, F. R. K., Leighton, F. T. and Rosenberg, A. L. (1989) Universal graphs for bounded-degree trees and planar graphs. *SIAM J. Discrete Math.* **2**(2) 145–155.

[7]   Bollobás, B. and Erdős, P. (1973) On the structure of edge graphs. *Bull. London Math. Soc.* **5** 317–321.

[8]   Chung, F. R. K. (1983) Unavoidable stars in 3-graphs. *J. Combin. Theory Ser. A* **35**(3) 252–262.

[9]   Chung, F. R. K. (1997) Open problems of Paul Erdős in graph theory. *J. Graph Theory* **25**(1) 3–36.

[10] Chung, F. R. K. and Erdős, P. (1983) On unavoidable graphs. *Combinatorica* **3**(2) 167–176.

[11] Chung, F. R. K. and Erdős, P. (1987) On unavoidable hypergraphs. *J. Graph Theory* **11**(2) 251–263.

[12] Chung, F. R. K., Erdős, P. and Graham, R. L. (1981) Minimal decompositions of graphs into mutually isomorphic subgraphs. *Combinatorica* **1**(1) 13–24.

[13] Chung, F. R. K. and Graham, R. L. (1979) On universal graphs. *Ann. New York Acad. Sci.* **319** 136–140.

[14] Chung, F. R. K. and Graham, R. L. (1983) On universal graphs for spanning trees. *J. London Math. Soc. (2)* **27**(2) 203–211.

[15] Chvátal, V. and Szemerédi, E. (1981) On the Erdös-Stone theorem. *J. London Math. Soc. (2)* **2** 207–214.

[16] Conlon, D., Ferber, A., Nenadov, R. and Škorić, N. (2017) Almost-spanning universality in random graphs. *Random Struct. Algorithms* **50**(3) 380–393.

[17] Duke, R. A. and Erdős, P. (1977) Systems of finite sets having a common intersection. In *Proceedings. 8th SE Conference on Combinatorics, Graph Theory and Computing*, pp. 247–252.

[18] Erdős, P. and Simonovits, M. (1966) A limit theorem in graph theory. *Studia Sci. Math. Hungar.* **1** 51–57.

[19] Erdös, P. and Stone, A. H. (1946) On the structure of linear graphs. *Bull. Amer. Math. Soc.* **52** 1087–1091.

[20] Füredi, Z. (1991) Turán type problems. In *Surveys in Combinatorics, (Guildford, 1991)*, Vol. 166, *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, pp. 253–300.

[21] Füredi, Z. and Simonovits, M. (2013) The history of degenerate (bipartite) extremal graph problems. In *Erdös Centennial*, Vol. 25 of *Bolyai Society Mathematical Studies*, János Bolyai Mathematical Society, Budapest, pp. 169–264.

[22] Hetterich, S., Parczyk, O. and Person, Y. (2016) On universal hypergraphs. *Electron. J. Combin.* **23**(4) P4.28.

[23] Keevash, P. (2011) Hypergraph Turán problems. In *Surveys in Combinatorics 2011*, *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, pp. 83–139.

[24] Montgomery, R. (2019) Spanning trees in random graphs. *Adv. Math.* **356** 1–92.

[25] Sidorenko, A. (1995) What we know and what we do not know about Turán numbers. *Graphs Comb.* **11**(2) 179–199.

[26] Sudakov, B. (2010) Recent developments in extremal combinatorics: Ramsey and Turán type problems. In *Proceedings of the International Congress of Mathematicians, Volume IV*, Hindustan Book Agency, New Delhi, pp. 2579–2606.

[27] Turán, P. (1941) On an extremal problem in graph theory. *Mat. Fiz. Lapok* **48** 436–452.