Removal Lemmas: Summer School 2025

3 The induced removal lemma

It is natural to consider a variant of the removal lemma for induced subgraphs. In this case, we allow both adding and removing edges, since adding edges may also be useful in order to make a graph induced H-free. Thus, an n-vertex graph G is said to be ε*-far* from induced H-free if one has to add/delete at least εn^2 edges in order to make G induced H-free. The following is the induced analogue of the removal lemma:

Theorem 3.1 (Induced removal lemma, Alon-Fischer-Krivelevich-Szegedy 2000)**.** *Let H be a fixed graph. For every $\varepsilon > 0$ there is $\delta = \delta_H(\varepsilon) > 0$ such that if an n-vertex graph G is ε-far from induced H-free, then G contains at least $\delta n^{v(H)}$ induced copies of H.*

The proof of Theorem 3.1 is significantly more complicated than that of Theorem 1.1. The natural approach is to take a regular partition and try to clean it, arguing that if after the cleaning there still remains an induced copy of H, then before the cleaning there are many such copies. For the cleaning, it is natural to delete all possible edges between V_i, V_j if $d(V_i, V_j)$ is close to 0, and to add all possible edges between V_i, V_j if $d(V_i, V_j)$ is close to 1. However, it is not clear how to handle the non-regular pairs (V_i, V_j) and the edges inside the sets V_i. Whereas in the non-induced case we could simply delete all such edges and thus make sure that any remaining H-copy uses none of them, it is not clear how to proceed in the induced case. We will need a more involved "regularity scheme". Let us now describe the structure that we need in order to prove Theorem 3.1. We only present the main ideas and avoid many of the details.

Regularity scheme for induced removal

The proof of Theorem 3.1 proceeds by finding an ε-regular partition $V_1, \dots, V_t$ of G, and disjoint $U_{i,1}, \dots, U_{i,h} \subseteq V_i$, where $h = |V(H)|$ (say), such that the following holds:

1. **All** pairs $(U_{i,k}, U_{j,\ell})$ for $(i,k) \neq (j,\ell)$ are ε-regular.
2. For every $1 \leq i < j \leq t$, all pairs $(U_{i,k}, U_{j,\ell})$ for $k, \ell \in [h]$ have the same density, up to ε.
3. For every $1 \leq i \leq t$, either $d(U_{i,k}, U_{i,\ell}) \geq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$, or $d(U_{i,k}, U_{i,\ell}) \leq \frac{1}{2}$ for all $1 \leq k < \ell \leq h$.
4. For all but εt^2 of the pairs $1 \leq i < j \leq t$, it holds that $|d(V_i, V_j) - d(U_{i,k}, U_{j,\ell})| \leq \varepsilon$ for all $k, \ell \in [h]$.

Cleaning the graph consists of the following:

(a) For every $1 \leq i < j \leq t$, if $d(U_{i,k}, U_{j,\ell}) \geq 1 - 2\varepsilon$ for all $k, \ell \in [h]$, then make (V_i, V_j) complete, and if $d(U_{i,k}, U_{j,\ell}) \leq 2\varepsilon$ for all $k, \ell \in [h]$, then make (V_i, V_j) empty. Else, make no changes.

(b) For every $1 \le i \le t$, if $d(U_{i,k}, U_{i,\ell}) \ge \frac{1}{2}$ for all $1 \le k < \ell \le h$ then make V_i a clique, and if $d(U_{i,k}, U_{i,\ell}) \le \frac{1}{2}$ for all $1 \le k < \ell \le h$ then make V_i an independent set.

Note that in Item (a), if we make no changes between (V_i, V_j) then $\varepsilon \le d(U_{i,k}, U_{j,\ell}) \le 1 - \varepsilon$ for all $k, \ell \in [h]$, meaning that we can embed both edges and non-edges in these pairs $(U_{i,k}, U_{j,\ell})$.

The sets $U_{i,k}$ are sometimes called *representatives* for V_i; the changes between V_i and V_j (or within V_i) are made according to their representatives $(U_{i,k}, U_{j,\ell})$. Items 2-3 require that these representatives are consistent. Item 4 is meant to ensure that the number of changes made when cleaning the graph (in Item (a)) is small. The proof proceeds by showing that if there is a copy of H in the cleaned graph, say (without loss of generality) between some sets $V_1, \dots, V_r$ and using a_i vertices from V_i for each $i \in [r]$, then we can find many H-copies in the original graph by taking a_i representative sets $U_{i,k}$ from V_i for each $1 \le i \le k$.

Very roughly speaking, the structure described in Items 1-4 is found as follows: First, apply the regularity lemma to find the partition $\mathcal{P} = \{V_1, \dots, V_t\}$. Then apply the regularity lemma again with a much smaller regularity parameter (which depends on t; much smaller than $1/t$ in fact) to find a partition $\mathcal{Q}$ refining $\mathcal{P}$, and sample a set $U_i \subseteq V_i$ randomly. With high probability, all pairs (U_i, U_j) are highly regular. Now apply the regularity lemma on $G[U_i]$ (this time with parameter ε again) to partition U_i into sets $U_{i,k}$, and apply Ramsey's theorem (preceded by Turán's theorem) on these sets to find $U_{i,1}, \dots, U_{i,h}$ satisfying Item 3. Items 1-2 hold because all pairs (U_i, U_j) are highly regular.

To satisfy Item 4, more is required, and in fact the above presentation is somewhat misleading: One applies the regularity lemma not just twice but repeatedly, obtaining a sequence of partitions $\mathcal{P}_i$ such that $\mathcal{P}_{i+1}$ refines $\mathcal{P}_i$ and is regular with a parameter appropriately defined in terms of $|\mathcal{P}_i|$. One stops when $q(\mathcal{P}_{i+1}) \le q(\mathcal{P}_i) + \varepsilon$, where $q(\cdot)$ is the mean square density. It is then possible to show that Item 4 holds. This argument proves the so-called *strong regularity lemma*, which we shall not go into. In the following section, we will see in more detail a variant of the above regularity scheme.

The infinite removal lemma

It is natural to ask for an analogue of the removal lemma for families of forbidden (induced) subgraphs. The more general result of this type was obtained by Alon and Shapira, and applies to *any* (possibly infinite) graph family.

Theorem 3.2 (Infinite removal lemma, Alon-Shapira 2005)**.** *Let $\mathcal{H}$ be a (possibly infinite) family of graphs. For every $\varepsilon > 0$ there exist $\delta = \delta_{\mathcal{H}}(\varepsilon) > 0$ and $m = m_{\mathcal{H}}(\varepsilon) \ge 1$ such that if an n-vertex graph G is ε-far from induced $\mathcal{H}$-freeness, then there is $H \in \mathcal{H}$ with $|V(H)| \le m$, such that G contains at least $\delta n^{v(H)}$ copies of H.*

Polynomial removal lemmas

A key question is to characterize the graph-families $\mathcal{H}$ for which $\delta_{\mathcal{H}}(\varepsilon)$ and $m_{\mathcal{H}}(\varepsilon)$ in Theorem 3.2 depend polynomially on ε (more precisely, on ε and $1/\varepsilon$, respectively).

Problem 3.3. *For which graph-families $\mathcal{H}$ is the induced $\mathcal{H}$-free removal lemma polynomial?*

A special case of this problem (for the property of not-necessarily-induced H-freeness) was handled in Section 2. Returning to the induced H-removal lemma for a single graph H, works of Alon-Shapira and Alon-Fox give the following almost complete characterization. We use P_k (resp. C_k) to denote the path (resp. cycle) with k vertices.

Theorem 3.4 (Alon-Shapira 2006, Alon-Fox 2015)**.**

1. *If* $H \in \{P_2, \overline{P_2}, P_3, \overline{P_3}, P_4\}$ *then the induced* H*-removal lemma is polynomial.*
2. *If* $H \notin \{P_2, \overline{P_2}, P_3, \overline{P_3}, P_4, C_4, \overline{C_4}\}$ *then the induced-*H *removal lemma is not polynomial.*

The only remaining case is $H = C_4$.[1] This case remains open, but an exponential bound is known:

Theorem 3.5 (Gishboliner-Shapira 2019)**.** *For the induced-*C_4 *removal lemma, we have*

$$\delta_{C_4}(\varepsilon) \geq 2^{-poly(1/\varepsilon)}.$$

Conjecture 3.6. *The induced-*C_4 *removal lemma is polynomial.*

4 VC-dimension and ultra-strong regularity

We begin by recalling the definition of VC-dimension, and then discuss its connection to graphs and regularity.

Definition 4.1 (Shattered set, VC-dimension)**.** *Let* $\mathcal{F}$ *be a family of subsets of a set* V*. A set* $S \subseteq V$ *is* shattered *by* $\mathcal{F}$ *if for every* $T \subseteq S$*, there exists* $F \in \mathcal{F}$ *with* $S \cap F = T$*. The* VC-dimension *of* $\mathcal{F}$ *is the maximum size of a shattered set.*

VC-dimension is a fundamental measure of complexity used in combinatorics and computer science. One of the basic facts about VC-dimension is the so-called Sauer-Shelah lemma, stating the following:

Theorem 4.2 (Sauer-Shelah lemma)**.** *Let* $\mathcal{F} \subseteq 2^{[n]}$ *be a family of subsets of* $[n]$ *with VC-dimension* d*. Then* $|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{n}{i}$*.*

Note that the bound in Theorem 4.2 is tight, because the set family consisting of all sets of size at most d has VC dimension d.

Proof of Theorem 4.2. We prove by induction on n that the number of sets shattered by $\mathcal{F}$ is at least $|\mathcal{F}|$. This suffices, because if $|\mathcal{F}| > \sum_{i=0}^{d} \binom{n}{i}$ then there must exist a shattered set of size larger than $d+1$, and hence the VC-dimension is larger than $d+1$.

Write $\mathcal{F}_0 = \{F \in \mathcal{F} : n \notin F\}$ and $\mathcal{F}_1 = \{F \setminus \{n\} : F \in \mathcal{F}, n \in F\}$. Then $\mathcal{F}_0, \mathcal{F}_1 \subseteq 2^{[n-1]}$, and $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$. By induction, $\mathcal{F}_i$ shatters at least $|\mathcal{F}_i|$ sets for $i = 0, 1$. Every set shattered by $\mathcal{F}_0$ or $\mathcal{F}_1$ is (trivially) shattered by $\mathcal{F}$. Also, it is easy to see that if S is shattered by both $\mathcal{F}_0, \mathcal{F}_1$, then $S \cup \{n\}$ is shattered by $\mathcal{F}$. This allows us to conclude that

$$\#\{\text{sets shattered by } \mathcal{F}\} \geq \#\{\text{sets shattered by } \mathcal{F}_0\} + \#\{\text{sets shattered by } \mathcal{F}_1\} \geq |\mathcal{F}_0| + |\mathcal{F}_1| = |\mathcal{F}|.$$

∎

In applications, the exact bound $\sum_{i=0}^{d} \binom{n}{i}$ from Theorem 4.2 is often not important; the crucial fact is that $|\mathcal{F}|$ is polynomial in n.

To use VC-dimension in the context of graphs, we consider the following set-family: Let G be a graph, and let $\mathcal{F} = \{N(v) : v \in V(G)\}$. The VC-dimension of G is defined as the VC-dimension of

[1] The case H and $\overline{H}$ are equivalent.

the set-family $\mathcal{F}$. What does it mean for a graph to have unbounded VC-dimension? It means that for every fixed $d \geq 1$, there is a set $S = \{x_1, \ldots, x_d\}$ shattered by $\mathcal{F}$. This in turn means that there are vertices $(y_I : I \subseteq [d])$, such that y_I is adjacent to x_i if and only if $i \in I$. In what follows, we will want to assume that $y_I \notin \{x_1, \ldots, x_d\}$ for every I. This can be achieved by taking a slightly larger shattered set $S = \{x_1, \ldots, x_d, x_{d+1}, \ldots, x_{d+k}\}$, where we think of $x_{d+1}, \ldots, x_{d+k}$ as "dummy vertices". Doing this supplies us with 2^k different vertices to play the role of y_I for each $I \subseteq [d]$, so if $2^k > d$ then we can choose such a y_I which is outside $\{x_1, \ldots, x_d\}$.

We now see that if G has unbounded VC-dimension, then it has a bi-induced copy of every fixed-size bipartite graph $H = (A, B)$ (recall Definition 1.6). Indeed, we can construct such a copy by first choosing a shattered set $S = \{x_1, \ldots, x_d\}$ to play the role of A, and then, for each $b \in B$, choosing y_I for I which corresponds to the neighborhood of b in A (namely, if $A = \{a_1, \ldots, a_d\}$, then I is the set of all $i \in [d]$ such that $a_i b \in E(H)$). By taking a slightly larger set S (of size $d + k$ for $2^k \geq |B|$, as above), we can make sure that we have enough vertices y_I with neighborhood $\{x_i : i \in I\}$ in $\{x_1, \ldots, x_d\}$, in case several vertices in B have the same neighborhood in H. We thus conclude the following:

Fact 4.3. *G has unbounded VC-dimension if and only if G contains a bi-induced copies of all fixed-size bipartite graphs.*[2]

The above is of course not a rigorous statement (because of the term "unbounded"), but it should be clear what it means. A rigorous statement would be that if G has VC-dimension at least d_1, then it contains bi-induced copies of all bipartite graphs of size d_2 (for some d_2 growing with d_1), and vice versa.

What can we say about a graph G which avoids bi-induced copies of some fixed bipartite H? Let us apply the regularity lemma to obtain an ε-regular equipartition $V_1, \ldots, V_t$. By Lemma 1.7, all regular pairs (V_i, V_j) have density at most γ or at least $1 - \gamma$, provided that $\varepsilon \ll \gamma$. Thus, there is an equipartition of $V(G)$ where all but γt^2 of the pairs (V_i, V_j) have density at most γ or at least $1 - \gamma$.[3] Such a partition is called *γ-homogeneous*. Thus, bounded VC-dimension implies the existence of γ-homogeneous partitions. However, the partition obtain in this way is very large, of tower-type size in $1/\gamma$. Can we do better? As we will now show, using the Sauer-Shelah lemma we can find a partition of size only polynomial in $1/\gamma$. The key fact we will need is as follows:

Lemma 4.4. *If G has VC-dimension d, then for every $\varepsilon > 0$, there are vertices $x_1, \ldots, x_t$, $t \leq (1/\varepsilon)^{O(d)}$, such that for every $x \in V(G)$ there is $i \in [t]$ with $|N(x) \triangle N(x_i)| \leq \varepsilon n$.*

Proof. Let $x_1, \ldots, x_t$ be a maximum collection of elements such that $|N(x_i) \triangle N(x_j)| > \varepsilon n$ for every $1 \leq i < j \leq t$. It suffices to show that $t < t_0 := (1/\varepsilon)^{d+1}$.[4] Suppose not, and suppose that $t = t_0$ (by disposing of the other x_i's). Sample a subset $U \subseteq V(G)$ of size $|U| = m := \frac{2\log(t)}{\varepsilon} = \tilde{O}(\frac{1}{\varepsilon})$. For $1 \leq i < j \leq t$, the probability that $N(x_i) \cap U = N(x_j) \cap U$ is at most

$$\frac{\binom{(1-\varepsilon)n}{m}}{\binom{n}{m}} \leq (1 - \varepsilon)^m \leq e^{-\varepsilon m} = t^{-2}.$$

[2] We only proved one direction, but the other direction is also easy: Take the $d \times 2^d$ incidence bipartite graph of $[d]$ versus subsets of $[d]$. If G contains a bi-induced copy of this bipartite graph, then its VC-dimension is at least d.

[3] Such a pair is called *γ-homogeneous*. Note that a being homogeneous is a stronger property than being regular: a γ-homogeneous pair is necessarily $\gamma^{1/3}$-regular. This is left as an exercise for the reader.

[4] As one can see from the proof, $(1/\varepsilon)^{d+1}$ can be replaced with $\tilde{O}(1/\varepsilon^d)$.

By the union bound, there is an outcome for U such that $N(x_i) \cap U \neq N(x_j) \cap U$ for every $1 \leq i < j \leq t$. But now $(N(x_i) \cap U : 1 \leq i \leq t)$ is a set system on U of size $t = (1/\varepsilon)^{d+1} > \sum_{i=0}^{d} \binom{|U|}{i}$, so by the Sauer-Shelah lemma (Theorem 4.2), it has VC-dimension larger than d, in contradiction to the assumption that G has VC-dimension d. ■

We will now use Lemma 4.4 to find a small ε-homogeneous equipartition of a graph G with bounded VC-dimension. As far as we know, this result is originally due to Lovász-Szegedy and Alon-Fischer-Newman.

Theorem 4.5 (Lovász-Szegedy 2010, Alon-Fischer-Newman 2007)**.** *If G has VC-dimension d, then it has an ε-homogeneous equipartition of size $(1/\varepsilon)^{O(d)}$.*

Proof. Let $x_1, \dots, x_t \in V(G)$, $t \leq (1/\gamma)^{O(d)}$, be the vertices given by Lemma 4.4, applied with parameter $\gamma = \text{poly}(\varepsilon) \ll \varepsilon$ to be chosen (implicitly) later. For each $i \in [t]$, let X_i be the set of all $x \in V(G)$ such that $|N(x) \triangle N(x_i)| \leq \gamma n$. So $X_1 \cup \cdots \cup X_t = V(G)$ by the guarantees of Lemma 4.4. The idea is to claim that $X_1, \dots, X_t$ is an ε-homogeneous partition. The partition $X_1, \dots, X_t$ is not an equipartition, so we need to adapt the definition of a ε-homogeneous partition to allow parts of different sizes: A partition $X_1, \dots, X_t$ is ε-homogeneous if the sum of $|X_i||X_j|$ over all pairs $1 \leq i, j \leq t$ with $\varepsilon < d(X_i, X_j) < 1 - \varepsilon$ is at most εn^2. This sum includes terms $i = j$.

Let us now show that $\{X_1, \dots, X_t\}$ is ε-homogeneous. Sample vertices $x, y \in V(G)$ uniformly at random and then a vertex x' belonging to the same part X_i as x. Let $\mathcal{A}$ be the event that $xy \in E(G)$ but $x'y \notin E(G)$, or vice versa. In other words, this is the event that $y \in N(x) \triangle N(x')$. We will show that due to the choice of $X_1, \dots, X_t$, $\mathbb{P}[\mathcal{A}] \leq 2\gamma$, and that this implies that $\{X_1, \dots, X_t\}$ is ε-homogeneous. First, condition on the choice of x, x'. Since x, x' are in the same part X_i, we have $|N(x) \triangle N(x')| \leq |N(x) \triangle N(x_i)| + |N(x') \triangle N(x_i)| \leq 2\gamma n$ (by the triangle inequality), so $\mathbb{P}[y \in N(x) \triangle N(x')] \leq 2\gamma$. It follows that $\mathbb{P}[\mathcal{A}] \leq 2\gamma$.

Now suppose by contradiction that $\{X_1, \dots, X_t\}$ is not ε-homogeneous. Fix a pair X_i, X_j with $\varepsilon < d(X_i, X_j) < 1 - \varepsilon$. We need the following claim:

Claim 4.6. *For disjoint vertex-sets U, V, if $\varepsilon < d(U, V) < 1 - \varepsilon$, then there are $\Omega(\varepsilon)|U|^2|V|$ triples $x, x' \in U, y \in V$ or $\Omega(\varepsilon)|V|^2|U|$ triples $x, x' \in V, y \in U$ satisfying $y \in N(x) \triangle N(x')$.*

The claim is left as an exercise for the reader.

By the claim, without loss of generality there are $\Omega(\varepsilon)|X_i|^2|X_j|$ triples $x, x' \in X_i, y \in X_j$ with $y \in N(x) \triangle N(x')$. The probability that the random vertices x, x', y form such a triple is

$$\frac{\Omega(\varepsilon)|X_i|^2|X_j|}{n^2|X_i|} = \frac{\Omega(\varepsilon)|X_i||X_j|}{n^2}.$$

Summing over all non-ε-homogeneous pairs (X_i, X_j) and using the assumption that $\{X_1, \dots, X_t\}$ is not ε-homogeneous, we get $\mathbb{P}[\mathcal{A}] \geq \Omega(\varepsilon^2) > 2\gamma$, provided that γ is small enough. This is a contradiction.

A last step, in order to obtain an equipartition, is to chop up each part X_i into equal-sized parts plus maybe one leftover part, then collect the leftover parts and partition them again into equal-sized parts. One can show that (if the part size is small enough) then the resulting partition is still β-homogeneous for β which depends polynomially on ε. (A refinement of a ε-homogeneous partition is $O(\sqrt{\varepsilon})$-homogeneous, which follows from Markov's inequality). ■

Alon, Fischer and Newman in fact proved a stronger statement: They show that it suffices to assume that G has few bi-induced copies of some fixed bipartite graph H (instead of assuming that

G has no such copies at all). This result, as well as Theorem 4.5, are sometimes called an *ultra-strong regularity lemma.*

Theorem 4.7 (Alon-Fischer-Newman 2007). *For every bipartite graph H and $\varepsilon > 0$, there is $\delta = poly(\varepsilon) > 0$, such that the following holds. If an n-vertex graph G has at most $\delta n^{v(H)}$ bi-induced copies of H, then G has an ε-homogeneous equipartition into at most $\frac{1}{\delta}$ parts.*

Let us now translate the condition of avoiding bi-induced copies of a bipartite graph to the condition of avoiding induced subgraphs of certain types. A *co-bipartite* graph is the complement of a bipartite graph. A *split* graph is a graph whose vertex-set can be partitioned into a clique and an independent set.

Lemma 4.8. *Let F_1 be a bipartite graph, F_2 be a co-bipartite graph, F_3 be a split graph. There is a bipartite graph $H = (A, B)$ such that if G has no induced copies of F_1, F_2, F_3, then it has no bi-induced copies of H.*

Note that the three graph types (bipartite, co-bipartite, split) capture all possibilities of partitioning the vertex-set into two homogeneous sets (cliques or independent sets).

Proof sketch of Lemma 4.8. We need to show that there is a bipartite $H = (A, B)$ such that no matter how we place edges inside A and inside B, we get a graph which contains an induced copy of F_1, F_2 or F_3. We show that a large enough random graph H satisfies this. Let the edges inside A and B be given. We can use Ramsey's theorem to partition almost all of A and of B into homogeneous sets of size k, where $k \geq |V(F_i)|$ for $i = 1, 2, 3$. Now, for such a partition, the probability that there is no induced copy of F_1, F_2, F_3 is at most $1 - 2^{-k^2}$. Thus, if $|A| = |B| = m$, the probability of failure is at most $(1 - 2^{-k^2})^{m^2/k^2} \leq e^{-\Omega_k(m^2)}$. On the other hand, the number of partitions is at most m^{2m}, so we can take a union bound. ■

By combining Theorem 4.7 and Lemma 4.8, one can prove the following:

Theorem 4.9 (Gishboliner-Shapira 2017). *If $\mathcal{H}$ is a finite graph-family containing a bipartite graph, a co-bipartite graph and a split graph, then the induced-$\mathcal{H}$ removal lemma has polynomial bounds.*

Proof sketch. The proof follows the scheme described in Section 3. Let $F_1, F_2, F_3 \in \mathcal{H}$ such that F_1 is bipartite, F_2 is co-bipartite and F_3 is split. Let H be the bipartite graph given by Lemma 4.8. Let G be an n-vertex graph which is ε-far from being induced-$\mathcal{H}$-free. Suppose first that G contains at least $\delta n^{v(H)}$ bi-induced copies of H.[5] Then by Lemma 4.8, there is $i = 1, 2, 3$ such that G contains at least $\frac{\delta}{3} n^{v(H)}$ vertex-sets of size $v(H)$ which contain an induced copy of F_i. On the other hand, each induced copy of F_i is in at most $n^{v(H)-v(F_i)}$ such vertex sets, so there are at least $\frac{\delta}{3} n^{v(F_i)}$ induced copies of F_i, as required.

From now on, suppose that G has less than $\delta n^{v(H)}$ bi-induced copies of H. We apply Theorem 4.7 to get an ε-homogeneous equipartition $\mathcal{P} = \{V_1, \dots, V_t\}$, and then apply Theorem 4.7 again to find an ε'-homogeneous equipartition Q which refines[6] $\mathcal{P}$, where ε' is small enough (but still polynomial) in terms of ε and t; i.e., $\varepsilon' = (\varepsilon/t)^C$ for a large constant C (depending on $\mathcal{H}$). Then, for each $1 \leq i \leq t$, sample $U_i \in \mathcal{Q}$ uniformly at random among all parts of $\mathcal{Q}$ which are contained in V_i. One can show that due to the choice of ε', the following holds with positive probability:

[5] For the sake of keeping the presentation simple, we will not choose δ explicitly. Rather, δ is the number given by Theorem 4.7; we will apply this theorem several times with different parameters, and δ is the minimum of the resulting numbers.

[6] While this is not part of the statement of Theorem 4.7, the theorem can be reproved to allow for a partition $\mathcal{P}$ as part of the input, such that the outputted equipartition refines $\mathcal{P}$.

(i) For every $1 \le i < j \le t$, (U_i, U_j) is ε'-homogeneous; i.e., $d(U_i, U_j) \le \varepsilon'$ or $d(U_i, U_j) \ge 1 - \varepsilon'$.

(ii) For all but $\sqrt{\varepsilon}t^2$ of the pairs $1 \le i < j \le t$, it holds that $|d(U_i, U_j) - d(V_i, V_j)| \le 10\sqrt{\varepsilon}$.

The next step is as follows: For each $i \in [t]$, apply Theorem 4.7 to $G[U_i]$ to get an ε-homogeneous equipartition $\mathcal{R}_i$ of U_i. Recall that this means that all but $\varepsilon|\mathcal{R}_i|^2$ of the pairs of parts in $\mathcal{R}_i$ are ε-homogeneous (have density at most ε or at least $1 - \varepsilon$). Apply Turán's theorem to pass to $\mathcal{R}'_i \subseteq \mathcal{R}_i$ of size roughly $|\mathcal{R}'_i| \approx \frac{1}{\varepsilon}$ such that any two parts in $\mathcal{R}'_i$ are ε-homogeneous, and then apply Ramsey's theorem to find $\mathcal{R}''_i = \{U_{i,1}, \dots, U_{i,h}\} \subseteq \mathcal{R}'_i$ such that either $d(U_{i,k}, U_{i,\ell}) \le \varepsilon$ for all $1 \le k < \ell \le h$ or $d(U_{i,k}, U_{i,\ell}) \ge 1 - \varepsilon$ for all $1 \le k < \ell \le h$.[7] Another important point is that since (U_i, U_j) is ε'-homogeneous and ε' is very small (but still polynomial) compared to ε, we have $|d(U_{i,k}, U_{j,\ell}) - d(U_i, U_j)| \le \varepsilon$ for all $1 \le i < j \le t$ and $k, \ell \in [h]$.

We now achieved the setting described by Items 1-4 in Section 3. Now clean the graph as described in that section. One then shows that if the cleaned graph has an induced copy of some $F \in \mathcal{H}$, then the original graph has many (i.e., $\delta n^{v(F)}$) such induced copies. ■

Gishboliner and Shapira also proved that if a finite $\mathcal{H}$ contains no bipartite graph or no co-bipartite graph, then the induced-$\mathcal{H}$ removal lemma is not polynomial (this proof uses similar constructions to those used in Section 2). The following remains open:

Problem 4.10. *Characterize the finite graph-families $\mathcal{H}$ for which the induced-$\mathcal{H}$ removal lemma is polynomial.*

This is of course a special case of Problem 3.3. The first open case is again $\mathcal{H} = \{C_4\}$. Note that C_4 is both bipartite and co-bipartite, but not split, so the aforementioned results of Gishboliner and Shapira do not apply.

[7] Here h can be chosen as $h = \max_{F \in \mathcal{F}} v(F)$.