ASYMMETRIC RESULTS ABOUT GRAPH HOMOMORPHISMS

LIOR GISHBOLINER, EOIN HURLEY, AND YUVAL WIGDERSON

ABSTRACT. Many important results in extremal graph theory can be roughly summarised as "if a triangle-free graph G has certain properties, then it has a homomorphism to a triangle-free graph Γ of bounded size". For example, bounds on homomorphism thresholds give such a statement if G has sufficiently high minimum degree, and the approximate homomorphism theorem gives such a statement for all G, if one weakens the notion of homomorphism appropriately.

In this paper, we study asymmetric versions of these results, where the assumptions on G and Γ need not match. For example, we prove that if G is a graph with odd girth at least 9 and minimum degree at least $\delta|G|$, then G is homomorphic to a triangle-free graph whose size depends only on δ. Moreover, the odd girth assumption can be weakened to odd girth at least 7 if G has bounded VC dimension or bounded domination number. This gives a new and improved proof of a result of Huang et al.

We also prove that in the asymmetric approximate homomorphism theorem, the bounds exhibit a rather surprising "double phase transition": the bounds are super-exponential if G is only assumed to be triangle-free, they become exponential if G is assumed to have odd girth 7 or 9, and become linear if G has odd girth at least 11.

Our proofs use a wide variety of techniques, including entropy arguments, the Frieze–Kannan weak regularity lemma, properties of the generalised Mycielskian construction, and recent work on abundance and the asymmetric removal lemma.

1. INTRODUCTION

1.1. **Background.** In this paper, we are concerned with questions of the following type: to what extent can the triangle-freeness of a large graph be "explained" by a small triangle-free graph? To make this vague question more formal, let us recall that a graph G is *homomorphic* to a graph Γ if there exists a function $V(G) \to V(\Gamma)$ which maps edges to edges. Equivalently, G is homomorphic to Γ if G is a subgraph of a blowup[1] of Γ. We write $G \to \Gamma$ if G is homomorphic to Γ.

Note that if $G \to \Gamma$ and Γ is triangle-free, then certainly G is triangle-free as well. If we think of G as large and of Γ as small (i.e. of constant size), then the existence of a homomorphism

[1]A *blowup* of Γ is obtained from Γ by replacing every vertex by an independent set, and replacing every edge by a complete bipartite graph.

$G \to \Gamma$ gives a constant-sized "explanation" of the triangle-freeness of G. That is, G is a subgraph of a blowup of the constant-sized Γ, and this structure already guarantees that G is triangle-free.

There are many results which state that under certain conditions, a large triangle-free graph G is homomorphic to a constant-sized triangle-free Γ. Perhaps the earliest such result is due to Andrásfai [3] (see also [4]), who proved that if G is a triangle-free graph with minimum degree greater than $\frac{2}{5}|G|$, then G is bipartite, i.e. homomorphic to K_2. The constant $\frac{2}{5}$ is tight in this result, as shown by the five-cycle C_5 as well as all of its balanced blowups. Extending Andrásfai's theorem, Häggkvist [19] proved that if G is a triangle-free graph with minimum degree greater than $\frac{3}{8}|G|$, then G is homomorphic to C_5. The constant $\frac{3}{8}$ is again best possible, as shown by by the 8-vertex Möbius ladder M_8, obtained by adding the four long diagonals to C_8, as well as all of its balanced blowups.

That is, above minimum degree $\frac{2}{5}$, there is only one triangle-free structure, given by the graph K_2. At the threshold $\frac{2}{5}$, another structure appears, namely that of C_5. These are the only two structures above the threshold $\frac{3}{8}$, at which point a new structure M_8 appears. Such a pattern continues for a while, and Jin [24] determined the next seven thresholds as well as the structures that appear in them.

However, as proved by Hajnal (quoted in [9]), this pattern cannot continue forever. Indeed, Hajnal constructed a family of triangle-free graphs G with minimum degree $(\frac{1}{3} - o(1))|G|$ whose chromatic number tends to infinity. In particular, this implies that these graphs G cannot be homomorphic to any constant-sized triangle-free graph. Rather remarkably, Łuczak [27] (extending work of Thomassen [36]) proved that the constant $\frac{1}{3}$ is best possible, in the following sense.

Theorem 1.1 (Łuczak [27])**.** *For every $\delta > 0$, there exists a triangle-free graph Γ such that every triangle-free graph G with minimum degree at least $(\frac{1}{3} + \delta)|G|$ is homomorphic to Γ.*

More recently, Brandt and Thomassé [5] gave a complete description of the family of all such graphs Γ; they are the so-called *Vega graphs*.

While the results of Andrásfai, Häggkvist, Jin, Łuczak, Brandt–Thomassé and many others are beautiful, they are too rigid to be useful in certain applications. We would often like to assume less about the graph G, and are willing to obtain less precise structural information. For example, a less restrictive notion than that discussed above is that of *approximate homomorphisms*.

Given two graphs G, Γ, a function $V(G) \to V(\Gamma)$ is said to be an *ε-approximate homomorphism* if it maps all but at most $\varepsilon|G|^2$ edges of G to edges of Γ. Equivalently, G has an ε-approximate homomorphism into Γ if it is ε-close[2] to a graph homomorphic to Γ.

The remarkable *approximate homomorphism theorem*, perhaps first explicitly observed by Tao [35], states that every triangle-free graph G has an ε-approximate homomorphism to a

---

[2]Two n-vertex graphs G_1, G_2 on the same vertex set are *ε-close* if one can be obtained from the other by adding/deleting at most εn^2 edges, and they are *ε-far* otherwise.

triangle-free graph Γ, whose size depends only on ε. In other words, up to some small noise, every large triangle-free graph G can be obtained from a *constant-sized* triangle-free graph Γ by blowing it up and then passing to a subgraph, where the constant size of Γ depends only on the amount of noise, and not on the size of G.

Moreover, the approximate homomorphism theorem holds much more generally, dealing not just with triangle-free graphs. To state it in full generality, let us say that a graph Γ is *F-hom-free* if there is no homomorphism from F to Γ. Equivalently, Γ is F-hom-free if every blowup of Γ is F-free. Note that if F is a triangle (or more generally a clique), then the properties of F-hom-freeness and F-freeness coincide. Clearly, if G is homomorphic to an F-hom-free graph Γ, then G is certainly F-hom-free as well. The general statement of the approximate homomorphism theorem gives a rough converse to this simple observation.

Theorem 1.2 (Approximate homomorphism theorem). *For every graph F and every $\varepsilon > 0$, there exists a constant $M_F(\varepsilon) > 0$ such that the following holds. If a graph G is F-hom-free, then it has an ε-approximate homomorphism to an F-hom-free graph Γ with $|\Gamma| \leqslant M_F(\varepsilon)$.*

We remark that we work with the property of F-hom-freeness, rather than the weaker F-freeness, as it is the more natural notion when working with approximate homomorphisms. But this is mostly for convenience, and it is not hard to show that Theorem 1.2 is equivalent to an analogous statement about F-free graphs (see e.g. [12]).

As we shortly discuss in greater detail, the approximate homomorphism theorem is very closely related to the famous *graph removal lemma* [2, 14, 31]. Indeed, if one examines the standard proof of the graph removal lemma using Szemerédi's regularity lemma, one immediately sees that it naturally constructs an approximate homomorphism to an F-hom-free graph of constant size. We remark that the connections between the graph removal lemma and minimum degree conditions such as those discussed above have recently been studied in [11, 17].

Both Theorems 1.1 and 1.2 are *symmetric* statements, in the sense that the structural property of G and Γ are the same: they are both triangle-free in Theorem 1.1, and they are both F-hom-free in Theorem 1.2. In this paper, we study *asymmetric* versions of these questions, where we impose different conditions on these two graphs. Quite surprisingly, this study is very subtle, and reveals a great deal of unexpected variation. We now turn to discuss these asymmetric statements, and our main results, in more detail.

1.2. **Asymmetric homomorphism thresholds.** Given a family $\mathcal{F}$ of graphs, a graph G is said to be $\mathcal{F}$-*free* if G does not contain any $F \in \mathcal{F}$ as a subgraph. The *homomorphism threshold* of $\mathcal{F}$, denoted $\delta_{\mathrm{hom}}(\mathcal{F})$, is defined as the infimum of all $\delta \in [0, 1]$ with the property that every $\mathcal{F}$-free graph G with minimum degree at least $\delta|G|$ is homomorphic to some $\mathcal{F}$-free graph Γ whose size depends only on δ. In case $\mathcal{F} = \{F\}$ consists of a single graph, we write $\delta_{\mathrm{hom}}(F)$ instead of $\delta_{\mathrm{hom}}(\{F\})$. In this language, Theorem 1.1 states that $\delta_{\mathrm{hom}}(C_3) \leqslant \frac{1}{3}$, and the construction of Hajnal discussed above yields a matching lower bound $\delta_{\mathrm{hom}}(C_3) \geqslant \frac{1}{3}$.

Recently, there has been a great deal of interest in determining the homomorphism thresholds of (families of) odd cycles. Let us write $\mathcal{C}_{2t+1}$ for the family $\{C_3, C_5, \ldots, C_{2t+1}\}$ of all odd cycles of length at most $2t + 1$. Letzter and Snyder [25] proved that $\delta_{\mathrm{hom}}(\mathcal{C}_5) = \frac{1}{5}$, and Ebsen and Schacht [8] extended this and showed that $\delta_{\mathrm{hom}}(\mathcal{C}_{2t+1}) = \frac{1}{2t+1}$ for all t. More recently, answering a question of Ebsen and Schacht, Sankar [32] proved that $\delta_{\mathrm{hom}}(C_{2t+1}) > 0$ for all t. This result is notable because it is the first lower bound on homomorphism thresholds that really uses the $\mathcal{F}$-freeness of Γ, as opposed to relying on chromatic number lower bounds. Indeed, Thomassen [37] proved that if $t \geqslant 2$, then every C_{2t+1}-free graph G with minimum degree at least $\delta|G|$ has constant chromatic number (depending only on δ and t).

Our first main result concerns an asymmetric variant of homomorphism thresholds.

Theorem 1.3. *Let $t \geqslant 1$ and $\delta > 0$, and let G be a $\mathcal{C}_{2t+5}$-free graph with minimum degree at least $\delta|G|$. Then G has a homomorphism to a $\mathcal{C}_{2t+1}$-free graph Γ with $|\Gamma| \leqslant (t+1)^{2/\delta}$.*

For example, setting $t = 1$, this result states that a graph with linear minimum degree and odd girth at least 9 is homomorphic to a triangle-free graph of constant size.

Let us define the *asymmetric homomorphsim threshold* $\delta_{\mathrm{hom}}(\mathcal{F}_1; \mathcal{F}_2)$ to be the infimum of all δ such that every $\mathcal{F}_1$-free graph G with minimum degree at least $\delta|G|$ is homomorphic to a constant-sized $\mathcal{F}_2$-free graph Γ. Note that if $\mathcal{F}_1 = \mathcal{F}_2$, this precisely recovers the earlier definition of the homomorphism threshold. In this language, Theorem 1.3 implies that $\delta_{\mathrm{hom}}(\mathcal{C}_{2t+5}; \mathcal{C}_{2t+1}) = 0$.

In fact, we believe that the inclusion of C_{2t+5} in Theorem 1.3 is unnecessary, and leave the following as a tantalizing open problem.

Conjecture 1.4. *If G is a $\mathcal{C}_{2t+3}$-free graph with minimum degree at least $\delta|G|$, then G is homomorphic to a $\mathcal{C}_{2t+1}$-free graph of constant size (depending only on δ). In other words, $\delta_{\mathrm{hom}}(\mathcal{C}_{2t+3}; \mathcal{C}_{2t+1}) = 0$.*

Note that, if true, Conjecture 1.4 would be best possible, as Ebsen and Schacht [8] proved that $\delta_{\mathrm{hom}}(\mathcal{C}_{2t+1}; \mathcal{C}_{2t+1}) = \frac{1}{2t+1} > 0$. As partial evidence towards Conjecture 1.4, we prove the following result, which gives the same conclusion if we replace the minimum degree assumption by the assumption of bounded domination number.

Theorem 1.5. *Let G be a $\mathcal{C}_{2t+3}$-free graph with domination number $\gamma(G)$. Then G has a homomorphism to a $\mathcal{C}_{2t+1}$-free graph Γ with $|\Gamma| \leqslant 3 \cdot (t+1)^{\gamma(G)-1} - 1$.*

Very recently, similar results to Theorems 1.3 and 1.5 were proved by Huang, Liu, Rong and Xu [23]. In particular, one of their results [23, Theorem 1.2] establishes Conjecture 1.4 under the added assumption that G has bounded VC dimension. Here, we recall that a set $S \subseteq V(G)$ is *shattered* if, for all $T \subseteq S$, there is a vertex adjacent to all vertices in T, but not adjacent to any vertex in $S \setminus T$, and that the *VC dimension* of G is defined as the maximum size of a shattered subset. VC dimension is a natural and widely studied notion of "bounded complexity" for a graph, and the result of Huang, Liu, Rong and Xu demonstrates that this assumption

allows one to prove Conjecture 1.4. As it turns out, if a graph has bounded VC dimension and linear minimum degree, then it has bounded domination number (see Lemma 2.6), hence Theorem 1.5 gives a short alternative proof of [23, Theorem 1.2].

Corollary 1.6. *Let G be a $\mathcal{C}_{2t+3}$-free graph with minimum degree at least $\delta|G|$ and VC dimension at most d. Then G is homomorphic to a $\mathcal{C}_{2t+1}$-free graph Γ with $|\Gamma| \leqslant 3 \cdot (t+1)^{\frac{8d}{\delta} \log \frac{8d}{\delta}}$.*

In addition to being substantially shorter, our proof of Corollary 1.6 yields much stronger quantitative bounds than that in [23]. Indeed, the proof in [23] produces such a Γ whose size is a tower of twos of height $t+1$, and whose top-most hyperexponent is $\delta^{-O(d)}$.

1.3. **Asymmetric approximate homomorphisms.** We now turn to our study of asymmetric versions of Theorem 1.2. Here, the most fundamental question does not have to do with minimum degree restrictions, but rather with the quantitative aspects of Theorem 1.2: how large is $M_F(\varepsilon)$ as a function of ε?

As mentioned previously, the standard proof of Theorem 1.2 follows the proof of the graph removal lemma, and these two results are in fact closely related. As such, we recall the statement of the graph removal lemma before proceeding.

Theorem 1.7 (Graph removal lemma [2, 14, 31]). *For every graph F and every $\varepsilon > 0$, there exists some $\delta > 0$ such that the following holds. If a graph G is ε-far from F-free, then G contains at least $\delta|G|^{|F|}$ copies of F.*

Despite its simple statement, this is a very deep result. All known proofs are rather complex, involving tools and ideas related to Szemerédi's regularity lemma [34]. This complexity is related to our best known bounds on δ as a function of ε: the best known proofs of the graph removal lemma [6, 10, 28] yield an upper bound on $1/\delta$ which is of tower type, with height $O_F(\log \frac{1}{\varepsilon})$. In the other direction, it is known [1, 31] that $1/\delta$ is at least super-polynomial in $1/\varepsilon$, namely $\frac{1}{\delta} \geqslant (\frac{1}{\varepsilon})^{\Omega_F(\log \frac{1}{\varepsilon})}$, whenever F is non-bipartite. These upper and lower bounds are extremely far apart, and it remains a major open problem to narrow the gap.

The standard proof of Theorem 1.7 using Szemerédi's regularity lemma also immediately yields Theorem 1.2. Unfortunately, this proof only supplies tower-type bounds on $M_F(\varepsilon)$ as a function of ε. In fact, as proved by Hoppen–Kohayakawa–Lang–Lefmann–Stagni [22] (for the upper bound) and by Fox–Zhao [12] (for the lower bound), the bounds in these two theorems are closely related. The formal results are somewhat technical, but roughly speaking, they imply that the best constant $M_F(\varepsilon)$ in Theorem 1.2 is exponential in $1/\delta_F(\varepsilon)$, where $\delta_F(\varepsilon)$ is the best constant in Theorem 1.7. In particular, the Fox–Zhao result [12] implies that if F is not bipartite, then

$$M_F(\varepsilon) \geqslant 2^{(1/\varepsilon)^{\Omega_F(\log(1/\varepsilon))}}, \tag{1}$$

that is, that $M_F(\varepsilon)$ is larger than any function of the form[3] $2^{1/\text{poly}(\varepsilon)}$.

---

[3]We use the notation $\text{poly}(\varepsilon)$ to denote any function f of the form $f(\varepsilon) = c\varepsilon^C$ for some absolute constants $c, C > 0$.

Here we are interested in an asymmetric version of this problem. Namely, we study the following function:

Definition 1.8. *For graphs F, H with $H \to F$ and for $\varepsilon > 0$, let $M_{F,H}(\varepsilon)$ be the smallest M such that every H-hom-free graph has an ε-approximate homomorphism to an F-hom-free graph on at most M vertices.*

The case $H = F$ corresponds to Theorem 1.2, and the existence of $M_{F,H}(\varepsilon)$ in general is an immediate consequence of Theorem 1.2.

Our results on this topic imply that the function $M_{F,H}(\varepsilon)$ can have a wide variety of interesting behaviours, even for very simple graphs F, H such as odd cycles. In particular, the following is a consequence of results that we shortly discuss in greater detail.

Theorem 1.9. *Let $\ell \geqslant 3$ be odd. We have that*

$$M_{K_3, C_\ell}(\varepsilon) = \begin{cases} superexponential & if\ \ell = 3, \\ 2^{\mathrm{poly}(1/\varepsilon)} & if\ \ell \in \{5, 7\}, \\ O(\frac{1}{\varepsilon}) & if\ \ell \geqslant 9. \end{cases}$$

The case $\ell = 3$ is simply the symmetric case where, as discussed in (1), superexponential bounds were proved by Fox–Zhao [12]. However, for $\ell \geqslant 5$, these results are new. In particular, we prove both upper and lower bounds of the form $2^{\mathrm{poly}(1/\varepsilon)}$ when $\ell \in \{5, 7\}$, and prove a linear upper bound when $\ell \geqslant 9$. It is not hard to prove, as we do in Proposition 5.3, that $M_{F,H}(\varepsilon) \geqslant \Omega((\frac{1}{\varepsilon})^{\frac{1}{2}})$ for all F, H with H non-bipartite[4], hence the linear upper bound for $\ell \geqslant 9$ is also close to best possible. Each of these bounds is actually a special case of a more general result, as we now discuss.

First, we remark that the upper bound on $M_{K_3, C_5}(\varepsilon)$ actually follows from Theorem 1.5. Indeed, it is not hard to show (as we do in Lemma 3.2) that from any graph G, we may delete at most $\varepsilon |G|^2$ edges to obtain a graph G' with domination number at most $3/\varepsilon$. In particular, this yields an ε-approximate homomorphism from G to a subgraph which has domination number at most $3/\varepsilon$. Since G' is also C_5-hom-free if G is, we may apply Theorem 1.5 and obtain that $M_{K_3, C_5}(\varepsilon) \leqslant 2^{O(1/\varepsilon)}$. The same argument shows that $M_{K_3, C_7}(\varepsilon) \leqslant 2^{O(1/\varepsilon)}$.

However, this argument is very special to the case of odd cycles, since our proof of Theorem 1.5 is restricted to this case. Because of this, we now discuss an alternative proof of the exponential upper bounds on $M_{K_3, C_\ell}(\varepsilon)$, which is far more general. The basic observation is that $M_{F,H}(\varepsilon)$ can be upper-bounded in terms of the parameter-dependence in the asymmetric (F, H) removal lemma, which we now introduce.

It is well-known and easy to see that the graph removal lemma implies an asymmetric version of the same result. It states that for any pair of graphs (F, H) with $H \to F$, if a graph G is ε-far from F-free, then it contains at least $\delta |G|^{|H|}$ copies of H, for some $\delta > 0$ depending only

---

[4]If H is bipartite then an H-hom-free graph has no edges, so trivially $M_{F,H}(\varepsilon) = 1$.

on F, H, and ε. The deduction of the asymmetric statement from Theorem 1.7 naturally incurs the same tower-type bounds that are the best known in Theorem 1.7. However, a recent line of work [7, 16, 18, 29] has demonstrated that in certain cases, the asymmetric removal lemma has *much* better bounds, namely in some cases polynomial bounds. Following [18], we say that H *is F-abundant* if there are polynomial bounds in the asymmetric (F, H) removal lemma, that is, if every graph G with is ε-far from F-free contains at least $\mathrm{poly}(\varepsilon)|G|^{|H|}$ copies of H.

With these definitions in place, we may state our bound connecting $M_{F,H}(\varepsilon)$ to the bounds in the asymmetric (F, H) removal lemma. This generalises a theorem of Hoppen–Kohayakawa–Lang–Lefmann–Stagni [22, Theorem 1.4], and follows from the same proof technique.

Proposition 1.10. *Let F, H be graphs and $\delta : (0, 1) \to (0, 1)$ be a function with the property that every graph G which is ε-far from F-hom-free contains at least $\delta(\varepsilon) \cdot |G|^{|H|}$ copies of H. Then $M_{F,H}(\varepsilon) \leqslant 2^{K^2}$, where*

$$K := \frac{5e(H)(2/\varepsilon)^{e(H)}}{\delta(\varepsilon/2)}. \tag{2}$$

If H is K_k-abundant, then the assumption of Proposition 1.10 holds with $\delta(\varepsilon) = \mathrm{poly}(\varepsilon)$, since being K_k-hom-free is the same as being K_k-free. Therefore, we obtain the following corollary of Proposition 1.10.

Corollary 1.11. *If H is K_k-abundant then $M_{K_k,H}(\varepsilon) \leqslant 2^{\mathrm{poly}(1/\varepsilon)}$.*

Gishboliner, Shapira, and Wigderson [18] proved that C_ℓ is K_3-abundant for all odd $\ell \geqslant 5$, hence Proposition 1.10 implies that $M_{K_3,C_\ell}(\varepsilon) \leqslant 2^{\mathrm{poly}(1/\varepsilon)}$, as claimed in Theorem 1.9. Note that this argument gives a weaker bound than the one using Theorem 1.5 (although both are of the form $2^{\mathrm{poly}(1/\varepsilon)}$), but this argument is much more general. For example, Girão, Hurley, Illingworth, and Michel [16] proved that the Petersen graph P is K_3-abundant, hence Proposition 1.10 also implies that $M_{K_3,P}(\varepsilon) \leqslant 2^{\mathrm{poly}(1/\varepsilon)}$.

Both Theorem 1.5 and Proposition 1.10 only yield exponential upper bounds on $M_{F,H}(\varepsilon)$ (and only in certain cases). However, in some cases, such as for $M_{K_3,C_\ell}(\varepsilon)$ with $\ell \geqslant 9$, we can obtain much stronger upper bounds. To state these results in full generality, we recall the following definition. Given a graph F, its *2-subdivision* $F^{\bullet\bullet}$ is obtained from F by adding two new vertices on every edge. Note that $F^{\bullet\bullet}$ is homomorphic to F for any graph F.

Theorem 1.12. *Let F, H be graphs with $H \to F^{\bullet\bullet}$. Then $M_{F,H}(\varepsilon) = O(\frac{1}{\varepsilon})$.*

Note that $K_3^{\bullet\bullet} = C_9$, and every odd cycle C_ℓ with $\ell \geqslant 9$ is homomorphic to C_9. Therefore, Theorem 1.12 in particular completes the upper bounds stated in Theorem 1.9. We remark that the proofs of both Theorem 1.12 and Proposition 1.10 yield efficient algorithms for constructing the F-hom-free graph Γ as well as the approximate homomorphism $G \to \Gamma$.

Therefore, all that remains to discuss is the exponential lower bound in Theorem 1.9. Namely, we need to show that for many graphs H, including some F-abundant ones, the exponential

dependence in Proposition 1.10 is necessary. For example, we will show that this is the case for $F = K_3$ and $H \in \{C_5, C_7\}$, completing the picture in Theorem 1.9. This lower bound comes from a variant of the aforementioned construction of Fox and Zhao [12]. In fact, we conjecture that this construction characterises the cases when $M_{F,H}(\varepsilon)$ is polynomial (and that in all other cases, this construction witnesses that $M_{F,H}(\varepsilon)$ is at least of the form $2^{\mathrm{poly}(1/\varepsilon)}$). To state this conjecture precisely, we now define the construction.

Construction 1.13. *Let F be a graph and let Δ be its maximum degree. For each vertex $v \in V(F)$, fix an ordering of the edges incident to v and label them $\{(v, 1), \ldots, (v, \deg(v))\}$.[5] When considering a copy of F in a graph G, we will (implicitly) fix an isomorphism from F to the copy, so that the copy is endowed with the above edge-labelling.*

Let G be an n-vertex graph where every edge is contained in a unique copy of F. Let $F_1, \ldots, F_m$ be the copies of F in G. The graph $G^\star$ is constructed as follows. The vertex set[6] of $G^\star$ is $V(G) \times \{1, \ldots, \Delta\}^m$. For every edge $e = uv \in E(G)$, let $k \in [m]$ be the unique index with $e \in E(F_k)$, and suppose that e has labels (u, i) and (v, j) (as an edge of F_k). Join all vertices $(u, \mathbf{x})$ and $(v, \mathbf{y})$ such that $\mathbf{x}_k = i$ and $\mathbf{y}_k = j$. These are the only edges in $G^\star$.

The following figure shows two examples of the construction $G^\star$ for $F = K_3$: first, when G is also a triangle, and second, when G is a "bowtie", consisting of two triangles sharing a vertex.

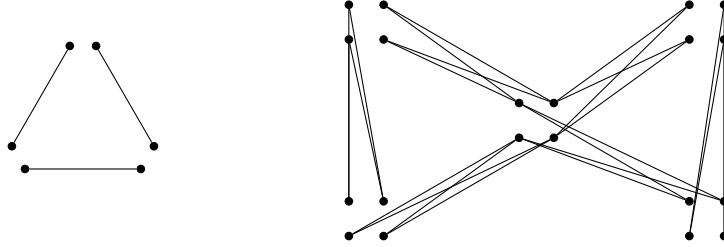
FIGURE 1. Illustration of Construction 1.13

Observe that each F-copy in G becomes a disjoint union of complete bipartite graphs in $G^\star$. Namely, for each vertex v in a copy F_k of F, the blowup set $\{v\} \times \{1, \ldots, \Delta\}^m$ is partitioned into Δ sets (according to the kth coordinate of the vector), and each of these sets participates in (at most) one complete bipartite graph corresponding to an edge of F (for a total of $\deg(v)$ such complete bipartite graphs). Moreover, these partitions for different F-copies are orthogonal, meaning that their Venn diagram has all possible regions. The main difference between Construction 1.13 and the construction of Fox and Zhao [12] is that in [12], each blowup set is partitioned into only two parts per F-copy, so the bipartite graphs arising from a single F-copy can overlap. We would like to avoid this situation in order to further restrict the types of subgraphs that can appear in the construction.

---

[5]Thus each edge receives two labels (one from each endpoint), and knowing one of them is enough to identify the edge.

[6]We denote a vertex of $G^\star$ as $(v, \mathbf{x})$, where $v \in V(G)$ and $\mathbf{x} \in \{1, \ldots, \Delta\}^m$.

By adapting the proof in [12], we will show the following.

Lemma 1.14. *For every graph F, there exists $c > 0$ such that the following holds. Let G be an n-vertex graph where every edge is contained in a unique copy of F, and let m denote the number of copies of F in G. Let $G^\star$ be given by Construction 1.13. Then for $\varepsilon < \frac{cm}{n^2}$, there is no ε-approximate homomorphism from $G^\star$ to an F-hom-free graph on at most $2^{cm/n}$ vertices.*

A graph is an *F-forest* if it can be obtained from an empty graph by repeatedly adding a copy of F with at least $|V(F)| - 1$ new vertices (hence the new copy shares at most one vertex with the current graph). The following figure shows an example of a K_3-forest.
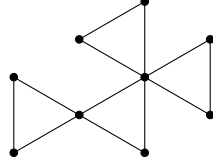

FIGURE 2. An example of a K_3-forest

Note that if F is 2-connected then every F-forest T satisfies the requirements of Construction 1.13, namely, every edge of T is contained in a unique copy of F. Recall that $T^\star$ denotes the graph given by that construction. Using Lemma 1.14, we will prove the following.

Theorem 1.15. *Let F, H be graphs, where F is 2-connected, and suppose that H is not homomorphic to $T^\star$ for any F-forest T with $T \to F$. Then for every small enough $\varepsilon > 0$, $M_{F,H}(\varepsilon) \geqslant 2^{(1/\varepsilon)^c}$, where $c > 0$ depends only on F, H.*

We remark that if F is vertex-transitive, then all F-forests are homomorphic to F. On the other hand, there are examples of non-vertex-transitive[7] F and F-forests T with $T \not\to F$.

As a corollary of Theorem 1.15 we obtain the following, implying the new lower bound in Theorem 1.9.

Corollary 1.16. $M_{K_3, C_\ell}(\varepsilon) \geqslant 2^{(1/\varepsilon)^c}$ *for $\ell \in \{5, 7\}$.*

We remark that if $H = C_\ell$ for $\ell \geqslant 9$, then in fact $H \to T^\star$ for some K_3-forest T. More generally, $F^{\bullet\bullet}$ is always homomorphic to $T^\star$ for some F-forest T[8], and thus Theorem 1.15 is consistent

---

[7]Indeed, if there exist $u, v \in V(F)$ such that there exist no homomorphisms $\phi, \psi : F \to F$ with $\phi(u) = \psi(v)$, then one can construct such an F-forest by taking two copies of F and gluing the vertex u in one copy to the vertex v in the other copy. This implies that if F is a core (see [21]), then every F-forest is homomorphic to F if and only if F is vertex-transitive.

[8]Indeed, consider an F-forest T consisting of one "central" copy F_0 of F, and for each vertex v of F_0, an F-copy F_v which intersects F_0 at v. Note that $F_0^\star$ contains the graph obtained from F_0 by replacing each $v \in V(F_0)$ with a set A_v of $\deg_F(v)$ different vertices and placing a perfect matching on $\bigcup_{v \in V(F)} A_v$ such that for every $uv \in E(F_0)$, there is an edge of the matching between A_u and A_v. Now, for each $v \in V(F_0)$, there is a vertex $w_v \in V(T^\star)$ which is adjacent to all vertices of A_v (this vertex comes from the F-copy F_v). This

with Theorem 1.12. In fact, we conjecture that the condition in Theorem 1.15 characterises the cases where $M_{F,H}(\varepsilon)$ is at least $2^{\mathrm{poly}(1/\varepsilon)}$, and in all other cases $M_{F,H}(\varepsilon)$ is polynomial.

Conjecture 1.17. *Let F, H be graphs, and suppose that there exists an F-forest T such that $T \to F$ and $H \to T^\star$. Then $M_{F,H}(\varepsilon) = \mathrm{poly}(1/\varepsilon)$.*

One upshot of Conjecture 1.17 would be that Construction 1.13 is "universal" for lower-bounding $M_{F,H}(\varepsilon)$, that is, that whenever this construction fails to witness an exponential lower bound, then in fact no such lower bound exists. In particular, Conjecture 1.17 would imply that $M_{F,H}(\varepsilon)$ cannot have any intermediate growth rate between polynomial and exponential.

1.4. **Organisation.** The rest of this paper is organised as follows. We prove our results on asymmetric homomorphism thresholds, Theorems 1.3 and 1.5, in Section 2. Our upper bounds on $M_{F,H}(\varepsilon)$, Proposition 1.10 and Theorem 1.12, are proved in Section 3. Finally, our lower bound on $M_{F,H}(\varepsilon)$ is proved over two sections: Section 4 contains the proof of the key Lemma 1.14, and the deduction of Theorem 1.15 is given in Section 5. We end with some concluding remarks and open problems in Section 6.

All logarithms in this paper are to base 2. We systematically omit floor and ceiling signs whenever they are not crucial.

2. Asymmetric homomorphism thresholds

In this section, we prove Theorems 1.3 and 1.5. Both are proved using the generalised Mycielskian construction, whose definition we now recall.

Construction 2.1. *Given an integer $t \geqslant 1$ and a graph Γ, its t-fold Mycielskian is the graph $M_t(\Gamma)$ defined as follows:*

(1) *$V(M_t(\Gamma)) = (V(\Gamma) \times \{1, \ldots, t+1\}) \cup \{r\}$. That is, the vertices of $M_t(\Gamma)$ are of the form (v, i) for some $v \in V(\Gamma)$ and $1 \leqslant i \leqslant t+1$, as well as a single distinguished vertex r.*
(2) *The induced subgraph on $V(\Gamma) \times \{t+1\}$ is a copy of Γ, and $V(\Gamma) \times \{i\}$ is an independent set for all $1 \leqslant i \leqslant t$.*
(3) *r is adjacent to all vertices in $V(\Gamma) \times \{1\}$ and to no other vertices of $M_t(\Gamma)$.*
(4) *For all $1 \leqslant i \leqslant t$, each vertex (v, i) is adjacent to all vertices $(w, i+1)$ with $vw \in E(\Gamma)$, and to no other vertices in $V(\Gamma) \times \{i+1\}$.*

Equivalently, one can obtain $M_t(\Gamma)$ from the tensor product of Γ with a path on t edges by inserting a copy of Γ inside the last fibre, and adding a new vertex r joined to all vertices of the first fibre.

We remark that generalised Mycielskians have been well-studied in the literature, going back at least to work of Stiebitz [33], and generalising the classical work of Mycielski [30]. The key

---

gives us a copy of $F^{\bullet\bullet}$ where $(w_v)_{v \in V(F)}$ play the roles of the vertices of F and $\bigcup_{v \in V(F)} A_v$ play the roles of the subdivision vertices.
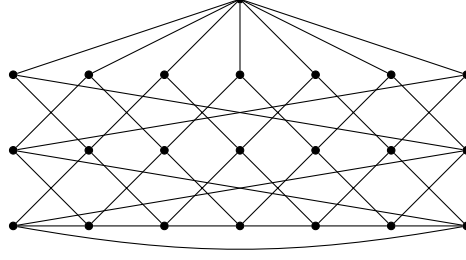
10

FIGURE 3. The 2-fold Mycielskian $M_2(C_7)$. Note that this graph is $\mathcal{C}_5$-free, as implied by Lemma 2.2.

property we need about t-fold Mycielskians is that they preserve odd girth, as stated in the following lemma, whose proof we provide for the sake of completeness. We recall that $\mathcal{C}_{2t+1}$ denotes the family $\{C_3, C_5, \ldots, C_{2t+1}\}$. Thus, being $\mathcal{C}_{2t+1}$-free is the same as having odd girth at least $2t + 3$.

Lemma 2.2. *If Γ is $\mathcal{C}_{2t+1}$-free, then $M_t(\Gamma)$ is $\mathcal{C}_{2t+1}$-free as well.*

Note that in case $t = 1$, this just says that the standard Mycielskian construction preserves triangle-freeness, which is the basic result used by Mycielski [30].

Proof of Lemma 2.2. We use the following simple facts: Every closed walk of length $2t + 1$ contains an odd cycle of length at most $2t + 1$, and any odd cycle of length at most $2t + 1$ gives rise to a closed walk of length $2t + 1$ (by traversing the cycle and then going back and forth along an edge). Therefore, it is sufficient to prove that $M_t(\Gamma)$ contains no closed walk of length $2t + 1$. Note that $M_t(\Gamma) \setminus \{r\}$ is homomorphic to Γ, via the homomorphism sending $(v, i) \mapsto v$. Therefore, as Γ is $\mathcal{C}_{2t+1}$-free, any closed walk of length $2t + 1$ in $M_t(\Gamma)$ must use the vertex r. Similarly, if we delete the layer $V(\Gamma) \times \{t + 1\}$ from $M_t(\Gamma)$, we obtain a bipartite graph, with the layer $V(\Gamma) \times \{t\}$ on one side of the bipartition. As a consequence, any subgraph containing at most one vertex from $V(\Gamma) \times \{t + 1\}$ is also bipartite (because in such a subgraph, the unique vertex from $V(\Gamma) \times \{t + 1\}$ only has neighbours in $V(\Gamma) \times \{t\}$). Hence, any walk of length $2t + 1$ must use at least two vertices of $V(\Gamma) \times \{t + 1\}$. But any walk from $V(\Gamma) \times \{t + 1\}$ to r must use at least t intermediate vertices, hence any closed odd walk in $M_t(\Gamma)$ must have length at least $2t + 3$. □

The next lemma is the key result underpinning Theorems 1.3 and 1.5. It allows us to extend a homomorphism $G[U] \to \Gamma$ to a homomorphism $G \to M_t(\Gamma)$, assuming that the structure of $V(G) \setminus U$ is sufficiently simple. Given a set X and an integer i, we denote by $N^i(X)$ the ith neighbourhood of X, namely the set of all vertices at distance i from some vertex of X, and at distance at least i from all vertices of X. Note that $N^0(X) = X$, $N^1(X) = N(X)$ is just the neighbourhood of X, and $N^i(X) \subseteq N(N^{i-1}(X))$ for all i.

Lemma 2.3. *Let G be a graph, and let $U \sqcup I$ be a partition of its vertex set. Suppose that $I, N(I), N^2(I), \ldots, N^t(I)$ are all independent sets in G. If $G[U]$ is homomorphic to some graph Γ, then G is homomorphic to $M_t(\Gamma)$.*

Proof. Let $\phi : G[U] \to \Gamma$ be a homomorphism, which exists by assumption. We define $\psi : G \to M_t(\Gamma)$ as follows.

(1) If $x \in I$, we set $\psi(x) = r \in V(M_t(\Gamma))$.
(2) If $x \in N^i(I)$ for some $1 \leqslant i \leqslant t$, then in particular $x \in U$, hence $\phi(x) = v$ for some $v \in V(\Gamma)$. We then set $\psi(x) = (v, i) \in V(M_t(\Gamma))$.
(3) Finally, if $x \in U \setminus (N^1(I) \cup \cdots \cup N^t(I))$, we set $\psi(x) = (\phi(x), t+1) \in V(M_t(\Gamma))$.

We claim that ψ is a homomorphism. So fix any $xy \in E(G)$. As I is an independent set, at most one of x, y is in I. If $x \in I$, then the fact that $xy \in E(G)$ implies that $y \in N^1(I)$, hence $\psi(y) \in V(\Gamma) \times \{1\}$. By the definition of $M_t(\Gamma)$, we conclude that $\psi(x) = r$ is adjacent to $\psi(y)$.

It thus remains to handle the case when both x and y are in U. If $x \in N^i(I)$ for some $1 \leqslant i \leqslant t$, then $y \in N^{i-1}(I) \cup N^{i+1}(I)$ (here we use the assumption that $N^i(I)$ is independent). We may assume without loss of generality that $y \in N^{i+1}(I)$ (in the other case, we interchange the roles of x and y and decrease i by one). Let $v = \phi(x)$ and $w = \phi(y)$, which are adjacent in Γ as ϕ is a homomorphism $G[U] \to \Gamma$. By the definition of ψ, we have $\psi(x) = (v, i)$ and $\psi(y) = (w, i+1)$, which are adjacent in $M_t(\Gamma)$ by definition. The final case is when both x and y are in $U \setminus (N^1(I) \cup \cdots \cup N^t(I))$. In this case, again denoting $v = \phi(x), w = \phi(y)$ as above, we have that $\psi(x) = (v, t+1)$ and $\psi(y) = (w, t+1)$. But as $V(\Gamma) \times \{t+1\}$ induces a copy of Γ in $M_t(\Gamma)$, we have that $\psi(x)$ and $\psi(y)$ are adjacent, since v and w are adjacent in Γ. $\square$

We are now ready to prove our results on asymmetric homomorphism thresholds, beginning with Theorem 1.3.

Proof of Theorem 1.3. Select a maximal collection $a_1, \ldots, a_k$ of vertices in G whose neighbourhoods $X_j := N(a_j)$ are pairwise disjoint. By the minimum degree assumption, we have $|X_j| \geqslant \delta|G|$ for all j, hence $k \leqslant \frac{1}{\delta}$. We partition the vertices in $V(G) \setminus \bigcup_{j=1}^k X_j$ into sets $Y_1, \ldots, Y_k$ as follows. For every vertex $v \in V(G) \setminus \bigcup_{j=1}^k X_j$, the maximality of the collection $a_1, \ldots, a_k$ implies that $N(v) \cap X_j \neq \varnothing$ for some j. We then set $v \in Y_j$ for the minimal such index j. Note that $V(G) = X_1 \cup \cdots \cup X_k \cup Y_1 \cup \cdots \cup Y_k$. Next, we prove the following:

Claim 2.4. *For every $1 \leqslant j \leqslant k$ and $0 \leqslant i \leqslant t$, the sets $N^i(X_j), N^i(Y_j)$ are independent in G.*

Proof. We first note that $N^i(a_j)$ is an independent set for all $1 \leqslant i \leqslant t+2$, as an edge between two vertices at distance i from a_j would yield a closed walk of length $2i + 1 \leqslant 2(t+2) + 1 = 2t + 5$, a contradiction to the assumption that G is $\mathcal{C}_{2t+5}$-free. It follows that the subgraph of G induced by $\bigcup_{i=0}^{t+2} N^i(a_j)$ is bipartite with parts $\bigcup_{i=0 \text{ even}}^{t+2} N^i(a_j)$ and $\bigcup_{i=0 \text{ odd}}^{t+2} N^i(a_j)$. Now fix any $0 \leqslant i \leqslant t$. Every vertex $v \in N^i(X_j)$ has a walk of length $i + 1 < t + 2$ to a_j, because v has a path of length i to $X_j = N(a_j)$. Hence, all vertices in $N^i(X_j)$ are in the same part of the

12

bipartition, implying that $N^i(X_j)$ is independent. Similarly, every vertex $v \in N^i(Y_j)$ has a walk of length $i + 2 \leqslant t + 2$ to a_j, because v has a path of length i to Y_j, and every vertex in Y_j has a walk of length 2 to a_j. Hence, again, all vertices in $N^i(Y_j)$ are in the same part of the bipartition and so $N^i(Y_j)$ is independent. □

We now define an increasing sequence of sets $U_1 \subseteq \cdots \subseteq U_{2k}$ by declaring

$$U_\ell = \bigcup_{j=1}^{\ell} X_j \qquad \text{and} \qquad U_{k+\ell} = U_k \cup \bigcup_{j=1}^{\ell} Y_j$$

for all $1 \leqslant \ell \leqslant k$. Let $G_j = G[U_j]$, and note that $G_{2k} = G$ since $U_{2k} = V(G)$. Moreover, if we set $I_{j+1} := U_{j+1} \setminus U_j \in \{X_1, \ldots, X_k, Y_1, \ldots, Y_k\}$, then by Claim 2.4, I_{j+1} and $N^i(I_{j+1})$ are independent in G for all $1 \leqslant i \leqslant t$, hence also independent in G_{j+1}.

Now let Γ_1 be the one-vertex graph, and let $\phi_1 : G_1 \to \Gamma_1$ be the constant function, which is a homomorphism since G_1 has no edges. Inductively, having defined $\phi_j : G_j \to \Gamma_j$, we let $\Gamma_{j+1} = M_t(\Gamma_j)$, and let $\phi_{j+1} : G_{j+1} \to \Gamma_{j+1}$ be the homomorphism given by Lemma 2.3, which we may apply since $V(G_{j+1}) = V(G_j) \sqcup I_{j+1}$ and since I_{j+1} and $N^i(I_{j+1})$ are independent sets for all $1 \leqslant i \leqslant t$. Note that by Lemma 2.2, each of the graphs Γ_j is $\mathcal{C}_{2t+1}$-free. Moreover, $|\Gamma_j| = (t+1)|\Gamma_{j-1}| + 1$, hence $|\Gamma_j| = \frac{(t+1)^j - 1}{t}$ holds by induction. At the end of this process, we have defined a homomorphism $\phi_{2k} : G_{2k} \to \Gamma_{2k}$, where $G_{2k} = G$ and Γ_{2k} is a $\mathcal{C}_{2t+1}$-free graph on $\frac{(t+1)^{2k} - 1}{t} \leqslant (t+1)^{2k} \leqslant (t+1)^{2/\delta}$ vertices, as claimed. □

Theorem 1.5 is proved in a very similar way; the only difference is that we do not need to pick out the maximal collection $a_1, \ldots, a_k$, and instead use the given dominating set of bounded size.

Proof of Theorem 1.5. We proceed by induction on $\gamma(G)$. The base case $\gamma(G) = 1$ is trivial as then G is a star, hence homomorphic to K_2. For the inductive step, let $a_1, \ldots, a_k$ be a dominating set of G with $k = \gamma(G) \geqslant 2$. Let $I = N(a_k)$ and let $U = V(G) \setminus I$. Note that $G_0 := G[U \setminus \{a_k\}]$ has domination number at most $k - 1$, as $a_1, \ldots, a_{k-1}$ is a dominating set of G_0. Moreover, as a subgraph of G, we have that G_0 is $\mathcal{C}_{2t+3}$-free. By the inductive hypothesis, there is a $\mathcal{C}_{2t+1}$-free graph Γ_0 on at most $3 \cdot (t+1)^{k-2} - 1$ vertices, as well as a homomorphism $\phi_0 : G_0 \to \Gamma_0$. Note that a_k is an isolated vertex in $G[U]$, hence we may extend ϕ_0 to a homomorphism $\phi : G[U] \to \Gamma_0$ by mapping a_k to an arbitrary vertex of Γ_0.

As in the proof of Theorem 1.3, one can show that $N^i(I)$ is independent for every $0 \leqslant i \leqslant t$, using the fact that $N^i(a_k)$ is independent for every $0 \leqslant i \leqslant t + 1$, which holds because G is $\mathcal{C}_{2t+3}$-free. Therefore, by Lemma 2.3, there is a homomorphism $\psi : G \to \Gamma$, where $\Gamma = M_t(\Gamma_0)$. Also, Γ is $\mathcal{C}_{2t+1}$-free by Lemma 2.2, and

$$|\Gamma| = (t+1)|\Gamma_0| + 1 \leqslant (t+1)(3 \cdot (t+1)^{k-2} - 1) + 1 = 3 \cdot (t+1)^{k-1} - t \leqslant 3 \cdot (t+1)^{k-1} - 1,$$

as claimed. □

Finally, we turn to the proof of Corollary 1.6, for which we need to recall a few results of VC theory. Let $\mathcal{F} \subseteq 2^V$ be a set system on a finite ground set V. We say that $\mathcal{F}$ *shatters* a set $S \subseteq V$ if, for all $T \subseteq S$, there exists $F \in \mathcal{F}$ with $F \cap S = T$. The *VC dimension* of $\mathcal{F}$ is then defined as the maximum size of a shattered set $S \subseteq V$. For a graph G, we define its VC dimension to be the VC dimension of the set system $\mathcal{F} := \{N(v) : v \in V(G)\} \subseteq 2^{V(G)}$. Note that this agrees with the definition of VC dimension of a graph discussed in the Introduction.

The key fact that we need about set systems of bounded VC dimension is the fundamental ε-net theorem of Haussler and Welzl [20]. A set $A \subseteq V$ is called an ε-*net* for $\mathcal{F} \subseteq 2^V$ if A intersects every $F \in \mathcal{F}$ with $|F| \geqslant \varepsilon |V|$. Haussler and Welzl proved the following.

Theorem 2.5 ([20, Theorem 3.8]). *If $\mathcal{F} \subseteq 2^V$ has VC dimension at most d, then $\mathcal{F}$ has an ε-net of size at most $\frac{8d}{\varepsilon} \log \frac{8d}{\varepsilon}$.*

For our purposes, Theorem 2.5 immediately implies that graphs of bounded VC dimension and linear minimum degree have bounded domination number.

Lemma 2.6. *If G is a graph with VC dimension at most d and minimum degree at least $\delta |G|$, then the domination number of G is at most $\frac{8d}{\delta} \log \frac{8d}{\delta}$.*

Proof. Let $\mathcal{F} := \{N(v) : v \in V(G)\} \subseteq 2^{V(G)}$. By assumption, $\mathcal{F}$ has VC dimension at most d. Therefore, Theorem 2.5 implies that there is a δ-net $A \subseteq V(G)$ for $\mathcal{F}$ of size at most $\frac{8d}{\delta} \log \frac{8d}{\delta}$. We claim that A is a dominating set of G. Indeed, fix some $w \in V(G)$. By the minimum degree condition, we have that $|N(w)| \geqslant \delta |G|$, hence there must exist some $v \in A$ such that $v \in N(w)$. But this exactly means that w is dominated by A, as claimed. $\square$

With this lemma in hand, Corollary 1.6 follows immediately.

Proof of Corollary 1.6. Let G be a $\mathcal{C}_{2t+3}$-free graph with VC dimension at most d and minimum degree at least $\delta |G|$. By Lemma 2.6, we have $\gamma(G) \leqslant \frac{8d}{\delta} \log \frac{8d}{\delta}$. Therefore, Theorem 1.5 implies that G is homomorphic to a $\mathcal{C}_{2t+1}$-free graph Γ with

$$|\Gamma| \leqslant 3 \cdot (t+1)^{\gamma(G)} \leqslant 3 \cdot (t+1)^{\frac{8d}{\delta} \log \frac{8d}{\delta}}.$$

$\square$

3. Approximate homomorphisms: upper bounds

In this section, we prove our upper bounds on $M_{F,H}(\varepsilon)$, beginning with Proposition 1.10.

3.1. Proof of Proposition 1.10. In the course of the proof of Proposition 1.10, we will need a consequence of the Frieze–Kannan weak regularity lemma [13] and the counting lemma for the cut distance. In order to state it, we will need to set up some notation. Recall that $t(H, G)$ denotes the *homomorphism density* from H to G, namely the probability that a random function $V(H) \to V(G)$ is a homomorphism. Let W be a weighted graph, that is, a symmetric function $W : V \times V \to [0, 1]$, where V is some finite set; note that we are allowing W to have

"loops", by allowing $W(x, x)$ to be non-zero. The homomorphism density $t(H, W)$ is then defined analogously, as

$$t(H, W) := \frac{1}{|V|^{|H|}} \sum_{v_1, \dots, v_h \in V} \prod_{ij \in E(H)} W(v_i, v_j).$$

Equivalently, this is the expected value of $\prod_{ij \in E(H)} W(v_i, v_j)$, where $v_1, \dots, v_h$ are independent uniformly random choices of vertices from V. Note that this definition recovers the earlier definition of $t(H, G)$ in case $W = G$ is a graph, that is, a $\{0, 1\}$-valued function. Finally, for a graph G and a partition $\mathcal{P} = A_1 \sqcup \cdots \sqcup A_M$ of $V(G)$, we denote by $G_{\mathcal{P}}$ the weighted graph whose vertex set is $[M]$, and where the edge weight of a pair (i, j) is given by[9] $d_G(A_i, A_j)$. Also, $\mathcal{P}$ is called *equitable* if $||A_i| - |A_j|| \leqslant 1$ for all i, j.[10]

The following result follows immediately by combining the statements of [26, Lemma 9.3 and Exercise 9.7] (the Frieze–Kannan weak regularity lemma) and [26, Lemma 10.23] (the counting lemma for the cut distance).

Lemma 3.1. *Let G be a graph and let $M \geqslant 1$ be an integer. There exists an equitable partition $\mathcal{P}$ of $V(G)$ into M parts such that, for any graph H, we have*

$$|t(H, G) - t(H, G_{\mathcal{P}})| \leqslant \frac{4e(H)}{\sqrt{\log M}}.$$

Using Lemma 3.1, it is straightforward to prove Proposition 1.10.

Proof of Proposition 1.10. Let $M = 2^{K^2}$, where K is as in (2), and let G be an n-vertex H-hom-free graph. Let $\mathcal{P} = V_1 \sqcup \cdots \sqcup V_M$ be the partition of $V(G)$ into M parts given by Lemma 3.1. Let Γ_0 be the simple graph on vertex set $[M]$ whose edges are those pairs (i, j) with $d_G(V_i, V_j) \geqslant \frac{\varepsilon}{2}$ (note that Γ_0 has no loops). Finally, let Γ be a maximum F-hom-free subgraph of Γ_0. Note that there is a natural map $\phi : V(G) \to V(\Gamma_0) = V(\Gamma)$, simply mapping each vertex v to the index i such that $v \in V_i$. Also, Γ is an F-hom-free M-vertex graph by construction, so it suffices to prove that ϕ is an ε-approximate homomorphism.

To prove this, we first claim that Γ_0 is $\frac{\varepsilon}{2}$-close to F-hom-free. Indeed, if not, then by the definition of δ, we find that Γ_0 contains at least $\delta(\frac{\varepsilon}{2}) \cdot M^{|H|}$ copies of H. Every such copy corresponds to $|H|$ distinct vertices $v_1, \dots, v_{|H|}$ in $V(\Gamma_0) = V(G_{\mathcal{P}})$, such that all the $e(H)$ pairs (v_i, v_j) corresponding to edges of H are present in Γ_0, and thus satisfy $d_G(V_{v_i}, V_{v_j}) \geqslant \frac{\varepsilon}{2}$. In particular, we have

$$\prod_{ij \in E(H)} G_{\mathcal{P}}(v_i, v_j) = \prod_{ij \in E(H)} d_G(V_{v_i}, V_{v_j}) \geqslant \left(\frac{\varepsilon}{2}\right)^{e(H)}.$$

---

[9] If A, B are vertex subsets of a graph G, we denote by $e_G(A, B)$ the number of pairs in $A \times B$ that are edges of G, and by $d_G(A, B) := e_G(A, B)/(|A||B|)$ the *edge density*.

[10] To keep the presentation simple, we will always assume that $|G|$ is divisible by M, so that $|A_i| = |A_j|$ for all i, j. This does not affect our asymptotic results.

Summing this up over all the copies of H in Γ_0, of which there are at least $\delta(\frac{\varepsilon}{2}) \cdot M^{|H|}$, we conclude that

$$t(H, G_{\mathcal{P}}) \geqslant \delta\left(\frac{\varepsilon}{2}\right) \cdot \left(\frac{\varepsilon}{2}\right)^{e(H)}.$$

But by Lemma 3.1, this implies that

$$
\begin{aligned}
t(H, G) \geqslant t(H, G_{\mathcal{P}}) - \frac{4e(H)}{\sqrt{\log M}} &\geqslant \delta\left(\frac{\varepsilon}{2}\right) \cdot \left(\frac{\varepsilon}{2}\right)^{e(H)} - \frac{4e(H)}{K} \\
&= \frac{1}{5}\delta\left(\frac{\varepsilon}{2}\right) \cdot \left(\frac{\varepsilon}{2}\right)^{e(H)} \\
&> 0,
\end{aligned}
$$

where the equality uses the choice of K in (2) This contradicts our assumption that G is H-hom-free, showing that Γ_0 is, as claimed, $\frac{\varepsilon}{2}$-close to F-hom-free. In particular, as Γ is a maximum F-hom-free subgraph of Γ_0, we conclude that

$$e(\Gamma) \geqslant e(\Gamma_0) - \frac{\varepsilon}{2}M^2. \tag{3}$$

We now prove that ϕ is an ε-approximate homomorphism. There are three types of edges of G that are mapped to non-edges of Γ: those edges within a part V_i, those edges between parts (V_i, V_j) with $d_G(V_i, V_j) < \frac{\varepsilon}{2}$, and those edges between parts (V_i, V_j) such that $ij \in E(\Gamma_0) \setminus E(\Gamma)$. Since the partition is equitable, there are at most $M\binom{n/M}{2} \leqslant \frac{n^2}{2M}$ edges of the first type. For the second type, there are at most $\binom{M}{2}$ such pairs, and each pair contributes at most $\frac{\varepsilon}{2}(\frac{n}{M})^2$ such edges, hence there are at most $\binom{M}{2} \cdot \frac{\varepsilon}{2}(\frac{n}{M})^2 \leqslant \frac{\varepsilon}{4}n^2$ edges of the second type. Finally, for the third type, by (3), there are at most $\frac{\varepsilon}{2}M^2$ such pairs (V_i, V_j), and each contributes at most $(\frac{n}{M})^2$ edges, so the number of edges of the third type is at most $\frac{\varepsilon}{2}n^2$. In total, the number of edges of G mapped to non-edges of Γ is at most

$$\frac{n^2}{2M} + \frac{\varepsilon}{4}n^2 + \frac{\varepsilon}{2}n^2 = \left(\frac{1}{2M} + \frac{3\varepsilon}{4}\right)n^2 \leqslant \varepsilon n^2,$$

since $M \geqslant \frac{2}{\varepsilon}$. Hence, ϕ is an ε-approximate homomorphism. $\qquad\square$

3.2. **Proof of Theorem 1.12.** In this section we prove Theorem 1.12. We need the following simple lemma.

Lemma 3.2. *Let G be an n-vertex graph and let $\varepsilon > 0$. There exist (not necessarily distinct)[11] vertices $v_1, \ldots, v_k \in V(G)$ and disjoint sets $S_1, \ldots, S_k \subseteq V(G)$ with the following properties.*

(1) *$|S_i| = \varepsilon n/3$ for all $1 \leqslant i \leqslant k$; hence $k \leqslant 3/\varepsilon$;*
(2) *for every $1 \leqslant i \leqslant k$, v_i is adjacent to all vertices in S_i; and*
(3) *letting $X := V(G) \setminus (S_1 \cup \cdots \cup S_k)$, there are at most $\varepsilon n^2/2$ edges incident to X.*

Proof. We run the following greedy algorithm to construct $v_1, \ldots, v_k, S_1, \ldots, S_k$.

---
[11]By slightly modifying the proof, we could also guarantee that the vertices $v_1, \ldots, v_k$ are pairwise distinct and $\{v_1, \ldots, v_k\}$ is disjoint from $S_1, \ldots, S_k$.

(1) Suppose we have already defined $v_1, \ldots, v_\ell, S_1, \ldots, S_\ell$. Let $S = S_1 \cup \cdots \cup S_\ell$ (in case $\ell = 0$, set $S = \varnothing$).

(2) Let $X = V(G) \setminus S$. If there are at most $\varepsilon n^2/2$ edges incident to X, terminate the algorithm.

(3) If there are at least $\varepsilon n^2/6$ edges inside X, then there is some vertex $v \in X$ which is adjacent to at least $\varepsilon n/3$ vertices of X. We set $v_{\ell+1} = v$ and let $S_{\ell+1}$ be an arbitrary set of $\varepsilon n/3$ neighbours of v in X. Then $S_{\ell+1}$ is disjoint from $S_1, \ldots, S_\ell$ since $S_{\ell+1} \subseteq X = V(G) \setminus (S_1 \cup \cdots \cup S_\ell)$.[12] Return to step 1.

(4) If there are at least $\varepsilon n^2/2$ edges incident to X but fewer than $\varepsilon n^2/6$ edges inside X, there must be at least $\varepsilon n^2/3$ edges between S and X. Hence, there exists $v \in S$ which is adjacent to at least $\varepsilon n/3$ vertices in X. We again let $v_{\ell+1} = v$ and let $S_{\ell+1}$ be an arbitrary set of $\varepsilon n/3$ neighbours of v in X. Again, $S_{\ell+1}$ is disjoint from $S_1, \ldots, S_\ell$. Return to step 1.

At the end of this process, we have found vertices $v_1, \ldots, v_k$ and sets $S_1, \ldots, S_k$. As discussed above, the sets $S_1, \ldots, S_k$ are pairwise disjoint and satisfy $|S_i| = \varepsilon n/3$ for all i. Additionally, by construction, we have that v_i is adjacent to all vertices in S_i, and that there are at most $\varepsilon n^2/2$ edges incident to $X = V(G) \setminus (S_1 \cup \cdots \cup S_k)$, since this is precisely the condition for terminating the algorithm. $\qquad \square$

Remark. As mentioned in the introduction, Lemma 3.2 can be used to quickly deduce the bound $M_{K_3,C_5}(\varepsilon) \leqslant 2^{O(1/\varepsilon)}$ from Theorem 1.5. Indeed, suppose that G is $\mathcal{C}_5$-free, let $(v_i, S_i)_{i=1}^k, X$ be given by Lemma 3.2, and set $G_0 = G - X$. Map $V(G)$ to $V(G_0)$ by mapping each $v \in V(G_0)$ to itself and mapping each vertex in X to an arbitrary vertex in G_0. This is an ε-approximate homomorphism from G to G_0, because only edges touching X are mapped to non-edges, and there are at most εn^2 edges touching X. Also, G_0 is $\mathcal{C}_5$-free and has domination number at most $k \leqslant 3/\varepsilon$. At this point we may apply Theorem 1.5 to G_0 to obtain the claimed result.

Next, we prove Theorem 1.12.

Proof of Theorem 1.12. Let G be an n-vertex $F^{\bullet\bullet}$-hom-free graph. We apply Lemma 3.2 to find vertices $v_1, \ldots, v_k$ and pairwise disjoint sets $S_1, \ldots, S_k$ with the properties given in Lemma 3.2. Let $X = V(G) \setminus (S_1 \cup \cdots \cup S_k)$.

Let Γ be the graph with vertex set $\{x, s_1, \ldots, s_k\}$ in which x is an isolated vertex, and where $s_i s_j \in E(\Gamma)$ if and only if $e(S_i, S_j) > 0$. Note that $|\Gamma| = 1 + k \leqslant 1 + 3/\varepsilon = O(\frac{1}{\varepsilon})$.

We define a map $\phi : V(G) \to V(\Gamma)$ by mapping every vertex in X to x and every vertex in S_i to s_i, for all $1 \leqslant i \leqslant k$. We first claim that ϕ is an ε-approximate homomorphism.

Note that the only edges of G mapped to non-edges of Γ by ϕ are edges incident to X and edges contained in some S_i. Indeed, every other edge of G goes between two vertices

---

[12]It may be that v has already appeared in the sequence $v_1, \ldots, v_\ell$, which is permissible since we do not require these vertices to be distinct.

$v \in S_i, v' \in S_j$ for some $1 \leqslant i < j \leqslant k$. As vv' is an edge between S_i and S_j, we see that $s_i s_j \in E(\Gamma)$, hence ϕ maps vv' to an edge of Γ. By Lemma 3.2, there are at most $\varepsilon n^2/2$ edges incident to X. Additionally, the number of edges contained in some S_i is at most

$$\sum_{i=1}^{k} \binom{|S_i|}{2} \leqslant \sum_{i=1}^{k} \frac{|S_i|^2}{2} = k \cdot \frac{(\varepsilon n/3)^2}{2} \leqslant \frac{\varepsilon n^2}{6},$$

using the upper bounds on k and $|S_i|$ given by Lemma 3.2. Since $\varepsilon n^2/2 + \varepsilon n^2/6 \leqslant \varepsilon n^2$, we conclude that ϕ is indeed an ε-approximate homomorphism.

It remains to check that Γ is F-hom-free, so suppose for contradiction that there is a homomorphism $\psi : F \to \Gamma$. Since x is an isolated vertex in Γ, we may assume that $x \notin \psi(V(F))$, since any vertex mapped to x must be isolated in F, and hence can be mapped to some s_i instead. Let the vertices of F be $u_1, \ldots, u_{|F|}$, and define $\ell_1, \ldots, \ell_{|F|}$ by $\psi(u_1) = s_{\ell_1}, \ldots, \psi(u_{|F|}) = s_{\ell_{|F|}}$.

For every edge $u_i u_j \in E(F)$, we have that $\psi(u_i)\psi(u_j) \in E(\Gamma)$, hence s_{ℓ_i} is adjacent to s_{ℓ_j} in Γ. By the definition of Γ, we conclude that G contains an edge between S_{ℓ_i} and S_{ℓ_j}. That is, we can pick some $x_{ij} \in S_{\ell_i}, x_{ji} \in S_{\ell_j}$ such that $x_{ij}x_{ji} \in E(G)$.

Now, we define a homomorphism $F^{\bullet\bullet} \to G$ as follows. For each vertex of $F^{\bullet\bullet}$ which corresponds to an original vertex u_i of F, we map it to v_{ℓ_i}. For the two new vertices added on an edge $u_i u_j$, we map them to x_{ij} and x_{ji}. This is indeed a homomorphism, since v_{ℓ_i} is adjacent to all vertices in S_{ℓ_i} by construction, and x_{ij} is adjacent to x_{ji} by the way we picked these vertices. This contradicts the assumption that G is $F^{\bullet\bullet}$-hom-free, and completes the proof. $\qquad\square$

4. Proof of Lemma 1.14

We now turn to proving our exponential lower bound on $M_{F,H}(\varepsilon)$. This section contains the proof of the key technical lemma, Lemma 1.14.

Following [12], the proof of Lemma 1.14 uses the language of entropy. We now recall the necessary notions. We also refer the reader to [15, Section 2] for additional information. Given a discrete random variable X, we use $H(X)$ to denote the binary entropy of X. For two random variables X, Y, $H(X \mid Y) := H(X,Y) - H(Y)$ is the conditional entropy of X given Y, and $I(X;Y) := H(X) - H(X \mid Y) = H(X) + H(Y) - H(X,Y)$ is the mutual information of X, Y. Note that $I(X;Y) = I(Y;X)$. We will use the fact that if $X_1, \ldots, X_m$ are independent random variables, then

$$\sum_{i=1}^{m} I(X_i; Y) \leqslant I((X_1, \ldots, X_m); Y). \tag{4}$$

This holds because $H(X_1, \ldots, X_m) = \sum_{i=1}^{m} H(X_i)$ for independent $X_1, \ldots, X_m$, and because $H((X_1, \ldots, X_m) \mid Y) \leqslant \sum_{i=1}^{m} H(X_i \mid Y)$, which is the subadditivity of (conditional) entropy.

We further define the function $H^{-1}(x)$ on $x \in [0,1]$ to be the unique value $p \in [0, 1/2]$ such that $H(\mathbf{Ber}(p)) = x$. Thus, for a Bernoulli random variable X with entropy at least h, we have $\mathbb{P}[X = 0], \mathbb{P}[X = 1] \geqslant H^{-1}(h)$. We need the following simple claim.

Lemma 4.1. *Let Z be a random variable taking values in $\{1, \ldots, k\}$ and suppose that $H(X) \geqslant \log k - \beta$. Then for every $a \in \{1, \ldots, k\}$, $\mathbb{P}[X = a] \geqslant H^{-1}(h)$, where $h = \log(\frac{k}{k-1}) - \beta$.*

Proof. Let J be the indicator random variable of the event $X = a$. We have

$$H(X) = H(X, J) = H(X \mid J) + H(J),$$

where the first equality is because X determines J. Now,

$$H(X \mid J) = \mathbb{P}[J = 1] \cdot H(X \mid J = 1) + \mathbb{P}[J = 0] \cdot H(X \mid J = 0).$$

Conditioned on $J = 1$, X is a constant random variable, so $H(X \mid J = 1) = 0$. Also, $H(X \mid J = 0) \leqslant \log(k - 1)$ because conditioned on $J = 0$, X takes at most $k - 1$ values. Combining all of the above, we get

$$\log k - \beta \leqslant H(X) = H(X \mid J) + H(J) \leqslant H(X \mid J = 0) + H(J) \leqslant \log(k - 1) + H(J).$$

So $H(J) \geqslant \log(\frac{k}{k-1}) - \beta = h$, which implies the required bound on $\mathbb{P}[X = a] = \mathbb{P}[J = 1]$. $\quad\square$

We now turn to proving Lemma 1.14. Before doing so, let us sketch the strategy. Fix an F-hom-free graph Γ with $|\Gamma| < 2^{cm/n}$ (where c will be chosen later) and a map $\phi : V(G^\star) \to V(\Gamma)$. We wish to prove that ϕ is not an ε-approximate homomorphism, so our goal is to show that more than $\varepsilon |G^\star|^2$ edges of $G^\star$ are not mapped to edges of Γ. Recall that Δ is the maximum degree of F, and that we denote the vertices of $G^\star$ as $(v, \mathbf{x})$, where $v \in V(G)$ and $\mathbf{x} \in \{1, \ldots, \Delta\}^m$.

For each $v \in V(G)$, let $\mathbf{x}^v$ be a uniformly random element of $\{1, \ldots, \Delta\}^m$, and let $\mathbf{y}^v = \phi(v, \mathbf{x}^v)$; that is, $\mathbf{y}^v$ is the image under ϕ of the vertex $(v, \mathbf{x}^v)$. We wish to analyse the mutual information $I(\mathbf{x}^v; \mathbf{y}^v)$ of these two random variables. We will show that our bound on $|\Gamma|$ implies that for all $v \in V(G)$, the mutual information $I(\mathbf{x}^v; \mathbf{y}^v)$ is at most $O(m/n)$. The number of F-copies in G per vertex $v \in V(G)$ is at least m/n on average, so this implies that the average "mutual information per F-copy" in G is at most some constant. Finally, we show that if the "mutual information of an F-copy F_i" in G is at most some constant, then the proportion of edges in the blowup of F_i which are not mapped to edges of Γ is also some constant. Calculations then show that this constant is larger than ε if c is sufficiently small.

We now turn to the formal proof.

Proof of Lemma 1.14. As discussed above, we fix any F-hom-free graph Γ with $|\Gamma| < 2^{cm/n}$, for a constant $c > 0$ to be chosen later, as well as any map $\phi : V(G^\star) \to V(\Gamma)$. For each $v \in V(G)$, let $\mathbf{x}^v$ be a uniformly random element of $\{1, \ldots, \Delta\}^m$, and let $\mathbf{y}^v = \phi(v, \mathbf{x}^v)$.

We first use our upper bound on $|\Gamma|$ to obtain an upper bound on $I(\mathbf{x}^v; \mathbf{y}^v)$ for every $v \in V(G)$. Indeed, we have $I(\mathbf{x}^v; \mathbf{y}^v) = H(\mathbf{y}^v) - H(\mathbf{y}^v \mid \mathbf{x}^v) = H(\mathbf{y}^v)$, where the first step is by definition, and the second uses that $H(\mathbf{y}^v \mid \mathbf{x}^v) = 0$ as $\mathbf{y}^v$ is determined by $\mathbf{x}^v$. Moreover, $H(\mathbf{y}^v) \leqslant \log(|\Gamma|) < c \cdot \frac{m}{n}$, because $\mathbf{y}^v$ takes values in $V(\Gamma)$ and $|\Gamma| \leqslant 2^{cm/n}$. Hence, for every $v \in V(G)$ we have

$$I(\mathbf{x}^v; \mathbf{y}^v) < c \cdot \frac{m}{n}. \tag{5}$$

We now consider the coordinates of $\mathbf{x}^v$ separately, that is, we consider $I(\mathbf{x}_k^v; \mathbf{y}^v)$ for $k \in [m]$. Using (4), we get

$$\sum_{k \in [m]} I(\mathbf{x}_k^v; \mathbf{y}^v) \leqslant I(\mathbf{x}^v; \mathbf{y}^v). \tag{6}$$

Finally we convert this into a bound on the mutual information per F-copy. Recall that G has m copies of F, denoted $F_1, \ldots, F_m$, and each edge of G is in exactly one of these copies. We have

$$\sum_{k \in [m]} \sum_{v \in V(F_k)} I(\mathbf{x}_k^v; \mathbf{y}^v) \leqslant \sum_{v \in V(G)} \sum_{k \in [m]} I(\mathbf{x}_k^v; \mathbf{y}^v) \leqslant \sum_{v \in V(G)} I(\mathbf{x}^v; \mathbf{y}^v) < cm, \tag{7}$$

where the first inequality holds simply because $V(F_k) \subseteq V(G)$ for every k; the second uses (6); and the third uses (5). By (7) and Markov's inequality, at least $\frac{m}{2}$ of the indices $k \in [m]$ satisfy

$$\sum_{v \in V(F_k)} I(\mathbf{x}_k^v; \mathbf{y}^v) \leqslant 2c. \tag{8}$$

In what follows, for $k \in [m]$, we denote by $F_k^\star$ the subgraph of $G^\star$ consisting of the vertices $V(F_k) \times \{1, \ldots, \Delta\}^m$ and of the edges $(u, \mathbf{x}), (v, \mathbf{y})$ with $uv \in E(F_k)$ (recall Construction 1.13). An edge of $G^\star$ is *bad* if it is not mapped by ϕ to an edge of Γ. The remainder of the proof consists in showing that if F_k satisfies (8), then $F_k^\star$ contains many bad edges. More precisely, we prove the following:

Claim 4.2. *For $k \in [m]$, if $\sum_{v \in V(F_k)} I(\mathbf{x}_k^v; \mathbf{y}^v) \leqslant 2c$ then $F_k^\star$ contains at least $2c\left(\frac{|G^\star|}{n}\right)^2$ bad edges.*

Let us first complete the proof of the lemma using Claim 4.2. As we saw above, at least $\frac{m}{2}$ of the indices $k \in [m]$ satisfy the condition of Claim 4.2. For each such k, there are at least $2c\left(\frac{|G^\star|}{n}\right)^2$ bad edges in $F_k^\star$. Since the F-copies in G are edge-disjoint, these bad edges are different for different F-copies. Thus, in total, $G^\star$ has at least $\frac{cm}{n^2}|G^\star|^2$ bad edges. So if $\varepsilon < \frac{cm}{n^2}$, then there are more than $\varepsilon|G^\star|^2$ bad edges, as required.

From now on our goal is to prove Claim 4.2. So fix any $k \in [m]$ and let $v_1, \ldots, v_f$ be the vertices of F_k. We assume that $V(F) = [f]$ and that $i \mapsto v_i$ is an isomorphism from F to F_k. Let $V_1, \ldots, V_f$ be the blowup-sets in $G^\star$ corresponding to $v_1, \ldots, v_f$, i.e., $V_i = \{v_i\} \times \{1, \ldots, \Delta\}^m$. To simplify the notation, let us set $X_i = \mathbf{x}_k^{v_i}$ and $Y_i = \mathbf{y}^{v_i}$ for $1 \leqslant i \leqslant f$. Thus X_i is a uniformly random element of $\{1, \ldots, \Delta\}$ and $Y_i = \phi(v_i, \mathbf{x}^{v_i})$ is a vertex of Γ. With this notation, the assumption of Claim 4.2 is that

$$\sum_{i=1}^f I(X_i; Y_i) \leqslant 2c.$$

Since mutual information is nonnegative, it follows that $I(X_i; Y_i) \leqslant 2c$ for every $1 \leqslant i \leqslant f$. Also, by the definition of mutual information, we have

$$H(X_i \mid Y_i) = H(X_i) - I(X_i; Y_i) = \log(\Delta) - I(X_i; Y_i) \geqslant \log(\Delta) - 2c. \tag{9}$$

Now, recall that by definition, $H(X_i \mid Y_i) = \mathbb{E}_{y_i}\big[H(X_i \mid Y_i = y_i)\big]$. Since $H(X_i \mid Y_i = y_i) \leqslant \log(\Delta)$ for every y_i, we get by (9) and Markov's inequality that

$$\mathbb{P}_{y_i}\Big[H(X_i \mid Y_i = y_i) < \log(\Delta) - 4fc\Big] \leqslant \frac{1}{2f}.$$

Hence, there is a set $R_i \subseteq V(\Gamma)$ with $\mathbb{P}[Y_i \in R_i] \geqslant 1 - \frac{1}{2f}$ and $H(X_i \mid Y_i = y_i) \geqslant \log(\Delta) - 4fc$ for every $y_i \in R_i$. By the union bound, the probability that $Y_i \in R_i$ for every $1 \leqslant i \leqslant f$ is at least $\frac{1}{2}$.

For $1 \leqslant i \leqslant f$ and $y_i \in V(\Gamma)$, define $U_i^{y_i} := \phi^{-1}(y_i) \cap V_i$. Also, for $1 \leqslant i < j \leqslant f$ and $y_i, y_j \in V(\Gamma)$, define $\delta_{i,j}^{y_i,y_j}$ to be the proportion of the pairs in $U_i^{y_i} \times U_j^{y_j}$ which are bad edges.

Claim 4.3. *For every (not necessarily distinct) $y_1, \ldots, y_f$ with $y_i \in R_i$ $(i = 1, \ldots, f)$, there exists an edge $ij \in E(F)$ such that $\delta_{i,j}^{y_i,y_j} \geqslant H^{-1}(h)^2$, where $h := \log(\frac{\Delta}{\Delta-1}) - 4fc$.*

Proof. As Γ is F-hom-free, there is an edge $ij \in E(F)$ such that $y_i y_j \notin E(\Gamma)$ (otherwise the map $i \mapsto y_i$ would be a homomorphism from F to Γ). Without loss of generality suppose that $i = 1, j = 2$, so that $v_1 v_2$ is an edge of F_k and $y_1 y_2 \notin E(\Gamma)$. Recall that each edge uv of F_k receives two labels of the form $(u, a), (v, b)$ (see Construction 1.13). So let $(1, a), (2, b)$ be the labels of the edge $v_1 v_2$.

Let us denote by X_i' the random variable X_i conditioned on $Y_i = y_i$. As X_1, X_2 are independent, so are X_1', X_2'. Also, for each $\ell \in \{1, \ldots, \Delta\}$, X_i' takes value ℓ with probability

$$\frac{|\{(v_i, \mathbf{x}) \in U_i^{y_i} : \mathbf{x}_k = \ell\}|}{|U_i^{y_i}|}.$$

Since $y_i \in R_i$, we have $H(X_i') \geqslant \log(\Delta) - \beta$ for $i = 1, 2$, where $\beta := 4fc$. By Lemma 4.1, this implies that $\mathbb{P}[X_1' = a, X_2' = b] = \mathbb{P}[X_1' = a] \cdot \mathbb{P}[X_2' = b] \geqslant H^{-1}(h)^2$ where $h = \log(\frac{\Delta}{\Delta-1}) - \beta$. But by the definition of $G^\star$, if $X_1 = a$ and $X_2 = b$ then the vertices $(v_1, \mathbf{x}^{v_1})$ and $(v_2, \mathbf{x}^{v_2})$ are adjacent in $G^\star$. It follows that at least an $H^{-1}(h)^2$-fraction of the pairs in $U_1^{y_1} \times U_2^{y_2}$ are adjacent and mapped to the non-edge $y_1 y_2$ of Γ. This proves the claim. $\qquad\square$

By Claim 4.3, we have

$$\mathbb{E}_{y_1,\ldots,y_f}\left[\sum_{ij \in E(F)} \delta_{i,j}^{y_i,y_j}\right] \geqslant \mathbb{P}[Y_i \in R_i \text{ for all } i] \cdot H^{-1}(h)^2 \geqslant \frac{1}{2}H^{-1}(h)^2. \tag{10}$$

On the other hand,

$$\mathbb{E}_{y_1,\ldots,y_f}\big[\delta_{i,j}^{y_i,y_j}\big] = \mathbb{E}_{y_i,y_j}\big[\delta_{i,j}^{y_i,y_j}\big], \tag{11}$$

which equals the proportion of pairs in $V_i \times V_j$ which are bad edges. Note that $|V_1| = \cdots = |V_f| = \frac{|G^\star|}{n}$. Combining (11) with (10) and using linearity of expectation, we see that $F_k^\star$ contains at least $c'(\frac{|G^\star|}{n})^2$ bad edges, where

$$c' := \frac{1}{2}H^{-1}(h)^2 = \frac{1}{2}H^{-1}\left(\log\left(\frac{\Delta}{\Delta-1}\right) - 4fc\right)^2.$$

Finally, observe that if c is sufficiently small with respect to F, then $c' > 2c$. Indeed, c' tends to the positive constant $\frac{1}{2}H^{-1}(\log(\frac{\Delta}{\Delta-1}))$ as $c \to 0$, whereas $2c \to 0$ as $c \to 0$, implying that

indeed $c' > 2c$ for c sufficiently small. This completes the proof of Claim 4.2 and hence of the lemma. □

5. Approximate homomorphisms: lower bounds

5.1. **Proof of Theorem 1.15.** We need a well-known fact about hypergraphs of large girth. Recall that a *Berge cycle* in a hypergraph is a sequence of distinct vertices $v_1, \ldots, v_k$ and distinct edges $e_1, \ldots, e_k$ ($k \geqslant 2$), such that $v_i, v_{i+1} \in e_i$ for every $1 \leqslant i \leqslant k$, with indices taken modulo k. A hypergraph is a *hyperforest* if it can be obtained from an empty hypergraph by repeatedly adding an edge which intersects the current hypergraph in at most one vertex, with all other vertices being new. The following fact is well-known, but we include a proof in Appendix A for completeness.

Lemma 5.1. *A hypergraph $\mathcal{G}$ is a hyperforest if and only if it has no Berge cycles.*

The following lemma follows from a standard probabilistic deletion argument. Again, for completeness, we include a proof in Appendix A.

Lemma 5.2. *For every $f, g \geqslant 2$ and every $n \geqslant f$, there is an n-vertex f-uniform, f-partite hypergraph with $\Omega(n^{1+1/g})$ edges and no Berge cycle of length at most g.*

With these preliminaries as well as Lemma 1.14, we are ready to prove Theorem 1.15.

Proof of Theorem 1.15. Set $f := |F|$, $h := |H|$. Let $\mathcal{G}$ be the hypergraph given by Lemma 5.2 with parameters $g := \max(f, h)$ and $n := c_0 \cdot (\frac{1}{\varepsilon})^{\frac{1}{1-1/g}}$, where $c_0 > 0$ is a small enough constant to be chosen later. Let G be the graph obtained from $\mathcal{G}$ by identifying the parts of $\mathcal{G}$ with the vertices of F, and accordingly placing a copy of F on each hyperedge of $\mathcal{G}$, so that G is a subgraph of a blowup of F. Letting m denote the number of F-copies in G, we have $m \geqslant e(\mathcal{G}) = \Omega(n^{1+1/g})$. Let $c = c(F)$ be the constant given by Lemma 1.14. Note that

$$\frac{cm}{n^2} = \frac{\Omega(n^{1+1/g})}{n^2} = \Omega(n^{-1+1/g}) = \Omega(c_0^{-1+1/g} \cdot \varepsilon) > \varepsilon,$$

provided that c_0 is small enough. Also, by definition, every edge of G is on a copy of F.

Consider any set $X \subseteq V(G) = V(\mathcal{G})$ with $|X| \leqslant g$, and let $e_1, \ldots, e_s$ be the hyperedges $e \in E(\mathcal{G})$ satisfying $|e \cap X| \geqslant 2$. Consider the hypergraph $\mathcal{G}_X = \{e_i \cap X : 1 \leqslant i \leqslant s\}$. Since $|X| \leqslant g$ and $\mathcal{G}$ has no Berge cycle of length at most g, we get that $\mathcal{G}_X$ has no Berge cycles. Therefore, $\mathcal{G}_X$ is a hyperforest by Lemma 5.1. In the case $|X| = f$, the graph $G[X]$ can contain a copy of F only if $\mathcal{G}_X$ has a hyperedge of size f, because F is 2-connected. Thus, the only F-copies in G are those corresponding to hyperedges of $\mathcal{G}$. As $\mathcal{G}$ has no Berge 2-cycles, it follows that every edge of G is on a unique copy of F. This in turn allows us to apply Construction 1.13.

So let $G^\star$ be the graph given by Construction 1.13. By Lemma 1.14, there is no ε-approximate homomorphism from $G^\star$ to any F-hom-free graph on at most $2^{cm/n}$ vertices. Note that $2^{cm/n} = 2^{\Omega(n^{1/g})} \geqslant 2^{(1/\varepsilon)^{1/g}}$ (provided that ε is small enough).

22

It remains to show that $G^\star$ is H-hom-free. So suppose for contradiction that there is a homomorphism $\psi : H \to G^\star$. For a vertex $u \in V(G)$, let $V_u := \{u\} \times \{1, \ldots, \Delta\}^m$ denote the blowup set of u in $G^\star$ (recall Construction 1.13). Let X be the set of all $u \in V(G)$ such that $V_u \cap \psi(V(H)) \neq \varnothing$; so $|X| \leqslant |H| = h \leqslant g$. Let $F_1, \ldots, F_m$ be an enumeration of the F-copies in G. We consider the F-copies in G which intersect X in at least 2 vertices; without loss of generality, these are $F_1, \ldots, F_s$. As we saw above, the hypergraph $\{V(F_i) \cap X : 1 \leqslant i \leqslant s\}$ is a hyperforest. Let T be the F-forest defined as follows: T has s copies of F, denoted $F_1', \ldots, F_s'$; $V(T)$ consists of the set X and the sets $V(F_i') \setminus X$ (for $i = 1, \ldots, s$), which are pairwise-disjoint; and $T[V(F_i') \cap X] = G[V(F_i) \cap X]$ as labelled graphs.[13] Then T is homomorphic to G via the homomorphism which maps F_i' isomorphically to F_i for every $i \in [s]$. As G is homomorphic to F, we get that T is homomorphic to F.

Let $T^\star$ be the graph obtained by applying Construction 1.13 to T. For a vertex $(v, \mathbf{x}) \in V(G^\star)$, let $\phi(v, \mathbf{x})$ be the result of projecting $\mathbf{x}$ on the first s coordinates, i.e., if $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$, then $\phi(v, \mathbf{x}) = (v, (\mathbf{x}_1, \ldots, \mathbf{x}_s))$. If $\mathbf{v} \in \psi(V(H))$ then $v \in X$, so $\phi(v, \mathbf{x}) \in V(T^\star)$. We claim that $(\phi \circ \psi) : V(H) \to V(T^\star)$ is a homomorphism. Indeed, let u_0, v_0 be adjacent vertices in H, and let $(u, \mathbf{x}) = \psi(u_0)$ and $(v, \mathbf{y}) = \psi(v_0)$. These two vertices are adjacent in $G^\star$ since ψ is a homomorphism. Hence, $uv \in E(G)$ by the definition of $G^\star$. Also, $u, v \in X$, so there exists $1 \leqslant k \leqslant s$ such that $uv \in E(F_k)$. Let $(u, i), (v, j)$ be the labels of the edge uv in F_k. By the definition of $G^\star$, the fact that $(u, \mathbf{x})$ and $(v, \mathbf{y})$ are adjacent means that $\mathbf{x}_k = i$ and $\mathbf{y}_k = j$. But $T^\star$ has the same adjacency criterion, meaning that $\phi(u, \mathbf{x})$ and $\phi(v, \mathbf{y})$ are adjacent in $T^\star$. Since this holds for arbitrary $u_0 v_0 \in E(H)$, we conclude that $\phi \circ \psi$ is indeed a homomorphism $H \to T^\star$. Thus, H is homomorphic to $T^\star$ for an F-forest T with $T \to F$. This contradicts the assumption of the theorem and concludes the proof. $\qquad\square$

We remark that by blowing up the graph $G^\star$ constructed in the above proof, we can obtain arbitrarily large H-hom-free graphs which do not have an ε-approximate homomorphism to an F-hom-free graph on at most $2^{(1/\varepsilon)^c}$ vertices.

Proof of Corollary 1.16. By Theorem 1.15, it is enough to prove that C_5, C_7 are not homomorphic to $T^\star$ for any K_3-forest T. So fix a K_3-forest T, and let us show that $T^\star$ has no odd cycles of length at most 7. This suffices because the homomorphic image of an odd cycle must contain an odd cycle. In this proof, we will use the notation introduced in Construction 1.13. Let $F_1, \ldots, F_m$ be an enumeration of the triangles in T. As before, for $u \in V(T)$, we denote by $V_u := \{u\} \times \{1, 2\}^m$ the corresponding blowup-set in $T^\star$. Let $\phi : V(T^\star) \to V(T)$ be the homomorphism sending V_u to u for every $u \in V(T)$. Let C be an odd cycle in $T^\star$. The image $\phi(C)$ must contain an odd cycle, and this cycle must be a triangle because the only odd cycles in a K_3-forest are triangles. So there are three vertices of C which are mapped by ϕ to a triangle

---

[13]Namely, the sets $V(F_i) \setminus X$ may intersect (in G), but we make these sets disjoint in T to make sure that T is an F-forest. We could avoid this technicality by increasing g to $\binom{h}{2}(f - 2)$, which would ensure that $G[\bigcup_{i=1}^s V(F_i)]$ is an F-forest.

u, v, w of T. Let us denote these three vertices by $\mathbf{u} = (u, \mathbf{x}), \mathbf{v} = (v, \mathbf{y}), \mathbf{w} = (w, \mathbf{z})$. Without loss of generality, $F_1 = (u, v, w)$. Observe that no two vertices among $\mathbf{u}, \mathbf{v}, \mathbf{w}$ have a path of length 2 (in $T^\star$) between them. Indeed, if such a path goes outside $V_u \cup V_v \cup V_w$ then it certainly has length more than 2, and there is no such path inside $V_u \cup V_v \cup V_w$ (see Figure 1). Also, we claim that if (say) $\mathbf{u}, \mathbf{v}$ are adjacent, then the distance of $\mathbf{w}$ to each of $\mathbf{u}, \mathbf{v}$ is at least 3, and moreover, the distance of $\mathbf{w}$ to one of these vertices is at least 4. To see this, suppose (without loss of generality) that the edge uv gets labels $(u, 1), (v, 2)$, the edge vw gets labels $(v, 1), (w, 2)$, and the edge wu gets labels $(w, 1), (u, 2)$. Then $\mathbf{x}_1 = 1, \mathbf{y}_1 = 2$. Now, $\mathbf{w}$ is not adjacent to $\mathbf{u}$ or $\mathbf{v}$ (because this would require $\mathbf{x}_1 = 2$ or $\mathbf{y}_1 = 1$, respectively), and we already saw that there is no path of length two from $\mathbf{w}$ to $\mathbf{u}$ or $\mathbf{v}$, so $\mathbf{w}$'s distance to $\mathbf{u}$ and $\mathbf{v}$ is at least 3. Now suppose without loss of generality that $\mathbf{z}_1 = 1$. We claim that $\mathrm{dist}(\mathbf{v}, \mathbf{w}) \geqslant 4$. Letting $V_u^2 := \{(u, \mathbf{x}') : \mathbf{x}'_1 = 2\}$ and $V_v^1 := \{(v, \mathbf{y}') : \mathbf{y}'_1 = 1\}$, observe that the set $S := V_u^2 \cup V_v^1$ separates $\mathbf{w}$ from $\mathbf{v}$, because S contains all vertices in $V_u \cup V_v$ which have a neighbour in V_w. Also, every vertex in V_v^1 is at distance at least 3 from $\mathbf{w}$ and at least 2 from $\mathbf{v}$, and every vertex in V_u^2 is at distance at least 1 from $\mathbf{w}$ and at least 3 from $\mathbf{v}$. It follows that $\mathrm{dist}(\mathbf{v}, \mathbf{w}) \geqslant 4$.

Summarising, either every pair among $\mathbf{u}, \mathbf{v}, \mathbf{w}$ is at distance at least 3, or two of these vertices are adjacent and the last vertex is at distance at least 3 from both of them and at least 4 from one of them. In either case $|C| \geqslant 9$ or $|C| \geqslant 1 + 3 + 4 = 8$, so $|C| > 7$. □

5.2. A polynomial lower bound.

In this section, we complete the picture in Theorem 1.9 by proving the following simple lower bound on $M_{F,H}(\varepsilon)$.

Proposition 5.3. *Let F, H be graphs with $H \to F$ such that H is not bipartite. We have $M_{F,H}(\varepsilon) \geqslant \Omega((\frac{1}{\varepsilon})^{1/2+\alpha})$, where $\alpha > 0$ depends only on F, H.*

Proof. Let g be the shortest length of an odd cycle in H, let $n = \frac{1}{\varepsilon}$, and let $p = n^{-1+1/g}/(4g)$. Let G be an n-vertex random graph with edge probability p. For every $3 \leqslant k \leqslant g$, the expected number of k-cycles in G is at most $p^k n^k \leqslant n/(4g)$. Hence, the total number of cycles of length at most g is, in expectation, at most $n/4$. By Markov's inequality, we conclude that G has at most $n/2$ cycles of length at most g with probability at least $1/2$. So by deleting one vertex from every cycle of length at most g, we obtain a subgraph G' with at least $n/2$ vertices and with girth at least $g + 1$. Then G' is H-hom-free, because every homomorphic image of H contains an odd cycle of length at most g.

Hence, it suffices to show that with positive probability, no induced subgraph of G on $n/2$ vertices has an ε-approximate homomorphism to an F-hom-free graph on k vertices, where $k := c\sqrt{p}n = \Omega(n^{\frac{1}{2}+\frac{1}{2g}}) \geqslant \Omega((\frac{1}{\varepsilon})^{\frac{1}{2}+\frac{1}{2g}})$, where $c > 0$ is a small enough constant. So fix $U \subseteq V(G)$ of size $|U| = n/2$, fix an F-hom-free graph Γ_0 on k vertices, and let $\phi : U \to V(\Gamma_0)$ be a map. Let G' be the graph on U consisting of all pairs $uv \in \binom{U}{2}$ with $\phi(u)\phi(v) \in E(\Gamma_0)$. Then ϕ is a homomorphism from G' to Γ_0. Hence, G' is F-hom-free because Γ_0 is. In particular, G' is

K_f-free, so by Turán's theorem, $e(G') \leqslant \frac{f-2}{f-1} \frac{|U|^2}{2}$. It follows that there are at least

$$\binom{|U|}{2} - \frac{f-2}{f-1} \frac{|U|^2}{2} = \frac{|U|^2}{2(f-1)} - \frac{|U|}{2} \geqslant \frac{|U|^2}{2f} = \frac{n^2}{8f}$$

pairs $uv \in \binom{U}{2}$ which are not mapped by ϕ to an edge of G_0. Let Z be the number of these pairs which are edges in G, so that Z stochastically dominates $\mathrm{Bin}(\frac{n^2}{8f}, p)$. By the Chernoff bound,

$$\mathbb{P}\left[Z \leqslant \frac{1}{2} \mathbb{E}[Z] \right] \leqslant e^{-\frac{\mathbb{E}[Z]}{8}} \leqslant e^{-\frac{pn^2}{64f}} = e^{-\Omega(n^{1+1/g})}.$$

There are at most 2^n choices for U, at most $2^{\binom{k}{2}} \leqslant 2^{c^2 pn^2}$ choices for G_0, and at most $k^n \leqslant n^n = e^{n \log n}$ choices for ϕ. So if c is small enough, then by the union bound, with high probability we have $Z \geqslant \Omega(n^{1+1/g}) > \varepsilon n^2$ for every choice of U, G_0, ϕ. If this happens, then no induced subgraph of G on $n/2$ vertices has an ε-approximate homomorphism to an F-free graph on k vertices, as required. $\qquad\square$

We remark that by taking blowups of G' (constructed in the proof of the proposition), one can in fact find a graph witnessing this lower bound on $M_{F,H}(\varepsilon)$ with any number of vertices. Note that the upper and lower bounds on $M_{K_3, C_\ell}(\varepsilon)$ for odd $\ell \geqslant 9$ are off by a factor of roughly $\varepsilon^{1/2}$, and it would be interesting to close this gap.

6. Concluding remarks

There remain a number of fascinating open problems related to asymmetric questions about graph homomorphisms. In particular, we again reiterate Conjectures 1.4 and 1.17; the first claims that asymmetric homomorphism thresholds of odd cycles are already zero for $\delta_{\mathrm{hom}}(C_{2t+3}; C_{2t+1})$, and the second would give a complete characterization of when one can get an exponential lower bound on $M_{F,H}(\varepsilon)$, showing that Construction 1.13 is universal for such lower bounds.

While excluding *all* odd cycles up to a certain length is very natural, the work of Sankar [32] suggests that fascinating structure may arise if one studies, for example, $\delta_{\mathrm{hom}}(C_5; C_3)$ or $\delta_{\mathrm{hom}}(C_7; C_5)$. In particular, we would be very interested to learn whether Sankar's topological techniques can be used to give positive lower bounds on these quantities.

When it comes to asymmetric approximate homomorphisms, it would be interesting to tighten the lower and upper bounds appearing in Theorem 1.9. While we obtain polynomial upper and lower bounds for $M_{K_3, C_\ell}(\varepsilon)$ for $\ell \geqslant 9$, and exponential upper and lower bounds for $M_{K_3, C_\ell}(\varepsilon)$ for $\ell \in \{5, 7\}$, these bounds do not agree on the correct exponent for ε.

References

[1] N. Alon, Testing subgraphs in large graphs, *Random Structures Algorithms* **21** (2002), 359–370. 5

[2] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, The algorithmic aspects of the regularity lemma, *J. Algorithms* **16** (1994), 80–109. 3, 5

[3] B. Andrásfai, Graphentheoretische Extremalprobleme, *Acta Math. Acad. Sci. Hungar.* **15** (1964), 413–438. 2

[4] B. Andrásfai, P. Erdős, and V. T. Sós, On the connection between chromatic number, maximal clique and minimal degree of a graph, *Discrete Math.* **8** (1974), 205–218. 2

[5] S. Brandt and S. Thomassé, Dense triangle-free graphs are four-colorable: A solution to the Erdős–Simonovits problem, 2011. Preprint available at `https://perso.ens-lyon.fr/stephan.thomasse/liste/vega11.pdf`. 2

[6] D. Conlon and J. Fox, Graph removal lemmas, in *Surveys in combinatorics* 2013, *London Math. Soc. Lecture Note Ser.*, vol. 409, Cambridge Univ. Press, Cambridge, 2013, 1–49. 5

[7] B. Csaba, Regular decomposition of the edge set of a graph with applications, *Australas. J. Combin.* **89** (2024), 249–267. 7

[8] O. Ebsen and M. Schacht, Homomorphism thresholds for odd cycles, *Combinatorica* **40** (2020), 39–62. 4

[9] P. Erdős and M. Simonovits, On a valence problem in extremal graph theory, *Discrete Math.* **5** (1973), 323–334. 2

[10] J. Fox, A new proof of the graph removal lemma, *Ann. of Math.* (2) **174** (2011), 561–579. 5

[11] J. Fox and Y. Wigderson, Minimum degree and the graph removal lemma, *J. Graph Theory* **102** (2023), 648–665. 3

[12] J. Fox and Y. Zhao, Removal lemmas and approximate homomorphisms, *Combin. Probab. Comput.* **31** (2022), 721–736. 3, 5, 6, 8, 9, 18

[13] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220. 14

[14] Z. Füredi, Extremal hypergraphs and combinatorial geometry, in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, Birkhäuser, Basel, 1995, 1343–1352. 3, 5

[15] D. Galvin, Three tutorial lectures on entropy and counting, 2014. Preprint available at arXiv:1406.7872. 18

[16] A. Girão, E. Hurley, F. Illingworth, and L. Michel, Abundance: asymmetric graph removal lemmas and integer solutions to linear equations, *J. Lond. Math. Soc.* (2) **110** (2024), Paper No. e70015, 26pp. 7

[17] L. Gishboliner, Z. Jin, and B. Sudakov, The minimum degree removal lemma thresholds, *J. Combin. Theory Ser. B* **166** (2024), 203–221. 3

[18] L. Gishboliner, A. Shapira, and Y. Wigderson, An efficient asymmetric removal lemma and its limitations, *Forum Math. Sigma* **13** (2025), Paper No. e38, 17pp. 7

[19] R. Häggkvist, Odd cycles of specified length in nonbipartite graphs, in *Graph theory* (*Cambridge*, 1981), *North-Holland Math. Stud.*, vol. 62, North-Holland, Amsterdam-New York, 1982, 89–99. 2

[20] D. Haussler and E. Welzl, ϵ-nets and simplex range queries, *Discrete Comput. Geom.* **2** (1987), 127–151. 14

[21] P. Hell and J. Nešetřil, The core of a graph, *Discrete Math.* **109** (1992), 117–126. 9

[22] C. Hoppen, Y. Kohayakawa, R. Lang, H. Lefmann, and H. Stagni, Estimating parameters associated with monotone properties, *Combin. Probab. Comput.* **29** (2020), 616–632. 5, 7

[23] X. Huang, H. Liu, M. Rong, and Z. Xu, Interpolating chromatic and homomorphism thresholds, 2025. Preprint available at arXiv:2502.09576. 4, 5

[24] G. P. Jin, Triangle-free graphs with high minimal degrees, *Combin. Probab. Comput.* **2** (1993), 479–490. 2

[25] S. Letzter and R. Snyder, The homomorphism threshold of $\{C_3, C_5\}$-free graphs, *J. Graph Theory* **90** (2019), 83–106. 4

[26] L. Lovász, *Large networks and graph limits*, *American Mathematical Society Colloquium Publications*, vol. 60, American Mathematical Society, Providence, RI, 2012. 15

[27] T. Łuczak, On the structure of triangle-free graphs of large minimum degree, *Combinatorica* **26** (2006), 489–493. 2

[28] G. Moshkovitz and A. Shapira, A sparse regular approximation lemma, *Trans. Amer. Math. Soc.* **371** (2019), 6779–6814. 5

[29] D. Mubayi and J. Solymosi, Many pentagons in triple systems, 2025. Preprint available at arXiv:2501.15861. 7

[30] J. Mycielski, Sur le coloriage des graphs, *Colloq. Math.* **3** (1955), 161–162. 10, 11

[31] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in *Combinatorics* (*Proc. Fifth Hungarian Colloq.*, *Keszthely*, 1976), *Vol. II*, *Colloq. Math. Soc. János Bolyai*, vol. 18, North-Holland, Amsterdam-New York, 1978, 939–945. 3, 5

[32] M. Sankar, Homotopy and the homomorphism threshold of odd cycles, 2022. Preprint available at arXiv:2206.07525. 4, 25

[33] M. Stiebitz, *Beiträge zur Theorie der färbungskritischen Graphen*, Ph.D. thesis, TU Ilmenau, 1985. 10

[34] E. Szemerédi, Regular partitions of graphs, in *Problèmes combinatoires et théorie des graphes* (*Colloq. Internat. CNRS*, *Univ. Orsay*, *Orsay*, 1976), *Colloq. Internat. CNRS*, vol. 260, CNRS, Paris, 1978, 399–401. 5

[35] T. Tao, A variant of the hypergraph removal lemma, *J. Combin. Theory Ser. A* **113** (2006), 1257–1280. 2

[36] C. Thomassen, On the chromatic number of triangle-free graphs of large minimum degree, *Combinatorica* **22** (2002), 591–596. 2

[37] C. Thomassen, On the chromatic number of pentagon-free graphs of large minimum degree, *Combinatorica* **27** (2007), 241–243. 4

APPENDIX A. LEMMAS FROM SECTION 5

Proof of Lemma 5.1. The "only if" direction is easy to prove by induction on the number of edges. We prove the "if" direction. The proof is by induction on $e(\mathcal{G})$, and the base case $e(\mathcal{G}) = 0$ is trivial, so suppose that $e(\mathcal{G}) \geqslant 1$. Consider the bipartite graph with sides $A = V(\mathcal{G})$ and $B = E(\mathcal{G})$ where v and e are adjacent if $v \in e$. A cycle in this graph gives a Berge cycle in $\mathcal{G}$, so this graph is a forest. Note that each $e \in B$ has degree $|e| \geqslant 2$ in the auxiliary graph, so a path ending in $e \in B$ can always be extended. We claim that there is $e \in B$ which has at most one non-leaf neighbour. Indeed, take a longest path P starting in A, let $v \in A$ be the first vertex of P, and let $e \in B$ be the neighbour of v on P. Then by the maximality of P, e has at most one non-leaf neighbour. This means that in $\mathcal{G}$, e has at most one vertex which is contained in another edge of $\mathcal{G}$. By induction, $\mathcal{G} - e$ is a hyperforest. Adding back e preserves this property. $\qquad\square$

Proof of Lemma 5.2. Consider a random f-uniform, f-partite hypergraph with parts of size n/f and edge probability $p = cn^{-f+1+1/g}$, where $c > 0$ is a small constant to be chosen later. The expected number of edges is $p(n/f)^f = \Omega(n^{1+1/g})$. For each $2 \leqslant k \leqslant g$, the expected number of Berge cycles of length k is at most $n^k n^{(f-2)k} p^k = n^{(f-1)k} p^k \leqslant \frac{1}{2g} \cdot p(n/f)^f$, using our choice of p and provided that c is a small enough constant. So the expected number of Berge cycles of length at most g is at most $\frac{1}{2}p(n/f)^f$. Deleting one edge from each Berge cycle of length at most g gives the result. $\qquad\square$

28