

# The multiplication table problem for bipartite graphs

Bhargav Narayanan, Julian Sahasrabudhe, and István Tomon

**ABSTRACT.** We investigate the following generalisation of the ‘multiplication table problem’ of Erdős: given a bipartite graph with  $m$  edges, how large is the set of sizes of its induced subgraphs? Erdős’s problem, posed in 1955, of estimating the number of distinct products  $ab$  with  $a, b \leq n$  is precisely the problem under consideration when the graph in question is the complete bipartite graph  $K_{n,n}$ . In this note, we prove that the set of sizes of the induced subgraphs of any bipartite graph with  $m$  edges contains  $\Omega(m/(\log m)^{12})$  distinct elements; furthermore, this is tight up to the exponent of the logarithmic factor in the estimate.

## 1. INTRODUCTION

For a bipartite graph  $G$ , we define, writing  $e(\cdot)$  for the number of edges of a graph, its *multiplication table*  $\mathcal{M}(G)$  by setting

$$\mathcal{M}(G) = \{e(H) : H \text{ is an induced subgraph of } G\};$$

our reasons for calling  $\mathcal{M}(G)$  the multiplication table of  $G$  will soon become evident. While the definition of  $\mathcal{M}(G)$  above is meaningful for any graph  $G$ , we shall mainly be concerned with bipartite graphs in this note.

It seems likely that the bipartite graphs with the smallest multiplication tables are the complete bipartite graphs; writing  $K_{n,n}$  for the complete bipartite graph between two disjoint sets of  $n$  vertices, we conjecture the following.

**Conjecture 1.1.** *Let  $n \in \mathbb{N}$  and suppose that  $G$  is a bipartite graph with  $e(G) = n^2$ . Then  $|\mathcal{M}(G)| \geq |\mathcal{M}(K_{n,n})|$ .*

In this note, our aim is to prove Conjecture 1.1 in a weak quantitative form. It turns out that  $|\mathcal{M}(K_{n,n})| = o(n^2)$ ; in fact, as was shown by Ford [14], there exists an absolute constant  $\delta \approx 0.086$  such that  $|\mathcal{M}(K_{n,n})| = n^2/(\log n)^{\delta+o(1)}$ . Our main result, stated below, gives a comparable bound for an arbitrary bipartite graph.

---

*Date:* 18 August 2014.

*2010 Mathematics Subject Classification.* Primary 05C35; Secondary 11B30.

**Theorem 1.2.** *If  $G$  is a bipartite graph with  $m$  edges, then*

$$|\mathcal{M}(G)| = \Omega\left(\frac{m}{(\log m)^{12}}\right).$$

Let us make a few remarks about Theorem 1.2. First, in the light of our earlier remarks about complete bipartite graphs, it is clear that we cannot do away with the logarithmic factor in our result; indeed, we cannot replace the exponent 12 in the statement of Theorem 1.2 by an exponent smaller than  $\delta \approx 0.086$ . Next, we should point out that it is trivial to prove that  $|\mathcal{M}(G)| = \Omega(m^{1/2})$  for any graph  $G$  with  $m$  edges; to see this, note that any such graph either contains a vertex of degree  $\Omega(m^{1/2})$  or an induced matching of size  $\Omega(m^{1/2})$ , and we are done in either case. If we do not insist that our graph is bipartite, this trivial argument can be seen to be essentially tight by considering, for example, the complete graph. Hence, to beat this trivial bound of  $m^{1/2}$ , it is necessary to exploit the fact that we are working with bipartite graphs. We urge the reader to pause for a moment and consider the question of beating this trivial lower bound of  $m^{1/2}$  for bipartite graphs; while Theorem 1.2 improves on this bound considerably, we do not have a short proof of even a lower bound of, say,  $m^{2/3}$ . Finally, while it might be possible to refine our methods to prove a bound of the form say,  $\Omega(m/(\log m)^9)$ , it seems unlikely that our proof can be adapted to prove Conjecture 1.1, or for that matter, to even prove a bound of  $\Omega(m/\log m)$ ; hence, we make no serious attempt to optimise the logarithmic factors in our proof.

Our main motivation for studying Conjecture 1.1 is because it is a natural combinatorial generalisation of a number-theoretic problem, now known colloquially as the ‘multiplication table problem’, posed by Erdős [8]. For  $n \in \mathbb{N}$ , write  $[n]$  for the set  $\{1, \dots, n\}$  and  $[n] \cdot [n]$  for the set of distinct products  $ab$  with  $a, b \leq n$ ; the multiplication table problem is then simple to state: how large is  $[n] \cdot [n]$ ? Observe that  $[n] \cdot [n]$  is precisely the set of sizes of the induced subgraphs of  $K_{n,n}$ ; consequently, the multiplication table problem can be rephrased as follows: how large is  $|\mathcal{M}(K_{n,n})|$ ? In this paper, we generalise this question and ask how large  $|\mathcal{M}(G)|$  is for an arbitrary bipartite graph  $G$  with a prescribed number of edges.

The multiplication table problem has received a great deal of attention over the past five decades. Erdős [8] showed, using the fact that almost all natural numbers less than  $n$  have about  $\log \log n$  distinct prime factors, that the cardinality of  $[n] \cdot [n]$  is  $o(n^2)$  as  $n \rightarrow \infty$ . Subsequently, better bounds were obtained, first by Erdős [9] and then by Tenenbaum [22]. Despite its innocuous appearance, the multiplication table problem has been settled only recently; a deep result of Ford [14] asserts that as  $n \rightarrow \infty$ ,

$$|[n] \cdot [n]| = \Theta\left(\frac{n^2}{(\log n)^\delta (\log \log n)^{3/2}}\right),$$

where  $\delta = 1 - (1 + \log \log 2)/\log 2 \approx 0.086$ . For results about higher dimensional analogues of the multiplication table problem, we refer the reader to the papers of Koukoulopoulos [17, 18]. Erdős also posed a number of related number-theoretic problems; see the book of Hall and Tenenbaum [15] and the section on the statistical theory of divisors in the comprehensive survey of Ruzsa [20].

The problem studied in this paper is also closely related to a number of combinatorial results about induced subgraph sizes that have been proved over the course of the last thirty years. Many of these questions and results about the sizes of induced subgraphs arise from trying to better understand the structure of *Ramsey graphs*; we discuss some of these problems below.

A subset of the vertices of a graph is said to be *homogeneous* if it either induces a clique or an independent set; let us write  $\text{hom}(G)$  for the size of the largest homogeneous set of vertices in a graph  $G$ . Alon and Bollobás [2] (see also [13]) proved that any graph without a large homogeneous set necessarily contains many distinct (non-isomorphic) induced subgraphs; in their proof, they distinguished between induced subgraphs using, amongst other parameters, their order and size. Subsequently, Erdős, Faudree and Sós [10, 12] conjectured that for every  $C > 0$ , there exists an  $\varepsilon = \varepsilon(C) > 0$  such that if  $G$  is an  $n$ -vertex graph with  $\text{hom}(G) \leq C \log n$ , then the number of distinct pairs  $(x, y)$  such that  $G$  has an induced subgraph on  $x$  vertices inducing  $y$  edges is at least  $\varepsilon n^{5/2}$ . While this conjecture still remains unresolved, a number of partial results have been proved; see the papers of Axenovich and Balogh [5], Alon and Kostochka [3], and Alon, Balogh, Kostochka and Samotij [1] for the state of the art.

Another conjecture with a similar flavour, due to Erdős and McKay [10, 11], also far from settled, asserts that for every  $C > 0$ , there exists an  $\varepsilon = \varepsilon(C) > 0$  such that if  $G$  is an  $n$ -vertex graph with  $\text{hom}(G) \leq C \log n$ , then  $G$  contains an induced subgraph with precisely  $y$  edges for every integer  $y$  between 0 and  $\varepsilon n^2$ ; the best known bounds for this problem are due to Alon, Krivelevich and Sudakov [4].

The sizes of the induced subgraphs of a random graph have also been investigated; we refer the reader to the paper of Calkin, Frieze and McKay [7] for details. Finally, let us also mention that the connection between the multiplication table problem and the sizes of induced subgraphs of complete bipartite graphs was exploited by the first author in [19] to construct ‘good’ colourings for a Ramsey-theoretic problem.

Returning to the question at hand, let us discuss, very briefly, one of the difficulties in proving Theorem 1.2. Note that a proof of Theorem 1.2 should also establish that  $|\llbracket n \rrbracket \cdot \llbracket n \rrbracket| = \Omega(n^2/(\log n)^{12})$ . To prove such a weak estimate for the size of  $\llbracket n \rrbracket \cdot \llbracket n \rrbracket$  is in itself not difficult. One could use the prime number theorem to show that one has many distinct products of the form  $ab$  in the set  $\llbracket n \rrbracket \cdot \llbracket n \rrbracket$  where both  $a$  and  $b$  are prime. Alternatively, one could use the fact that the set  $\llbracket n \rrbracket$  has small ‘additive doubling’ to conclude, using a beautiful theorem of Solymosi [21], that  $\llbracket n \rrbracket \cdot \llbracket n \rrbracket$  is large. These are

however the only ways, to the best of our knowledge, of proving a reasonable lower bound for  $|\llbracket n \rrbracket \cdot \llbracket n \rrbracket|$  without resorting to somewhat involved divisor estimates, and neither of these methods would appear to generalise easily to the setting of bipartite graphs. Hence, to prove Theorem 1.2, we shall require, in addition to graph-theoretic techniques, a few additive-combinatorial and number-theoretic tools; some of these might be of independent interest.

The rest of this paper is organised as follows. In Section 2, we establish some notation and then prove some straightforward number-theoretic estimates. We then describe our strategy for proving Theorem 1.2 in Section 3. After establishing a key partitioning lemma in Section 4, we prove Theorem 1.2 in Section 5. We conclude by discussing some problems in Section 6.

## 2. PRELIMINARIES

In this section, we establish some notation and collect together some number-theoretic estimates that we shall make use of when proving our main result.

**2.1. Notation.** Given a set  $S$  and  $r \in \mathbb{N}$ , we write  $S^{(r)}$  for the family of subsets of  $S$  of cardinality  $r$ . Given  $A, B \subset \mathbb{Z}$ , we write  $A + B$  and  $A \cdot B$  respectively for the set of distinct sums  $a + b$  and products  $ab$  with  $a \in A$  and  $b \in B$ .

It will help to have some notation in place for working with finite sequences. Given a sequence of integers  $\mathbf{a} = (a_i)_{i=1}^n$ , we define  $\mathcal{S}(\mathbf{a})$ , its *set of sums*, by setting

$$\mathcal{S}(\mathbf{a}) = \left\{ \sum_{i \in I} a_i : I \subset [n] \right\}.$$

Given two sequences of integers  $\mathbf{a}$  and  $\mathbf{b}$ , we write  $\mathcal{S}(\mathbf{a}, \mathbf{b})$  for the set of sums of the *concatenation of  $\mathbf{a}$  and  $\mathbf{b}$* ; equivalently,  $\mathcal{S}(\mathbf{a}, \mathbf{b}) = \mathcal{S}(\mathbf{a}) + \mathcal{S}(\mathbf{b})$ . We write  $k \circ a$  to denote the sequence of length  $k$  each of whose terms is  $a$ . So for example, the set  $\mathcal{S}(k \circ a, l \circ b)$  consists of those integers which can be written as  $ax + by$  for some  $0 \leq x \leq k$  and  $0 \leq y \leq l$ .

Our conventions for asymptotic notation are largely standard; however, we feel obliged to point out that when we write, say  $\Omega_k(\cdot)$ , we mean that the constant suppressed by the asymptotic notation is allowed to depend on (but is completely determined by) the parameter  $k$ . Occasionally, we shall find it convenient to switch to Vinogradov's notation: given functions  $f$  and  $g$ , we write  $f \ll g$  if  $f = O(g)$  and  $f \gg g$  if  $g = O(f)$ .

We use standard graph-theoretic notation and refer the reader to [6] for terms and notation not defined here. To keep the exposition uncluttered, we omit floors and ceilings whenever they are not crucial.

**2.2. Number-theoretic estimates.** We now collect together a few easy number-theoretic estimates; for the sake of completeness, we shall prove them.

**Lemma 2.1.** *Let  $\mathbf{a} = (a_i)_{i=1}^n$  be a sequence of positive integers and let  $b \in \mathbb{N}$  be a positive integer such that  $\gcd(a_i, b) \leq g$  for  $1 \leq i \leq n$ . Then for any  $k \in \mathbb{N}$ ,*

$$\mathcal{S}(\mathbf{a}, k \circ b) \geq k \min\{b/g, n\}.$$

*Proof.* Writing  $\mathcal{S}_b(\mathbf{a})$  for the set of residues modulo  $b$  attained by the elements of  $\mathcal{S}(\mathbf{a})$ , it is clearly sufficient to show that  $|\mathcal{S}_b(\mathbf{a})| \geq \min\{b/g, n\}$ ; we shall prove this by induction on  $n$ .

The result is trivial if  $n = 1$ , so suppose that  $n > 1$ . Consider  $\mathbf{a}' = (a_i)_{i=1}^{n-1}$ , and assume inductively that  $|\mathcal{S}_b(\mathbf{a}')| \geq \min\{b/g, n-1\}$ . We are done if  $|\mathcal{S}_b(\mathbf{a}')| \geq n$  since then,  $|\mathcal{S}_b(\mathbf{a})| \geq |\mathcal{S}_b(\mathbf{a}')| \geq n \geq \min\{b/g, n\}$ . Also, if  $n-1 \geq b/g$ , then we are done once again since  $\min\{b/g, n\} = \min\{b/g, n-1\} = b/g$ .

Hence, we may assume that  $\mathcal{S}_b(\mathbf{a}')$  contains exactly  $n-1$  distinct residues modulo  $b$  and also that  $n-1 < b/g$ . Let  $t = \lceil b/g \rceil - 1$  and note that since  $n-1 < b/g$ , it is also true that  $n-1 \leq t$ . Choose  $s \in \mathcal{S}_b(\mathbf{a}')$  and observe that the numbers

$$s, s + a_n, s + 2a_n, \dots, s + ta_n$$

are all distinct modulo  $b$  since  $\gcd(a_n, b) \leq g$ . Also, as  $|\mathcal{S}_b(\mathbf{a}')| = n-1 \leq t$ , one of

$$s + a_n, s + 2a_n, \dots, s + ta_n$$

is not in  $\mathcal{S}_b(\mathbf{a}')$  since these numbers are, modulo  $b$ , all distinct and distinct from  $s \in \mathcal{S}_b(\mathbf{a}')$ . Now choose the minimal  $1 \leq l \leq t$  such that  $s + la_n \notin \mathcal{S}_b(\mathbf{a}')$ . By the minimality of  $l$ , we have  $s + (l-1)a_n \in \mathcal{S}_b(\mathbf{a}')$ , so  $s + (l-1)a_n + a_n = s + la_n \in \mathcal{S}_b(\mathbf{a})$ . Consequently,  $s + la_n \in \mathcal{S}_b(\mathbf{a}) \setminus \mathcal{S}_b(\mathbf{a}')$  and we are done.  $\square$

Let us record here, for convenience, a special case of Lemma 2.1

**Lemma 2.2.** *Let  $a, b \in \mathbb{N}$  be positive integers such that  $\gcd(a, b) \leq g$ . Then for any  $k, l \in \mathbb{N}$ ,*

$$|\mathcal{S}(k \circ a, l \circ b)| \geq k \min\{a/g, l\}. \quad \square$$

We need the following easy consequence of the prime number theorem; see [16], for example.

**Proposition 2.3.** *For every  $\varepsilon > 0$ , the number of primes in the interval  $[n, (1+\varepsilon)n]$  is  $\Omega_\varepsilon(n/\log n)$  as  $n \rightarrow \infty$ .*  $\square$

Indeed, it follows from the prime number theorem that the number of primes in the interval  $[n, (1+\varepsilon)n]$  is asymptotic to  $\varepsilon n/\log n$  for any fixed  $\varepsilon > 0$ . However, the weaker estimate above will be sufficient for our purposes.

**Lemma 2.4.** *If  $A \subset \mathbb{N}$  is a set of positive integers and  $b \in \mathbb{N}$  is a positive integer such that  $\max A < b^3/16$ , then*

$$|A \cdot [b]| \gg \frac{|A|b}{\log b}.$$

*Proof.* Let  $P_b$  denote the set of primes in the interval  $[b/2, b]$ ; we know that  $|P_b| \gg b/\log b$  by Proposition 2.3. Note that any  $k \in A \cdot [b]$  has at most three distinct prime factors in  $P_b$ . Indeed, if not, then  $k$  is divisible by at least four distinct primes each of which is at least  $b/2$ , whence  $k \geq b^4/16$ , which contradicts the fact that  $k \leq b \max A < b^4/16$ . Consequently, the number of distinct ordered pairs  $(a, p) \in A \times P_b$  such that  $ap = k$  is at most  $3 \times 2 = 6$ . Hence,

$$|A \times [b]| \geq \frac{|A||P_b|}{6} \gg \frac{|A|b}{\log b}. \quad \square$$

**Lemma 2.5.** *Let  $a, b, d, k \in \mathbb{N}$  be positive integers. Then*

$$\min_{0 \leq i < k} \gcd(a, b - id) \ll a^{1/k} dk^2.$$

*Furthermore, if  $a \neq b$ , then*

$$\min_{0 \leq i < k} \gcd(a - id, b - id) \ll |a - b|^{1/k} dk^2.$$

*Proof.* Let  $g = \gcd(a, b, d)$  and  $f_i = \gcd(a, b - id)$ . We claim that

$$\prod_{i=0}^{k-1} f_i \mid ad^k \prod_{q < k} q^{\lceil k/q \rceil},$$

where in the above,  $q$  ranges over the set of prime powers less than  $k$ . To check this claim, consider any prime  $p$ : it suffices to show that the largest power of  $p$  dividing  $\prod_{i=0}^{k-1} f_i$  also divides  $ad^k \prod_{q < k} q^{\lceil k/q \rceil}$ . Given  $p$ , fix  $0 \leq j \leq k-1$  so that the largest power of  $p$  dividing  $f_j$  is the greatest amongst  $f_0, \dots, f_{k-1}$ ; since  $f_j \mid a$ , we have accounted for the contribution from  $f_j$ . Next, note that the largest power of  $p$  dividing  $f_i$  for any  $i \neq j$  is the same as the largest power of  $p$  dividing  $\gcd(f_i, f_j)$ . Observe that

$$\gcd(f_i, f_j) = \gcd(a, b - id, b - jd) = \gcd(a, b - id, (i - j)d),$$

whence it is clear that

$$\gcd(f_i, f_j) \mid \gcd(a, b - id, d) \gcd(a, b - id, i - j) \mid g(i - j) \mid d(i - j).$$

Consequently, the largest power of  $p$  dividing  $\prod_{i \neq j} f_i$  also divides  $d^{k-1} \prod_{i \neq j} (i - j)$ . It suffices to account for the largest power of  $p$  dividing  $\prod_{i \neq j} (i - j)$ . But note that for any prime power  $q < k$ , the number of indices  $i \neq j$  such that  $q \mid (i - j)$  is at most  $\lceil k/q \rceil$ ; the claim then follows.

An old result of Mertens asserts that

$$\sum_{p < k} \frac{\log p}{p} = (1 + o(1)) \log k,$$

where in the above,  $p$  ranges over the set of primes less than  $k$ . It follows that

$$\prod_{q < k} q^{\lceil k/q \rceil} = k^{k+o(k)} \ll k^{2k}.$$

Hence, there exists an  $i \leq k - 1$  such that

$$f_i \leq a^{1/k} d \left( \prod_{q < k} q^{\lceil k/q \rceil} \right)^{1/k} \ll a^{1/k} d k^2.$$

To finish the proof of the lemma, note that the second assertion follows from the first since  $\gcd(a - id, b - id) = \gcd(a - b, b - id)$ .  $\square$

### 3. OVERVIEW OF OUR APPROACH

To illustrate our approach, we make an easy observation. Let  $G = (X, Y; E)$  be a bipartite graph and suppose that both vertex classes of  $G$  have the same size. Assume that  $G$  is *half-regular*; in other words, assume that the vertices on one side of the bipartition, say  $X$ , all have the same degree  $d$ . Assume also that there exists a vertex  $v \in Y$  with  $n/\log n < d(v) < n - n/\log n$ , where  $n = |X| = |Y|$ . If we remove  $v$  from  $G$ , then each vertex of  $X$  has degree either  $d$  or  $d - 1$  in the resulting graph; moreover, there are at least  $n/\log n$  vertices with each of these two degrees. By considering induced subgraphs of the form  $G[X' \cup Y \setminus \{v\}]$  where  $X' \subset X$ , we see that  $\mathcal{S}(k \circ d, l \circ (d - 1)) \subset \mathcal{M}(G)$  for some pair of natural numbers  $k, l \geq n/\log n$ . Since  $d$  and  $d - 1$  are coprime, it follows from Lemma 2.2 that

$$\begin{aligned} |\mathcal{M}(G)| &\geq |\mathcal{S}(k \circ d, l \circ (d - 1))| \\ &\geq \frac{n}{\log n} \min \left\{ d, \frac{n}{\log n} \right\} \\ &\geq \frac{nd}{(\log n)^2} \geq \frac{m}{(\log m)^2}. \end{aligned}$$

To prove Theorem 1.2, we first reduce the problem of bounding the size of the multiplication table of a general bipartite graph to the situation above, namely that of bounding the size of the multiplication table of a half-regular bipartite graph. We shall show (see Lemma 4.1) that given any bipartite graph  $G = (X, Y; E)$ , it is possible find a reasonably large subset of  $X$  that can be partitioned into many groups  $X_1, \dots, X_k$  such that the sums of the degrees of the vertices in each of these groups is the same. If we now imagine contracting each such set  $X_i$  into a single vertex, we obtain, after discarding  $X \setminus (\bigcup_i X_i)$  from our graph, a half-regular bipartite *multigraph*.

If we can then find a vertex in  $Y$  whose degree is neither too large nor too small, then we finish the proof using a variant of the argument sketched above. However, it might be the case that there is no such vertex in  $Y$ . If  $Y$  contains only a few vertices of very large degree, then our graph is somewhat sparse, and in this case, we use Lemma 5.1 to complete the proof. If it turns out that many vertices of  $Y$  have very large degrees, then we show that there is a reasonably dense induced subgraph within which we can find many distinct subgraph sizes; the argument in this case is number-theoretic in nature.

#### 4. A PARTITIONING LEMMA

In this section, we shall prove a partitioning lemma for multisets of positive integers that will play an important part in the proof of the main result. This lemma allows us to select a large proportion of the vertices of a vertex class of a bipartite graph and partition the selected vertices into many small, disjoint sets of equal size and such that the sum of the degrees of the vertices in each of these sets is the same.

**Lemma 4.1.** *Let  $\mathbf{a} = (a_i)_{i=1}^n$  be a sequence of positive integers with  $\sum_{i=1}^n a_i = m$ . There exist positive integers  $k, r, d \in \mathbb{N}$  and pairwise disjoint sets  $I_1, \dots, I_k \in [n]^{(r)}$  such that*

- (1)  $1 \leq r \leq \log m$ ,
- (2)  $kd \geq m/(\log m)^3$ , and
- (3)  $\sum_{i \in I_j} a_i = d$  for each  $1 \leq j \leq k$ .

Let us fix, for the rest of this section,  $a_1, \dots, a_n$  and  $m = \sum_{i=1}^n a_i$ . For  $i \in [n]$ , we shall think of  $a_i$  as the *weight on  $i$* , and given a hypergraph  $\mathcal{F}$  on  $[n]$ , we define the *weight covered by  $\mathcal{F}$*  to be the sum of the weights of those vertices contained in at least one edge of  $\mathcal{F}$ . Given  $r, d \in \mathbb{N}$ , it is natural to consider the  $r$ -uniform hypergraph  $\mathcal{F}(r, d)$  on  $[n]$  whose edges are the sets  $I \in [n]^{(r)}$  with  $\sum_{i \in I} a_i = d$ . In this language, Lemma 4.1 tells us that we can find appropriate  $r, d \in \mathbb{N}$  so that the hypergraph  $\mathcal{F}(r, d)$  contains a matching, i.e., a set of independent edges, covering a  $1/(\log m)^3$  proportion of the total weight, namely  $m$ .

Our first two propositions together show that one of the hypergraphs  $\mathcal{F}(r, d)$  must necessarily contain a matching with many edges; using this, we then show that one of these large matchings must cover a  $1/(\log m)^3$  proportion of the total weight.

**Proposition 4.2.** *If  $\mathcal{F}(r, d)$  contains  $l$  edges, then there exist positive integers  $r' \leq r$  and  $d' \leq d$  for which  $\mathcal{F}(r', d')$  contains a matching of size at least  $l^{1/r}/r$ .*

*Proof.* We prove the claim by induction on  $r$ . If  $r = 1$ , then it clearly suffices to take  $r' = r = 1$  and  $d' = d$ . Now suppose that  $r > 1$ . If some  $i \in [n]$  is contained in at least  $l^{(r-1)/r}$  edges of  $\mathcal{F}(r, d)$ , we proceed inductively as follows. By looking at the edges of  $\mathcal{F}(r, d)$  containing  $i$ , we see that  $\mathcal{F}(r-1, d-a_i)$  has at least  $l^{(r-1)/r}$  edges. From the inductive hypothesis applied to  $\mathcal{F}(r-1, d-a_i)$ , we conclude that there exist



$r' \leq r - 1 < r$  and  $d' \leq d - a_i < d$  such that  $\mathcal{F}(r', d')$  contains a matching of size at least

$$\frac{(l^{(r-1)/r})^{1/r-1}}{r-1} = \frac{l^{1/r}}{r-1} \geq \frac{l^{1/r}}{r}.$$

Now suppose that each  $i \in [n]$  is contained in at most  $l^{(r-1)/r}$  edges of  $\mathcal{F}(r, d)$ . In this case, we claim that  $\mathcal{F}(r, d)$  contains a matching of size  $l^{1/r}/r$ ; in fact, we claim that any maximal matching of  $\mathcal{F}(r, d)$  contains at least  $l^{1/r}/r$  edges. Indeed, if  $\mathcal{M}$  is a maximal matching of  $\mathcal{F}(r, d)$ , then each edge of  $\mathcal{F}(r, d)$  meets at least one edge of  $\mathcal{M}$ . However, each edge of  $\mathcal{M}$  meets at most  $rl^{(r-1)/r}$  edges of  $\mathcal{F}(r, d)$ . It follows that the number of edges in  $\mathcal{M}$  is at least  $l/r l^{(r-1)/r} = l^{1/r}/r$ .  $\square$

**Proposition 4.3.** *For any  $r \in \mathbb{N}$ , there exist positive integers  $1 \leq r' \leq r$  and  $d' \in \mathbb{N}$  for which  $\mathcal{F}(r', d')$  contains a matching of size at least  $n/(r^2 m^{1/r})$ .*

*Proof.* As  $a_1, \dots, a_n$  are positive integers whose sum is  $m$ , it is clear that  $\sum_{i \in I} a_i \in [m]$  for all  $I \in [n]^{(r)}$ . Consequently, there exists a  $d \in [m]$  for which  $\mathcal{F}(r, d)$  contains at least  $\binom{n}{r}/m$  edges. It follows from Proposition 4.2 that there exist positive integers  $r' \leq r$  and  $d' \leq d$  for which  $\mathcal{F}(r', d')$  contains a matching of size at least

$$\frac{1}{r} \left( \frac{\binom{n}{r}}{m} \right)^{1/r} \geq \frac{1}{r} \left( \frac{n^r}{r^r m} \right)^{1/r} = \frac{n}{r^2 m^{1/r}}. \quad \square$$

We are now ready to prove Lemma 4.1.

*Proof of Lemma 4.1.* To prove the lemma, we greedily apply Proposition 4.3 so as to cover  $[n]$  with  $(\log m)^3$  matchings, where each matching is from one of the hypergraphs  $\mathcal{F}(r, d)$ . We can then conclude that one of these matchings covers a  $1/(\log m)^3$  proportion of the total weight, whence follows the lemma.

Define a collection of sets  $A_1 \supset A_2 \supset \dots$  recursively as follows. First set  $A_1 = [n]$ . Assume that we have defined  $A_t \subset [n]$ , the set of uncovered points after  $t$  steps. We know from Proposition 4.3 that we can find  $1 \leq r_t \leq \log m$ ,  $k_t \geq |A_t|/(\log m)^2$  and pairwise disjoint sets  $I_{t,1}, \dots, I_{t,k_t} \in A_t^{(r_t)}$  such that, for some  $d_t \in \mathbb{N}$ ,  $\sum_{i \in I_{t,j}} a_i = d_t$  for each  $1 \leq j \leq k_t$ . Now define

$$A_{t+1} = A_t \setminus \bigcup_{j=1}^{k_t} I_{t,j}.$$

We claim that it takes at most  $(\log m)^3$  steps before we cover all the elements of  $[n]$ . To see this, simply note that at stage  $t+1$ , the number of elements we remove from  $A_t$  to form  $A_{t+1}$  is  $k_t r_t \geq k_t \geq |A_t|/(\log m)^2$  and hence,

$$|A_{t+1}| \leq \left( 1 - \frac{1}{(\log m)^2} \right)^t n < e^{-t/(\log m)^2} n.$$

Since  $n \leq m$ , it follows that  $A_{(\log m)^3+1} = \emptyset$ . It follows that there exists a  $t \leq (\log m)^3$  for which the weight covered by the matching  $I_{t,1}, \dots, I_{t,k_t}$  is at least  $m/(\log m)^3$ , thus concluding the proof of the lemma.  $\square$

## 5. PROOF OF THE MAIN RESULT

As we remarked in Section 3, to prove Theorem 1.2, we shall find it useful to work with multigraphs. Before we proceed further, we set out some notation.

Let  $G = (V, E)$  be a multigraph. We say that two vertices are neighbours in  $G$ , or are adjacent to each other in  $G$ , if they are joined by *at least one edge* in  $G$ . Given  $v \in V$ , we write  $d(v)$  for the number of edges incident to  $v$ , and  $\Gamma(v)$  for the set of neighbours of  $v$ ; as  $d(v)$  is not necessarily equal to the cardinality of  $\Gamma(v)$  in a multigraph, we write  $\gamma(v) = |\Gamma(v)|$  for the number of distinct neighbours of  $v$ . Given  $v \in V$  and  $t \in \mathbb{N}$ , we write  $\Gamma_t(v)$  for the set of vertices joined to  $v$  by exactly  $t$  parallel edges and define  $\gamma_t(v) = |\Gamma_t(v)|$ . So for example,  $\Gamma_0(v)$  is the set of vertices not adjacent to  $v$ ,  $\Gamma(v) = \bigcup_{t \geq 1} \Gamma_t(v)$ ,  $\gamma(v) = \sum_{t \geq 1} \gamma_t(v)$ , and  $d(v) = \sum_{t \geq 0} t\gamma_t(v)$ .

Given  $U \subset V$ , we denote the subgraph of  $G$  induced by  $U$  by  $G[U]$ . For a subgraph  $H$  of  $G$ , we shall write  $e(H)$  for the number of edges of  $H$  counted with multiplicity, and for  $U \subset V$ , we write  $e(U)$  to denote  $e(G[U])$ . Finally, we define the multiplication table  $\mathcal{M}(G)$  of  $G$  as we did for simple graphs by setting

$$\mathcal{M}(G) = \{e(U) : U \subset V\}.$$

If  $G = (X, Y; E)$  is a bipartite graph with vertex classes  $X$  and  $Y$ , then note that for any  $X' \subset X$  and  $Y' \subset Y$ , we have

$$e(X' \cup Y') = \sum_{x \in X'} d(x) - e(X' \cup (Y \setminus Y'));$$

we shall make use of this simple observation repeatedly in the proof of Theorem 1.2.

The next lemma will be useful when dealing with sparse graphs in the proof of Theorem 1.2; an analogous proposition for *simple* bipartite graphs appears in [4].

**Lemma 5.1.** *Let  $G = (X, Y; E)$  be a bipartite multigraph with at most  $r \geq 1$  parallel edges between any pair of vertices and suppose that each vertex of  $G$  has positive degree. If  $|X| = n$ , then*

$$|\mathcal{M}(G) \cap [l]| \geq \frac{l}{2r}$$

for each  $1 \leq l \leq n$ .

*Proof.* If a vertex  $y \in Y$  is such that each of its neighbours in  $X$  has two or more distinct neighbours in  $Y$ , we delete  $y$  from  $Y$ . Doing this repeatedly if necessary, we assume that every  $y \in Y$  has a neighbour  $x_y \in X$  with  $\Gamma(x_y) = \{y\}$ . Note that even after these deletions, every vertex still has positive degree.

First, suppose that there exists a  $y \in Y$  with  $\gamma(y) > n/2r$ . Then if  $X_1 \subset \dots \subset X_{l/2r} \subset \Gamma(y)$  is a strictly monotone increasing sequence of sets with  $|X_i| = i$ , then we see that the sizes of the subgraphs induced by the sets  $X_i \cup \{y\}$  are all distinct and contained in  $[l]$ .

Hence, suppose that  $\gamma(y) \leq n/2r$  for every  $y \in Y$ . We shall construct a sequence  $X_0, \dots, X_{n/2r}$  of subsets of  $X$  with  $|X_i| \leq i$  with the property that the sequence  $(e_i)_{i=0}^{n/2r}$  defined by  $e_i = e(X_i \cup Y)$  satisfies  $0 < e_{i+1} - e_i \leq r$  for each  $i \geq 0$ . If we can do this, then we are done since

$$e_1, \dots, e_{l/2r} \in \mathcal{M}(G) \cap [l]$$

for each  $l \in [n]$ .

We build the sets  $X_i$  recursively. We begin by setting  $X_0 = \emptyset$ . Having constructed  $X_i$ , we construct  $X_{i+1}$  as follows. If there exists a  $y \in Y$  with  $x_y \notin X_i$ , we take  $X_{i+1} = X_i \cup \{x_y\}$  in which case it is clear that  $|X_{i+1}| \leq i+1$ ; we also have  $0 < e_{i+1} - e_i \leq r$  since  $y$  and  $x_y$  are joined by at least 1 and at most  $r$  edges.

Now suppose that  $x_y \in X_i$  for every  $y \in Y$ . Since we have assumed that  $\gamma(y) \leq n/2r$  for each  $y \in Y$ , and since any two vertices of  $G$  are joined by at most  $r$  parallel edges, it follows that  $e(G) \leq n|Y|/2$ . Since  $|X_i| \leq i \leq n/2$ , we conclude by double counting that there is a vertex  $x \in X \setminus X_i$  such that  $d(x) \leq |Y|$ . Choose some  $k$  vertices  $y_1, \dots, y_k \in Y$  so that

$$0 < d(x) - \sum_{j=1}^k d(x_{y_j}) \leq r.$$

This is possible since  $\sum_{y \in Y} d(x_y) \geq |Y| \geq d(x)$  and since, for each  $y \in Y$ ,  $d(x_y) \leq r$  as  $y$  is the sole neighbour of  $x_y$ . Now define

$$X_{i+1} = \left( X_i \setminus \bigcup_{j=1}^k x_{y_j} \right) \cup \{x\}.$$

Clearly,  $|X_{i+1}| \leq i+1$ . Since we have assumed that  $x_y \in X_i$  for each  $y \in Y$ , we also see that

$$e_{i+1} = e_i + d(x) - \sum_{j=1}^k d(x_{y_j}).$$

We have shown how to construct  $X_{i+1}$  with all the requisite properties; this completes the proof.  $\square$

We are now in a position to prove Theorem 1.2.

*Proof of Theorem 1.2.* Let  $G = (X, Y; E)$  be a bipartite graph with  $m$  edges; all the inequalities in our proof will hold when  $m$  is sufficiently large.

The first step in proving Theorem 1.2 is to pass from  $G$  to an induced subgraph of  $G$  within which we have better control over the vertex degrees, while at the same time retaining a large fraction of the edges of  $G$  in this induced subgraph.

Let  $|X| = n_1$  and let  $x_1, \dots, x_{n_1}$  be the vertices of  $X$ . Applying Lemma 4.1 to the sequence  $d(x_1), \dots, d(x_{n_1})$ , we see that it is possible to find positive integers  $k_1, r_1, d_1 \in \mathbb{N}$  and pairwise disjoint sets  $X_1, \dots, X_{k_1} \in X^{(r_1)}$  such that  $r_1 \leq \log m$ ,  $k_1 d_1 \geq m/(\log m)^3$ , and

$$\sum_{x \in X_j} d(x) = d_1$$

for  $1 \leq j \leq k_1$ . We delete the vertices  $X \setminus (\bigcup_{j=1}^{k_1} X_j)$  from  $G$  and also discard any vertices of  $Y$  which subsequently become isolated; note that our graph still has  $k_1 d_1 \geq m/(\log m)^3$  edges.

If  $k_1$  is small, we need to work a bit harder. Suppose that  $k_1 < m^{1/2}/(\log m)^4$ . Let  $|Y| = n_2$  and let  $y_1, \dots, y_{n_2}$  be the vertices of  $Y$ . We again apply Lemma 4.1, but on this occasion to the sequence  $d(y_1), \dots, d(y_{n_2})$ , to find positive integers  $k_2, r_2, d_2 \in \mathbb{N}$  and pairwise disjoint sets  $Y_1, \dots, Y_{k_2} \in Y^{(r_2)}$  such that  $r_2 \leq \log m$ ,  $k_2 d_2 \gg m/(\log m)^6$ , and

$$\sum_{y \in Y_j} d(y) = d_2$$

for  $1 \leq j \leq k_2$ . We then delete the vertices  $Y \setminus (\bigcup_{j=1}^{k_2} Y_j)$  from  $G$  and discard any vertices of  $X$  which subsequently become isolated. Observe that  $G$  still has  $k_2 d_2 \gg m/(\log m)^6$  edges.

Notice that after these deletions,  $|X| \leq k_1 r_1 \leq k_1 \log m$  and  $|Y| = k_2 r_2 \leq k_2 \log m$ . Since  $G$  still has  $\Omega(m/(\log m)^6)$  edges, it follows that

$$k_1 k_2 (\log m)^2 \gg \frac{m}{(\log m)^6}.$$

Consequently, if  $k_1 < m^{1/2}/(\log m)^4$ , then  $k_2 \gg (m^{1/2}/(\log m)^4)$ .

Relabelling  $X$  and  $Y$  if necessary, note that  $G$  now has the property that there exist positive integers  $k, r, d \in \mathbb{N}$  such that  $k \gg m^{1/2}/(\log m)^4$ ,  $r \leq \log m$ ,  $e(G) = kd \gg m/(\log m)^6$ , and there exists a partition

$$X = \bigcup_{j=1}^k X_j$$

of  $X$  into  $k$  sets each of cardinality  $r$  with the property that

$$\sum_{x \in X_j} d(x) = d$$

for  $1 \leq j \leq k$ .

Let  $H = (X_H, Y_H; E_H)$  be the half-regular bipartite multigraph obtained from  $G$  by contracting the vertices of each  $X_j$  into a single vertex for  $1 \leq j \leq k$ . Clearly,

$$\mathcal{M}(H) \subset \mathcal{M}(G),$$

so it suffices to bound  $|\mathcal{M}(H)|$  from below. Henceforth, we shall work exclusively with  $H$ , so in what follows, all vertex degrees, neighbourhoods, etc. will be with respect to the multigraph  $H$ . For easy reference, let us list the properties of  $H$  that we shall require in the rest of the proof.

- (1)  $H$  has no isolated vertices.
- (2) There are at most  $r \leq \log m$  parallel edges between any two vertices of  $H$ .
- (3) For each  $x \in X_H$ ,  $d(x) = d$ .
- (4)  $|X_H| = k \gg m^{1/2}/(\log m)^4$ .
- (5)  $e(H) = kd \gg m/(\log m)^6$ .

Our goal now is to establish that  $|\mathcal{M}(H)| \gg m/(\log m)^{12}$ . We may assume that  $d \geq (\log m)^4$ . If not, then  $k \gg m/(\log m)^{10}$  since  $kd \gg m/(\log m)^6$ . But then  $|\mathcal{M}(H)| \geq k \gg m/(\log m)^{10}$  since clearly,  $d, 2d, \dots, kd \in \mathcal{M}(H)$ .

We claim that we are also done if there exists a vertex  $y \in Y_H$  and  $0 \leq a < b \leq r$  such that  $\gamma_a(y), \gamma_b(y) \geq k/2(\log m)^2$ . Indeed, if such a  $y$  exists, choose  $V_1 \subset \Gamma_a(y)$  and  $V_2 \subset \Gamma_b(y)$  and note that

$$e(V_1 \cup V_2 \cup (Y_H \setminus \{y\})) = (d-a)|V_1| + (d-b)|V_2|,$$

from which it follows that

$$\mathcal{S}(\gamma_a(y) \circ (d-a), \gamma_b(y) \circ (d-b)) \subset \mathcal{M}(H).$$

Note that  $\gcd(d-a, d-b) \leq b-a \leq \log m$ . Since  $k \gg m^{1/2}/(\log m)^4$ ,  $d \geq (\log m)^4$  and  $kd \gg m/(\log m)^6$ , it follows from Lemma 2.2 that

$$\begin{aligned} |\mathcal{M}(H)| &\geq \frac{k}{2(\log m)^2} \min \left\{ \frac{k}{2(\log m)^2}, \frac{d - \log m}{\log m} \right\} \\ &\gg \min \left\{ \frac{k^2}{(\log m)^4}, \frac{kd}{(\log m)^3} \right\} \\ &\gg \min \left\{ \frac{m}{(\log m)^{12}}, \frac{m}{(\log m)^9} \right\} \gg \frac{m}{(\log m)^{12}}. \end{aligned}$$

Hence, in what follows, we shall assume that  $d \geq (\log m)^4$  and that for each  $y \in Y_H$ , there is at most one  $0 \leq \tau \leq r$  such that  $\gamma_\tau(y) \geq k/2(\log m)^2$ .

Since  $\sum_{i=0}^r \gamma_i(y) = k$  and  $r \leq \log m$ , a consequence of assuming there is at most one  $0 \leq \tau \leq r$  for which  $\gamma_\tau(y) \geq k/2(\log m)^2$  is that there in fact exists a unique  $\tau$  for which  $\gamma_\tau(y) \geq k(1 - 1/2 \log m)$ ; we call this unique value  $\tau$  the *type of  $y$*  and say that

the vertex  $y$  is of *type*- $\tau$ . In the rest of the proof, each  $y \in Y_H$  will be assumed to have unique type  $\tau \leq \log m$ .

Note that a vertex  $y \in Y_H$  of type-0 only has a few distinct neighbours in  $X_H$ . We shall distinguish two cases depending on the number of type-0 vertices of  $Y_H$ . We first deal with the case where most vertices are of type-0.

**Case 1: All but at most  $d/2 \log m$  vertices of  $Y_H$  are of type-0.** Note that since each  $x \in X_H$  has at least  $d/r \geq d/\log m$  distinct neighbours in  $Y_H$ , each  $x \in X_H$  is adjacent to at least one vertex of type-0 in  $Y_H$  as at most  $d/2 \log m$  vertices of  $Y_H$  are of nonzero type. Next, observe that if  $y \in Y_H$  is of type-0, then

$$\gamma(y) = \sum_{t=1}^r \gamma_t(y) < \frac{rk}{2(\log m)^2} \leq \frac{k}{2 \log m}.$$

Consequently, we can greedily construct a set  $U \subset Y_H$  of type-0 vertices such that the set

$$S = \bigcup_{y \in U} \Gamma(y) \subset X_H$$

satisfies, writing  $s = |S|$ ,

$$\frac{k}{2} - \frac{k}{2 \log m} \leq s \leq \frac{k}{2}.$$

Let  $F$  be the subgraph of  $H$  induced by  $S \cup U$ . We conclude from Lemma 5.1 that

$$\mathcal{M}(F) \cap [d] \geq \min \left\{ \frac{k}{6 \log m}, \frac{d}{2 \log m} \right\}.$$

The inequality above follows directly from Lemma 5.1 when  $s \geq d$ . If  $s < d$ , then since  $s \geq k/2 - k/2 \log m > k/3$ , we have  $\mathcal{M}(F) \cap [k/3] \subset \mathcal{M}(F) \cap [d]$  and the claimed inequality once again follows from Lemma 5.1. We conclude that  $\mathcal{M}(F)$  contains  $\Omega(\min\{k, d\}/\log m)$  different values modulo  $d$ .

Now, if  $S' \subset S$ ,  $X' \subset X_H \setminus S$  and  $U' \subset U$ , then

$$e(S' \cup X' \cup (Y_H \setminus U')) = d|X'| + d|S'| - e(S' \cup U').$$

As we have already observed, there exist  $\Omega(\min\{k, d\}/\log m)$  choices of  $S'$  and  $U'$  for which the quantities  $d|S'| - e(S' \cup U')$  are all distinct modulo  $d$ ; since  $|X'|$  can be any integer between 0 and  $k - s \geq k/2$ , we see that

$$|\mathcal{M}(H)| \gg k \min \left\{ \frac{k}{\log m}, \frac{d}{\log m} \right\} = \min \left\{ \frac{k^2}{\log m}, \frac{kd}{\log m} \right\} \gg \frac{m}{(\log m)^9}.$$

**Case 2: At least  $d/2 \log m$  vertices of  $Y_H$  are not of type-0.** Set  $p = \log m$  and  $q = d/4 \log m$ . As  $d \geq (\log m)^4$ , we can find  $1 \leq \tau \leq \log m$  and a set  $U \subset Y_H$  of  $p$

vertices all of which are of type- $\tau$ . Let

$$S = \bigcap_{y \in U} \Gamma_\tau(y) \subset X_H;$$

since  $\gamma_\tau(y) \geq k(1 - 1/2 \log m)$  for each  $y \in U$ , note that  $s = |S| \geq k/2$ .

Now choose  $q$  vertices from  $Y_H \setminus U$ , say  $y_1, \dots, y_q$ , so that each of these vertices is of nonzero type; this is possible since there are at least  $d/(2 \log m) - p > q$  vertices of nonzero type in  $Y_H \setminus U$ . For  $1 \leq i \leq q$ , let the type of  $y_i$  be  $\tau_i > 0$ , let  $Y_i = Y_H \setminus \{y_1, \dots, y_i\}$  and let  $H_i$  be the subgraph of  $H$  induced by  $S \cup Y_i$ . We say that  $H_i$  is *good* if at least  $2s/3$  vertices of  $S$  have the same degree in  $H_i$ . We take  $H_0 = H$ ; clearly,  $H_0$  is good since every vertex of  $S$  has degree  $d$  in  $H_0$ .

**Case 2A:  $H_1, \dots, H_q$  are all good.** Since  $H_i$  is good, we know that there are at least  $2s/3$  vertices of  $S$  with the same degree in  $H_i$ ; let  $\alpha_i$  be this common degree and let  $S_i \subset S$  be the set of those vertices with degree  $\alpha_i$  in  $H_i$ . Clearly,  $\alpha_0 = d$ . We claim that  $\alpha_i = \alpha_{i-1} - \tau_i < \alpha_{i-1}$  for each  $1 \leq i \leq q$ . To see this, first note that every vertex of  $S_{i-1} \cap \Gamma_{\tau_i}(y_i)$  has degree  $\alpha_{i-1} - \tau_i$  in  $H_i$ . Recall that  $s = |S| \geq k/2 \gg m^{1/2}/(\log m)^4$  and  $\gamma_{\tau_i}(y_i) \geq k(1 - 1/2 \log m)$ ; since  $H_{i-1}$  is good,

$$|S_{i-1} \cap \Gamma_{\tau_i}(y_i)| \geq |S_{i-1}| - \frac{k}{2 \log m} \geq \frac{2s}{3} - \frac{k}{2 \log m} > \frac{s}{2}.$$

As we have assumed that  $H_i$  is good, we know that if more than  $s/2$  vertices of  $S$  have the same degree in  $H_i$ , then these vertices must all belong to  $S_i$  and hence,  $\alpha_i = \alpha_{i-1} - \tau_i$ .

If  $S'_i \subset S_i$ , then note the  $e(S'_i \cup Y_i) = \alpha_i |S'_i|$ . Hence, writing  $A = \{\alpha_1, \dots, \alpha_q\}$ , we see that

$$A \cdot [2s/3] \subset \mathcal{M}(H)$$

Clearly,  $\max A \leq d \leq m$  while  $s^3 \gg k^3 \gg m^{3/2}/(\log m)^{12}$ , so by Lemma 2.4,  $|A \cdot [2s/3]| \gg sq/\log s$ . It follows that

$$|\mathcal{M}(H)| \gg \frac{sq}{\log s} \gg \frac{kd}{(\log m)^2} \gg \frac{m}{(\log m)^8}.$$

**Case 2B: One of  $H_1, \dots, H_q$  is not good.** Let  $1 \leq l \leq q$  be the minimal index for which  $H_l$  is not good. Since  $H_0, \dots, H_{l-1}$  are all good, we can, arguing as in the previous case, find  $s/2$  vertices of  $S$  which all have the same degree  $\alpha$  in  $H_l$  with  $\alpha \geq d - l \log m \geq d - q \log m = 3d/4$ ; let  $S_\alpha \subset S$  be this set of vertices. Also, as  $H_l$  is not good, we know that  $S_\beta = S \setminus S_\alpha$  contains at least  $s/3$  vertices.

For  $x \in S_\beta$ , let  $\beta_x \neq \alpha$  denote the degree of  $x$  in  $H_l$ . For each  $x \in S_\beta$ , there exists, by Lemma 2.5, an  $0 \leq f_x \leq \log m$  such that

$$\begin{aligned} \gcd(\alpha - f_x \tau, \beta_x - f_x \tau) &\ll |\alpha - \beta_x|^{1/\log m} \tau (\log m)^2 \\ &\ll m^{1/\log m} \tau (\log m)^2 \ll (\log m)^3. \end{aligned}$$

So there exists  $0 \leq f \leq \log m$  and a set  $S_f \subset S_\beta$  of size at least  $|S_\beta|/\log m \geq s/3 \log m \geq k/6 \log m$  such that for each  $x \in S_f$ ,

$$\gcd(\alpha - f\tau, \beta_x - f\tau) \ll (\log m)^3.$$

Recall that  $p = \log m \geq f$ , and that  $S = \bigcap_{y \in U} \Gamma_\tau(y)$ , where  $U \subset Y_H$  is a set of at least  $p$  vertices of type- $\tau$ . Fix a subset of  $U$  of size  $f$ , say  $U_f$ . We shall only consider the induced subgraphs of  $H[S \cup (Y_l \setminus U_f)]$ . If  $S'_\alpha \subset S_\alpha$  and  $S'_f \subset S_f$ , then note that

$$e(S'_\alpha \cup S'_f \cup (Y_l \setminus U_f)) = (\alpha - f\tau)|S'_\alpha| + \sum_{x \in S'_f} (\beta_x - f\tau).$$

Hence, writing  $\beta$  for the sequence  $(\beta_x - f\tau)_{x \in S_f}$ , we see from the arguments above that

$$\mathcal{S}(\beta, s/2 \circ (\alpha - f\tau)) \subset \mathcal{M}(H)$$

since  $|S_\alpha| \geq s/2$ .

As  $\gcd(\alpha - f\tau, \beta_x - f\tau) \ll (\log m)^3$  for each  $x \in S_f$  and since  $\alpha - f\tau \geq 3d/4 - (\log m)^2 > d/2$ , we use Lemma 2.1 to deduce that

$$\begin{aligned} |\mathcal{M}(H)| &\geq |\mathcal{S}(\beta, s/2 \circ (\alpha - f\tau))| \\ &\gg \frac{s}{2} \min \left\{ \frac{\alpha - f\tau}{(\log m)^3}, |S_f| \right\} \\ &\gg \frac{k}{4} \min \left\{ \frac{d}{(\log m)^3}, \frac{k}{\log m} \right\} \\ &\gg \min \left\{ \frac{kd}{(\log m)^3}, \frac{k^2}{\log m} \right\} \gg \frac{m}{(\log m)^9}. \end{aligned}$$

This concludes the proof of Theorem 1.2.  $\square$

## 6. CONCLUSION

There are a number of problems related to the question studied in this paper worth investigating of which Conjecture 1.1 is perhaps the most natural. We discuss a few other related questions below.

Let  $\mathcal{M}(m)$  denote the minimum value of  $|\mathcal{M}(G)|$  taken over all bipartite graphs  $G$  with  $m$  edges. Trivially,  $\mathcal{M}(m) \leq m$ , and here, we have shown that  $\mathcal{M}(m) \gg m/(\log m)^{12}$ . The question of determining the correct order of magnitude of  $\mathcal{M}(m)$  still remains.

**Problem 6.1.** *Determine the asymptotic order of magnitude of  $\mathcal{M}(m)$ .*

We suspect Problem 6.1 might be difficult. For example, it is not at all clear that  $\mathcal{M}(m)$  is an increasing function; indeed, we believe otherwise. We propose the following question as a possible first step towards settling Problem 6.1.

**Problem 6.2.** *Is  $\mathcal{M}(m) = o(m)$  for every  $m \in \mathbb{N}$ ?*



We know that  $\mathcal{M}(n^2) = o(n^2)$  and in general, we suspect that the exact value of  $\mathcal{M}(m)$  depends a great deal on how close  $m$  is to a number with a reasonably ‘balanced’ factorisation. Let us say that  $m$  is *k-balanced* if there exist positive integers  $a, b \geq k$  such that  $m = ab$ . The set of positive integers  $m$  such that  $m$  is  $(\log m)$ -balanced has asymptotic density 1 in  $\mathbb{N}$ . If  $m = ab$  is a  $(\log m)$ -balanced factorisation of  $m$ , then as noted in [19], one can show that  $\mathcal{M}(m) = o(m)$  by considering  $K_{a,b}$ , the complete bipartite graph between two disjoint sets of size  $a$  and  $b$ , and using Ford’s estimates for the size of the set  $[a] \cdot [b]$ . It would be interesting to decide if one can say something similar for all sufficiently large positive integers.

Finally, it would be interesting to determine the structure of extremal graphs. Recall that Conjecture 1.1 asserts that the amongst all bipartite graphs with  $n^2$  edges,  $K_{n,n}$  has the smallest multiplication table. Of course one could, and should, ask what the extremal graphs are when the number of edges is no longer a square. We believe that if  $|\mathcal{M}(G_m)| = \mathcal{M}(m)$  for some bipartite graph  $G_m$  with  $m$  edges, then  $G_m$  must necessarily contain a large subgraph that ‘resembles’ a complete bipartite graph. For example, a natural conjecture is that  $\mathcal{M}(n(n+1)) = |\mathcal{M}(K_{n,n+1})|$ ; in general however, we have no precise guesses for what the extremal graphs are.

#### ACKNOWLEDGEMENTS

The problem of proving Conjecture 1.1 in a weak quantitative form was proposed by the first author in August 2013 at the 5<sup>th</sup> Emléktábla Workshop in Budapest; also, some of the ideas used here to prove Theorem 1.2 were developed while the authors were visitors at the IMT Institute for Advanced Studies Lucca. The authors are grateful to Guido Caldarelli and the other members of the Complex Networks Group at IMT Lucca for their hospitality and in addition, the first author is grateful to the organisers of the Emléktábla Workshop for their hospitality.

#### REFERENCES

1. N. Alon, J. Balogh, A. V. Kostochka, and W. Samotij, *Sizes of induced subgraphs of Ramsey graphs*, Combin. Probab. Comput. **18** (2009), 459–476. 3
2. N. Alon and B. Bollobás, *Graphs with a small number of distinct induced subgraphs*, Discrete Math. **75** (1989), 23–30. 3
3. N. Alon and A. V. Kostochka, *Induced subgraphs with distinct sizes*, Random Structures Algorithms **34** (2009), 45–53. 3
4. N. Alon, M. Krivelevich, and B. Sudakov, *Induced subgraphs of prescribed size*, J. Graph Theory **43** (2003), 239–251. 3, 10
5. M. Axenovich and J. Balogh, *Graphs having small number of sizes on induced k-subgraphs*, SIAM J. Discrete Math. **21** (2007), 264–272. 3

6. B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1998. [4](#)
7. N. Calkin, A. Frieze, and B. D. McKay, *On subgraph sizes in random graphs*, *Combin. Probab. Comput.* **1** (1992), 123–134. [3](#)
8. P. Erdős, *Some remarks on number theory*, *Riveon Lematematika* **9** (1955), 45–48. [2](#)
9. ———, *An asymptotic inequality in the theory of numbers*, *Vestnik Leningrad. Univ.* **15** (1960), 41–49. [2](#)
10. ———, *Some of my favorite problems in various branches of combinatorics*, *Matematiche (Catania)* **47** (1992), 231–240. [3](#)
11. ———, *Some of my favourite problems in number theory, combinatorics, and geometry*, *Resenhas* **2** (1995), 165–186. [3](#)
12. ———, *Some recent problems and results in graph theory*, *Discrete Math.* **164** (1997), 81–85. [3](#)
13. P. Erdős and A. Hajnal, *On the number of distinct induced subgraphs of a graph*, *Discrete Math.* **75** (1989), 145–154. [3](#)
14. K. Ford, *The distribution of integers with a divisor in a given interval*, *Ann. of Math.* **168** (2008), 367–433. [1](#), [2](#)
15. R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1988. [3](#)
16. G. J. O. Jameson, *The prime number theorem*, London Mathematical Society Student Texts, Cambridge University Press, Cambridge, 2003. [5](#)
17. D. Koukoulopoulos, *Localized factorizations of integers*, *Proc. Lond. Math. Soc.* **101** (2010), 392–426. [3](#)
18. ———, *On the number of integers in a generalized multiplication table*, *J. Reine Angew. Math.* **689** (2014), 33–99. [3](#)
19. B. Narayanan, *Exactly  $m$ -coloured complete infinite subgraphs*, *J. Combin. Theory Ser. B* **106** (2014), 163–173. [3](#), [17](#)
20. I. Z. Ruzsa, *Erdős and the integers*, *J. Number Theory* **79** (1999), 115–163. [3](#)
21. J. Solymosi, *Bounding multiplicative energy by the sumset*, *Adv. Math.* **222** (2009), 402–408. [3](#)
22. G. Tenenbaum, *Sur la probabilité qu’un entier possède un diviseur dans un intervalle donné*, *Compositio Math.* **51** (1984), 243–263. [2](#)

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UK

*Email address:* `b.p.narayanan@dpmms.cam.ac.uk`

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF MEMPHIS, MEMPHIS TN 38152, USA

*Email address:* `julian.sahasra@gmail.com`

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, UNIVERSITY OF CAMBRIDGE, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, UK

*Email address:* `i.tomon@dpmms.cam.ac.uk`