Induced subgraphs of different sizes of bipartite graphs

Mantas Baksys, Xuanang Chen

September 5th, 2021

Abstract

Throughout this summer research placement, we have investigated the size of the number of induced subgraphs of bipartite graphs. Given a bipartite graph with m edges, we proved that the size of induced subgraph has $\Omega(m/(\log m)^{10})$ edges, optimizing the results from Narayanan, Sahasrabudhe and Tomon. We also introduce the concept of *C-bipartite-Ramsey*, proving that in 'most' cases, these graphs have multiplication tables of $\Omega(e(G))$ in size, which gives evidence to the conjecture that $K_{n,n}$ is the minimiser of the multiplication table on n^2 edges. While reading [2] and [10] we found a small oversight in them and we successfully fixed it afterwards. We also tried to tackle the conjecture that given m edges, one can always find a bipartite graph with a multiplication table of size $o(m)$ as $m \to \infty$.

We would like to thank Dr. Aled Walker for his supervision, encouragement and guidance during this Summer Research Project.

Contents

1 Introduction

This is a write-up for the Summer Research Project undertaken during the 8-week period June 23rd - August 15th 2021. The main purpose of this write-up is to serve as a record of the ideas considered and the results derived.

In 1960, Erdős introduced the multiplication table problem, which asks about the asymptotic order of the size of the set $M(n) = \{x \in \mathbb{N} : x = ab, 0 \leq a, b \leq n, a, b \in \mathbb{N}\}$. Intuitively, one might conjecture that $M(n) = \Omega(n^2)$. However, using the Hardy-Ramanujan theorem (for the exact statement and proof see [11]), Erdős in [12] showed that $M(n) = o(n^2)$. Much later, in 2008, the asymptotic order of magnitude of $M(n)$ was settled by Kevin Ford in [3] using number theoretic techniques. Specifically, we have by [Corollary 3, [3]] that $M(n) = \Theta\left(\frac{n^2}{(\log n)^{\delta}(\log\log n)^{3/2}}\right)$, where $\delta = 1 - (1 + \log\log 2)/\log 2 \approx 0.086$.

The following generalisation of the multiplication table problem is due to Narayanan, Sahasrabudhe and Tomon in [1]. If for any simple graph G we define $\mathcal{M}(G) = \{e(H) : H \text{ is an induced subgraph of } G\}$, then we have $\mathcal{M}(K_{n,n}) = M(n)$, which creates a connection between the graph-theoretic nature of $\mathcal{M}(G)$ and number-theoretic nature of $M(n)$. Now, write $e(G)$ for the number of vertices in a graph G, and $\Phi(G) = |\mathcal{M}(G)|$. It is natural to ask which bipartite graphs G are extremal in terms of the behaviour of $\Phi(G)$. One way to concretely raise this question is by fixing the number of edges in the bipartite graphs under consideration, which leads to the following conjecture, which is formulated in [1]:

Conjecture 1.1. *Let $n \in \mathbb{N}$ and G be a bipartite graph with $e(G) = n^2$. Then $\Phi(G) \geq \Phi(K_{n,n})$.*

In this note, we partially prove the above conjecture for G being a *C-bipartite-Ramsey* graph, which we define as follows:

Definition 1.2. For a bipartite graph $G(X, Y)$, G is *C-bipartite-Ramsey* if it does not contain $K_{a,b}$ or $\overline{K_{a,b}}$ as an induced subgraph for any $a \geq C\log(|X|), b \geq C\log(|Y|)$.

The idea to come up with this definition stems from the definition of a *C-Ramsey* graph, which states that a simple graph G on n vertices is *C-Ramsey* if it does not contain K_m or $\overline{K_m}$ as an induced subgraph for $m \geq C\log n$. This set of graphs is well-studied in the literature and is well-known to satisfy certain quasi-randomness properties. Recently, Kwan and Sudakov set out to prove in [2] the following Theorem:

Theorem 1.3. *For any fixed $C > 0$, and any n-vertex C-Ramsey graph G, we have $|\Phi(G)| = \Omega(n^2)$.*

While reading the aforementioned paper, we found a small oversight in [Lemma 4, [2]], which is later on pivotal in their proof of Theorem 1.3. We were successful at fixing this oversight and a detailed explanation of this correction can be found in Section 4. We also note that our correction is sufficient to fix a very similar issue in another paper [10] by Kwan and Sudakov.

Taking inspiration from Theorem 1.3, we make the following Conjecture:

Conjecture 1.4. *For any fixed $C > 0$, all C-bipartite-Ramsey graphs $G(X, Y)$ with $|X||Y| = m$, we have $\Phi(G) = \Omega(m)$.*

Using similar methods as Kwan and Sudakov, in Section 2, we will prove the following weaker statement of Conjecture 1.4:

Theorem 1.5. *For any C-bipartite-Ramsey graph $G(X, Y)$, if there exists a constant $\alpha > 0$ such that $|X||Y| = m$ and $|X|, |Y| \geq m^{\alpha}$, where $|X|$ and $|Y|$ are dependent on m, then $\Phi(G) = \Omega(m)$.*

Note that $\Phi(G) = \Omega(m)$ is the best result we can get, because of the trivial inequality $\Phi(G) \leq e(G)$, giving that $\Phi(G) = O(m)$ because $e(G) = O(m)$. The last statement follows from the graph G having positive density independent of m, which we'll later prove in Section 2. We also derive a partial density result when $|X|$ is sub-polynomial in m.

Let us give a few remarks about Theorem 1.5. In general, for $m \in \mathbb{N}$, $d \leq \sqrt{m}$, $d|m$, we have $\Phi(K_{d,m/d}) = \Theta\left(\frac{m}{(\log d)^{\delta}(\log\log d)^{3/2}}\right)$. To see this, consider $H(x,y,z)$, the number of positive integers $n \in N$ such that $n \leq x$ and there exists $d \in \mathbb{N}$ such that $d|n$ and $y < d \leq z$. First set $x = m/4, y = d/4, z = d/2$, then set $x = m/2^k, y = d/2^{k+1}, z = d/2^k$ for non-negative integer k. The result follows by applying [Theorem 1(v), [3]] into the inequality

$$H(\frac{m}{4},\frac{d}{4},\frac{d}{2}) \leq \Phi(K_{d,m/d}) \leq \sum_{k\geq 0} H(\frac{m}{2^k},\frac{d}{2^{k+1}},\frac{d}{2^k}),$$

and noting that u is defined in [3] as $y^{1+u} = z$, which means that $u = \log 2/\log y$ in our cases.

Thus, if a bipartite graph G is *C-bipartite-Ramsey* and satisfies the condition in Theorem 1.5, then we can deduce that $\Phi(G)$ is almost surely not the minimiser of those graphs with the same number of edges as G, due to the fact that most positive integers x have a divisor in the range $(\log x, \sqrt{x}]$. This is because [Theorem 3, [6]] states that for $1 \leq y \leq z \leq x$, $H(x,y,z) = x\left(1 + O\left(\log y/\log z\right)\right)$. Now letting $y = \log x$, $z = \sqrt{x}$, we have that $O\left(\log y/\log z\right) = O\left(2\log\log x/\log x\right) = o(1)$, which means that $H(x, \log x, \sqrt{x}) = x(1 + o(1))$. Therefore the numbers $n \in \mathbb{N}$ such that $n \leq x$ and n having a divisor between $\log x$ and $\sqrt{x}$ have asymptotic density 1. From this, the claim that the set of numbers $x \in \mathbb{N}$ with a divisor between $(\log x, \sqrt{x}]$ have asymptotic density 1 is straight-forward, as desired.

The reason why we are interested in *C-bipartite-Ramsey* is because 'most' random bipartite graphs are *C-bipartite-Ramsey*. To illustrate this idea, we give a specific example below. Consider a random bipartite graph $G(n,n)$ with probability $\frac{1}{2}$ of connecting an edge. It is almost surely *C-bipartite-Ramsey* for $C > 5$ as $n \to \infty$: let A be the event that G has $K_{5\log n, 5\log n}$ or $\overline{K_{5\log n,5\log n}}$ as an induced subgraph, let $k = 5\log n$, then

$$\mathbb{P}(A) \leq 2\binom{n}{k}^2\left(\frac{1}{2}\right)^{k^2} \leq \frac{2\left(n^{5\log n}\right)^2}{2^{25(\log n)^2}} \leq 2(\sqrt{e}/2)^{25(\log n)^2} \to 0 \ as\ n \to \infty,$$

where we implicitly used the union bound.

In Section 3, we will give some insights to a conjecture raised in [1] regarding the asymptotic magnitude of the size of the minimal multiplication table on graphs with m edges. We reformulate this conjecture in terms of density of so-called *balanced* numbers and give heuristic arguments regarding the validity of the reformulated question. We also give a concrete construction based on generalized Hardy-Littlewood conjecture on linear configurations of primes.

In Section 5, we list a few miscellaneous results and ideas that we have. We come up with an easy counting example to show that random graphs almost surely has $\Phi(G) \gg n^{3/2-\varepsilon}$ for any $\varepsilon > 0$. Also, we managed to improve the coefficient from 12 to 10 in [Theorem 1.2, [1]] by Narayanan, Sahasrabudhe and Tomon.

1.1 Notation and Basic Definitions

We use standard notation for asymptotic estimates. Floor and ceiling symbols are generally omitted, as we are concerned about the asymptotic magnitude of the functions considered to which floor and ceiling symbols do not have any impact. We also use standard graph theoretic notations. Besides, we write the density of a graph $d(G) = e(G)/\binom{|G|}{2}$ and the density of the bipartite graph $G(X,Y)$ by $d(G) = e(G)/|X||Y|$, where $e(G)$ is the number of edges inside G, and X, Y are two sides of it. Apart from Section 3, we use $|X| = n_1, |Y| = n_2$ or $|X| = f(m) \leq \sqrt{m}, |Y| = m/f(m)$, depending on which look nicer in different occasions. We write $N_U(x) = N(x) \cap U$ to be the neighbourhood of vertex x into vertex set U, and let $d_U(x) = |N_U(x)|$. For a pair of vertices $\boldsymbol{v} = \{v_1, v_2\}$, write $d(\boldsymbol{v}) = d(v_1) + d(v_2)$ to be the size of $N(\boldsymbol{v}) = N(v_1) \cup N(v_2)$, where the union here denotes the multiset union. Also, we work with the symmetric difference of two (multi)sets, defined for (multi)sets A, B by $A \triangle B = (A \setminus B) \cup (B \setminus A)$. Lastly, let $hom(G)$ to be the largest homogeneous set (either independent or complete set) of graph G. The definition of a *C-Ramsey* graph is standard. Say a graph G with m vertices is $C - Ramsey$ if $hom(G) \leq C \log m$.

Also, we say a number $n \in \mathbb{N}$ is k-smooth if the largest prime divisor of n, denoted by $P(n)^+$, satisfies $P(n)^+ \leq k$. Say a number $n \in \mathbb{N}$ is k-balanced if there exists $a, b \geq k$ such that $ab = n$.

2 C-Bipartite-Ramsey Graphs

The main goal of this section is to give a proof of Theorem 1.5. For this, we split this section into a short introduction and three sub-sections. In the first sub-section we give an overview of the proof of Theorem 1.5 and the previous work on which we build on. In the second part, we investigate the density of *C-bipartite-Ramsey* graphs, in the cases when vertex sets have sizes that are either sub-polynomial or polynomial in m. Lastly, in the third sub-section, we finish the proof of Theorem 1.5.

Recall our definition for *C-bipartite-Ramsey* graph given as in Definition 1.2. Our main conjecture of this section (and also this write-up) is Conjecture 1.1. For simplicity, throughout this section we will always assume that for a bipartite graph $G(X,Y)$ we have $|X| \leq |Y|$.

One important immediate observation, which shows the distinction between *C-Ramsey* graph theory and *C-bipartite-Ramsey* theory is the fact that we will allow our vertex sets to vary in size relative to each other. As opposed to *C-Ramsey* graphs, where each vertex can be connected to all of the other vertices, we place the condition of the graphs under consideration to be bipartite, which will severely restrict the degrees of vertices based on their vertex set. It is for this reason, that the balance of sizes of vertex sets X and Y will be important for the proof of Theorem 1.5. Specifically, one reason for this is that our proof strategy will involve constructing many induced subgraphs, whose sizes are sufficiently well-spaced in size, using probabilistic techniques. Note that each vertex $x \in X$ can potentially be connected to $\Omega(|Y|)$ other vertices in Y. Because of this, as the sizes of the vertex sets become more imbalanced, the variance in number of edges, while generating a random vertex set U with expected number of edges l will also increase. This will create additional difficulty in proving Conjecture 1.4, making us instead prove the weaker Theorem 1.5, which nevertheless takes care of 'most' sizes of vertex sets X and Y. Another reason why the imbalance in the sizes of vertex sets complicates our proof is because it becomes harder to prove that these graphs $G(X,Y)$ have density in the interval $(\varepsilon, 1-\varepsilon)$ for some absolute constant $\varepsilon > 0$.

Now, we will introduce some new definitions, adapted to bipartite graphs, which will help us to attain our result. Showing that *C-bipartite-Ramsey* graphs satisfy these definitions will be our way to quantify the quasi-randomness properties of the aforementioned set of graphs. Firstly, we will require a condition stating that 'most' neighbourhoods of the bipartite graph are very different.

Definition 2.1. For $c, \delta > 0$, a bipartite graph $G(X,Y)$ is (c,δ)-bipartite-diverse if for each $x_1 \in X$, there exists at most $|X|^\delta$ vertices $x_2 \in X$ such that $|N(x_1) \triangle N(x_2)| < c|Y|$ and if for each $y_1 \in Y$, there exists at most $|Y|^\delta$ vertices $y_2 \in Y$ such that $|N(y_1) \triangle N(y_2)| < c|X|$.

Secondly, we will also require a very similar condition to the one above, saying that if a pair of vertices has very different neighbourhoods, then 'most' other disjoint pairs of vertices will also have very different neighbourhoods to the first pair.

Definition 2.2. For $\alpha, \delta, \varepsilon > 0$, a bipartite graph $G(X,Y)$ is $(\alpha,\delta,\varepsilon)_2$-bipartite-diverse if for each $\boldsymbol{x} = \{x_1, x_2\} \in \binom{X}{2}$ such that $|N(x_1) \triangle (\overline{N(x_2)} \cap Y)| \geq \alpha|Y|$ there exists at most $|X|^\delta$ other disjoint pairs $\boldsymbol{x'} \in \binom{X}{2}$ such that $|N(\boldsymbol{x}) \triangle N(\boldsymbol{x'})| < \varepsilon|Y|$ and the corresponding statement holds with the roles of vertex sets X and Y reversed.

Lastly, we will require another, slightly more general definition of richness for bipartite graphs, which will be the most important quasi-randomness property of *C-bipartite-Ramsey* graphs. One can think of it as a succinct condition, which will allow us to deduce other diversity results. We will do exactly that later on in Lemma 2.9.

Definition 2.3. For $\gamma, \delta, \varepsilon > 0$, a bipartite graph $G(X,Y)$ is $(\gamma,\delta,\varepsilon)$-bipartite-rich if for each $W_X \subseteq X$ such that $|W_X| \geq \gamma|X|$, there exists at most $|Y|^\delta$ vertices $y \in Y$ such that $|N(y) \cap W_X| < \varepsilon|X|$ or

$|\overline{N(y)} \cap W_X| < \varepsilon|X|$ and the corresponding statement holds with the roles of vertex sets X and Y reversed.

It must be noted that the definitions diversity and richness for *C-bipartite-Ramsey* graphs follow very closely from the definitions of diversity and richness for *C-Ramsey* graphs by Kwan and Sudakov in [2]. The first key distinction is that we adjust the definitions to the bipartite setting by separating the conditions to each of the vertex sets. The second key distinction is that we introduce a third coefficient in our definition of richness, which fixes a small oversight by Kwan and Sudakov in [2]
(look to Section 4 for more details).

Throughout this section, we will often use common probabilistic inequalities, namely Markov's inequality, Chebyshev's inequality, Azuma's inequality and Hoeffding's inequality. The statements and proofs of all of these can be found in [13].

Also, we will make frequent use of Turan's Theorem, and so we will give two different statements of it here:

Theorem 2.4 (Turan's Theorem)**.** *Let G be a graph on n vertices with average degree d_G, then G has an independent set of size* $\dfrac{n}{1+d_G}$.

In practice, it will be hard to find out the average degree of G, so we will instead use the following Corollary of Theorem 2.4:

Corollary 2.5. *Let G be a graph on n vertices with maximum degree $\triangle(G)$, then G has an independent set of size* $\dfrac{n}{1+\triangle(G)}$.

The proof of Corollary 2.5 follows from simply noting that $\triangle(G) \geq d_G$.

Sometimes, we will not have a good understanding of the maximum degree of the given graph G, but we will know the number of edges $e(G)$ within G, in which case we will use the second corollary of Turan's theorem:

Corollary 2.6. *Let G be a graph on n vertices, then G has an independent set of size at least* $\dfrac{n^2}{2e(G)+n}$

Again, Corollary 2.6 can be easily deduced from Theorem 2.4 by multiplying both the numerator and denominator by n and noting that $2e(G) \geq nd_G$.

Part I: Overview of Our Proof

Our proof of Theorem 1.5 based on the following key construction. We claim, we can always construct sufficiently many sets of induced subgraphs, which are well-separated in size, using probabilistic methods. Specifically, it will be sufficient to show that for a given function $g(m)$, in each set of $O(g(m))$ consecutive positive integers in the range $[0, e(G)]$, we can construct $\Omega(g(m))$ induced subgraphs with different sizes in this range. For the immediate implication of the Theorem 1.5, we will need to prove that $e(G) = \Omega(m)$ for *C-bipartite-Ramsey* graphs with vertex sets of polynomial size in m and the result will follow by picking $\Omega(m/g(m))$ different sets of $O(g(m))$ consecutive integers in the aformentioned range, which do not intersect pairwise. In our proof, we will prove this exact claim for a carefully chosen function $g(m)$.

In [2], Kwan and Sudakov proved a very similar claim, although concerning *C-Ramsey* graphs G on n vertices. They showed, that one can construct $\Omega(n^{3/2})$ subgraphs on different sizes in intervals of length $O(n^{3/2})$, which do not intersect pairwise. A well, known fact in *C-Ramsey* theory is that $e(G) = \Omega(n^2)$, which gives the result that for all *C-Ramsey* graphs, $\Phi(G) = \Omega(n^2)$. Kwan and Sudakov followed a construction, based on randomly generating a vertex set U, such that $e(U)$ is not too far away from it's expected value. Then, using richness and diversity of *C-Ramsey* graphs, they construct another vertex set W, whose vertices have similar degree, yet there exist many vertices whose degrees are well-separated. It is exactly these properties of W, which will enable Kwan and Sudakov to construct sufficiently many induced subgraphs with different sizes by considering the values of $e(U \cup Z)$, for different choices of $Z \subseteq W$.

In our proof, we largely follow the same construction as Kwan and Sudakov in [2], yet do this for a different set of graphs. First, it is important to make the trivial observation that bipartite graphs on n vertices have an independent set of size at least $n/2$, which immediately implies that they are not *C-Ramsey*. However, one would expect that given a similar condition as *C-Ramsey*, it should be possible to show that bipartite graphs have 'many' induced subgraphs of different sizes. This is exactly the intuition, which leads us to define the *C-bipartite-Ramsey* property with the hope of proving Theorem 1.5.

As remarked before, the proof strategy remains largely the same as in Kwan and Sudakov, yet many definitions of richness, diversity and etc. need to be non-trivially altered, since we're dealing with a significantly different set of graphs. For our proof, we will also require a statement that *C-bipartite-Ramsey* graphs have density bounded away from 0 and 1, whose proof will rest on Bipartite Ramsey theory, rather than classical Ramsey theorems. Most importantly, as remarked above, we will allow vertex sets of bipartite graphs considered be imbalanced, which will slightly complicate the proof as it will cause higher variance in our probabilistic method of generating a vertex set U with a set number of edges. This will mean that, the other vertex set W will need to be chosen significantly larger in size then using the methods of [2], in order for us to construct sufficiently many subgraphs of different sizes.

Initially, we thought that we could decrease the variance, while generating U by taking all $x \in X$ to be in U in order to make the variance small. But then we realised that this stops us from using any probabilistic arguments later, which involve vertices in Y and we cannot use probabilistic inequalities as we would want. So instead, we introduce a new definition $(\gamma, \delta, \varepsilon)$-bipartite-rich, and choose δ to be sufficiently small. This will not decrease the variance, but will allow us to generate W to be as large as required.

In practice, we will generate W as the union of disjoint vertex sets S, T and Z, all of which will serve different purposes. We will require a condition, that all vertices in W have similar degree into U for better control, yet all degrees of vertices in S into U will be sufficiently larger then degrees of all vertices in T into U. By picking different subsets of $S \cup T$ and all vertices in U, we will already be able to generate many induced subgraphs of different sizes. Finally, we will require Z to have many vertices of different degrees, which will allow us to finish the proof of Theorem 1.5 by adding different vertices of Z to previously generated induced subgraphs. It is important to note that we will only allow vertices of Y to be in W, since Y will generally be much bigger in size. Since our graph G is bipartite, this will give us an extremely valuable condition that there are no edges between vertices in W, which will lead to significant simplification in the proof of Theorem 1.5, when compared to the approach of [2].

Also, note that the method used in the original paper by Kwan and Sudakov [2] only considers the set of vertices with unique degrees into U in order to construct W, which would contain vertices, which are well-separated in degree. We realised that this condition can be significantly relaxed in order to acquire W significantly larger in size.

Lastly, it is important to point out that to keep the size of W as large as possible, sometimes W will need to be taken as a set of disjoint pairs of vertices. This will require a separate definition of diversity, but will not cause any other major changes in the proof. Also, the reader should think of richness and diversity, as special properties of *C-bipartite-Ramsey* graphs. It is precisely these properties that enable us to use probabilistic arguments in order to efficiently find many sets of induced subgraphs, well-separated in size. In the case, when for all $\alpha > 0$, $|X| < m^{\alpha}$ our methods break down because of the presence of the logarithm in the definition of *C-bipartite-Ramsey*. We choose to continue the use of the logarithm from *C-Ramsey* theory, as choosing functions of higher asymptotic magnitude would cause major problems, while deducing the quasi-randomness properties, necessary for our proof. Although, we expect that other methods could be used in this case, when $|X|$ is sub-polynomial to give similar results. Ths is because we expect $\Phi(G)$ to increase in size, as vertex sets become more imbalanced, keeping the edge count of G fixed. One can see this heuristically by noting that as the size of $|X|$ decreases, each vertex in Y is on average connected to a smaller number of vertices in X, allowing for smaller differences in size between induced subgraphs.

Part II: Density of C-bipartite-Ramsey graphs

In [7], Erdős proved the density for *C-Ramsey* graph G is such that $d(G) \in (\varepsilon, 1-\varepsilon)$ for some $\varepsilon > 0$. For *C-bipartite-Ramsey* graph, the proof becomes harder, partially because of the possible imbalance from two sides. We have the following result:

Proposition 2.7. For *C-bipartite-Ramsey* graph $G = G(X, Y)$ with $|X||Y| = m$ and $|X|, |Y|$ dependent on m, it has density between $(\varepsilon, 1-\varepsilon)$ for some $\varepsilon = \varepsilon(C, \log m/\log|X|) > 0$, if:

1. $|X|, |Y| \geq m^\alpha$ for some $\alpha > 0$ independent of m, or
2. $|X| = o(\exp\{\sqrt{\log m}\})$

For convenience, in the proof, let $|X| = n_1, |Y| = n_2$, and $n_1 n_2 = m$. The first condition is equivalent to $\log n_2/\log n_1 \leq M$ for some constant M independent of n_1, n_2. The second condition is equivalent to $(\log n_1)^2 = o(\log n_2)$. We first prove the first case.

Proof of Proposition 2.7.1. When $\log n_2/\log n_1$ is bounded, it is sufficient to prove the statement in equal-sized case. To see this, note that there exists M such that $n_2 < n_1^M$, i.e. $\log n_2 < M \log n_1$. So for any induced subgraph $G(X, Y)$ which has n_1 vertices on both sides, apply the equal-sized result (assume it is true) with $C' = CM$, we know that this induced subgraph has density within $(\varepsilon, 1-\varepsilon)$ for some $\varepsilon = \varepsilon(C', \log m/\log|X|) > 0$. As the density is within $(\varepsilon, 1-\varepsilon)$ for all such induced subgraph, we know the density of $G(X, Y)$ is between $(\varepsilon, 1-\varepsilon)$ as well.

So we now focus on equal-sized case. Consider $G(X, Y)$ with $|X| = |Y| = n$. If $e(G) > \dfrac{n^2}{40}$, the theorem is proved. If not, we have the following claim:

Lemma 2.8. *For any $k \geq 40$, let $G(n, n)$ be a bipartite graph with n vertices on both sides and $e(G) \leq \dfrac{n^2}{k}$, then it contains either $K_{a,a}$ or $\overline{K_{a,a}}$ for $a \gg \dfrac{k\log(n)}{\log(k)}$.*

Proof of Lemma 2.8. Let X and Y be two sides of the bipartite graph. First, pick vertices in X and Y which have degree no more than $5n/k$, call them as $X' = \{x_1, x_2, ..., x_{m_1}\}$ and $Y' = \{y_1, y_2, ..., y_{m_2}\}$. Let $G' = G(X' \cup Y')$ be the subgraph spanned by X' and Y'. By a simple counting argument, we have $m_1, m_2 \geq n/2$.

Let H be the largest independent graph $\overline{K_{l,l}}$ for which l is the largest among all the equal-sized independent subgraph of G'. We can assume that $l < c_1 k \log n/\log k$ for some $c_1 < 1/240$, since otherwise the proof is complete. Let $H = \{x_1, ..., x_l\} \cup \{y_1, ..., y_l\}$ after relabelling. We have $\sum_{i=1}^{l} d(x_i) < 5nl/k < 10m_1 l/k$ and an analogous statement holds for $\{y_1, y_2, ..., y_l\}$. By the same counting argument, at least $m_1/2$ vertices in X' connected to $s = 20l/k$ or less vertices in $\{y_1, y_2, ..., y_l\}$ and a respective statement holds for Y'.

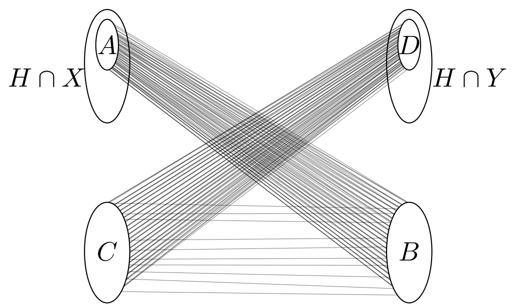
We have $s < c\log n/\log k$, for $c < 1/6$. Let $A_s := \#\{A \subset \{y_1, y_2, ..., y_l\} : |A| < s\}$ be the number of subsets of $\{y_1, y_2, ..., y_l\}$ with fewer than s elements. We then have:

$$A_s \le s\binom{l}{s-1} < s^2\left(\frac{l}{s}\right)^s e^s < s^2\left(\frac{ek}{20}\right)^{20l/k} < s^2 k^{c\log n/\log k} < n^{1/3},$$

where in the second inequality we used Stirling's approximation and later made use of the fact that $c < 1/6$, $s \ll \log n$ and $s - 1 \le l/2$.

By Pigeonhole Principle, there exist $A = \{a_1, a_2, ..., a_{p_1}\} \subset \{x_1, x_2, ..., x_l\}$ with $p_1 < s$ and $B = \{b_1, b_2, ..., b_{q_1}\} \subset Y'\backslash\{y_1, y_2, ..., y_l\}$ with $q_1 > \dfrac{m_2/2 - l}{n^{1/3}} > n^{1/2}$, such that each vertex in B is connected to all vertices in A but no other vertices in $\{x_1, x_2, ..., x_l\}\backslash A$. Similarly, obtain the set $C \subset X'\backslash\{x_1, x_2, ..., x_l\}$ of size $q_2 > n^{\frac{1}{2}}$ and set $D \subset \{y_1, y_2, ..., y_l\}$ of size $p_2 < s$ by reversing the roles of X' and Y' in the previous argument.

Let $G'' = G(B \cup C)$ be the graph spanned by B and C. Then G'' does not have $\overline{K_{s,s}}$ as an independent subset, since if it does, then consider $\overline{K_{s,s}} \cup (\overline{H} \setminus (A \cup D))$ which is an independent subgraph of G with both sides of size more than l, contradicting to the fact that l is the largest size among all the equal-sized independent subgraph of G'.



No induced $\overline{K_{s,s}}$ in $G'' = G(B \cup C)$

Now, we claim that if a bipartite graph $G(\sqrt{n}, \sqrt{n})$ with $\sqrt{n}$ vertices on both sides doesn't contain $\overline{K_{s,s}}$, then it must contain $K_{m,m}$ for some $m \gg \dfrac{ck\log(n)}{\log(k)}$.

From [8], Hattingh and Henning define the Bipartite Ramsey Number to be the smallest integer $b = b(p, q)$ such that any bipartite graph with b vertices on both sides will contain either $K_{p,p}$ or $\overline{K_{q,q}}$. They proved that $b(p, q) \le \binom{p+q}{p}$. By setting $p = ck\log(n)/\log(k)$, $q = c\log(n)/\log(k)$ and applying Stirling's Formula, we have

$$\binom{p+q}{p} = \Theta\left(\frac{\left(\frac{ck\log(n)}{\log(k)}\right)^{\frac{c\log(n)}{\log(k)}} e^{\frac{c\log(n)}{\log(k)}}}{\sqrt{2\pi\frac{c\log(n)}{\log(k)}}\left(\frac{c\log(n)}{\log(k)}\right)^{\frac{c\log(n)}{\log(k)}}}\right) = O\left((ke)^{\frac{c\log(n)}{\log(k)}}\right) = O(n^{c+\frac{c}{\log(k)}}) = o(\sqrt{n}) \qquad (1)$$

Thus, if $G(n, n)$ contains $\overline{K_{a,a}}$ as induced subgraph for $a \gg \dfrac{k\log n}{\log k}$, we're done. If not, we can find its induced subgraph $G'(\sqrt{n}, \sqrt{n})$ which doesn't contain $\overline{K_{s,s}}$ as induced subgraph for $s \le c\log n/\log k$ and any $c < 1/6$. Then by the argument above, for $m \ge ck\log n/\log k$ we have $b(s, m) \le \binom{s+m}{s} = o(\sqrt{n})$, so $G(n, n)$ must contain $K_{m,m}$ as induced subgraph. We're done in both cases. $\square$

From Lemma 2.8 and the argument before, the first claim of Proposition 2.7 is proved. $\square$

Proof of Proposition 2.7.2. Now, we finish the proof of the second claim of Proposition 2.7. When $\log n_2/\log n_1$ is unbounded and $\log n_2/\log n_1 > (\log n_1)^{1+\varepsilon}$ for any $\varepsilon > 0$, applying essentially the same argument in Lemma 2.8, replacing n by n_1, n_2 appropriately, and taking $c = 1/20 < 1/6$ in the proof, what we are left to prove is:

Claim. *If $G(X, Y)$ is a bipartite graph with $|X| = \sqrt{n_1}, |Y| = \sqrt{n_2}$, which does not have $\overline{K_{\log n_1/20\log k, \log n_2/20\log k}}$ as induced subgraph, then it contains $K_{C\log n_1, C\log n_2}$ as induced subgraph.*

To prove the claim, we use a different method. Choose k to be the smallest positive integer such that $k/\log k > C$. From (1), taking $n = n_1$, $c = 1/20$, then $c + c/\log k < n^{1/10}$, so for any induced subgraph $H \subseteq G(X, Y)$ with $n_1^{1/10}$ vertices on both sides, H either contains induced $\overline{K_{s,s}}$ or $K_{m,m}$, for $s = \log n_1/20\log k$, $m = k\log n_1/20\log k$.

There are at least $\sqrt{n_2}/n_1^{1/10}$ many such H in $G(X, Y)$. Firstly, suppose $\sqrt{n_2}/2n_1^{1/10}$ of them contain $K_{m,m}$ as an induced subgraph. By pigeon-hole principle, at least $\log n_2/\log n_1$ such H will have the same vertex set in X among $\log n_2/\log n_1\binom{n_1^{1/10}}{\log n_1/20\log k} + 1 \le n_1^{k\log(n_1)/10\log k}\log n_2/\log n_1$ many of them. Their union contains $K_{k\log n_1/20\log k, k\log n_2/20\log k}$ as induced subgraph, therefore it has $K_{C\log n_1, C\log n_2}$ as an induced subgraph, since $k/\log k > C$. To get $n_1^{k\log(n_1)/10\log k}\log n_2/\log n_1$ many H, the number of vertices in Y should be

$$
\begin{aligned}
n_1^{k\log(n_1)/10\log(k)+1/10}\log(n_2)/\log(n_1) &= \exp\left((\log n_1)^2\frac{k}{10\log k} + \frac{\log(n_2)}{\log(n_1)} + \frac{1}{10}\log(n_1)\right)\\
&\le \exp\left(3\max\left\{(\log n_1)^2\frac{k}{10\log k}, \frac{\log(n_2)}{\log(n_1)}, \frac{1}{10}\log(n_1)\right\}\right)\\
&\le \exp\left(\frac{1}{4}\log n_2\right)\\
&\le \frac{\sqrt{n_2}}{2}
\end{aligned}
$$

The inequality holds as $(\log n_1)^2 = o(\log n_2)$, and $k = k(C)$ is a constant. This is a contradiction to the assumption that $G(X, Y)$ does not have $K_{C\log n_1, C\log n_2}$ as induced subgraph. For the other case, there are at least $\sqrt{n_2}/2n_1^{1/10}$ many H has $K_{m,m}$ as induced subgraph. Apply the same argument again, we would get the result that $G(X, Y)$ has $\overline{K_{\log n_1/20\log k, \log n_2/20\log k}}$ as an induced subgraph. We are done in both cases. □

We are unable to show the similar density result when $\exp\{\sqrt{\log m}\} \ll f(m) \ll m^\alpha$, yet we suspect the result will also hold in this case.

Part III: Proof of Theorem 1.5

As promised before, we now deduce other diversity conditions from richness.

Lemma 2.9. *Assume a bipartite graph $G(X, Y)$ is $(\gamma, \delta, \varepsilon)$-rich and $\gamma < 1/2$. Then, the following statements all hold:*

1. *G is $(\delta, \varepsilon/2)$-bipartite-diverse.*

2. *G is $(\alpha\varepsilon/2, \delta, \alpha)_2$-bipartite-diverse for $\alpha \ge 2\gamma$.*

3. *There are at most $|Y|^{1+\delta}$ disjoint pairs of vertices $\boldsymbol{y} = \{y_1, y_2\} \in \binom{Y}{2}$ such that the inequality $|N(y_1)\triangle(\overline{N(y_2)} \cap X)| < \varepsilon|X|/2$ holds, and there are at most $|X|^{1+\delta}$ disjoint pairs of vertices $\boldsymbol{x} = \{x_1, x_2\} \in \binom{X}{2}$ such that $|N(x_1)\triangle(\overline{N(x_2)} \cap Y)| < \frac{\varepsilon|Y|}{2}$ holds.*

Proof. Without loss of generality, we prove each statement only for vertices of X only, as we can conclude the full proof by exchanging the roles of X and Y. For the first statement, note that for each $x_1 \in X$, either $|N(x_1)| \geq |Y|/2$ or $|\overline{N(x_1)} \cap Y| \geq |Y|/2$. Since $\gamma < 1/2$, in the former case, there are at most $|X|^\delta$ vertices $x_2 \in X$ such that $|\overline{N(x_2)} \cap N(x_1)| < \varepsilon|N(x_1)| \leq \varepsilon|Y|/2$ and in the latter case, there are at most $|X|^\delta$ vertices $x_2 \in X$ such that $|N(x_2) \cap (\overline{N(x_1)} \cap Y)| < \varepsilon|\overline{N(x_1)} \cap Y| \leq \varepsilon|Y|/2$. In each case, there are at most $|X|^\delta$ vertices $x_2 \in X$ such that $|N(x_1) \triangle N(x_2)| < \varepsilon|Y|/2$, since $\overline{N(x_2)} \cap N(x_1) \subseteq N(x_1) \triangle N(x_2)$ and $N(x_2) \cap (\overline{N(x_1)} \cap Y) \subseteq N(x_1) \triangle N(x_2)$.

For the second statement, let $\boldsymbol{x} = \{x_1, x_2\}$ and note that if $|N(x_1) \triangle (\overline{N(x_2)} \cap Y)| \geq \alpha|Y|$, then either $|N(x_1) \cap N(x_2)| \geq \alpha|Y|/2$ or $|\overline{N(x_1)} \cap \overline{N(x_2)} \cap Y| \geq \alpha|Y|/2$. Without loss of generality, assume that $|N(x_1) \cap N(x_2)| \geq \alpha|Y|/2 \geq \gamma|Y|$, since $\alpha \geq 2\gamma$. For contradiction, suppose that there exists a set Z of at least $|X|^\delta$ disjoint pairs $\boldsymbol{x'} \in \binom{X}{2}$ such that $|N(\boldsymbol{x}) \triangle N(\boldsymbol{x'})| < \alpha\varepsilon|Y|/2$. Then, for each vertex x' in a given $\boldsymbol{x'} \in Z$, we have $|\overline{N(x')} \cap N(x_1) \cap N(x_2)| \leq |N(\boldsymbol{x}) \triangle N(\boldsymbol{x'})| \leq \alpha\varepsilon|Y|/2 \leq \varepsilon|N(x_1) \cap N(x_2)|$, which contradicts $(\gamma, \delta, \varepsilon)$-bipartite-richness, as $|N(x_1) \cap N(x_2)| \geq \gamma|Y|$. Hence, the claim follows.

For the third statement, we will prove that for each of $|X|$ choices in $x_1 \in X$ there are at most $|X|^\delta$ vertices $x_2 \in X$ for which $|N(x_1) \triangle (\overline{N(x_2)} \cap Y)| < \varepsilon|Y|/2$. Again, note that either $|N(x_1)| \geq |Y|/2$ or $|\overline{N(x_1)} \cap Y| \geq |Y|/2$. In the former case, for at most $|X|^\delta$ vertices $x_2 \in X$ such that $|N(x_1) \cap \overline{N(x_2)}| < \varepsilon|N(x_1)| \leq \varepsilon|Y|/2$ and in the latter case, there are again at most $|X|^\delta$ vertices $x_2 \in X$ such that $|\overline{N(x_1)} \cap \overline{N(x_2)} \cap Y| < \varepsilon|\overline{N(x_1)} \cap Y| \leq \varepsilon|Y|/2$. Now note that $N(x_1) \cap \overline{N(x_2)} \subseteq N(x_1) \triangle (\overline{N(x_2)} \cap Y)$ and $\overline{N(x_1)} \cap \overline{N(x_2)} \cap Y \subseteq N(x_1) \triangle (\overline{N(x_2)} \cap Y)$ to have that in each case, the required claim follows. ☐

Now, we will prove that *C-bipartite-Ramsey* graphs satisfy our richness definition.

Lemma 2.10. *For any C-bipartite-Ramsey graph $G(X, Y)$, with $|X| = f(m)$ and $|Y| = m/f(m)$ such that for some $\alpha > 0$, $m^\alpha \leq f(m) \leq \sqrt{m}$, then for any $\delta > 0$, there exists $\gamma = \gamma(C, \delta, \alpha) > 0$ with $\varepsilon = 4\gamma$ such that G is $(\gamma, \delta, \varepsilon) - bipartite - rich$.*

Proof. Suppose, for contradiction that G fails to be $(\gamma, \delta, 4\gamma)$-bipartite-rich for all $\gamma > 0$. As before, it will suffice to prove richness in one direction, and the other direction will be obtained by exchanging the roles of X and Y and making the obvious adjustments. Hence, we can assume that there exists a set $W_X \subseteq X$ such that $|W_X| \geq \gamma|X| \geq \gamma f(m)$ and there exist a set $Z_Y \subseteq Y$ such that $|Z_Y| \geq |Y|^\delta$ contradicting $(\gamma, \delta, 4\gamma)$-bipartite-richness, which means that $\forall v \in Z_Y, |N(v) \cap W_X| < 4\gamma|W_X|$ or $|\overline{N(v)} \cap W_X| < 4\gamma|W_X|$. Without loss of generality, assume that for $T_Y \subseteq Z_Y$ of size $|T_Y| \geq |Y|^\delta/2$ such that $\forall v \in T_Y, |N(v) \cap W_X| < 4\gamma|W_X|$. Now let $h(m) = \min(\gamma m^\alpha, (m/f(m))^\delta/2)$ and note that $h(m) \geq m^\beta$ for some $\beta = \beta(\alpha, \delta) > 0$. By a simple counting argument, there exist subsets $S_X \subset W_X$ and $S_Y \subset Z_Y$ such that $|S_X| = |S_Y| = \sqrt{h(m)}$ for m large enough such that $d(S_X, S_Y) < 8\gamma$. Then, by Lemma 2.8 $G[S_X, S_Y]$ contains either $K_{a,a}$ or $\overline{K_{a,a}}$ as a subgraph with $a \geq \frac{\beta \log m}{3840\gamma \log(1/8\gamma)}$ and choosing γ dependent on C and $\beta = \beta(\delta, \alpha)$ one can find a $\gamma = \gamma(C, \delta, \alpha)$ small enough, so that $a \geq C \log m$, which contradicts the *C-bipartite-Ramsey* property as $G[S_X, S_Y]$ is a subgraph of G and G does not contain $K_{C\log(f(m)), C\log(m/f(m))}$ or $\overline{K_{C\log f(m), C\log(m/f(m))}}$ as subgraph, which gives the desired contradiction. ☐

It will be convenient to deduce Theorem 1.5 from the following Lemma:

Lemma 2.11. *Let $G(X, Y)$ be an C-bipartite-Ramsey graph with $|X| = f(m)$ and $|Y| = m/f(m)$, such that there exists $\alpha > 0$ for which $m^\alpha \leq f(m) \leq \sqrt{m}$. Then there exist $c = c(C, \alpha) > 0$ such that for any l with $cm \leq l \leq 2cm$, there exist disjoint subsets $U \subseteq X \cup Y$ and $W \subseteq Y$, for which $|e(U) - l| = O(\frac{m}{\sqrt{f(m)}})$ and $|W| = O(m/f(m)^{\frac{3}{2}})$. Also, $|\{e(U \cup W') : W' \subseteq W\}| = \Omega(\frac{m}{\sqrt{f(m)}})$.*

Given Lemma 2.11, the proof of our theorem is easy, since each degree of a vertex in W is at most $f(m)$, so $|e(U \cup W) - e(U)| = O(\frac{m}{\sqrt{f(m)}})$. This means that $\Omega(\frac{m}{\sqrt{f(m)}})$ values in the multiplication table are within $O(\frac{m}{\sqrt{f(m)}})$ centred at $e(U)$, which means that since $|e(U) - l| = O(\frac{m}{\sqrt{f(m)}})$, we can pick $\Omega(\sqrt{f(m)})$ values of l, to get $\Omega(m) = \Omega(e(G))$ different values in the multiplication table. □

To prove Lemma 2.11, we will need to construct an elaborate construction by first randomly generating the set U such that for a given l such that $cm \leq l \leq 2cm$, one has $|e(U) - l| = O(\frac{m}{\sqrt{f(m)}})$ and secondly, for each such U, we will need to construct the set $W$of size at most $O(m/f(m)^{\frac{3}{2}})$ such that $|\{e(U \cup W') : W' \subseteq W\}| = \Omega(\frac{m}{\sqrt{f(m)}})$.

To start the construction, we will need the following lemma:

Lemma 2.12. *For a C-bipartite-Ramsey graph G and any l with $cm \leq l \leq 2cm$, there exist sets U, S, T, Z satisfying the following claims:*

1. $|e(U) - 4l| = O(\frac{m}{\sqrt{f(m)}})$ *and U is an ordinary set of vertices.*
2. *S, T, Z are disjoints sets of vertices, or disjoint sets of pairs of vertices, also disjoint from U. Also, $S \cup T \cup Z = O(m/f(m)^{\frac{3}{2}})$*
3. $\forall \boldsymbol{v} \in S \cup T \cup Z, d_U(\boldsymbol{v}) = d + O(\sqrt{f(m)})$, *where* $d = \Theta(f(m))$.
4. $\min_{\boldsymbol{x} \in S} d_U(\boldsymbol{x}) - \max_{\boldsymbol{x} \in T} d_U(\boldsymbol{x}) = \Omega(\sqrt{f(m)})$.
5. *The degrees from Z into U are distinct (that is $\forall \{\boldsymbol{z_1}, \boldsymbol{z_2}\} \in Z$, $d_U(\boldsymbol{z_1}) \neq d_U(\boldsymbol{z_2})$ if $\boldsymbol{z_1} \neq \boldsymbol{z_2}$).*
6. $|Z| = \Omega(\sqrt{f(m)}), |S| = \Omega(\sqrt{f(m)}), |T| = \Omega(m/f(m)^{\frac{3}{2}})$ *or* $|S| = \Omega(\sqrt{f(m)}), |T| = \Omega(m/f(m)^{\frac{3}{2}})$.

Proof. We prove them in a certain order. Firstly, from Lemma 2.10, there exists $\gamma = \gamma(C, \alpha, \delta)$ such that for n large enough, G is $(\gamma, \delta, 4\gamma)$-bipartite-rich, with $\delta = \alpha/5$. Hence, from Lemma 2.9, our graph G is $(\delta, 2\gamma)$-bipartite-diverse and $(4\gamma^2, \delta, 2\gamma)_2$-bipartite-diverse. Now consider the $\Omega(m^2/f(m)^2)$ sums $d(y_1) + d(y_2)$ for $\{y_1, y_2\} \in \binom{Y}{2}$, all of which lie in the range $[0, 2f(m)]$. By Pigeonhole principle, there exists some $d' = \Theta(f(m))$ such that there exists $W \subset \binom{Y}{2}$ of size $|W| = \Omega(m^2/f(m)^{\frac{5}{2}})$ such that for all pairs $\{w_1, w_2\} \in W$, we have $d(w_1) + d(w_2) = d' + O(\sqrt{f(m)})$. By Lemma 2.10, we can delete at most $|Y|^{1+\delta} = (\frac{m}{f(m)})^{1+\delta} = o(m^2/f(m)^{\frac{5}{2}})$ pairs of $\{y_1, y_2\} \in \binom{Y}{2}$ from W to obtain $W' \subseteq W$, where $W' = \Omega(m^2/f(m)^{\frac{5}{2}})$ and for $\{w_1, w_2\} \in W$ we have $|N(w_1) \cap \overline{N(w_2)}| < 2\gamma|X|$. Note that $|Y|^{1+\delta} = (\frac{m}{f(m)})^{1+\delta} = o(m^2/f(m)^{\frac{5}{2}})$ since $f(m) \leq \sqrt{m}$ and $\delta \leq \frac{1}{10}$. Now interpret W' as a graph on the vertex set Y and hence have that either there exists a vertex with degree $\Omega(m/f(m)^{\frac{5}{4}})$ or there exists a matching of size $\Omega(m/f(m)^{\frac{5}{4}})$. In the former case, let this vertex be denoted by $y \in Y$. Then, set $d'' = d' - d_G(y)$ and let L be the set of neighbours of y in W'. In the latter case, let L be the set of pairs comprising the matching and $d'' = d'$. Then in both cases $|L| = \Omega(m/f(m)^{\frac{5}{4}})$ and $\forall \boldsymbol{y} \in L, d(\boldsymbol{y}) = d'' + O(\sqrt{f(m)})$. Next, let $F \subseteq \binom{L}{2}$ be the set of $\{\boldsymbol{y_1}, \boldsymbol{y_2}\} \in \binom{L}{2}$ such that $|N(\boldsymbol{y_1}) \triangle N(\boldsymbol{y_2})| < 4\gamma^2|X|$ (note that we can assume $\gamma < \frac{1}{4}$). Then, in each case using a different diversity assumption, if we interpret L as a graph with edges given by F, the graph has maximal degree $|Y|^\delta$ and so using Turan's Theorem (Corollary 2.5), our graph has an independent set A of size $\Omega(\frac{|L|}{|Y|^\delta}) = \Omega(\frac{m}{f(m)^{5/4}} \frac{f(m)^\delta}{m^\delta}) = \Omega(m/f(m)^{\frac{3}{2}})$, for all $\delta \leq \frac{\alpha}{4(1-\alpha)} \leq \frac{\alpha}{5}$. Hence in the original graph, for every $\{\boldsymbol{y_1}, \boldsymbol{y_2}\} \in \binom{A}{2}$, $|N(\boldsymbol{y_1}) \triangle N(\boldsymbol{y_2})| = |\Omega(f(m))|$.

Let $c = c(C, \alpha) > 0$ be such that $e(G) \geq 800cm$ and let $cm \leq l \leq 2cm$, while letting $p = \sqrt{\frac{4l}{e(G)}} \in (0, 0.1)$ be the probability of picking a vertex in $X \cup Y$ independently to be in U.

Claim. *The following each hold with probability greater than* 0.8.

1. $|e(U) - 4l| = O(\frac{m}{\sqrt{f(m)}})$*;*

2. There is $Q \subseteq A$ *involving no vertices of* U *with* $|Q| \geq \dfrac{2}{3}|A|$*;*

3. There is $R \subseteq A$ *with* $|R| \geq \dfrac{2}{3}|A|$ *and* $d_U(\boldsymbol{x}) = pd'' + O(\sqrt{f(m)})$ *for each* $\boldsymbol{x} \in R$*;*

4. $|N_U(\boldsymbol{x}) \triangle N_U(\boldsymbol{y})| = \Omega(f(m))$ *for each* $\{\boldsymbol{x}, \boldsymbol{y}\} \in \binom{A}{2}$*;*

5. The equality $d_U(\boldsymbol{x}) = d_U(\boldsymbol{y})$ *holds for* $O(m^2/f(m)^{\frac{7}{2}})$ *pairs* $\{\boldsymbol{x}, \boldsymbol{y}\} \in \binom{A}{2}$

Proof. 1. Note that $\mathbb{E}(e(U)) = 4l$ and $\mathrm{Var}(e(U)) = O(m^2/f(m))$, as each edge shares endpoints with at most $O(m/f(m))$ other edges. By Chebyshev's inequality, $|e(U) - 4l| = O(\frac{m}{\sqrt{f(m)}})$ with probability at least 0.99 for large enough implied constant in the $O(\frac{m}{\sqrt{f(m)}})$ term.

2. This follows directly from our definition of picking U: $\mathbb{E}(|\{r \in A, r \notin U\}|) \geq (1-p)^2|A|$, variance is $O(|A|)$. Recall that $(1-p)^2 \geq 0.81$, the result follows from Chebyshev's inequality.

3. For each $\boldsymbol{x} \in A$ we have $\mathbb{E}d_U(\boldsymbol{x}) = pd'' + O(\sqrt{f(m)})$ and the variance is $O(f(m))$, by Chebyshev with probability at least 0.999 we have $d_U(\boldsymbol{x}) = pd'' + O(\sqrt{f(m)})$. Let R be the set of $\boldsymbol{x}$ satisfying the condition as in the claim, using Markov's inequality we acquire the result. Now set $d = pd''$ and have that $d = \Theta(f(m))$.

4. Recall that $|N(\boldsymbol{x}) \triangle N(\boldsymbol{y})| = \Omega(f(m))$. Since $|N_U(\boldsymbol{x}) \triangle N_U(\boldsymbol{y})| = |(N(\boldsymbol{x}) \triangle N(\boldsymbol{y})) \cap U|$, we know that $|N_U(\boldsymbol{x}) \triangle N_U(\boldsymbol{y})|$ has a binomial distribution with parameters $|N(\boldsymbol{x}) \triangle N(\boldsymbol{y})|$ and p. By Chernoff bound, $\mathbb{P}(|N_U(\boldsymbol{x}) \triangle N_U(\boldsymbol{y}| < (p/2)|N(\boldsymbol{x}) \triangle N(\boldsymbol{y}|) = e^{-\Omega(f(m))} \to 0$ as $f(m) \to \infty$, and the desired result follows directly.

5. Since $|N(\boldsymbol{x}) \triangle N(\boldsymbol{y})| = \Omega(f(m))$, we let $|N(\boldsymbol{x}) \setminus N(\boldsymbol{y})| = a$, $|N(\boldsymbol{y}) \setminus N(\boldsymbol{x})| = b$. Without loss of generality assume that $a \leq b$ and note that $a + b \geq c'f(m)$, for some $c' > 0$, such that $b \geq \dfrac{c'f(m)}{2}$. Then,

$$\begin{aligned}
\mathbb{P}(d_U(\boldsymbol{x}) = d_U(\boldsymbol{y})) &= \mathbb{P}(|N_U(\boldsymbol{x})/N_U(\boldsymbol{y})| = |N_U(\boldsymbol{y})/N_U(\boldsymbol{x})|) \\
&= \sum_{k=1}^{a} p^k (1-p)^{a-k} \binom{a}{k} p^k (1-p)^{b-k} \binom{b}{k} \\
&\leq \max_{i \in \{1,2,\ldots,b\}} p^i (1-p)^{b-i} \binom{b}{i} \\
&= O\left(p^{pb+\lambda}(1-p)^{b-pb-\lambda} \binom{b}{pb+\lambda}\right)
\end{aligned}$$

for some $\lambda \in (-1, 1)$, since the mode of a binomial distribution with parameters b and p is $pb + \lambda$. Then, $O(p^{pb+\lambda}(1-p)^{b-pb-\lambda}\binom{b}{pb+\lambda}) = O(\frac{1}{\sqrt{p(1-p)b}}) = O(1/\sqrt{f(m)})$ by applying Stirling's formula. Now, the expected number of pairs $\boldsymbol{x}, \boldsymbol{y} \in \binom{A}{2}$ such that $d_U(\boldsymbol{x}) = d_U(\boldsymbol{y})$ is $O(m^2/f(m)^{\frac{7}{2}})$ and the desired result follows from Markov's inequality.

□

Now fix an outcome of U satisfying all 5 of the above properties, arbitrarily divide $R \cap Q$ which has size at least $\dfrac{|A|}{3}$ into two subsets H and P of size $\Omega(m^2/f(m)^{\frac{3}{2}})$. Consider the new graphs on the vertex sets H and P, where $\boldsymbol{xy}$ is an edge if and only if $d_U(\boldsymbol{x}) = d_U(\boldsymbol{y})$. In both cases, this new graph has $O(m/f(m)^{\frac{7}{2}})$ edges, so by Turan's theorem (Corollary 2.6) H has an independent set B and P has an independent set Z both of size $\Omega(\sqrt{f(m)})$. Now order the degrees $d_U(\boldsymbol{b})$ for $\boldsymbol{b} \in B$ in an increasing order. If the upper $|B|/2$ elements in ordering of B have $\Omega(m/f(m)^{\frac{3}{2}})$ neighbours in H, then we let T be the vertex set of this upper half elements and the vertices in H of the same degrees as them, let S to be the lower $|B|/3$ elements in B. If not, then the lower $|B|/2$ elements in B has $\Omega(m/f(m)^{\frac{3}{2}})$ many neighbours in H, then we let S be the vertex set of this lower half elements and the vertices in H of the same degrees as them, let T to be the upper $|B|/3$ elements in B. In either case, we will get T, S of size $\{\Omega(m/f(m)^{\frac{3}{2}}), \Omega(\sqrt{f(m)})\}$ (in some order), and $\min_{\boldsymbol{x}\in T} d(\boldsymbol{x}) - \max_{\boldsymbol{x}\in S} d(\boldsymbol{x}) \geq |B|/6 = \Omega(\sqrt{f(m)})$, as required.

□

Given the construction of Lemma 2.12, we make use of it to define the following. Without loss of generality, assume that $|S| = \Omega(m/f(m)^{\frac{3}{2}})$ and $|T| = \Omega(\sqrt{f(m)})$ and fix the orderings of their elements. Now let $c'' > 0$ be such that

$$\min_{\boldsymbol{y}\in T} d_U(\boldsymbol{y}) - \max_{\boldsymbol{y}\in S} d_U(\boldsymbol{y}) \geq 8c''\sqrt{f(m)},$$
$$|T| \geq c''\sqrt{f(m)} \textit{ and } |S| \geq 2c''m/f(m)^{\frac{3}{2}}$$

Then, let $\mathcal{P}$ be the set of pairs $(k,i) \in \mathbb{Z}^2$ such that $c'm/f(m)^{\frac{3}{2}} \leq k \leq 2c'm/f(m)^{\frac{3}{2}}$ and $0 \leq i \leq c'\sqrt{f(m)}$. Then, for each $(k,i) \in \mathcal{P}$, define $Q_{k,i}$ as the union of the first $(k-i)$ elements of S and the first i elements of T. For ease of notation, we will let $U_{k,i} = U \cup Q_{k,i}$.

We already have that for all k and i in the appropriate specified range, where all of the implied constants are positive:

$$e(U \cup Q_{k,0}) - e(U \cup Q_{k-1,0}) = \Theta(f(m)) \tag{1}$$
$$e(U \cup Q_{k,i}) - e(U \cup Q_{k,i-1}) = \Theta(\sqrt{f(m)}) \tag{2}$$

The statement in (1) trivially comes from the fact that for all $\boldsymbol{y} \in S$, $d_U(\boldsymbol{y}) = \Theta(f(m))$. For (2) we use that $\min_{\boldsymbol{y}\in T} d_U(\boldsymbol{y}) - \max_{\boldsymbol{y}\in S} d_U(\boldsymbol{y}) = \Theta(\sqrt{f(m)})$ and observe that $e(Q_{k,i}) = 0$ for all pairs $(k,i) \in \mathcal{P}$, since G is bipartite and S, T only contain vertices of Y. Now letting k and i range through $\mathcal{P}$, we already have $\Omega(|S|\sqrt{f(m)}) = \Omega(\frac{m}{f(m)})$ unique values in $\Phi(G)$. To prove Theorem 1.5, we will need to improve this number to $\Omega(\frac{m}{\sqrt{f(m)}})$, as remarked before.

Therefore, the final building block of the proof of Lemma 2.11 will be to show that we can induce $\Omega(\sqrt{f(m)})$ more subgraphs in magnitude, which are of different sizes for each U and W. This will be done by separating out the values of $e(U_{k,i})$ by at least $Q\sqrt{f(m)}$, for Q being a sufficiently large constant and using that Z is a set of $\Omega(\sqrt{f(m)})$ vertices of different degrees into U, which will allow us to achieve the exact improvement.

Proof of Lemma 2.11. First, let $\mathcal{P}' \subseteq \mathcal{P}$, where $(k,i) \in \mathcal{P}$ is in $\mathcal{P}'$ if $e(U_{k+1,0}) - e(U_{k,i}) = \Omega(\sqrt{f(m)})$ for some positive implied constant. Now by (1) and (2), we have that for each k in the range $[c'm/f(m)^{\frac{3}{2}}, 2c'm/f(m)^{\frac{3}{2}} - 1]$, there are $\Omega(\sqrt{f(m)})$ such i that $(k,i) \in \mathcal{P}'$ and so $|\mathcal{P}'| = \Omega(m/f(m))$. Now, let $\mathcal{I} = \{e(U_{k,i}) : (k,i) \in \mathcal{P}'\}$ to have that $|\mathcal{I}| = \Omega(m/f(m))$ and each pair of elements of $\mathcal{I}$ are separated by at least $\Omega(\sqrt{f(m)})$. Now let $Q > 0$ be a constant such that for every $\boldsymbol{x} \in W$, $d_U(x) \in [d - Q\sqrt{f(m)}, d + Q\sqrt{f(m)}]$. Then, order $\mathcal{I}$ in an increasing way and form $\mathcal{I}' \subseteq \mathcal{I}$ by picking every qth element in $\mathcal{I}$ for some constant $q = q(Q) > 0$, so that every pair of elements in $\mathcal{I}'$ are separated by $2Q\sqrt{f(m)}$. Note that we still have that $|\mathcal{I}'| = \Omega(m/f(m))$ because the gaps between consecutive elements of $\mathcal{I}$ are $\Omega(\sqrt{f(m)})$. Now for each $\boldsymbol{z} \in Z$, $d_{U_{k,i}}(\boldsymbol{z}) = d_U(\boldsymbol{z})$, as $\{\boldsymbol{z}\}, Q_{k,i} \subset Y$ and G is bipartite. Hence, for each $(k,i) \in \mathcal{P}'$ and each $\boldsymbol{z} \in Z$, $e(U_{k,i} \cup \boldsymbol{z})$ are distinct, which means that we've constructed $\Omega(m/\sqrt{f(m)})$ induced subgraphs of different sizes, as required. □

3 Minimisers of the multiplication table

First, let $\mathcal{M}(m) = \min\{\Phi(G) : G \textit{ is a bipartite graph with } m \textit{ edges}\}$ be the minimiser function of the multiplication table for $m \in \mathbb{N}$. Using Ford's estimates in [3], for $a|m, a \leq \sqrt{m}$, one can show that $\mathcal{M}(m) \ll \dfrac{n}{\log(a)^\delta \log\log(a)}$. Thus, if m has a divisor $a \leq \sqrt{m}$ such that $a \geq g(m)$ for some function g such that $g(m) \to \infty$ as $m \to \infty$, then we can assert $\mathcal{M}(m) = o(m)$. We naturally have the following conjecture, which is formulated by Narayanan, Sahasrabudhe and Tomon in [1]:

Conjecture 3.1. *$\mathcal{M}(m) = o(m)$ for all $m \in \mathbb{N}$.*

We are yet unsure if the conjecture is true or false, but our idea of tackling this conjecture might give some light to it. As stated in the introduction, $\log n$-balanced numbers have asymptotic density 1 as $n \to \infty$, meaning that the density of n satisfying the conjecture is 1. Instead of working with $\log n$-balanced numbers, we instead chose to work with n^ε-balanced numbers for convenience.

Our idea is: given any $n \in \mathbb{N}$ and $\varepsilon > 0$, if we can find n_1 such that $|n - n_1| = O((\varepsilon \log n)^\delta)$ and n_1 is an n^ε-balanced number, then we're done. To see this, without loss of generality, assume $n_1 < n$. Then consider the graph G by trivially adding $(n - n_1)$ edges in addition to the complete bipartite graph with $n_1$vertices. We have $\Phi(G) \leq (\varepsilon \log(n))^\delta \Phi(K_{n_1^\varepsilon, n_1^{1-\varepsilon}}) = O(\dfrac{n}{(\log\log n^\varepsilon)^{2/3}}) = o(n)$, since adding some edges independently can at most increase the multiplication table by a factor of its size, as required. Thus, we want to ask about how large the gaps between consecutive n^ε balanced numbers can be. We wish to find some $\varepsilon > 0$ such that there exists an n^ε-balanced number in $[n, n + (\varepsilon \log n)^\delta]$ for all large n, so that the conjecture will have been shown to hold.

However, we provide a heuristic argument to why we believe that not every interval $[n, n + (\varepsilon \log n)^\delta]$ has an n^ε-balanced number. Firstly, observe that for any fixed $\varepsilon > 0$, x^ε-balanced $\Leftrightarrow$ $x^{1-\varepsilon}$-smooth. In 1949, V. Ramaswami in [4] studied the Dickman function $\psi(x, x^\gamma)$ for $\gamma \in (0, 1)$, which counts the number of x^γ-smooth numbers up to x. Ramaswami proved that $\psi(x, x^\gamma) = x\rho(\frac{1}{\gamma}) + O(\frac{x}{\log x})$, where $\rho(u)$ is the Dickman function satisfying the delay differential equation $u\rho'(u) + \rho(u - 1) = 0$ and the initial condition $\rho(u) = 1$ for $u \in [0, 1]$.

Now, consider the interval $(n, 2n]$, assume that each number in this interval is $n^{1-\varepsilon}$-smooth independently with probability $q = \rho(\frac{1}{1-\varepsilon})$. For $1 \leq 1/\varepsilon \leq 2$, we have $q = \rho(\frac{1}{1-\varepsilon}) = 1 + \log(1 - \varepsilon)$ by solving the above differential equation.

We are interested in the longest consecutive successes in n consecutive Bernoulli trials (with $p = 1 - q$). Denote by X, the Bernoulli random variable measuring the number of $x^{1-\varepsilon}$-smooth numbers in the range $(n, 2n]$. Then, $\mathbb{E}[X] = nq$ and so $\mathbb{P}(\text{there exists a run of } l \text{ non-}x^{1-\varepsilon}\text{-smooth numbers}) \approx nqp^l$ by viewing this run as nq independent geometric random variables with parameter p. And since we are interested in the case, when the run of length l occurs for all n large enough, setting the above probability to 1 gives that the longest run is almost surely of length $l \sim \frac{\log(nq)}{\log(\frac{1}{p})}$. We would want this number to be smaller than $(\log(n^\varepsilon))^\delta$, but by a simple computation one can check that is not the case for all fixed $\varepsilon > 0$. By setting $q = 1 + \log(1 - \varepsilon)$ and $p = -\log(1 - \varepsilon)$, we know that the longest run $l = \Theta(\log n)$ almost surely.

Then, a natural question to ask is how long can the sequence of consecutive not k-balanced numbers be deterministically.

Proposition 3.2. Assuming the generalized Hardy-Littlewood k-tuple Conjecture as [Theorem 1.2,[5]], for all $\varepsilon > 0$, we can find arbitrary long sequence of consecutive integers such that all of them are not $x^{1-\varepsilon}$-smooth numbers.

Proof. For fixed $k \in \mathbb{N}$, consider the set $A_n = \{k!n+1, k!n/2+1, ..., k!n/k+1\}$. This set is admissible for the Hardy-Littlewood Conjecture, as for each prime p, letting $p \mid n$, we know A_n is not a covering system of p. By the Hardy-Littlewood Conjecture, $|\{n \leq x, \text{all numbers in } A_n \text{ are prime}\}| = (C + o(1))\frac{x}{(\log x)^k}$

for some $C > 0$. For $n \in \mathbb{N}$ such that A_n is made up of only primes, consider the set of consecutive numbers $\{k!n+1, ..., k!n+k\}$. Now set $x = k!n+1$ and clearly have that all numbers in this set are not $x^{1-\varepsilon}$-smooth for x large enough. Since one can choose k arbitrarily large, the claim follows. $\square$

Using exactly the same technique, if we don't fix ε, we have: For any $k \in \mathbb{N}$, we can find a large x and corresponding $\varepsilon > 0$ such that there are consecutive $\varepsilon \log(x)$ numbers, which are non-$x^{1-\varepsilon}$-smooth, starting from x. For this, we set $x = k!n+1$, and restrict ε such that $k < x^{\varepsilon}$. Then, we want $k > \varepsilon \log(x)$, which means that ε must satisfy $\frac{\log(k)}{\log(x)} < \varepsilon < \frac{k}{\log(x)}$, which clearly yields a solution. Yet, because the $o(1)$ term is present in the Conjecture, we cannot find an explicit upper-bound for x such that this occurs.

4 C-Ramsey Graphs

In this section, we fix a small oversight in the paper [4] by Kwan and Sudakov. After reading a closely related paper [10], also by Kwan and Sudakov, we find that the same issue persists, but thankfully our fix is sufficient here as well.

The authors define a graph on n vertices to be (δ, ε)-rich if for each $W \subseteq V$ such that $|W| > \delta n$, there exists at most n^δ vertices $v \in V$ such that $|N(v) \cap W| < \varepsilon n$ or $|\overline{N(v)} \cap W| < \varepsilon n$. Then, the authors give the following statement in [Lemma 4, [4]]:

Lemma 4.1. *For any $C, \delta > 0$, there exist $\varepsilon = \varepsilon(C) > 0$ and $c = c(C, \delta) > 0$ such that every n-vertex C-Ramsey graph contains a (δ, ε)-rich induced subgraph on at least cn vertices.*

This lemma is used twice in Lemma 7 of [2]. The first instance is used to set $\delta = \varepsilon/4$ and the second is to show that $O(n^{1+\delta}) = o(n^{5/4})$. In the lemma above, ε is claimed to be independent of δ, which makes the aforementioned two applications possible. However, this is incorrect, as in the proof of Lemma 4 of [2], it is clear that ε must depend on δ. This problem stems from the fact that δ is used twice in the definition of richness.

Following the proof of Lemma 4 from [2], Kwan and Sudakov, construct the set S of size $|S| = \frac{K(cn)^\delta}{2}$, such that $d(S) = 4\varepsilon + \frac{2}{K}$. But then, they claim that for $\varepsilon > 0$ small enough, S has too little density to not contain a $C \log n$ sized homogeneous subgraph. Here, the authors implicitly use Theorem 2 from [7]. To keep this explanation self-contained, we will give the statement of this theorem:

Theorem 4.2. *Let G be a graph on n vertices such that for $k \geq 40$, $e(G) \leq \frac{n^2}{k}$, then $hom(G) \geq \frac{c' k \log n}{\log k}$, where c' is a positive constant satisfying $c' < 1/240$.*

Now applying Theorem 4.2, assuming that $4\varepsilon + \frac{2}{K} < \frac{1}{40}$, we have that:

$$hom(S) \geq \frac{\dfrac{c'}{4\varepsilon + \frac{2}{K}} \log |S|}{\log\left(\dfrac{1}{4\varepsilon + \frac{2}{K}}\right)} = \frac{c'(\log(\frac{K}{2}) + \delta \log(cn))}{-(4\varepsilon + \frac{2}{K}) \log(4\varepsilon + \frac{2}{K})} \tag{2}$$

It is here, where the argument $hom(G) \geq hom(S) \geq C \log n$ would come in to finish the proof, however it not quite precise. Already, from (2), we see that if $\varepsilon = \varepsilon(C)$, then we can always choose $\delta > 0$ (say $\delta = -\dfrac{C}{2c'}(4\varepsilon + \frac{2}{K}) \log(4\varepsilon + \frac{2}{K})$), such that the expression on the right-hand side of (2) is less than $C \log n$, since $c < 1$. This gives a contradiction for the fact that ε and δ can be chosen independently using this argument. This does not yet disprove that using the argument by Kwan and Sudakov, we could set $\delta = \varepsilon/4$.

For simplicity, we will set $c = 1$, as in reality c is much smaller than 1 but the argument still fails in this case. Also, set $\delta = \dfrac{\varepsilon}{4}$, as desired and now we're left with showing that there exists $\varepsilon = \varepsilon(C)$, such that:

$$\frac{c'(\log(\frac{K}{2}) + \dfrac{\varepsilon}{4} \log n)}{-(4\varepsilon + \frac{2}{K}) \log(4\varepsilon + \frac{2}{K}))} \geq C \log n \tag{3}$$

Since, $K = K(C)$, we can ignore the $\log(\frac{K}{2})$ term in the numerator and after cancelling terms we're left with the following inequality:

$$-(16 + \frac{8}{K\varepsilon}) \log(4\varepsilon + \tfrac{2}{K}) \leq \frac{c'}{C}$$

If the above inequality is satisfied, then since $K, \varepsilon > 0$, we have $-16 \log(4\varepsilon + \frac{2}{K}) \leq \dfrac{c'}{C}$, which gives $d(S) = 4\varepsilon + \frac{2}{K} \geq e^{-c'/16C}$. Note that $c' < 1/240$ and can assume $C > 1/2$, as no graph G is *C-Ramsey* for $C \leq 1/2$ by a classical result due to Erdős and Szekeres in [9]. Hence, we have the inequality $d(S) \geq e^{-c'/16C} \geq e^{-1/1920} \geq 0.999$, which clearly leads to a contradiction if ε is small.

We fix this, by easing the definition of richness and for clarity, we state it as a definition:

Definition 4.3. Any graph G on n vertices is $(\gamma, \delta, \varepsilon)$-rich if for every $W \subseteq V$ such that $|W| > \gamma n$, there exists at most n^δ vertices $v \in V$ such that $|N(v) \cap W| < \varepsilon n$ or $|\overline{N(v)} \cap W| < \varepsilon n$.

Note that we can choose $\delta = \frac{1}{5}$ as this will be sufficient to fix the proof of [4] and [10]. Under this construction, one can show that γ and ε can be chosen to satisfy $\gamma = \varepsilon/4$ and this would still induce a linear sized rich subgraph.

Lemma 4.4. *For any C-Ramsey graph G on n vertices, there exist $\gamma, \varepsilon, c > 0$, which are all dependent on C such that $\varepsilon = 4\gamma$ and for large enough n, G has an induced subgraph of size cn, which is $(\gamma, \frac{1}{5}, \varepsilon)$-rich.*

Proof. For contradiction, suppose that every induced subgraph on cn vertices fails to be $(\gamma, 1/5, \varepsilon)$-rich for any $\gamma > 0$ with $\varepsilon = 4\gamma$. Then, we will construct a sequence of induced subgraphs $G = G[U] \supseteq G[U_1] \supseteq ... \supseteq G[U_K]$, where for all i, $|U_i| \geq cn$ and a sequence of disjoint vertex sets $S_1, S_2, ..., S_K$ such that for each i, $|S_i| = \dfrac{(cn)^{\frac{1}{5}}}{2}$ and $S_i \subset U_{i-1}$ and either for all $j > i$, $d(S_i, S_j) < 4\varepsilon$ or $d(S_i, S_j) > 1 - 4\varepsilon$. Without loss of generality, assume that the former holds for at least half of i and let this set be I. Then set $S = \bigcup_{i \in I} S_i$ and have that $|S| \geq \dfrac{K(cn)^{\frac{1}{5}}}{4}$ and clearly $d(S) < 4\varepsilon + \frac{2}{K}$. Let $C' = 240C$, $K = 1800C'^2$, $\gamma = \dfrac{1}{K} = \dfrac{1}{1800C'^2}, \varepsilon = 4\gamma = \dfrac{1}{450C'^2}$. $c = (\dfrac{\gamma}{4})^K$. By Theorem 2 in [7], for every graph G on n vertices and $k \geq 40$ such that $d(G) \leq \frac{1}{k}$, then $hom(G) \geq \frac{k \log n}{240 \log k}$. Applying this to the subgraph S, we have:

$$
\begin{aligned}
hom(G) \geq hom(S) &\geq \frac{\dfrac{1}{4\varepsilon + \frac{2}{K}} \log |S|}{240 \log \dfrac{1}{4\varepsilon + \frac{2}{K}}} \\
&\geq \frac{K(\log \frac{K}{4} + \frac{1}{5} \log c + \frac{1}{5} \log n)}{240 \log \frac{K}{18}} \\
&\geq \frac{K(-\dfrac{1}{10} \log n + \dfrac{1}{5} \log n)}{240 \log \dfrac{K}{18}} \\
&\geq \frac{\sqrt{\dfrac{K}{18}} \log n}{2400} \\
&= C \log n
\end{aligned}
$$

These inequalities hold since $\log(\frac{K}{4}) > 0$, $\frac{\log c}{\log n} \leq \dfrac{1}{2}$ for n large enough and $\frac{K}{\log(K/18)} \geq \frac{K}{18} / \sqrt{\frac{K}{18}} = \sqrt{\frac{K}{18}}$.

But by assumption, $hom(G) < C \log n$, which gives the required contradiction.

Now, we complete the construction as described above. Let $U = V(G)$ and assume $U, U_1, ..., U_{i-1}$ and $S_1, S_2, ..., S_{i-1}$ have been constructed. Then, for $c = (\dfrac{\gamma}{4})^K$ we have $|U_{i-1}| \geq cn$, so there exists a set $W \subseteq U_{i-1}$ such that $|W| \geq \gamma |U_{i-1}|$ and a set $Y \subseteq U_{i-1}$ with $|Y| \geq (cn)^{\frac{1}{5}}$ such that for all $v \in Y, |N(x) \triangle W| < \varepsilon |W|$ or $|\overline{N(x)} \triangle W| < \varepsilon |W|$, contradicting $(\gamma, \frac{1}{5}, \varepsilon)$-richness. Then, without loss of generality, can assume that for $S_i \subset Y$ and $|S_i| = \dfrac{(cn)^{\frac{1}{5}}}{2}$ such that for all $v \in S_i$, we have $|N(x) \triangle W| < \varepsilon |W|$. Then, let $U = W \setminus S_i$ and have $|U| \geq \dfrac{|W|}{2}$ for n large enough. Then, let $U_i \subseteq U$ such that for all $v \in U_i$, we have $d(\{v\}, S_i) < 4\varepsilon$. Now, it suffices to show that $|U_i| \geq \dfrac{\gamma}{4} |U_{i-1}|$. First, note that for all $s \in S_i$, we have that $d(U_i, s) < \frac{\varepsilon|W|}{\frac{|W|}{2}} = 2\varepsilon$ and then observe

$$4\varepsilon|U \setminus U_i| = \sum_{v \in U \setminus U_i} d(\{v\}, S_i) \leq \frac{e(U, S_i)}{|S_i|} = \frac{|U|}{|S_i|} \sum_{s \in S_i} d(U, \{s\}) \leq 2\varepsilon|U|$$

which gives that $|U_i| \geq \frac{|U|}{2} \geq \frac{|W|}{4} \geq \frac{\gamma}{4}|U_{i-1}|$, as required. □

Similarly, the paper [10] by Kwan and Sudakov appears to have the same issue. In this comment, we will borrow the notation of the aforementioned paper. Firstly, the same problem persists in Lemma 3.1, which is later used in Lemma 3.3 and Lemma 4.2 in their paper. Still, by using the $(\gamma, 1/5, \varepsilon)$-rich definition as in Definition 4.3, one can fix $\gamma = \varepsilon^K$ for some large absolute K, as in [10] proof of Lemma 4.2. Then, following exactly the same argument as in Lemma 4.4, we can find $\varepsilon = \varepsilon(C, K)$, such that a linear sized subgraph of G is $(\gamma, 1/5, \varepsilon)$-rich, which fixes the proof.

5 Miscellaneous attempts and results

Regarding Conjecture 1.1, one of the most natural questions to ask at first, is whether $\Phi(K_{a,b}) \geq \Phi(K_{n,n})$ for any $a, b \in \mathbb{N}$ such that $ab = n^2$. This question seems to be number-theoretic in nature because of the trivial representation $\Phi(K_{a,b}) = |\{pq : 0 \leq p \leq a, 0 \leq q \leq b\}|$. At first, we conjectured that, in general, if $a < c \leq d < b$ and $ab = cd = n^2$, then $\Phi(K_{a,b}) \geq \Phi(K_{c,d})$, however this is not the case. Although the above statement holds in 'most' cases, we were able to find counterexamples, where a and b are sufficiently close. Specifically, one can consider for example $\Phi(K_{16,36}) = 256 < 258 = \Phi(K_{18,32}) = 258$ or $\Phi(K_{48,147}) = 2782 < 2783 = \Phi(K_{49,144})$, which are both clear contradictions. Furthermore, the existence of counterexamples does not seem to vanish for larger n. Specifically, for $n \leq 300$, we were able to compute that 37 values of n such that one is able to find specific counterexamples. To better understand the differences in size of multiplication tables of complete graphs, we tried to find sufficiently nice characterisations of the set differences $\mathcal{M}(K_{a,b}) \setminus \mathcal{M}(K_{c,d})$ and $\mathcal{M}(K_{c,d}) \setminus \mathcal{M}(K_{c,d})$, however this turned out to be difficult and often the best we could do would rely on counting the number of primes in the interval $[d, b]$, denoted by $P_{[d,b]}$ to estimate $|\mathcal{M}(K_{a,b}) \setminus \mathcal{M}(K_{c,d})| \geq p_{[d,b]}a$. As for a non-trivial upper bound for $\mathcal{M}(K_{a,b}) \setminus \mathcal{M}(K_{c,d})$, we did not come up with any precise estimates. Given the computational data described above, we turned to one of the extremal cases, which would be necessary to prove that $K_{n,n}$ is the minimiser of the multiplication table for complete bipartite graphs of n^2 edges. Namely, we decided to compare $\Phi(K_{n^2,(n+1)^2})$ and $\Phi(K_{n(n+1),n(n+1)})$. In the case, when there is a prime number in the range $(n(n+1), (n+1)^2)$, the claim is trivial, as then

$$|\mathcal{M}(K_{n^2,(n+1)^2}) \setminus \mathcal{M}(K_{n(n+1),n(n+1)})| \geq n^2 \geq |\mathcal{M}(K_{n(n+1),n(n+1)}) \setminus \mathcal{M}(K_{n^2,(n+1)^2})|,$$

where the second inequality follows trivially from the fact that only numbers $m \in \mathbb{N}$, which have factors $a, b \in \mathbb{N}$ such that $n^2 + 1 \leq a \leq b \leq n(n+1)$ and $m = ab$ can contribute the set difference on the right. However, Legendre's Conjecture, which has certainly stood the test of time, stating that there always is a prime in the interval $(n^2, (n+1)^2)$ has neither been proved or disproved for more than 100 years. Hence, trying to prove that there always exists a prime in the even shorter interval $(n(n+1), (n+1)^2)$ seems like a too difficult problem to tackle.

Now shifting our focus to any bipartite graph G with m edges, trivially we have $\Phi(G) = \Omega(\sqrt{m})$; to see this, note that any such graph either contains a vertex of degree $\Omega(\sqrt{m})$ or an induced matching of size $\Omega(\sqrt{m})$. Yet there's no simple argument for $\Phi(G) = \Omega(m^{3/4})$. We found a simple counting method showing that almost surely a random bipartite graph G with $|X| = |Y| = n$ has $\Phi(G) = \Omega(n^{3/2-\varepsilon})$ for any $\varepsilon > 0$.

Example 5.1. Almost surely, for any random bipartite graph $G(n, n)$ with probability 1/2 of connecting an edge, $\Phi(G) \gg n^{\frac{3}{2}-\varepsilon}$ for any $\varepsilon > 0$.

To prove this, note that almost surely $d(x) = \frac{n}{2} + O(n^{\frac{1}{2}+\varepsilon})$ by Chebyshev's inequality, where X, Y are two sides of G of size n and $x \in X$. Clearly, almost surely, there exists $x_1 \in X$ such that for all $\varepsilon > 0$ s.t. $d(x_1) > n^{1-\varepsilon}$. Then separate Y into $N'(x_1)$ and $Z := Y \setminus N'(x_1)$, where $N'(x_1) \subset N(x_1)$ is of cardinality $n^{1-\varepsilon}$. From now on we will show that in every interval of length $O(n)$, we can find at least $n^{1-\varepsilon}$ entries in $\Phi(G)$, which together contribute to the exact cardinality we claimed. Pick $X_{2k+1} \subset X$ consisting of $x_1, ..., x_{2k+1}$, then almost surely we have

$$|\mathcal{M}(X_{2k+1} \cup Y) \cap [kn - 2kn^{\frac{1}{2}+\varepsilon}, kn + n/2 + (2k+1)n^{\frac{1}{2}+\varepsilon}]| \geq n^{1-\varepsilon},$$

since we can fix $(X_{2k+1} \setminus x) \cup (Y \setminus N'(x))$, and then adding x and each $N'(x)$ one by one creating non-overlapped-increasing elements in $\mathcal{M}(G)$. Finally, we have to stop this iterative when $kn^{\frac{1}{2}+\varepsilon} = \Omega(n)$, in which case adding two vertices to our set does not necessarily ensure that there is no overlap in the multiplication table. This means that as long as $k = o(n^{\frac{1}{2}-\varepsilon})$, we are allowed to continue the iteration. Now counting the number of induced subgraphs of different sizes before the stop, there are in total at least $n^{\frac{1}{2}-2\varepsilon} \times n^{1-\varepsilon} = n^{\frac{3}{2}-3\varepsilon}$ in $\mathcal{M}(G)$.

Despite the seeming difficulty in dealing with general cases, Narayanan, Sahasrabudhe and Tomon in [Theorem 1.2,[1]] proved that $\Phi(G) = \Omega(\frac{m}{(\log m)^{12}})$ for any bipartite G. We notice that the coefficient can be improved from 12 to 10 by a simple argument:

Theorem 5.2. *If G is a bipartite graph with m edges, then*

$$\Phi(G) = \Omega\Big(\frac{m}{(\log m)^{10}}\Big)$$

Proof. Firstly, it is important to point out that we are using the notation, as used in [1]. In the second paragraph of page 15 of [1], the authors assume that there exists $y \in V$ such that there exist $0 \leq a < b \leq r$ such that $\lambda_a(y), \lambda_b(y) \geq k/2(\log m)^2$. We use the following technique to improve this. By pigeonhole principle since $\sum_{i=0}^{r} \lambda_i(y) = k$ and $r \leq \log m$, without loss of generality we can assume that there exists some a such that $\lambda_a(y) = \Omega(k/\log m)$. If for all $0 \leq x \leq r$ and $x \neq a$, we have $\lambda_x(y) = o(k/\log m)$, then $\sum_{i=0}^{r} \lambda_i(y) = o(k) + \lambda_a(y) = k$, so $\lambda_a(y) = \Omega(k)$. Hence, either there exists $0 \leq a < b \leq r$ such that $\lambda_a(y), \lambda_b(y) = \Omega(k/\log m)$ or there exists distinct a and b such that $\lambda_a(y) = \Omega(k)$ and $\lambda_b(y) = \Omega(k/(\log m)^2)$. In each case, substituting into Lemma 2.2 of [1], we get the desired result under the first assumption. Note that in all other cases, the authors obtain a bound of at least $\Omega(\frac{m}{(\log m)^{10}})$, giving the result. □

Yet, we should point out that it seems there's no way to prove Conjecture 1 by optimizing the above coefficients.

6 Conclusion

During the 8 weeks we spent working on this project, we have proposed a new Conjecture, namely Conjecture 1.4. We have given a proof of a weaker version of it in the form of Theorem 1.5. After reading related papers and doing some computations, we feel that Conjecture 1.1 is true, but it seems that it's proof would involve a combination of graph theoretic and number theoretic results in a major way. A good starting point would be to prove that the density is within $(\varepsilon, 1-\varepsilon)$ for some $\varepsilon > 0$ for all sub-polynomial *C-Bipartite-Ramsey* graphs.

As for Conjecture 3.1, after doing some heuristic arguments, we suspect that it might be false. However, it seems that to disprove it, one should have a better understanding of what the minimiser graphs on m edges for any $m \in \mathbb{N}$ are. Yet, we are nowhere near proving Conjecture 1.1, which is a specific instance of this problem.

We have also derived some other cute and small results during the project. They're probably not that significant, but they witness the days when we were struggling to find a direction to do further research.

All in all, this was a wonderful project. For each of us, it was enjoyable to work in a pair and be able to always share our ideas. We feel like this project has given us a huge amount of experience and has greatly set us up for future research placements.

References

[1] B. Narayanan, J. Sahasrabudhe, and I. Tomon, *The multiplication table problem for bipartite graphs,* Combinatorica. **37** (2017), no. 5, 991–1010.

[2] M. Kwan and B. Sudakov, *Ramsey graphs induce subgraphs of quadratically many sizes*, Int. Math. Res. Not. IMRN (2020), no. 6, 1621–1638.

[3] K. Ford, *The distribution of integers with a divisor in a given interval,* Ann. of Math. (2) **168** (2008), no. 2, 367–433.

[4] V. Ramaswami, *On the number of positive integers less than x and free of prime divisors greater than x^c*, Bull. Amer. Math. Soc. **55** (1949), 1122–1127.

[5] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math., **171** (2010), no.2, 1753–1850.

[6] G. Tenenbaum, *Sur la probabilitè qu'un entier possède un diviseur dans un intervalle donnè*, Compositio Math. **51** (1984), 243–263.

[7] P. Erdős and A. Szemerédi, *On a Ramsey type theorem*, Period. Math. Hungar. **2** (1972), 295– 299, Collection of articles dedicated to the memory of Alfréd Rényi, I.

[8] J.H. Hattingh and M.A. Henning, *Bipartite Ramsey theory,* Util. Math. **53** (1998), 217–230.

[9] P. Erdős and G. Szekeres, *A combinatorial problem in geometry,* Compositio Math. **2** (1935), 463–470.

[10] M. Kwan and B. Sudakov, *Proof of a conjecture on induced subgraphs of Ramsey graphs,* Trans. Amer. Math. Soc. **372** (2019) no. 8, 5571–5594.

[11] P.Turán, *On a Theorem of Hardy and Ramanujan,* Journal of the London Mathematical Society **4** (194), 274-276

[12] P. Erdős, *An asymptotic inequality in the theory of numbers,* Vestnik Leningrad Univ. **15** (1960), 41–49 (Russian).

[13] N. Alon and J. H. Spencer, *The probabilistic method,* fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley Sons, Inc., Hoboken, NJ, 2016