ON EXISTENTIALLY COMPLETE TRIANGLE-FREE GRAPHS

SHOHAM LETZTER AND JULIAN SAHASRABUDHE

Abstract. For a positive integer k, we say that a graph is k-existentially complete if for every $0 \leqslant a \leqslant k$, and every tuple of distinct vertices $x_1, \ldots, x_a$, $y_1, \ldots, y_{k-a}$, there exists a vertex z that is joined to all of the vertices $x_1, \ldots, x_a$ and to none of the vertices $y_1, \ldots, y_{k-a}$. While it is easy to show that the binomial random graph $G_{n,1/2}$ satisfies this property with high probability for $k = (1 - o(1)) \log n$, little is known about the "triangle-free" version of this problem: does there exist a finite triangle-free graph G with a similar "extension property"? This question was first raised by Cherlin in 1993 and remains open even in the case $k = 4$.

We show that there are no k-existentially complete triangle-free graphs on n vertices with $k > \frac{8 \log n}{\log \log n}$, for n sufficiently large. This gives the first non-trivial, non-existence result on this "old chestnut" of Cherlin. We believe that this result breaks through a natural barrier in our understanding of the problem.

1. Introduction

If one constructs a graph on vertex set $\mathbb{N}$ by flipping a fair, independent coin for each possible edge $\{i, j\}$ then one has constructed, with probability 1, a unique graph (up to isomorphism) which is known as the *Rado graph*. This curious object, of interest to logicians and combinatorialists alike [1, 4, 11], has the following important "universal property": the Rado graph is the unique countable graph G into which any countable graph H can be "greedily" embedded[1].

This property is best thought of as a consequence of the fact that the Rado graph is the unique countable graph with the *k-extension property* for all k. For an integer $k \in \mathbb{N}$, say that a graph has the k-extension property if for every $0 \leqslant a \leqslant k$ and every tuple of distinct vertices $x_1, \ldots, x_a$, $y_1, \ldots, y_{k-a}$ there exists a vertex adjacent to all of $x_1, \ldots, x_a$ and none of $y_1, \ldots, y_{k-a}$.

Interestingly, the Rado graph can be "approximated" by finite graphs in the sense that for every $k \in \mathbb{N}$, there exist finite graphs that have the k-extension property. Indeed, for $p \in (0, 1)$, we define the *binomial random graph* $G_{n,p}$ to be the probability space defined on all graphs with vertex set $[n]$, where the edge $\{i, j\}$ is included with probability p, independently of all other edges. It is not hard to see that a graph G sampled from $G_{n,1/2}$ has the k-extension property with $k = (1 - o_n(1)) \log_2 n$, with probability $1 - o_n(1)$, as n tends to infinity[2].

A fascinating analogue of the Rado graph is the Rado graph for the class of triangle-free graphs (this graph sometimes sports the title "the universal homogenous triangle-free graph"). More technically, there is a unique countable, triangle-free graph G into which every countable, triangle-free graph H can be "greedily" embedded. While a simple "random" construction is not available to us, the construction of the triangle-free Rado graph is easy; the graph is built up in stages, starting from a single vertex $\{v_0\} = G_0$ we define $G_{i+1} \subseteq G_i$ by adding a vertex with neighbourhood $I \subseteq G_i$, for all independent sets I in G_i. Now define $G = \cup_{i \geqslant 1} G_i$.

[1]This means that if a finite number of vertices of a countable graph H have been embedded into the Rado graph, one can always find further vertices to extend the embedding to all of H.

[2]Here we use the notation $o_n(1)$ to denote a quantity that tends to 0 as n tends to infinity.

Again, the key behind this special embedding property is a similar extension property: say that a graph has the *k-triangle-free extension property* if for every $0 \leqslant a \leqslant k$ and every tuple of distinct vertices $x_1, \ldots, x_a$, $y_1, \ldots, y_{k-a}$ there exists a vertex adjacent to all of $x_1, \ldots, x_a$ and none of $y_1, \ldots, y_{k-a}$, *provided* $x_1, \ldots, x_a$ form an independent set. In analogy with the Rado graph, this graph has the k extention property for all k. We will say a graph with the k-triangle-free extension property is called *k-existentially complete triangle-free* (and henceforth k-ECTF).

The question of whether there exist *finite* graphs that approximate the *triangle-free* Rado graph was raised and studied by Cherlin in 1993 [2, 3] in the context of logic and model theory and has recently made its way over to combinatorics by way of Even-Zohar and Linial [8]. More preceisley, Cherlin asked if there exist finite k-ECTF graphs for every fixed $k \in \mathbb{N}$. To date, this problem remains poorly understood [3] and the state-of-the-art can be summarized as follows. The case $k = 1$ is trivial; a graph is 2-ECTF if and only if it is maximal triangle-free, twin-free and not a cycle on five vertices or a single edge; there are various (non-trivial) constructions for 3-ECTF graphs [2, 3, 8, 9]; and the case $k = 4$ is open.

Our belief is along the lines of Even-Zohar and Linial, who have conjectured that no such graphs exist for $k \geqslant k_0$, where $k_0 \in \mathbb{N}$. In the present paper we take a step towards this conjecture by giving a non-trivial restriction on the maximum possible value of k, relative to n, the number of vertices in the graph. To this end, let $f(n)$ be the largest integer k for which there exists a k-ECTF graph on n vertices. We first note that an easy argument reveals that $f(n) \leqslant \log n$, for sufficiently large n. Indeed, if G is k-ECTF with $k > \log n$, let I be an independent set in G of size $\ell = \min\{k, \lceil \log n \rceil\}$ (such a set always exists in a triangle-free graph - see Lemma 4) then for every subset $S \subseteq I$ there must exist a vertex v_S in G so that v_S is joined to all vertices in S and no vertices in $I \setminus S$. Each such vertex v must be distinct and thus $2^\ell \leqslant n$.

Our main result gives an asymptotic improvement over this estimate, thereby giving a first non-trivial restriction on $f(n)$.

Theorem 1. *Let $n \in \mathbb{N}$ be sufficiently large. There do not exist k-ECTF graphs on n vertices, with $k > \frac{8 \log n}{\log \log n}$. That is, $f(n) = O\left(\frac{\log n}{\log \log n}\right)$.*

One might interpret Theorem 1 as giving the first concrete evidence that the triangle-free version of the problem is substantially different than the problem without the restriction on triangles. Indeed recall that, with high probability, G sampled from $G_{n,1/2}$ is k-existentially complete with $k = (1-o_n(1)) \log n$ and thus essentially matches the trivial bound of $\log n$, which can be proved as above (here it suffices to pick an arbitrary set, rather than an independent one, of size $\min\{k, \lceil \log n \rceil\}$). Theorem 1 also makes a concrete step towards showing the non-existence of finite k-ECTF graphs. We should mention that there have been other non-existance results [3] for k-ECTF, but these have only been shown for graphs possessing a strong symmetry property - so called "strongly-regular graphs".

We point out that a related "extension property" for triangle-free graphs was raised and studied by Erdős and Fajtlowicz [5] and later by Pach [9]. In particular, they studied graphs with the property that every independent set of size at most k has a common neighbour, a one-sided version of the k-TFEC property. While it is conjectured that such graphs should have strong structural charateristics, little is known except in this case where k is large: Pach [9] gave a classification of triangle-free graphs where *all* independent sets have a common neighbour. This direction was furthered by Erdős and Pach [6] who showed that if G is a triangle-free graph with the property that every independent of size $k \leqslant \log n$ has a common neighbor then G has minimum degree at least $\frac{n+1}{3}$.

2. Proof of Main Theorem

2.1. **Proof motivation and Sketch.** As one might be lead to believe from the coin-flipping construction of the Rado graph, we proceed with the vague intuition that a k-ECTF graph must look random-like (in a sense).

Indeed, if we knew that our graph really looked locally like the binomial random graph, we could argue as follows (we intentionally use the word "locally" rather vaguely here) . Given a k-ECTF graph with large k, we start by finding a bipartite graph $H = (A, B, E)$ in G with the property that for every $1 \leqslant a \leqslant k$, and every distinct $x_1, \dots, x_a$ $y_1, \dots, y_{k-a} \in A$ there is a vertex in B that is joined to all of $x_1, \dots, x_a$ and none of $y_1, \dots, y_{k-a}$. So while the k-tuples in A are "taken care of", we turn our attention to how the neighborhoods of the graph cover "cross independent sets", independent sets of the form $A' \cup B'$, where $A' \subset A$ and $B' \subset B$. Now, if it were the case that A, B were roughly of the same size and the graph between A and B looked random, then we should expect to find many cross independent sets of size k that cannot be extended by much. That is, we could find lots of k-tuples $A' \cup B'$ for which there are no largeish sets $A'' \supset A'$ and $B'' \supset B'$ for which $A'' \cup B''$ is also independent. We now observe that if a vertex $v \in V(G) \setminus V(H)$ covers our cross independent k-tuple $A' \cup B'$ it cannot cover too many more such tuples by the restriction on triangles. We would now conclude that it is impossible for G to be k-ECTF for there are not enough vertices in the graph to cover all such cross independent sets of size k.

Now, this is what we would do if things really *did* look random between A and B, but in reality, we have little control over the relative sizes of A and B, and little control over the local densities (as one has in standard notions of pseudo-randomness). The idea here is to find a more subtle notion of the "size" (or rather of the *measure*) of a subset in the bipartite graph H. In particular, we define a measure on subsets of B that will give large weight to sets that cover many k-tuples in A.

Beyond the definition of our special measure, there are two main ingredients, captured in Lemmas 2 and 3 that go into the proof of Theorem 1. Lemma 2 is ultimately used to say that "large" neighborhoods are needed to cover many k-tuples. In fact, this notion of "large" is generalized to an arbitrary probability measure, which we will apply to our special measure. The second ingredient, Lemma 3, says that if a set has large measure (with respect to our special measure), then it must expand quite a bit, in the sense of having many neighbors.

We can now sketch the proof. Given our bipartite graph $H = (A, B, E)$ as above, we have sets $B'' \subseteq B$, that have large mass in our covering measure. But there are still many independent sets (for reasons we do not go into here) of size k which have the form $A' \cup B'$ and $A' \subseteq A$, $B' \subseteq B''$. Now a vertex v which contains $A' \cup B'$ in its neighbourhood cannot cover too many more such cross independent sets as the edges of B'' are expanding and so v cannot join to many vertices in A. The conclusion is then the same as in the toy problem (when we were assuming everything to be random like): we arrive at a contradiction as the graph would need more than n vertices to simultaneously cover all these cross independent sets.

2.2. **A few lemmas.** Given a finite set X, we say that μ is a *probability measure on* X if $\mu : \mathcal{P}(X) \to [0, 1]$ where $\mu(A) = \sum_{x \in A} \mu(\{x\})$, for all $A \subset X$ and $\mu(X) = 1$.

For a graph $G = (V, E)$, and disjoint subsets $X, Y \subseteq V$, let $G[X, Y]$ denote the *induced bipartite graph* on vertex set $X \cup Y$, with bipartition $\{X, Y\}$, and $x \in X$ adjacent to $y \in Y$ if and only if $xy \in E$.

Let G be a bipartite graph with vertex partition $\{A, B\}$. For $s, t \in \mathbb{N}$, we say G is *(s,t)-separating for A* if, for every pair of disjoint subsets $S, T \subseteq A$ with $|S| \leqslant s$ and $|T| \leqslant t$, there exists a vertex $v \in B$ so that v is joined to all the vertices in S and none of the vertices in T.

It is easy to see that if $k \in \mathbb{N}$ and $G = (A, B, E)$ is a bipartite graph which is (k,k)-separating for A, where $|A| \geqslant k$, then $|B| \geqslant 2^k$. The following lemma, gives a strengthened bound when we impose a restriction on the neighbourhoods of vertices in B.

Lemma 2. *For $k \in \mathbb{N}$, let G be a bipartite graph with bipartition $\{A, B\}$ with $|A|, |B| \geqslant 1$, and let μ be a probability measure on A. If G is $(k,0)$-separating for A and $\mu(N(x)) < \varepsilon$ for each $x \in B$, then $|B| > 1/\varepsilon^k$.*

Proof. Sample the points $x_1, \dots x_k \in A$ independently at random and according to the distribution μ. Then

$$\begin{aligned} 1 = {} & \mathbb{P}\left(x_1, \dots, x_k \in N(x) \text{ for some } x \in B\right) \\ \leqslant {} & \sum_{x \in B} \mathbb{P}\left(x_1, \dots, x_k \in N(x)\right) \\ = {} & \sum_{x \in B} \mu(N(x))^k < |B|\varepsilon^k, \end{aligned}$$

thus completing the proof. □

For $s, t \in \mathbb{N}$, let $G = (A, B, E)$ be a bipartite graph that is (s,t)-separating for A. We now define a measure on B that measures how well a given subset of B covers the s-tuples of A. In particular, define the *covering measure* $\mu_{G,s,A}$, with respect to G, by defining a way of sampling it: first sample $X_1, \dots, X_s \in A$ independently and uniformly from A. Then, uniformly at random, choose a vertex among all vertices $v \in B$ so that $X_1, \dots, X_s \in N(v)$. A key property of this measure is that for every $B' \subseteq B$, we have that

$$\mu_{G,s,A}(B') \leqslant \mathbb{P}(X_1, \dots, X_s \in N(x), \text{ for some } x \in B'). \tag{1}$$

Here $\mathbb{P}$ denotes the uniform measure on A for the $X_1, \dots, X_s$. The following lemma says that if $G = (A, B, E)$ is $(s,0)$-separating for A and a set $B' \subset B$ is given large mass by $\mu_{G,s,A}$, then the neighbourhoods of $x \in B'$ "expand" and collectively cover many vertices of A.

Lemma 3. *For $k \in \mathbb{N}$, let $G = (A, B, E)$ be a bipartite graph which is $(k,0)$-separating for A and let $\mu = \mu_{G,k,A}$ be the covering measure defined on B. If $B' \subseteq B$ has $\mu(B') > \varepsilon$ for some $\varepsilon > 0$, then*

$$\left| \bigcup_{x \in B'} N(x) \right| \geqslant \left(1 - \frac{1}{k}\log\left(\varepsilon^{-1}\right)\right)|A|.$$

Proof. Write $\left|\bigcup_{x \in B'} N(x)\right| = (1-\eta)|A|$ for some $0 < \eta < 1$. Then if $X_1, \dots, X_k$ are sampled independently and uniformly from A, we have

$$\begin{aligned} & \mathbb{P}(X_1, \dots, X_k \in N(x) \text{ for some } x \in B') \\ \leqslant {} & \mathbb{P}\left(X_1, \dots, X_k \in \bigcup_{x \in B'} N(x)\right) \\ \leqslant {} & (1-\eta)^k \leqslant e^{-k\eta}. \end{aligned} \tag{2}$$

Now apply the observation at (1) to (2) to obtain the inequality

$$\varepsilon < \mu(B') \leqslant \mathbb{P}\left(X_1, \dots, X_k \in N(x) \text{ for some } x \in B'\right) \leqslant e^{-k\eta}.$$

Taking logarithms gives $\eta < \frac{1}{k}\log\left(\varepsilon^{-1}\right)$, as desired. □

We also require a basic fact about triangle-free graphs, which is a special case of the quantitative form of Ramsey's theorem [10], first obtained by Erdős and Szekeres [7].

Lemma 4. *Every triangle-free graph on n vertices contains an independent set of size $\geqslant \lfloor\sqrt{n}\rfloor$*

Proof. If G contains a vertex of degree at least $\lfloor\sqrt{n}\rfloor$ then the neighbourhood of this vertex is an independent set and we are done. Otherwise, all neighbourhoods are of size at most $\lfloor\sqrt{n}\rfloor - 1$. In this latter case we may greedily construct an independent set of size $\sqrt{n}$. □

2.3. **Proof of Theorem 1.** We are now in a position to give the proof of our main theorem. For a vertex $x \in V(G)$, we shall use $N(x) = \{y : xy \in E(G)\}$ to denote the set ofvertices adjacent to x and for a subset $B \subseteq V(G)$ we denote $N_B(x) = B \cap N(x)$. Our logarithms are always taken in base 2.

Proof of Theorem 1. Suppose that G is a $2k$-ECTF graph on n vertices with $k \geqslant \frac{4\log n}{\log\log n}$. To reduce clutter, let $\ell = \lceil \frac{2\log n}{\log\log n}\rceil$ and let ε be such that $\log\varepsilon^{-1} = \frac{\log\log n}{4}$ so that $\frac{1}{\varepsilon^k} = n$. Fix an independent set $I \subseteq V(G)$ with $|I| \geqslant \lfloor\sqrt{n}\rfloor$ and choose $x_0 \in I$. Then set $J = I \setminus \{x_0\}$. We define a procedure that will discover a collection of more than n distinct vertices in G, thus giving a contradiction. Let us set $\alpha = \frac{4}{\ell}\log\varepsilon^{-1}$ and note that

$$\alpha = \frac{4}{\ell}\log\varepsilon^{-1} = (1+o(1))\frac{(\log\log n)^2}{2\log n}.$$

From this we derive the inequality

$$\alpha^{-\ell} > n. \tag{3}$$

To see this, take a logarithm of the left-hand-side

$$\begin{aligned}\ell\log\alpha^{-1} &= \frac{2\log n}{\log\log n}\log\left((1+o(1))\frac{2\log n}{(\log\log n)^2}\right)\\ &= (2-o(1))\log n,\end{aligned}$$

which is at least the logarithm of the right-hand-side, for sufficiently large n. We also note the inequality

$$\frac{\alpha}{2} + \frac{\ell}{\sqrt{n}-2} \leqslant \alpha, \tag{4}$$

which holds for n sufficiently large.

We prove the following statement by induction on $t \in [0, n+1]$: for each $t \in [0, n+1]$ we may find vertices $w_1, \ldots, w_t \in V(G)$ and a set $L_t \subseteq J^\ell$ so that the following conditions hold.

(1) The vertices $w_1, \ldots, w_t$ are distinct.
(2) If $(v_1, \ldots, v_\ell) \in L_t$, then $\{v_1, \ldots, v_\ell\}$ is not contained in any of the neighbourhoods $\{N(w_i)\}_{i=1}^t$. That is,
$$(v_1, \ldots, v_\ell) \notin \bigcup_{i=1}^{t}(N(w_i))^\ell.$$
(3) We have $|L_t| \geqslant \left(1 - t\alpha^\ell\right)|J|^\ell$.

For the basis step ($t = 0$), set $L_0 = J^\ell$. In this case, Items (1) and (2) of the induction hypothesis vacuously hold while Item (3) holds by definition. Now assume that $t \geqslant 1$ and that we have defined distinct vertices $w_1, \dots, w_{t-1}$ and a set L_{t-1} satisfying the above. We show that we may find appropriate w_t and L_t.

Note that $|L_{t-1}| \geqslant 1$, as $|L_{t-1}| \geqslant |J|^\ell(1-(t-1)\alpha^\ell) \geqslant |J|^\ell(1-n\alpha^\ell) > 0$, as $\alpha^{-\ell} > n$, by the inequality at (3). So we may fix $y_1, \dots, y_\ell \in J$ so that $(y_1, \dots, y_\ell) \in L_{t-1}$. Define $B \subseteq V(G)$ to be the collection of vertices in G that are adjacent to x_0 and not adjacent to any of $y_1, \dots, y_\ell$. Note that since each vertex in B joins to x_0, B is an independent set. Now put $A = I \setminus \{x_0, y_1, \dots, y_\ell\}$ and consider $G[A, B]$ (see Figure 2.3 for a depiction of the sets mentioned here). Observe that $G[A, B]$ is (ℓ, ℓ)-separating for A; indeed, for any choice of distinct $a_1, \dots, a_\ell, b_1, \dots, b_\ell \in A$, there is a vertex in G that is joined to all of $x_0, a_1, \dots, a_\ell$ and to none of $b_1, \dots, b_\ell, y_1, \dots, y_\ell$ (because G is $2k$-ECTF, and $2k \geqslant 3\ell + 1$), and such a vetex is in B by definition. Let $\mu = \mu_{G[A,B],\ell,A}$ be the covering measure defined on B, with respect to the bipartite graph $G[A, B]$.

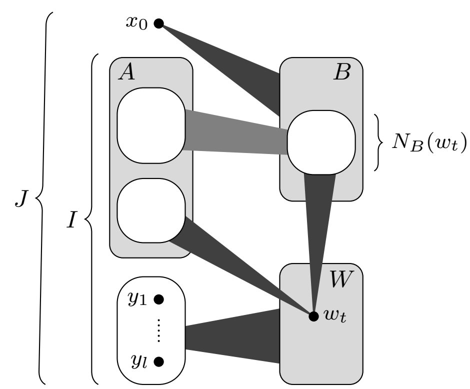

FIGURE 1. Picking w_t

Define W to be the set of vertices in G that *are* joined to *all* of $y_1, \dots, y_\ell$. Note that the graph $G[B, W]$ is (ℓ, ℓ)-separating for B, as there are no edges between $y_1, \dots, y_\ell$ and B and B is an independent set in G. We now claim that there exists a vertex $w \in W$ with $\mu(N_B(w)) > \varepsilon^2$. Suppose to the contrary that $\mu(N_B(x)) < \varepsilon^2$ for all $x \in W$. Since $G[B, W]$ is (ℓ, ℓ)-separating for B, we may apply Lemma 2 to learn that $|W| > \frac{1}{\varepsilon^k} = n$, which is a contradiction.

So we may choose some $w \in W$ with $\mu(N_B(w)) \geqslant \varepsilon^2$ and apply Lemma 3 to learn that

$$\left| \bigcup_{x \in N_B(w)} N_A(x) \right| \geqslant \left(1 - \frac{2}{\ell}\log\left(\varepsilon^{-1}\right)\right)|A|. \tag{5}$$
$$= (1 - \alpha/2)\,|A|.$$

The key here is that w is not adjacent to any of the vertices in the union on the left hand side of (5), as this would create a triangle. Thus, (5) tells us that w is adjacent to at most $\alpha|A|/2$ vertices in A and thus w is adjacent to at most $\alpha|A|/2 + \ell$ vertices in J. Thus the number of ℓ-tuples that

w covers in J is at most

$$(\alpha|A|/2+\ell)^\ell = |J|^\ell \left(\frac{\alpha|A|}{2|J|}+\frac{\ell}{|J|}\right)^\ell \leqslant |J|^\ell \left(\frac{\alpha}{2}+\frac{\ell}{\sqrt{n}-2}\right)^\ell \leqslant (\alpha|J|)^\ell \tag{6}$$

Here we have used the inequality $|J| = |I|-1 \geqslant \lfloor\sqrt{n}\rfloor - 1$ and the inequality at (4). So we define $w_t = w$ and set

$$L_t = L_{t-1} \setminus \{(v_1,\ldots,v_\ell) : v_1,\ldots,v_\ell \in N_J(w)\}\,.$$

By induction and the bound at (6) we have $|L_t| \geqslant |J|^\ell\left(1-t\alpha^\ell\right)$. Finally, we note that w_t must be distinct from $w_1,\ldots,w_{t-1}$ as w_t is joined to all of $y_1,\ldots,y_\ell$ which is not true of any of the $w_1,\ldots,w_{t-1}$, by the fact that $(y_1,\ldots,y_\ell) \in L_{t-1}$ and Item (2) in the induction hypothesis.

So, by induction, we have constructed $n+1$ distinct vertices in a n-vertex graph; a contradiction. This implies that there are no t-ECTF graphs with $t = 2k \geqslant \frac{8\log n}{\log\log n}$, thus completing the proof of Theorem 1. □

3. Acknowledgements

We should like to thank Béla Bollobás for introducing us to the problem of Cherlin. The second named author would like to thank Trinity College Cambridge for support. He would also like to thank the Cambridge Combinatorics group for their hospitality. We should also like to thank the anonymous referee for thoughtful comments that improved the presentation of this paper.

References

1. P.J. Cameron, *The random graph*, The Mathematics of Paul Erdős II, Springer, 1997, pp. 333–351.
2. G.L. Cherlin, *Combinatorial problems connected with finite homogeneity*, Contemporary Mathematics **131** (1993), 3–30.
3. ———, *Two problems on homogeneous structures, revisited*, Contemporary Mathematics **558** (2011), 319–416.
4. R. Diestel, I. Leader, A. Scott, and S. Thomassé, *Partitions and orientations of the Rado graph*, Trans. Am. Math. Soc. **5** (2007), 2395–2405.
5. P. Erdős and S. Fajtlowicz, *Maximum degree in graphs of diameter 2*, Networks **1** (1980), 87–90.
6. P. Erdős and J. Pach, *Remarks on stars and independent sets*, Aspects of Topology: In Memory of Hugh Dowker 1912-1982 **93** (1985), 307.
7. P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Mathematica **2** (1935), 463–470 (eng).
8. C. Even-Zohar and N. Linial, *Triply existentially complete triangle-free graphs*, J. Graph Theory **78** (2015), 26–35.
9. J. Pach, *Graphs whose every independent set has a common neighbour*, Discrete Math. **37** (1981), 217–228.
10. F. P. Ramsey, *On a problem of formal logic*, Proceedings of the London Mathematical Society **s2-30** (1930), no. 1, 264–286.
11. J. Spencer, *The strange logic of random graphs*, 1st ed., Springer Publishing Company, Incorporated, 2010.