Holes in graphs

Yuejian Peng

Department of Mathematics and Computer Science
Emory University, Atlanta, USA
peng@mathcs.emory.edu

Vojtech Rödl*

Department of Mathematics and Computer Science
Emory University, Atlanta, USA
rodl@mathcs.emory.edu

Andrzej Ruciński †

Department of Discrete Mathematics
Adam Mickiewicz University, Poznań, Poland
rucinski@amu.edu.pl

Submitted: November 7, 2000; Accepted: October 14, 2001.
MR Subject Classifications: 05C35

Abstract

The celebrated Regularity Lemma of Szemerédi asserts that every sufficiently large graph G can be partitioned in such a way that most pairs of the partition sets span ϵ-regular subgraphs. In applications, however, the graph G has to be dense and the partition sets are typically very small. If only one ϵ-regular pair is needed, a much bigger one can be found, even if the original graph is sparse. In this paper we show that every graph with density d contains a large, relatively dense ϵ-regular pair. We mainly focus on a related concept of an (ϵ, σ)-dense pair, for which our bound is, up to a constant, best possible.

1 Introduction

Szemerédi's Regularity Lemma is one of the most powerful tools in extremal graph theory. It guarantees an ϵ-regular partition of every graph G with n vertices, but the size of each

*Research supported by NSF grant DMS 9704114.

†Research supported by KBN grant 2 P03A 032 16. Part of this research was done during the author's visit to Emory University.

ϵ-regular pair is at most n/T, where T is the tower of 2's of height $(1/\epsilon)^{\frac{1}{16}}$ ([4]). However, in some applications, only one pair is needed. That was already observed and explored by Komlós (see [8]) and Haxell [6]. The goal of this paper is to estimate the size of the largest such pair that can be found in any graph of given size and density. The density may decay to 0 with $n \to \infty$.

The density of a bipartite graph $G = (V_1, V_2, E)$ is defined as

$$d(G) = \frac{|E|}{|V_1||V_2|},$$

and the density of a pair (U_1, U_2), where $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$, is defined as

$$d(U_1, U_2) = \frac{e(U_1, U_2)}{|U_1||U_2|},$$

where $e(U_1, U_2)$ is the number of edges of G with one endpoint in U_1 and the other in U_2.

Definition 1.1 *Let $G = (V_1, V_2, E)$ be a bipartite graph and $0 < \epsilon < 1$. A pair (U_1, U_2), where $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$, is called ϵ-regular if for every $W_1 \subseteq U_1$ and $W_2 \subseteq U_2$ with $|W_1| \geq \epsilon|U_1|$ and $|W_2| \geq \epsilon|U_2|$, we have*

$$(1-\epsilon)d(U_1, U_2) \leq d(W_1, W_2) \leq (1+\epsilon)d(U_1, U_2).$$

Our first result states that in every bipartite graph one can find a reasonably large and relatively dense ϵ-regular pair.

Theorem 1.1 *Let $0 < \epsilon, d < 1$. Then every bipartite graph $G = (V_1, V_2, E)$ with $|V_1| = |V_2| = n$ and $d(G) = d$ contains an ϵ-regular pair (U_1, U_2) with density not smaller than $(1 - \frac{\epsilon}{3})d$ and $|U_1| = |U_2| \geq \frac{n}{2}d^{c/\epsilon^2}$, where c is an absolute constant.*

The constant c in Theorem 1.1 is determined by inequality (40). For instance, one can take $c = 50$.

In most applications the whole strength of ϵ-regular pairs is not used. Instead, it is only required that $d(W_1, W_2)$ is not much smaller than $d(U_1, U_2)$ whenever $W_1 \subseteq U_1$ and $W_2 \subseteq U_2$ are large enough. This observation leads to the following definition.

Definition 1.2 *Let $G = (V_1, V_2, E)$ be a bipartite graph and $0 < \epsilon, \sigma < 1$. A pair (U_1, U_2), where $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$, is called (ϵ, σ)-dense if for every $W_1 \subseteq U_1$ and $W_2 \subseteq U_2$ with $|W_1| \geq \epsilon|U_1|$ and $|W_2| \geq \epsilon|U_2|$, we have $e(W_1, W_2) \geq \sigma|W_1||W_2|$. The graph G itself is called (ϵ, σ)-dense if (V_1, V_2) is an (ϵ, σ)-dense pair.*

Now, let us consider the following problem. For a bipartite graph G with n vertices in each color class and density d, we want to find an $(\epsilon, d/2)$-dense pair as large as possible. (The choice of $\sigma = d/2$ is not essential here.)

Definition 1.3 *For any given* $0 < \epsilon, d < 1$ *and a positive integer* n*,* $f(\epsilon, d, n)$ *is the largest integer* f *such that every bipartite graph* G *with* n *vertices in each color class and density at least* d *contains an* $(\epsilon, d/2)$*-dense subgraph with* f *vertices in each color class.*

As for $\epsilon \leq 2 - \sqrt{2.5}$, every ϵ-regular pair with density at least $(1-\epsilon/3)d$ is $(\epsilon, d/2)$-dense, Theorem 1.1 immediately implies that $f(\epsilon, d, n) \geq \frac{n}{2} d^{c/\epsilon^2}$.

In 1991, Komlós stated the following lower bound for $f(\epsilon, d, n)$.

Theorem 1.2 *[8] For all* $0 < \epsilon \leq \epsilon_0$*,* $0 < d < 1$ *and for all integers* n*,*

$$f(\epsilon, d, n) \geq n d^{(3/\epsilon) \ln(1/\epsilon)}.$$

In Section 2 of this paper we prove a different bound which is better for small values of ϵ.

Theorem 1.3 *For all* $0 < \epsilon < 1$*,* $0 < d < 1$*, and for all integers* n*,*

$$f(\epsilon, d, n) \geq \frac{1}{2} n d^{12/\epsilon}.$$

We also prove the following upper bound on $f(\epsilon, d, n)$, which shows that, up to a constant, Theorem 1.3 is best possible.

Theorem 1.4 *For all* $0 < \epsilon \leq \epsilon_0$ *and* $0 < d \leq d_0$*, there exists* $n_0 < (1/d)^{1/(12\epsilon)}$ *such that for all* $n \geq n_0$*,*

$$f(\epsilon, d, n) < 4 n d^{c/\epsilon},$$

where c *is an absolute constant.*

In fact, we prove a stronger result than Theorem 1.4.

Definition 1.4 *Let* $G = (V_1, V_2, E)$ *be a bipartite graph and* $0 < \epsilon < 1$*. A pair* (U_1, U_2)*, where* $U_1 \subseteq V_1$*,* $U_2 \subseteq V_2$*, is said to contain an* ϵ*-hole if there exist* $W_1 \subseteq U_1$ *and* $W_2 \subseteq U_2$ *with* $|W_1| \geq \epsilon |U_1|$ *and* $|W_2| \geq \epsilon |U_2|$ *such that* $e(W_1, W_2) = 0$*.*

By definition, if a pair contains an ϵ-hole, then it cannot be (ϵ, σ)-dense for any $\sigma > 0$.

Definition 1.5 *For any given* $0 < \epsilon, d < 1$ *and a positive integer* n*, let* $h(\epsilon, d, n)$ *be the largest integer* h *such that, every bipartite graph* G *with* n *vertices in each color class and density at least* d *contains a subgraph with* h *vertices in each color class and with no* ϵ*-hole.*

Clearly, $f(\epsilon, d, n) \leq h(\epsilon, d, n)$.

Theorem 1.5 *For all* $0 < \epsilon \leq \epsilon_0$ *and* $0 < d \leq d_0$ *there exists* $n_0 < (1/d)^{1/(12\epsilon)}$ *such that for all* $n \geq n_0$,

$$h(\epsilon, d, n) < 4nd^{c/\epsilon},$$

where c *is an absolute constant.*

With no effort to optimize, it follows from the proofs of Theorems 1.4 and 1.5 that the constant c appearing in them can be equal to 1/2000.

2 Lower bound

In this section we prove the lower bound given in Theorem 1.3. That is, we show that any bipartite graph $G = (V_1, V_2, E)$ with n vertices in each color class and density d contains an $(\epsilon, d/2)$-dense bipartite subgraph with at least $\frac{1}{2}nd^{c_1/\epsilon}$ vertices in each color class. We then show that Theorem 1.1 the proof of which is a refinement of the proof of Theorem 1.3.

Before giving the proof of Theorem 1.3, we prove the following claim which plays a crucial role.

Claim 2.1 *Every bipartite graph* $H = (V_1^H, V_2^H, E)$ *with* $|V_1^H| = |V_2^H| = m$ *contains a pair* (U_1, U_2) *satisfying one of the following conditions:*

1. (U_1, U_2) *is an* $(\epsilon, d(H)/2)$*-dense pair and* $|U_1| = |U_2| \geq m/2$,
2. $|U_1| = |U_2| \geq m/4$ *and* $d(U_1, U_2) \geq (1 + \epsilon/8)d(H)$.

Proof: Assuming that H contains no pair satisfying condition *1*, we are going to prove that H contains a pair satisfying condition *2*. For simplicity, we assume that $1/\epsilon$ is an integer.

Since, in particular, H itself is not $(\epsilon, d(H)/2)$-dense, there exist $A_1' \subset V_1^H, B_1' \subset V_2^H$ with $|A_1'| = |B_1'| \geq \epsilon m$ and $e(A_1', B_1') < \frac{d(H)}{2}|A_1'||B_1'|$. By an averaging argument, we can take $A_1 \subset A_1', B_1 \subset B_1'$ satisfying $|A_1| = |B_1| = \frac{\epsilon}{2}m$ and $e(A_1, B_1) < \frac{d(H)}{2}|A_1||B_1|$. (For simplification, we assume that $\frac{\epsilon}{2}m$ is an integer. Later we will make similar assumption which are not essential but simplify our presentation.) Let F_1 be the graph obtained by removing A_1 from V_1^H and B_1 from V_2^H.

By the assumption, F_1 is not an $(\epsilon, d(H)/2)$-dense graph, and we apply the same argument as above to F_1.

In general, after l steps, $l < 1/\epsilon$, we define l disjoint pairs $(A_1, B_1), \cdots, (A_l, B_l)$ of size $|A_i| = |B_i| = \frac{\epsilon}{2}m$ for $1 \leq i \leq l$. Assume that F_l is obtained by removing $\bigcup_{j=1}^{l} A_j$ from V_1^H and $\bigcup_{j=1}^{l} B_j$ from V_2^H. By assumption, F_l is not $(\epsilon, d(H)/2)$-dense, therefore there exists

$A'_{l+1} \subset V_1^H \setminus \bigcup_{j=1}^{l} A_j, B'_{l+1} \subset V_2^H \setminus \bigcup_{j=1}^{l} B_j$ of size $|A'_{l+1}| = |B'_{l+1}| \geq \epsilon\,(1 - l\epsilon/2)\,m \geq \frac{\epsilon}{2}m$ and $e\left(A'_{l+1}, B'_{l+1}\right) < \frac{d(H)}{2}|A'_{l+1}||B'_{l+1}|$. Take $A_{l+1} \subset A'_{l+1}, B_{l+1} \subset B'_{l+1}, |A_{l+1}| = |B_{l+1}| = \frac{\epsilon}{2}m$ and $e\left(A_{l+1}, B_{l+1}\right) < \frac{d(H)}{2}|A_{l+1}||B_{l+1}|$.

After $1/\epsilon$ steps the sets $\bigcup_{j=1}^{1/\epsilon} A_j$ cover a half of V_1^H, and the sets $\bigcup_{j=1}^{1/\epsilon} B_j$ cover a half of V_2^H. Denote $\bar{V}_1 = \bigcup_{j=1}^{1/\epsilon} A_j$ and $\bar{V}_2 = \bigcup_{j=1}^{1/\epsilon} B_j$. Set $e_0 = e\left(\bar{V}_1, \bar{V}_2\right), e_1 = e\left(\bar{V}_1, V_2^H \setminus \bar{V}_2\right), e_2 = e\left(V_1^H \setminus \bar{V}_1, \bar{V}_2\right), e_3 = e\left(V_1^H \setminus \bar{V}_1, V_2^H \setminus \bar{V}_2\right).$

Now we claim that there exists a pair satisfying condition *2*. Indeed, if

$$e_0 \leq (1 - 3\epsilon/8)\,d(H)m^2/4,$$

then

$$e_1 + e_2 + e_3 = d(H)m^2 - e_0 \geq 3\left(1 + \frac{\epsilon}{8}\right)d(H)\frac{m^2}{4}.$$

Therefore, there exists $i \in \{1, 2, 3\}$ satisfying

$$e_i \geq \left(1 + \frac{\epsilon}{8}\right)d(H)\frac{m^2}{4}$$

and we find a pair satisfying condition *2*.

If $e_0 > (1 - 3\epsilon/8)\,d(H)m^2/4$, we define $e_{ij} = e\left(A_i, B_j\right)$. Then

$$\sum_i \sum_{j \neq i} e_{ij} = e_0 - \sum_{i=1}^{1/\epsilon} e\left(A_i, B_i\right) > \left(1 - \frac{3\epsilon}{8}\right)d(H)\frac{m^2}{4} - \frac{1}{\epsilon}\frac{d(H)}{2}\left(\frac{\epsilon m}{2}\right)^2 = \left(1 - \frac{7}{8}\epsilon\right)d(H)\frac{m^2}{4}.$$

For any $I \subset \{1, \ldots, 1/\epsilon\}$ of size $|I| = 1/(2\epsilon)$, we define

$$e\left(I\right) = \sum_{i \in I} \sum_{j \in \{1, \ldots, 1/\epsilon\} \setminus I} e_{ij}.$$

Then $\sum_I e\left(I\right)$ counts each e_{ij} exactly $\binom{1/\epsilon - 2}{1/(2\epsilon) - 1}$ times, where $i \neq j$. Thus, there exists I_0 such that

$$e(I_0) \geq \frac{\sum_I e\left(I\right)}{\binom{1/\epsilon}{1/(2\epsilon)}} = \frac{\binom{1/\epsilon - 2}{1/(2\epsilon) - 1}}{\binom{1/\epsilon}{1/(2\epsilon)}} \sum_i \sum_{j \neq i} e_{ij} > \frac{(1 - 7\epsilon/8)\,d(H)m^2/4}{4\,(1 - \epsilon)} \geq \left(1 + \frac{\epsilon}{8}\right)d(H)\frac{m^2}{16},$$

and consequently the pair $(\bigcup_{i \in I_0} A_i, \bigcup_{j \in \{1, \ldots, 1/\epsilon\} \setminus I_0} B_j)$ satisfies condition *2*. ■

Proof of Theorem 1.3. Let $G = (V_1, V_2, E)$ be any bipartite graph with n vertices in each color class and density d. If G contains a pair satisfying condition *1* in Claim 2.1, then we are done. Otherwise, by Claim 2.1, there exists an induced subgraph $G_1 \subset G$ with at least $n/4$ vertices in each color class and $d(G_1) \geq (1 + \epsilon/8)d$. Applying Claim 2.1 to G_1, if G_1 contains a pair satisfying condition *1* in Claim 2.1, then we have an

$(\epsilon, d(G_1)/2)$-dense pair, which is also an $(\epsilon, d/2)$-dense pair, with at least $n/8$ vertices in each color class, and we are done again. Otherwise we find an induced subgraph $G_2 \subset G_1$ with at least $n/16$ vertices in each color class and $d(G_2) \geq (1+\epsilon/8)^2 d$.

Suppose we have iterated this process s times, obtaining a subgraph G_s of G with at least $n/4^s$ vertices in each color class and density at least $(1+\epsilon/8)^s d$. If the $(s+1)$-th iteration cannot be completed, it means that G_s contains an $(\epsilon, d/2)$-dense subgraph with at least $n/(2 \cdot 4^s)$ vertices in each color class. Because the density of any graph is not larger than 1, we can only iterate this process at most t times, where t is the smallest integer such that

$$\left(1+\frac{\epsilon}{8}\right)^{t+1} d > 1.$$

Hence, at some point an $(\epsilon, d/2)$-dense subgraph with at least $n/(2 \cdot 4^t)$ vertices in each color class must be found. It remains to estimate t from above. By the choice of t, we have $(1+\epsilon/8)^t d \leq 1$, or, equivalently,

$$t \leq \frac{\log_2(1/d)}{\log_2(1+\epsilon/8)},$$

and so

$$4^t = 2^{2t} \leq (1/d)^{\frac{2}{\log_2(1+\epsilon/8)}}.$$

Notice that $\log_2(1+\epsilon/8) \geq \epsilon/6$ for $0 < \epsilon < 1$. Indeed, it follows from the facts that $g(x) = \log_2(1+\epsilon/8) - \epsilon/6$ is concave in $[0,1]$, $g(0) = 0$ and $g(1) > 0$. Therefore

$$\frac{1}{2}\frac{n}{4^t} \geq \frac{1}{2} n d^{12/\epsilon},$$

and consequently we have proved the existence of an $(\epsilon, d/2)$-dense subgraph of G with at least $\frac{1}{2} n d^{12/\epsilon}$ vertices in each color class. This completes the proof of Theorem 1.3. ■

Proof of Theorem 1.1 (Sketch). The proof of Theorem 1.1 is similar to the proof of Theorem 1.3; the only modification is to replace Claim 2.1 by Claim 2.3 below.

The first alternative of Claim 2.3, rather than asking for a large ϵ-regular pair, demands a stronger property which is however easier to analyze.

Definition 2.1 *Let $G = (V_1, V_2, E)$ be a bipartite graph, $0 < \epsilon < 1$. A pair (U_1, U_2), where $U_1 \subseteq V_1$ and $U_2 \subseteq V_2$, is called (ϵ, G)-regular if for every $W_1 \subseteq U_1$ and $W_2 \subseteq U_2$ with $|W_1| \geq \epsilon |U_1|$ and $|W_2| \geq \epsilon |U_2|$, we have*

$$(1-\epsilon/3) d(G) \leq d(W_1, W_2) \leq (1+\epsilon/3) d(G). \tag{1}$$

Fact 2.2 *Every (ϵ, G)-regular pair (U_1, U_2) is ϵ-regular.*

Claim 2.3 *Every bipartite graph $H = (V_1^H, V_2^H, E)$ with $|V_1^H| = |V_2^H| = m$ contains a pair (U_1, U_2) satisfying one of the following conditions:*

1. $|U_1|, |U_2| \geq m/2$ *and* (U_1, U_2) *is* (ϵ, H)*-regular,*
2. $|U_1|, |U_2| \geq m\epsilon/2$ *and* $d(U_1, U_2) \geq (1 + \epsilon/3)d(H)$,
3. $|U_1|, |U_2| \geq m/4$ *and* $d(U_1, U_2) \geq (1 + \epsilon^2/12)d(H)$.

Assuming that H contains no pair satisfying conditions *1* or *2*, and using the same technique as in the proof of Claim 2.1, we can prove that H must contain a pair satisfying condition *3*.

Applying Claim 2.3, one can prove Theorem 1.1 in the same way as we derived Theorem 1.3 from Claim 2.1 (see the Appendix for details). Note that the obtained ϵ-regular pair (U_1, U_2) has density at least $(1 - \epsilon/3)d$. ∎

3 Upper bound

In this section we prove the upper bound for $h(\epsilon, d, n)$ given in Theorem 1.5. To prove that $h(\epsilon, d, n) < u$, we need to find a bipartite graph G with n vertices in each color class and density at least d such that every subgraph of G with u vertices in each color class contains an ϵ-hole. The following construction will be central for the proof.

Let k and t be positive integers, and $[t]$ denote $\{1, 2, \ldots, t\}$. Let $G(k, t) = (V_1, V_2, E)$ be the bipartite graph with

$$V_1 = \{\mathbf{x} = (x_1, x_2, \ldots, x_t) : 1 \leq x_s \leq k, 1 \leq s \leq t.\},$$

$$V_2 = \{\mathbf{y} = (y_1, y_2, \ldots, y_t) : 1 \leq y_s \leq k, 1 \leq s \leq t.\},$$

and $\mathbf{xy} \in E$ if and only if $x_s \neq y_s$ for each $s \in [t]$, where $\mathbf{x} = (x_1, x_2, \ldots, x_t) \in V_1$ and $\mathbf{y} = (y_1, y_2, \ldots, y_t) \in V_2$.

Observe that $G(k, t)$ is a bipartite graph with k^t vertices in each color class and density $\left(\frac{k-1}{k}\right)^t$. For $G(k, t)$ we prove the following property. From now on we set $n_1 = k^t$.

Lemma 3.1 *Let k and t be positive integers and let $0 < \epsilon \leq 1/4k$. For every $U_1 \subseteq V_1$, $U_2 \subseteq V_2$ such that*

$$\min\{|U_1|, |U_2|\} \geq n_1 \left(\frac{e}{2\epsilon k}\right)^{2\epsilon k t} \left(\frac{1 + 4\epsilon k}{\sqrt{2}}\right)^{2t},$$

there exists an ϵ-hole in the subgraph of $G(k, t)$ induced by the sets U_1 and U_2.

Proof: Suppose that there is no ϵ-hole in the subgraph of $G(k,t)$ induced by the sets U_1, U_2. We will estimate $\min\{|U_1|, |U_2|\}$ from above.

For each $s = 1, 2, \ldots, t$, the integer $i \in [k]$ is called *rare* with respect to s in U_1 if

$$|\{\mathbf{x} \in U_1 : x_s = i\}| < \epsilon |U_1|.$$

Otherwise i is called *frequent* with respect to s. Let R_s^1 be the set of all rare values $i \in [k]$ with respect to s in U_1 and F_s^1 be the set of all frequent values $i \in [k]$ with respect to s in U_1. Similarly, let F_s^2 be the set of all frequent values $i \in [k]$ with respect to s in U_2. Note that $F_s^1 \cap F_s^2 = \emptyset$ for each $s \in [t]$, since otherwise the vertices $\mathbf{x} \in U_1$ and $\mathbf{y} \in U_2$ with $x_s = y_s = i \in F_s^1 \cap F_s^2$ would form an ϵ-hole between U_1 and U_2.

Next we are going to prove that more than half of the vertices in U_1 have each less than $2\epsilon k$ rare coordinates. At the same time we give an upper bound on the number of such vertices which enables us to estimate $|U_1|$.

For every $\mathbf{x} = (x_1, \ldots, x_s, \ldots, x_t) \in V_1$, define $S_{\mathbf{x}} = \{s : x_s \in R_s^1\}$. Let $V_1' = \{\mathbf{x} \in V_1 : |S_{\mathbf{x}}| < 2\epsilon kt\}$ and $U_1' = U_1 \cap V_1'$.

Claim 3.2

$$|U_1'| > \frac{1}{2}|U_1|, \tag{2}$$

$$|U_1'| \leq |V_1'| \leq 2\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{2\epsilon k^2 t + \sum_{s=1}^t |F_s^1|}{t}\right)^t. \tag{3}$$

Proof of Claim 3.2: To prove (2), we use a standard double counting argument. Consider an auxiliary bipartite graph $M = (U_1, [t], E(M))$ in which a pair $\{\mathbf{x}, s\} \in E(M)$ if and only if $x_s \in R_s^1$, where $\mathbf{x} = (x_1, x_2, \ldots, x_t) \in U_1$ and $s \in [t]$. By the definition of R_s^1, it is easy to see that $deg_M(s) < \epsilon k |U_1|$ for any $s \in [t]$. Therefore there are fewer than $\frac{1}{2}|U_1|$ vertices $\mathbf{x} \in U_1$ which satisfy $|S_{\mathbf{x}}| = deg_M(\mathbf{x}) \geq 2\epsilon kt$.

Now we prove (3). Let $L \subset [t]$ with $|L| < 2\epsilon kt$. Then by the definition of $S_{\mathbf{x}}$,

$$|\{\mathbf{x} \in V_1 : S_{\mathbf{x}} = L\}| \leq \prod_{q \in L} |R_q^1| \prod_{s \in [t] \setminus L} |F_s^1|.$$

Hence

$$|V_1'| \leq \sum_{L \subset [t], |L| < 2\epsilon kt} k^{|L|} \prod_{s \in [t] \setminus L} |F_s^1|. \tag{4}$$

Since the geometric mean is not larger than the arithmetic mean, we obtain

$$|U_1'| \leq |V_1'| \leq \sum_{l < 2\epsilon kt} \binom{t}{l} \left(\frac{kl + \sum_{s=1}^t |F_s^1|}{t}\right)^t. \tag{5}$$

Since $l < 2\epsilon kt \leq t/2$, we have $\binom{t}{l} \leq \binom{t}{2\epsilon kt} \leq \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt}$, and

$$|V_1'| \leq 2\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{2\epsilon k^2 t + \sum_{s=1}^t |F_s^1|}{t}\right)^t, \tag{6}$$

which completes the proof of the claim. ■

Now we continue the proof of Lemma 3.1. By Claim 3.2

$$|U_1| < 2|U_1'| \le 2|V_1'| \le 4\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{2\epsilon k^2 t + \sum_{s=1}^{t} |F_s^1|}{t}\right)^t. \tag{7}$$

Similarly,

$$|U_2| < 4\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{2\epsilon k^2 t + \sum_{s=1}^{t} |F_s^2|}{t}\right)^t. \tag{8}$$

Since $F_s^1 \cap F_s^2 = \emptyset$ for each $s \in [t]$, we have $\sum_{s=1}^{t} |F_s^1| \le \frac{tk}{2}$ or $\sum_{s=1}^{t} |F_s^2| \le \frac{tk}{2}$. Therefore,

$$\begin{aligned}\min\{|U_1|, |U_2|\} &< 4\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{2\epsilon k^2 t + \frac{kt}{2}}{t}\right)^t & (9)\\ &= 4\epsilon kt \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} (1+4\epsilon k)^t \left(\frac{k}{2}\right)^t. & (10)\end{aligned}$$

Applying the inequality $4\epsilon kt < (1+4\epsilon k)^t$, we finally obtain that

$$\begin{aligned}\min\{|U_1|, |U_2|\} &< \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} (1+4\epsilon k)^{2t} \left(\frac{k}{2}\right)^t & (11)\\ &= n_1 \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}, & (12)\end{aligned}$$

which completes the proof. ■

Now for any $n \ge n_1$, let r and q, where $0 \le q < n_1$, be the positive integers such that $n = rn_1 + q$. We "blow up" the graph $G(k,t)$ in the following sense: fix any q vertices in each color class, and replace each of them by $r+1$ new vertices. At the same time replace every other vertex by r new vertices. Finally, replace every edge of $G(k,t)$ by the corresponding complete bipartite graph ($K_{r,r}$, $K_{r+1,r}$, or $K_{r+1,r+1}$). Denote this new graph by $G_n(k,t) = (V_1^n, V_2^n, E)$. It is easy to see that

$$\frac{r}{r+1}\left(\frac{k-1}{k}\right)^t \le d(G_n(k,t)) \le \frac{r+1}{r}\left(\frac{k-1}{k}\right)^t. \tag{13}$$

For this graph we now prove the following lemma which is very similar to Lemma 3.1. Recall that $n_1 = k^t$.

Lemma 3.3 *Let k and t be positive integers and let $0 < \epsilon \le 1/4k$. For every $n \ge n_1$, and for all $U_1 \subseteq V_1^n$, $U_2 \subseteq V_2^n$ such that*

$$\min\{|U_1|, |U_2|\} \ge 2n \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t},$$

there exists an ϵ-hole in the subgraph of $G_n(k,t)$ induced by the sets U_1 and U_2.

Proof: Assume that there is no ϵ-hole in the subgraph of $G_n(k,t)$ induced by the sets U_1, U_2. For each $s \in [t]$, define rare and frequent values $i \in [k]$ with respect to s, for U_1 and U_2, in the same way as in the proof of Lemma 3.1. We follow the lines of the proof of Lemma 3.1. The only novelty is to multiply the right hand side of equations (4) – (11) by $r+1$. Therefore, we have

$$\min\{|U_1|,|U_2|\} < (r+1)n_1 \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}.$$

Since $(r+1)/r \leq 2$, and thus $(r+1)n_1 \leq 2rn_1 \leq 2(rn_1+q) = 2n$, we obtain

$$\min\{|U_1|,|U_2|\} < 2n \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}. \tag{14}$$

■

The goal of blowing up $G(k,t)$ was to obtain graphs with more vertices than n_1 and still having ϵ-holes in large subgraphs. Next we consider a random "contraction" of $G(k,t)$ to obtain graphs with fewer than n_1 vertices and with the same property.

From now on, to make our description simpler, we set

$$\alpha = \log_k \frac{k}{k-1}, \qquad \delta = \log_k 2 - 2\epsilon k \log_k \frac{e}{2\epsilon k} - 2\log_k(1+4\epsilon k), \qquad n_0 = \max\{n_1^{3\alpha/2}, n_1^{3\delta/2}\}.$$

Note that $n_0 \leq n_1$ when $k \geq 3$, and under this notation,

$$n_1^{-\alpha} = d(G(k,t)) = \left(\frac{k-1}{k}\right)^t$$

and

$$n_1^{-\delta} = \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}.$$

Lemma 3.4 *Let $k \geq 3$ be a positive integer, $0 < \epsilon \leq 1/4k$, and $t > t_0 = t_0(k,\epsilon)$. Then, for every $n_0 \leq n < n_1$, there exists a graph $G_n = (V_1^n, V_2^n, E^n)$ with n vertices in each color class such that*

$$\frac{k-1}{k} n_1^{-\alpha} \leq d(G_n) \leq \frac{k}{k-1} n_1^{-\alpha}, \tag{15}$$

and for all $U_1 \subseteq V_1^n$, $U_2 \subseteq V_2^n$ with

$$\min\{|U_1|,|U_2|\} \geq 4n \left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt} \left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t},$$

there exists an ϵ-hole in the subgraph of G_n induced by the sets U_1 and U_2 .

Proof: We define a random subgraph $G^*(k,t) = (V_1^*, V_2^*, E^*)$ of $G(k,t)$ by choosing uniformly an n-element subset V_1^* of V_1, and independently, an n-element subset V_2^* of V_2, and including to E^* all edges of $G(k,t)$ with one endpoint in V_1^* and the other in V_2^*.

For each $v \in V_1$, let $N(v)$ denote the neighborhood of v in $G(k,t)$. Then $|N(v) \cap V_2^*|$ is a random variable with hypergeometric distribution of expectation $\mathbb{E}(|N(v) \cap V_2^*|) = nn_1^{-\alpha}$. Applying Chernoff's inequality ([7], page 27, formula (2.9)),

$$Prob\left(\exists v \in V_1 : \big||N(v) \cap V_2^*| - nn_1^{-\alpha}\big| > \frac{1}{k} nn_1^{-\alpha}\right) \le 2n_1 e^{-nn_1^{-\alpha}/3k^2}. \tag{16}$$

Define

$$\mathcal{F} = \{\pi = (F_1, \dots, F_t) \text{ where } F_i \subset [k], \quad i = 1, \dots, t\}.$$

Clearly, $|\mathcal{F}| = 2^{kt} = n_1^{k\ln 2/\ln k}$. For every $\pi \in \mathcal{F}$ and $\mathbf{x} = (x_1, \dots, x_s, \dots, x_t) \in V_i$, $i = 1,2$, define $S_{\mathbf{x}}^{\pi} = \{s : x_s \in [k] \setminus F_s\}$ and $V_i(\pi) = \{\mathbf{x} : |S_{\mathbf{x}}^{\pi}| < 2\epsilon kt\}$.

For each $\pi \in \mathcal{F}$, and $i = 1,2$, $|V_i(\pi) \cap V_i^*|$ is a random variable with hypergeometric distribution. If $|V_i(\pi)| < n_1^{1-\delta}$, then

$$\mathbb{E}(|V_i(\pi) \cap V_i^*|) = \frac{n}{n_1}|V_i(\pi)| < nn_1^{-\delta} \tag{17}$$

Therefore, by Chernoff's inequality ((([7], page 28, formula (2.10)),

$$Prob\left(\exists \pi \in \mathcal{F} \text{ and } \exists i \in \{1,2\} : |V_i(\pi)| < n_1^{1-\delta} \text{ but } |V_i(\pi) \cap V_i^*| > 2nn_1^{-\delta}\right) \le \frac{2n_1^{k\ln 2/\ln k}}{e^{c' nn_1^{-\delta}}}, \tag{18}$$

where $c' = \ln 2 - 1/2$.

Since $nn_1^{-\delta} \ge \max\{n^{\delta/2}, n^{\alpha/2}\}$ and δ, α do not depend on t, for sufficiently large t the right hand side of (16) and (18) are each smaller than $1/2$, yielding the existence of an induced subgraph $G_n = G(k,t)[V_1^n, V_2^n]$ of $G(k,t)$ with $|V_1^n| = |V_2^n| = n$, which satisfies (15) and such that

$$\forall \pi \in \mathcal{F}, i = 1,2 : |V_i(\pi)| \ge n_1^{1-\delta} \text{ or } |V_i(\pi) \cap V_i^n| \le 2nn_1^{-\delta}. \tag{19}$$

Now take any $U_1 \subset V_1^n, U_2 \subset V_2^n$ with no ϵ-hole between U_1 and U_2. These two sets determine, as in the proof of Lemma 3.1, two sequences π^1 and π^2 of sets of frequent values F_s^1 and F_s^2 such that $F_s^1 \cap F_s^2 = \emptyset$, $s = 1, \dots t$. Let $U_i' = |V_i(\pi^i) \cap V_i^n|$ be defined as in the proof of Lemma 3.1. Then, as it was shown in that proof, $|U_i| < 2|U_i'|$, and

$$\min\{|V_1(\pi^1)|, |V_2(\pi^2)|\} < n_1^{1-\delta}.$$

Hence, by (19),

$$\begin{aligned} |U_i| &< 2|U_i'| = 2|V_i(\pi^i) \cap V_i^n| \le 4nn_1^{-\delta} \\ &= 4n\left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt}\left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t} \end{aligned}$$

for $i = 1$ or $i = 2$, a contradiction. ■

Now we are ready to prove Theorem 1.5.

Proof of Theorem 1.5: Fix any $0 < \epsilon \leq \epsilon_0$, and let $k \geq 3$ be the integer such that

$$\frac{1}{25(k+1)} < \epsilon \leq \frac{1}{25k}. \tag{20}$$

Fix any $0 < d \leq d_0 \leq 1/8$, and let t be the integer such that

$$\frac{1}{2}\left(\frac{k-1}{k}\right)^{t+1} < d \leq \frac{1}{2}\left(\frac{k-1}{k}\right)^{t}. \tag{21}$$

Observe that $k \leq t$, since otherwise

$$\frac{1}{2}\left(\frac{k-1}{k}\right)^{t+1} \geq \frac{1}{2}\left(\frac{k-1}{k}\right)^{k} \geq \frac{1}{8} \geq d,$$

a contradiction.

Now recall that $n_0 = \max\{n_1^{3\delta/2}, n_1^{3\alpha/2}\}$ and consider two separate cases.

Case 1. $n \geq n_1 = k^t$

Take the blown-up graph $G_n(k,t)$. By (13), we have

$$d(G_n(k,t)) \geq \frac{r}{r+1}\left(\frac{k-1}{k}\right)^{t} \geq \frac{1}{2}\left(\frac{k-1}{k}\right)^{t} \geq d.$$

Thus, by Lemma 3.3

$$h(\epsilon, d, n) < 2n\left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt}\left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}. \tag{22}$$

Case 2. $n_0 \leq n < n_1$

Take the graph G_n satisfying Lemma 3.4. By (15), we have, again,

$$d(G_n) > \frac{k-1}{k} n_1^{-\alpha} \geq d.$$

Thus, by Lemma 3.3

$$h(\epsilon, d, n) < 4n\left(\frac{e}{2\epsilon k}\right)^{2\epsilon kt}\left(\frac{1+4\epsilon k}{\sqrt{2}}\right)^{2t}. \tag{23}$$

Combining these two cases, we conclude that (23) holds for every $n \geq n_0$. By reshaping the right hand side of (23), we arrive at

$$h(\epsilon, d, n) < 4n\left(\frac{1}{2}\left(\frac{k-1}{k}\right)^{t+1}\right)^{\phi} \tag{24}$$

$$< 4nd^{\phi}, \tag{25}$$

where

$$\phi = \frac{t\left(\ln 2 - 2\epsilon k \ln \frac{e}{2\epsilon k} - 2\ln\left(1+4\epsilon k\right)\right)}{\ln 2 + (t+1)\ln \frac{k}{k-1}}\ . \tag{26}$$

In what follows we will be relying on (20) and the well-known inequalities

$$x/2 \le \ln(1+x) \le x \tag{27}$$

valid for $0 \le x \le 1$. First notice that

$$\ln \frac{k}{k-1} \le \frac{1}{k-1} \le \frac{3}{k+1} < 75\epsilon \tag{28}$$

and

$$\ln 2 + (t+1)\ln \frac{k}{k-1} \le 1 + 75(t+1)\epsilon < 100(t+1)\epsilon. \tag{29}$$

Also

$$\ln 2 - 2\epsilon k \ln \frac{e}{2\epsilon k} - 2\ln\left(1+4\epsilon k\right) > \frac{1}{10} \tag{30}$$

when $\epsilon \le 1/25k$. Indeed, $q(x) = \ln 2 - x\ln\frac{e}{x} - 2\ln(1+2x)$ is decreasing when $x < 1$ and $q(2/25) > 1/10$.

Combining (25), (26), (29), (30) and the fact that $t/(t+1) \ge 1/2$, we have

$$h(\epsilon, d, n) < 4nd^{1/2000\epsilon}. \tag{31}$$

It remains to estimate $n_0 = \max\{n_1^{3\delta/2}, n_1^{3\alpha/2}\}$. Observe that $n_1^{\delta} = \left(\frac{2\epsilon k}{e}\right)^{2\epsilon k t}\left(\frac{\sqrt{2}}{1+4\epsilon k}\right)^{2t}$ and $n_1^{\alpha} = \left(\frac{k}{k-1}\right)^{t}$. We have

$$\begin{aligned} n_1^{3\delta/2} &= \left(\frac{2\epsilon k}{e}\right)^{3\epsilon k t}\left(\frac{\sqrt{2}}{1+4\epsilon k}\right)^{3t} && (32)\\ &= \left[2\left(\frac{k}{k-1}\right)^{t}\right]^{\eta} && (33)\\ &\le (1/d)^{\eta}\,, && (34) \end{aligned}$$

where

$$\eta = \frac{3t\left[\ln 2 - 2\epsilon k \ln\frac{e}{2\epsilon k} - 2\ln(1+4\epsilon k)\right]}{2t\ln\frac{k}{k-1} + 2\ln 2}.$$

Applying (27) and (20), we have

$$\eta < \frac{3\ln 2}{2\ln(k/k-1)} < 3k\ln 2 \le \frac{1}{12\epsilon}.$$

So, by (34), we obtain

$$n_1^{3\delta/2} < (1/d)^{1/(12\epsilon)}\,. \tag{35}$$

We also have

$$n_1^{3\alpha/2} = \left(\frac{k}{k-1}\right)^{3t/2} < \left[2\left(\frac{k}{k-1}\right)^t\right]^{3/2} \le (1/d)^{3/2}. \tag{36}$$

Comparing (35) and (36), it is easy to see that $n_1^{3\delta/2} \ge n_1^{3\alpha/2}$, since $1/(12\epsilon) \ge 3/2$ when $\epsilon \le 1/50$. Hence, $n_0 < (1/d)^{1/(12\epsilon)}$. ■

4 Applications

As an immediate application of our Theorem 1.3, we improve slightly an upper bound on the cycle partition number of an r-edge-colored $K_{n,n}$ discussed in [6]. The cycle partition number of an r-edge-colored graph G is defined to be the minimum k such that whenever the edges of G are colored with r colors, the vertices of G can be covered by at most k vertex-disjoint monochromatic cycles. Erdös, Gyárfás, and Pyber ([3]) proved that the cycle partition number of r-colored complete graphs K^n is $O(r^2 \ln r)$. They also raised the question whether the cycle partition number for the complete bipartite graph $K_{n,n}$ is independent of n. In [6], Haxell proved that the upper bound on the cycle partition number for an r-edge-colored $K_{n,n}$ is $O((r \ln r)^2)$ ([6]). Replacing Lemma 2 from [6] by our Theorem 1.3, this can be improved to $O(r^2 \ln r)$. We omit the details.

We conclude the paper with another application leading to what we believe is an interesting problem. Let $\mathcal{B}(m,\Delta)$ be the family of all bipartite graphs with m vertices in each color class and maximum degree at most Δ. We say that a graph G is (m,Δ)-universal if G contains a copy of H for every $H \in \mathcal{B}(m,\Delta)$. In [1] and [2] the problem of finding minimum $M = M(m)$ for which *there exists* an (m,Δ)-universal graph with M edges is investigated. Here we apply Theorems 1.3 and 1.4 to a related problem.

Given $\Delta \ge 1$, $0 < d < 1$ and n, let $g(\Delta, d, n)$ be the largest integer m such that *every* bipartite graph G with n vertices in each color class and at least dn^2 edges is (m,Δ)-universal.

Proposition 4.1 *For all $\Delta \ge \Delta_0$ and $d \le d_0$, there is n_0 such that for all $n \ge n_0$,*

$$\frac{1}{2} n d^{c_1 (d/2)^{-\Delta}} \le g(\Delta, d, n) \le 4 n d^{c_2 \Delta / \ln \Delta},$$

where c_1 and c_2 are absolute constants.

Proof: For the proof of the upper bound we need to find, for every $n \ge n_0$, a bipartite graph G with n vertices in each color class and $d(G) \ge d$, as well as a graph $H_0 \in \mathcal{B}(m,\Delta)$

such that $H_0 \not\subseteq G$. As G we will use the graph considered in the proof of Theorem 1.5 which is known to contain an ϵ-hole in every m by m subgraph, where $m = 4nd^{c/\epsilon}$.

With this approach, a natural candidate for H_0 is then a graph with no large holes. By considering a random bipartite graph with $2m$ vertices in each color class and with edge probability $\Delta/(4m)$, a standard application of the first moment method yields the existence of a graph $H_0 \in \mathcal{B}(m, \Delta)$ which contains no $9 \ln \Delta/\Delta$-hole. Setting $\epsilon = 9 \ln \Delta/\Delta$, this proves the upper bound with $c_2 = c/9$, where c is the constant appearing in Theorem 1.5.

For the lower bound, in addition to Theorem 1.3, we use the following embedding result.

Lemma 4.2 ([5], Lemma 2) *Every bipartite, $(\sigma^{\Delta}, \sigma)$-dense graph F with at least $\sigma^{-\Delta}m$ vertices in each color class is (m, Δ)-universal.* ■

Given Δ, d and n, set $\epsilon = (d/2)^{\Delta}$ and

$$m = \frac{1}{2}n(d/2)^{\Delta}d^{12/\epsilon} \geq \frac{1}{2}nd^{14/\epsilon}.$$

By Theorem 1.3, every bipartite graph G with n vertices in each color class and at least dn^2 edges contains an $(\epsilon, d/2)$-dense subgraph F with at least $\frac{1}{2}nd^{12/\epsilon} = (d/2)^{-\Delta}m$ vertices in each color class. By Lemma 4.2 with $\sigma = d/2$, F is (m, Δ)-universal and so is G. This proves the lower bound with $c_1 = 14$. ■

It seems to be a challenging problem to narrow the gap between the lower and upper bound in Proposition 4.1. We believe that the upper bound is closer to the true value of $g(\Delta, d, n)$. The proof of this fact, however, would require an essential strengthening of the current graph embedding methods.

It is interesting to note that the nonbipartite version of graph $G(k, t)$ which serves as a basis for constructing a counterexample in Theorem 1.5, and consequently in the right hand side of Proposition 4.1, was proved in [1] to be (k^t, Δ)-universal if only $k = k(\Delta)$ is sufficiently large.

Acknowledgment. We thank Andrzej Dudek and an anonymous referee for careful reading of the manuscript.

References

[1] N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński, E. Szemerédi, *Universality and tolerance*, In Proceedings of the 41st IEEE Annual Symposium on FOCS (2000), 14-21.

[2] N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński, E. Szemerédi, *Near-optimum universal graphs for graphs with bounded degrees*, APPROX-RANDOM 2001, LNCS 3139 (2001) 170-180

[3] P. Erdös, A. Gyárfás, and L. Pyber, *Vertex coverings by monochromatic cycles and trees*, J. Combin. Theory Ser. B 51 (1991), 90-95.

[4] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, GAFA, Geom. Funct. Anal. 7 (1997), 322-337.

[5] R. L. Graham, V. Rödl and A. Ruciński, *On bipartite graphs with linear Ramsey numbers*, Combinatorica, 21 (2001), 199-209.

[6] P. E. Haxell, *Partitioning complete bipartite graphs by monochromatic cycles*, Journal of Combinatorial Theory, Series B 69, (1997), 210-218.

[7] S. Janson, T. Łuczak, A. Ruciński, Random Graphs, John Wiley and Sons, New York, 2000.

[8] J. Komlós, M.Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics, Paul Erdős is Eighty (Volume 2), Keszthely (Hungary), 1993, Budapest (1996), 295-352.

[9] E. Szemerédi, *Regular partitions of graphs* in Problèmes combinatoires et théorie des graphes, Orsay 1976, J.-C. Bermond, J.-C. Fournier, M. Las Vergnas, D. Sotteau, eds., Colloq. Internat. CNRS 260, Paris, 1978, 399–401.

5 Appendix

At the end of Section 2, we sketched how to prove Theorem 1.1. Here we give all the details of that proof.

Proof of Claim 2.3. Assuming that H contains no pair satisfying conditions *1* or *2*, we will prove that H must contain a pair satisfying condition *3*.

Since, in particular, the pair (V_1^H, V_2^H) is not (ϵ, H)-regular, there exist $A_1' \subset V_1^H, B_1' \subset V_2^H$ with $|A_1'| = |B_1'| \geq \epsilon m$ satisfying either

$$d(A_1', B_1') > (1 + \epsilon/3)d(H), \tag{37}$$

or

$$d(A_1', B_1') < (1 - \epsilon/3)d(H). \tag{38}$$

If (37) holds, then we have a pair satisfying condition *2*. So (38) holds, and by an averaging argument, we can take $A_1 \subset A_1', B_1 \subset B_1'$ satisfying $|A_1| = |B_1| = \frac{\epsilon}{2}m$ and $d(A_1, B_1) \leq d(A_1', B_1') < (1 - \epsilon/3)d(H)$. Let F_1 be the graph obtained by removing A_1 from V_1^H and B_1 from V_2^H.

We apply the same argument to F_1, and in general, after l steps, $l < 1/\epsilon$, we define l disjoint pairs $(A_1, B_1), \cdots, (A_l, B_l)$ of size $|A_i| = |B_i| = \frac{\epsilon}{2}m$ such that $d(A_i, B_i) < (1 - \epsilon/3)d(H)$, $1 \leq i \leq l$. Assume that F_l is obtained by removing $\bigcup_{j=1}^{l} A_j$ from V_1^H

and $\bigcup_{j=1}^{l} B_j$ from V_2^H. By our assumption that H does not contain a pair satisfying conditions *1* or *2*, there exist $A'_{l+1} \subset V_1^H \setminus \bigcup_{j=1}^{l} A_j, B'_{l+1} \subset V_2^H \setminus \bigcup_{j=1}^{l} B_j$ of size $|A'_{l+1}| = |B'_{l+1}| \geq \epsilon\,(1 - l\epsilon/2)\,m \geq \frac{\epsilon}{2}m$ satisfying $d(A'_{l+1}, B'_{l+1}) < (1-\epsilon/3)d(H)$, and again we can find $A_{l+1} \subset A'_{l+1}$ such that $B_{l+1} \subset B'_{l+1}, |A_{l+1}| = |B_{l+1}| = \frac{\epsilon}{2}m$ and $d(A_{l+1}, B_{l+1}) < (1-\epsilon/3)d(H)$.

After $1/\epsilon$ steps the sets $\bigcup_{j=1}^{1/\epsilon} A_j$ cover a half of V_1^H, and the sets $\bigcup_{j=1}^{1/\epsilon} B_j$ cover a half of V_2^H. Set $\bar{V_1} = \bigcup_{j=1}^{1/\epsilon} A_j$, $\bar{V_2} = \bigcup_{j=1}^{1/\epsilon} B_j$, $e_0 = e\left(\bar{V_1}, \bar{V_2}\right), e_1 = e\left(\bar{V_1}, V_2^H \setminus \bar{V_2}\right), e_2 = e\left(V_1^H \setminus \bar{V_1}, \bar{V_2}\right)$, and $e_3 = e\left(V_1^H \setminus \bar{V_1}, V_2^H \setminus \bar{V_2}\right)$.

We claim that there exists a pair (U_1, U_2) satisfying condition *3*. Indeed, if

$$e_0 \leq \left(1 - \epsilon^2/4\right) d(H)m^2/4,$$

then

$$e_1 + e_2 + e_3 = d(H)m^2 - e_0 \geq 3\left(1 + \frac{\epsilon^2}{12}\right) d(H)\frac{m^2}{4},$$

and therefore, there exists $i \in \{1,2,3\}$ such that

$$e_i \geq \left(1 + \frac{\epsilon^2}{12}\right) d(H)\frac{m^2}{4}.$$

Let (U_1, U_2) be the pair defining e_i. It is easy to see that (U_1, U_2) satisfies condition *3*.

If $e_0 > (1 - \epsilon^2/4)\, d(H)m^2/4$, we define $e_{ij} = e\left(A_i, B_j\right)$. By the choice of pairs (A_i, B_i), we have $e_{ii} < (1 - \frac{\epsilon}{3})d(H)\left(\frac{\epsilon m}{2}\right)^2$ for every $i \leq 1/\epsilon$. Therefore

$$\begin{aligned}\sum_i \sum_{j \neq i} e_{ij} = e_0 - \sum_{i=1}^{1/\epsilon} e\left(A_i, B_i\right) &> \left(1 - \frac{\epsilon^2}{4}\right) d(H)\frac{m^2}{4} - \frac{1}{\epsilon}\left(1 - \frac{\epsilon}{3}\right) d(H)\left(\frac{\epsilon m}{2}\right)^2 \\ &> \left(1 - \epsilon + \frac{\epsilon^2}{12}\right) d(H)\frac{m^2}{4}.\end{aligned}$$

For any $I \subset \{1, \ldots, 1/\epsilon\}, |I| = 1/(2\epsilon)$, we define

$$e\left(I\right) = \sum_{i \in I} \sum_{j \in \{1,\ldots,1/\epsilon\} \setminus I} e_{ij}.$$

Then $\sum_I e\left(I\right)$ counts each e_{ij} $\binom{1/\epsilon - 2}{1/(2\epsilon) - 1}$ times, where $i \neq j$. Therefore, there exists I such that

$$e\left(I\right) \geq \frac{\sum_I e\left(I\right)}{\binom{1/\epsilon}{1/(2\epsilon)}} = \frac{\binom{1/\epsilon-2}{1/(2\epsilon)-1}}{\binom{1/\epsilon}{1/(2\epsilon)}} \sum_i \sum_{j \neq i} e_{ij} > \frac{\left(1 - \epsilon + \frac{\epsilon^2}{12}\right) dm^2/4}{4\left(1 - \epsilon\right)} \geq \left(1 + \frac{\epsilon^2}{12}\right) d(H)\frac{m^2}{16}.$$

Set $U_1 = \bigcup_{i \in I} A_i, U_2 = \bigcup_{j \notin I} B_j$. Then (U_1, U_2) is a pair satisfying condition *3*. ■

Proof of Theorem 1.1. Let $G = (V_1, V_2, E)$ be a bipartite graph with $|V_1| = |V_2| = n$ and density d. If G contains a pair (U_1, U_2) satisfying condition *1* in Claim 2.3, then, due to Fact 2.2, (U_1, U_2) is an ϵ-regular pair with $|U_1| = |U_2| \geq n/2$.

Assuming that G contains no pair satisfying condition *1* in Claim 2.3, and applying Claim 2.3 to G, we can find either a subgraph $G_{(1,0)} \subset G$ with at least $\epsilon n/2$ vertices in each color class and density at least $(1 + \epsilon/3)d$, or a subgraph $G_{(0,1)} \subset G$ with at least $n/4$ vertices in each color class and density at least $(1 + \epsilon^2/12)\, d$.

Without loss of generality, we may assume that the former is true. If $G_{(1,0)}$ contains a pair satisfying condition *1* in Claim 2.3, then this pair is ϵ-regular. So, again, assuming that $G_{(1,0)}$ contains no pair satisfying condition *1* in Claim 2.3, and applying Claim 2.3 to $G_{(1,0)}$, we can find either a subgraph $G_{(2,0)}$ of $G_{(1,0)}$ with at least $n(\epsilon/2)^2$ vertices in each color class and density at least $(1 + \epsilon/3)^2 d$, or a subgraph $G_{(1,1)}$ of $G_{(1,0)}$ with at least $n\epsilon/8$ vertices in each color class and density at least $(1 + \epsilon/3)(1 + \epsilon^2/12)d$.

Suppose we have iterated this process (s_1, s_2) times, where s_1 is the number of times of obtaining pairs satisfying condition *2* in Claim 2.3, and s_2 is the number of times obtaining pairs satisfying condition *3* in Claim 2.3. Then we obtain a subgraph $G_{(s_1,s_2)}$ of G with at least $n(\epsilon/2)^{s_1}(1/4)^{s_2}$ vertices in each color class and density at least $(1 + \epsilon/4)^{s_1}(1 + \epsilon^2/12)^{s_2} d$. Because the density of no graph is larger than 1, this process has to stop in finite times. Let (t_1, t_2) be the number of times we iterate before the process stops, then $(1 + \epsilon/3)^{t_1}(1 + \epsilon^2/12)^{t_2} d \leq 1$.

At this point, we obtain an ϵ-regular pair with at least $\frac{n}{2}(\epsilon/2)^{t_1}(1/4)^{t_2}$ vertices in each color class. It remains to estimate t_1 and t_2 from above. By the choice of t_1, t_2, we have $(1 + \epsilon/3)^{t_1} d \leq 1$, and $(1 + \epsilon^2/12)^{t_2} d \leq 1$, thus

$$t_1 \leq \frac{\ln(1/d)}{\ln(1 + \epsilon/3)}$$

and

$$t_2 \leq \frac{\ln(1/d)}{\ln(1 + \epsilon^2/12)}.$$

Hence,

$$\frac{n}{2}(\epsilon/2)^{t_1}(1/4)^{t_2} \geq \frac{n}{2} d^{\phi}. \tag{39}$$

where

$$\phi = \frac{\ln(2/\epsilon)}{\ln(1 + \epsilon/3)} + \frac{\ln 4}{\ln(1 + \epsilon^2/12)}.$$

Notice that $\ln(1 + x) \geq x/2$ holds for $0 \leq x \leq 1$. Therefore

$$\phi \leq \frac{6\ln(2/\epsilon)}{\epsilon} + \frac{48 \ln 2}{\epsilon^2} \leq \frac{c}{\epsilon^2}, \tag{40}$$

where c is an absolute constant. Applying (40) to (39), we have

$$\frac{n}{2}(\epsilon/2)^{t_1}(1/4)^{t_2} \geq \frac{n}{2} d^{c/\epsilon^2},$$

and consequently we have proved the existence of an ϵ-regular pair in G with at least $\frac{1}{2} n d^{c/\epsilon^2}$ vertices in each color class. This completes the proof of Theorem 1.1. ■