Polynomial Removal Lemma for Ordered Matchings

Lior Gishboliner[a,b]   Borna Šimić[a]

Abstract

We prove that for every ordered matching H on t vertices, if an ordered n-vertex graph G is ε-far from being H-free, then G contains $\mathrm{poly}(\varepsilon)n^t$ copies of H. This proves a special case of a conjecture of Tomon and the first author. We also generalize this statement to uniform hypergraphs.

Mathematics Subject Classifications: 05C35

1   Introduction

The graph removal lemma is a fundamental result in extremal graph theory, stating that for every fixed graph H and $\varepsilon > 0$, if an n-vertex graph G is ε-far from being H-free, in the sense that εn^2 edges must be deleted in order to turn G into an H-free graph, then G contains at least $\delta n^{|V(H)|}$ copies of H, where $\delta = \delta(H, \varepsilon) > 0$. This was proved in a seminal work of Ruzsa and Szemerédi [18]. The removal lemma was subsequently generalized to many other combinatorial structures, notably induced subgraphs [4], hypergraphs [16, 17, 20] and ordered graphs [3]. Removal lemmas are also closely related to graph property testing in the dense graph model, where they correspond to testing algorithms with constant query complexity, see the book [15].

A drawback of the known proofs of the removal lemma (and its many generalizations) is that all such proofs rely on Szemerédi's regularity lemma [19] or a generalization thereof. This results in weak quantitative bounds; for example, for the graph removal lemma stated above, the best known bound [9] is that $1/\delta \leqslant \mathrm{tower}(O(\log(1/\varepsilon)))$, where $\mathrm{tower}(x)$ is a tower of x exponents. This situation has led to research on the problem of characterizing the cases where the removal lemma has polynomial bounds, namely, where δ depends polynomially on ε. By now there are many works of this type [1, 2, 5, 6, 7, 11, 10, 12, 13, 14].

Here we focus on ordered graphs. An ordered graph is a graph with a linear ordering on its vertices. A copy of an ordered graph H in an ordered graph G is an injection

---

[a] Department of Mathematics, ETH, Zürich, Switzerland
  (lior.gishboliner@math.ethz.ch, bosimic@student.ethz.ch).

$\varphi : V(H) \to V(G)$ which preserves the vertex order and satisfies that $\varphi(x)\varphi(y) \in E(G)$ for every $xy \in E(H)$.

In an important work [3], Alon, Ben-Eliezer and Fischer proved an ordered analogue of the graph removal lemma. They further asked to study cases where the ordered removal lemma has polynomial bounds. Addressing this question, Tomon and the first author [14] characterized the ordered graphs H for which the induced H-removal lemma has polynomial bounds. They also studied the non-induced case, and conjectured that the (non-induced) H-removal lemma has polynomial bounds if and only if the ordered core[1] of H is an (ordered) forest. As observed in [14], to prove this conjecture it suffices to show that the H-removal lemma has polynomial bounds for every ordered forest H. Here we make progress on this conjecture by proving it for every ordered matching H. We also generalize this to s-uniform hypergraphs, $s \geqslant 3$. An (ordered) n-vertex s-uniform hypergraph G is said to be ε-far from being H-free if one has to delete at least εn^s edges to turn G into an H-free hypergraph. Our main result is as follows.

Theorem 1. *For every $t \geqslant s \geqslant 2$, there exists $C = C(t)$ such that the following holds. Let $\varepsilon > 0$, let H be an ordered s-uniform matching on t vertices, and let G be an ordered s-uniform hypergraph on n vertices. If G is ε-far from being H-free, then G contains at least $(\varepsilon/C)^C n^t$ copies of H.*

Proving removal-type statements for ordered structures tends to be considerably more difficult than for their unordered counterparts. For example, the proof of the ordered removal lemma in [3] is substantially more involved than the original proof of the removal lemma in [18]. The key difficulty is to find copies of H which respect the vertex order. To deal with this difficulty, our proof uses a novel argument of considering nested partitions of the vertex-set (with each partition refining the previous one) and "cleaning" the graph with respect to each of these levels.

2 Proof of Theorem 1

Assume G is as in the statement of the theorem, with vertex set $[n]$. For two subsets $A, B \subseteq [n]$, we write $A < B$ to mean that $a < b$ for all $a \in A, b \in B$, i.e., all elements of A are smaller than all elements of B. We begin by partitioning[2] $[n]$ into $k := \frac{1}{\varepsilon}$ intervals $I_1 < \cdots < I_k$ of length εn each, and delete all edges with at least two vertices inside one of these intervals. Let $G_0 \subseteq G$ be the resulting hypergraph. This step deletes less than

$$\frac{1}{\varepsilon}\binom{\varepsilon n}{2} \cdot \binom{n-2}{s-2} < \frac{\varepsilon}{2}n^s$$

---

[1] The ordered core of H is defined as follows. Recall that a graph homomorphism from a graph G to a graph G' is a map $\varphi : V(G) \to V(G')$ such that $\varphi(x)\varphi(y) \in E(G')$ for every $xy \in E(G)$. For ordered graphs G, G', an ordered homomorphism from G to G' is a graph homomorphism which also preserves the vertex order. The ordered core of G is defined as the smallest (in terms of number of vertices) subgraph of G such that there is an homomorphism from G to G'. One can show that the ordered core is unique up to ordered isomorphism.

[2] Here and throughout the proof, for the sake of simplicity, we omit floor and ceiling signs by assuming that n is divisible by an appropriate (polynomial) function of ε.

edges, so G_0 is still $\frac{\varepsilon}{2}$-far from being H-free.

Set $\gamma := \frac{\varepsilon}{4t}$ (recall that $t = |V(H)|$). For each $1 \leqslant \ell \leqslant k$, we define t nested partitions of I_ℓ as follows. Set $\mathcal{J}_{\ell,1} = \{I_\ell\}$. For $j = 2, \ldots, t$ and for each $J \in \mathcal{J}_{\ell,j-1}$, split J into intervals of length $\gamma|J|$ and add these intervals to $\mathcal{J}_{\ell,j}$. Note that for each $1 \leqslant j \leqslant t$, $\mathcal{J}_{\ell,j}$ forms a partition of I_ℓ into intervals of size $\gamma^{j-1}|I_\ell|$. Put $\mathcal{J}_\ell := \bigcup_{j=1}^{t} \mathcal{J}_{\ell,j}$. For each $v \in I_\ell$ and $1 \leqslant j \leqslant t$, it will be convenient to denote by $J_j(v)$ the interval in $\mathcal{J}_{\ell,j}$ containing v, so that

$$v \in J_t(v) \subseteq J_{t-1}(v) \subseteq \cdots \subseteq J_1(v) = I_\ell.$$

Set $\beta := 2\gamma = \frac{\varepsilon}{2t}$. We now perform a sequence of k cleaning steps and define k corresponding hypergraphs $G_0 \supseteq G_1 \supseteq \cdots \supseteq G_k$, where G_ℓ is the hypergraph obtained after the ℓth cleaning step ($1 \leqslant \ell \leqslant k$). At step ℓ we clean with respect to the interval I_ℓ, as follows: For every choice of $s - 1$ vertices $v_1 < v_2 < \cdots < v_{s-1}$ outside of I_ℓ, and for every interval $J \in \mathcal{J}_\ell$, let $L_\ell(v_1, v_2, \ldots, v_{s-1}, J)$ denote the leftmost $\beta|J|$ vertices $w \in J$ such that $\{v_1, \ldots, v_{s-1}, w\} \in E(G_{\ell-1})$, if there are at least $\beta|J|$ such vertices, and else let $L_\ell(v_1, v_2, \ldots v_{s-1}, J)$ be the set of all such vertices w. Delete all edges $\{v_1, \ldots, v_{s-1}, w\} \in E(G_{\ell-1})$ with $w \in L_\ell(v_1, v_2, \ldots, v_{s-1}, J)$. The resulting hypergraph is G_ℓ. By definition, for every given $(s-1)$-tuple $v_1, v_2, \ldots, v_{s-1}$ and for every interval $J \in \mathcal{J}_\ell$, this operation deletes at most $\beta|J|$ edges of the form $\{v_1, \ldots, v_{s-1}, w\}$ with $w \in J$. Since the intervals in $\mathcal{J}_{\ell,j}$ form a partition of I_ℓ (for every $1 \leqslant j \leqslant t$), we delete at most $\beta|I_\ell|$ edges when considering these intervals. Summing over $1 \leqslant j \leqslant t$, this gives a total of at most $t\beta|I_\ell|$ edge deletions for each of the less than n^{s-1} choices of $v_1, \ldots, v_{s-1}$. Therefore, $e(G_{\ell-1}) - e(G_\ell) < t\beta n^{s-1}|I_\ell|$. Summing over $\ell = 1, \ldots, k$, we get that

$$e(G_0) - e(G_k) < \sum_{\ell=1}^{k} t\beta n^{s-1}|I_\ell| = t\beta n^s = \frac{\varepsilon}{2} n^s,$$

using our choice of β. As G_0 is $\frac{\varepsilon}{2}$-far from being H-free, G_k must contain a copy of H. We will use this fact later on. The key property guaranteed by the cleaning procedure is the following.

Lemma 2. *Let $1 \leqslant \ell \leqslant k$ and $1 \leqslant m \leqslant t$, let $w_1 < \cdots < w_m$ be vertices in I_ℓ, and let $v_{i,j} \in [n] \backslash I_\ell$, where $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant s-1$, such that $\{v_{i,1}, v_{i,2}, \ldots v_{i,s-1}, w_i\} \in E(G_\ell)$ for every $i = 1, \ldots, m$. Then there are at least $\left(\frac{\varepsilon}{4t}\right)^{m(t+1)} n^m$ different m-tuples of vertices $w_1' < \cdots < w_m'$ in I_ℓ, such that $\{v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, w_i'\} \in E(G_{\ell-1})$ for every $i = 1, \ldots, m$.*

Note that by Lemma 2, if there is a copy H_ℓ of H in G_ℓ having m vertices in I_ℓ, then in $G_{\ell-1}$ there are at least $\left(\frac{\varepsilon}{4t}\right)^{m(t+1)} n^m$ copies of H which agree with H_ℓ on the vertices outside of I_ℓ.

Proof of Lemma 2. Define sets $L_1, \ldots, L_m \subseteq I_\ell$ as follows:

$$L_i := L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i)) \setminus J_{i+1}(w_i)$$

for $1 \leqslant i < m$, and

$$L_m := L_\ell(v_{m,1}, v_{m,2}, \ldots, v_{m,s-1}, J_m(w_m)).$$

For every $1 \leqslant i \leqslant m$ and $w_i' \in L_i$, it holds that $\{v_{i,1}, v_{i,2}, \ldots v_{i,s-1}, w_i'\} \in E(G_{\ell-1})$, by the definition of $L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))$.

Claim 3. $L_1 < \cdots < L_m$.

Proof. Fix any $1 \leqslant i < m$, and let us show that $L_i < L_{i+1}$. We begin by proving that $L_i < w_i$, namely, that all vertices of L_i are to the left of w_i. Indeed, first note that $L_i \subseteq L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))$. Recall that $L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))$ is a set of leftmost vertices $w \in J_i(w_i)$ satisfying $\{v_{i,1}, v_{i,2}, \ldots v_{i,s-1}, w\} \in E(G_{\ell-1})$, and the edges $\{v_{i,1}, v_{i,2}, \ldots v_{i,s-1}, w\}$ for $w \in L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))$ are deleted when obtaining G_ℓ from $G_{\ell-1}$, while the edge $\{v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, w_i\}$ is still present in G_ℓ. This shows that $L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))$ is to the left of w_i, implying that $L_i < w_i$.

Next, note that L_i is disjoint from $J_{i+1}(w_i)$ by definition. It follows that $L_i < J_{i+1}(w_i)$; indeed, for each $w \in L_i$, we have $J_{i+1}(w) \leqslant J_{i+1}(w_i)$ as $w < w_i$, and also $J_{i+1}(w) \neq J_{i+1}(w_i)$ because $w \notin J_{i+1}(w_i)$, hence $w < J_{i+1}(w_i)$. We conclude that $L_i < J_{i+1}(w_i) \leqslant J_{i+1}(w_{i+1})$, using that $w_i < w_{i+1}$. As $L_{i+1} \subseteq J_{i+1}(w_{i+1})$, we get that $L_i < L_{i+1}$, as required. $\qquad\square$

Claim 4. $|L_i| \geqslant \left(\frac{\varepsilon}{4t}\right)^{t+1} n$ *for all* $1 \leqslant i \leqslant m$.

Proof. Recall that for every $w \in I_\ell$, we have

$$|J_t(w)| = \gamma|J_{t-1}(w)| = \cdots = \gamma^{t-1}|J_1(w)| = \gamma^{t-1}|I_\ell| = \gamma^{t-1}\varepsilon n > \left(\frac{\varepsilon}{4t}\right)^t n.$$

Now, observe that $|L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))| \geqslant \beta|J_i(w_i)|$, because otherwise we would have deleted all edges of the form $\{v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, w\}$ with $w \in J_i(w_i)$ when obtaining G_ℓ from $G_{\ell-1}$, but the edge $\{v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, w_i\}$ is still present in G_ℓ. Using that $|J_{i+1}(w_i)| = \gamma|J_i(w_i)| = \frac{\beta}{2}|J_i(w_i)|$, we get by the definition of L_i that

$$|L_i| \geqslant |L_\ell(v_{i,1}, v_{i,2}, \ldots, v_{i,s-1}, J_i(w_i))| - |J_{i+1}(w_i)| \geqslant \frac{\beta}{2}|J_i(w_i)| \geqslant \left(\frac{\varepsilon}{4t}\right)^{t+1} n. \qquad\square$$

We now complete the proof of Lemma 2. As we saw above, for every $1 \leqslant i \leqslant m$ and $w_i' \in L_i$, it holds that $\{v_{i,1}, v_{i,2}, \ldots v_{i,s-1}, w_i'\} \in E(G_{\ell-1})$. Thus, the lemma follows from Claims 3 and 4. $\qquad\square$

Recall that G_k contains a copy of H; denote it H_k. We can now use this initial copy and Lemma 2 to construct the required number of distinct H-copies in G_0. This will be done in the following lemma. For $1 \leqslant \ell \leqslant k$, let m_ℓ be the number of vertices of H_k in the interval I_ℓ. For convenience, put $\delta := \left(\frac{\varepsilon}{4t}\right)^{t+1}$.

Lemma 5. *For every* $\ell = k, \ldots, 0$, *there are at least* $(\delta n)^{m_{\ell+1} + \cdots + m_k}$ *copies of* H *in* G_ℓ *which have* m_i *vertices in* I_i *for every* $1 \leqslant i \leqslant k$, *and have the same vertices as* H_k *in* $I_1 \cup \cdots \cup I_\ell$.

Proof. The proof is by reverse induction on ℓ. The base case $\ell = k$ holds trivially, because H_k is a copy of H in G_k satisfying the required properties, and $(\delta n)^0 = 1$. For the induction step, let $0 < \ell \leqslant k$. By the induction hypothesis, there is a collection $\mathcal{H}_\ell$ of at least $(\delta n)^{m_{\ell+1} + \cdots + m_k}$ copies of H in G_ℓ which have m_i vertices in I_i for every $1 \leqslant i \leqslant k$, and have the same vertices as H_k in $I_1 \cup \cdots \cup I_\ell$. If $m_\ell = 0$ then there is nothing to prove, so suppose that $m_\ell \geqslant 1$. Fix any $H_\ell \in \mathcal{H}_\ell$. Note that every edge of H_ℓ touching I_ℓ has exactly one vertex in I_ℓ, because every edge of G_0 has at most one vertex in each of the intervals $I_1, \ldots, I_k$ (by the definition of G_0). Namely, every edge $e \in E(H_\ell)$ with $e \cap I_\ell \neq \emptyset$ is of the form $\{v_1, \ldots, v_{s-1}, w\}$ with $w \in I_\ell$ and $v_1, \ldots, v_{s-1} \in [n] \setminus I_\ell$. Let $w_1 < \cdots < w_{m_\ell}$ be the vertices of H_ℓ in I_ℓ. For each $1 \leqslant i \leqslant m_\ell$, let $v_{i,1}, \ldots, v_{i,s-1} \in [n] \setminus I_\ell$ such that $\{v_{i,1}, \ldots, v_{i,s-1}, w_i\} \in E(H_\ell)$. By Lemma 2, we can replace $w_1, \ldots, w_{m_\ell}$ in $(\delta n)^{m_\ell}$ ways to obtain copies of H in $G_{\ell-1}$. Doing this for different $H_\ell, H'_\ell \in \mathcal{H}_\ell$ gives different copies of H, because H_ℓ, H'_ℓ differ on vertices outside I_ℓ (as they both agree with H_k on I_ℓ), and we do not change the vertices of H_ℓ, H'_ℓ which are outside I_ℓ. Thus, doing the above for each $H_\ell \in \mathcal{H}_\ell$ gives the required $(\delta n)^{m_\ell} |\mathcal{H}_\ell| \geqslant (\delta n)^{m_\ell + \cdots + m_k}$ copies of H in $G_{\ell-1}$. This completes the induction step. $\qquad\square$

For $\ell = 0$, Lemma 5 gives

$$(\delta n)^{m_1 + \cdots + m_k} = (\delta n)^t = \left(\frac{\varepsilon}{4t}\right)^{t(t+1)} n^t$$

copies of H in G_0 (and so in G), as required. This proves the theorem.

3   Concluding remarks

We proved that ordered matchings admit a polynomial removal lemma. It would be interesting to extend this result to other families of ordered forests, such as ordered paths.

Our proof of Theorem 1 shows that one can take $C(t) = O(t^2)$. It would be interesting to improve this to $C(t) = O(t)$. For fixed s and growing t, this would be tight, as shown by the following proposition:

Proposition 6. *Let H be an ordered s-uniform hypergraph with t vertices and m edges, let $\varepsilon > 0$ be small enough, and let $n \geqslant n_0(\varepsilon)$. Then there exists an n-vertex s-uniform ordered hypergraph G which is ε-far from being H-free but contains only $O(n^t \varepsilon^m)$ copies of H.*

If H is an ordered matching then $e(H) = \frac{t}{s}$, so the proposition implies that $C(t) \geqslant \frac{t}{s}$. We give a proof sketch of the proposition.

Proof sketch of Proposition 6. Take $G \sim \mathbb{H}_s(n, \varepsilon)$ to be the random (binomial) ordered s-uniform hypergraph; namely, each of the $\binom{n}{s}$ potential edges is included in G with probability ε, independently. The expected number of copies of H in G is $\Theta(n^t \varepsilon^m)$, because there are $\Theta(n^t)$ potential copies, and each appears with probability ε^m. Similarly, a given edge of G appears in expectation in $O(n^{t-s} \varepsilon^{m-1})$ copies of H. Using Azuma's

inequality (see e.g. [8, Chapter 7]), one can show that w.h.p. every edge appears in $O(n^{t-s}\varepsilon^{m-1})$ copies of H, and the total number of copies of H is $\Theta(n^t\varepsilon^m)$. Thus, one has to delete $\Omega\left(\frac{n^t\varepsilon^m}{n^{t-s}\varepsilon^{m-1}}\right) = \Omega(\varepsilon n^s)$ edges to destroy all copies of H, meaning that G is $\Omega(\varepsilon)$-far from being H-free. $\qquad\square$

References

[1] N. Alon. Testing subgraphs in large graphs. *Random Structures & Algorithms*, 21(3-4):359–370, 2002.

[2] N. Alon and O. Ben-Eliezer. Efficient removal lemmas for matrices. *Order*, 37(1):83–101, 2020.

[3] N. Alon, O. Ben-Eliezer, and E. Fischer. Testing hereditary properties of ordered graphs and matrices. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 848–858. IEEE, 2017.

[4] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[5] N. Alon, E. Fischer, and I. Newman. Efficient testing of bipartite graphs for forbidden induced subgraphs. *SIAM Journal on Computing*, 37(3):959–976, 2007.

[6] N. Alon and A. Shapira. Testing subgraphs in directed graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 700–709, 2003.

[7] N. Alon and A. Shapira. A characterization of easily testable induced subgraphs. *Combinatorics, Probability and Computing*, 15(6):791–805, 2006.

[8] N. Alon and J. H. Spencer. **The Probabilistic Method**. John Wiley & Sons, 2016.

[9] J. Fox. A new proof of the graph removal lemma. *Annals of Mathematics*, pages 561–579, 2011.

[10] L. Gishboliner and A. Shapira. Efficient removal without efficient regularity. *Combinatorica*, 39(3):639–658, 2019. Also in Proc of ITCS 2018, 1-15.

[11] L. Gishboliner and A. Shapira. Removal lemmas with polynomial bounds. *International Math Research Notices (IMRN)*, pages 14409–14444, 2021. Also in Proc of STOC 2017, 510-522.

[12] L. Gishboliner and A. Shapira. Hypergraph removal with polynomial bounds. arXiv:2202.07567, 2022.

[13] L. Gishboliner and I. Tomon. On 3-graphs with no four vertices spanning exactly two edges. *Bulletin of the London Mathematical Society*, 54:2117–2134, 2022.

[14] L. Gishboliner and I. Tomon. Polynomial removal lemmas for ordered graphs. *Combinatorial Theory*, 2, 2022.

[15] O. Goldreich. Contemplations on testing graph properties. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 547–554. Springer, 2011.

[16] B. Nagle, V. Rödl, and M. Schacht. The counting lemma for regular k-uniform hypergraphs. *Random Structures & Algorithms*, 28(2):113–179, 2006.

[17] V. Rödl and J. Skokan. Applications of the regularity lemma for uniform hypergraphs. *Random Structures & Algorithms*, 28(2):180–194, 2006.

[18] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai*, 18:939–945, 1978.

[19] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[20] T. Tao. A variant of the hypergraph removal lemma. *Journal of Combinatorial Theory, Series A*, 113(7):1257–1280, 2006.