# Lower bounds for superpatterns and universal sequences

Zachary Chroman[*]        Matthew Kwan[†]        Mihir Singhal[‡]

## Abstract

A permutation $\sigma \in S_n$ is said to be *k-universal* or a *k-superpattern* if for every $\pi \in S_k$, there is a subsequence of $\sigma$ that is order-isomorphic to $\pi$. A simple counting argument shows that $\sigma$ can be a $k$-superpattern only if $n \geq (1/e^2 + o(1))k^2$, and Arratia conjectured that this lower bound is best-possible. Disproving Arratia's conjecture, we improve the trivial bound by a small constant factor. We accomplish this by designing an efficient encoding scheme for the patterns that appear in $\sigma$. This approach is quite flexible and is applicable to other universality-type problems; for example, we also improve a bound by Engen and Vatter on a problem concerning $(k + 1)$-ary sequences which contain all $k$-permutations.

## 1  Introduction

We say that a permutation $\sigma \in S_n$ *contains* a permutation $\pi \in S_k$ *as a pattern* when there exists some subset of indices $t_1 < \ldots < t_k$ such that $\sigma(t_i) < \sigma(t_j)$ if and only if $\pi(i) < \pi(j)$ (that is to say, $\sigma$ has a subsequence which is order-isomorphic to $\pi$). Starting with some foundational work by MacMahon [14] and later Knuth [13], the subject of permutation patterns has become quite vast over the years; see for example the books [11, 5], the surveys [20, 21], and Tenner's *Database of Permutation Pattern Avoidance* [22].

If a permutation $\sigma$ contains all $k!$ patterns in $S_k$, it is said to be *k-universal* or a *k-superpattern*. Following some early work in the case $k = 3$ by Simion and Schmidt [19], superpatterns were first introduced as a general notion by Arratia in 1999 [4]. Arratia raised perhaps the most fundamental question about superpatterns: how short can a $k$-superpattern be? He described a $k$-superpattern of length $k^2$, and observed that any $k$-superpattern $\sigma \in S_n$ must have length at least $n \geq (1/e^2 + o(1))k^2$. This lower bound arises from the trivial inequality $\binom{n}{k} \geq k!$, which holds because a permutation $\sigma \in S_n$ has only $\binom{n}{k}$ different subsequences (and therefore patterns) of length $k$. Arratia conjectured that this trivial lower bound of $(1/e^2 + o(1))k^2$ is in fact best-possible.

This conjecture has remained open for the last twenty years, but various improvements have been made to the upper bound. Following earlier work by Eriksson, Eriksson, Linusson and Wastlund [8], a clever construction due to Miller [15] achieves a $k$-superpattern of length $(k^2 + k)/2$, and this was later improved slightly to $(k^2 + 1)/2$ by Engen and Vatter [7]. Actually, the authors of [8] made a conjecture (contradicting Arratia's conjecture) that $(1/2 + o(1))k^2$ is the true minimum length of a $k$-superpattern.

Arratia's conjecture may at first seem a little unmotivated, but when placed in a wider context, it is quite natural. Generally speaking, a mathematical structure is said to be *universal* if it contains all possible substructures, in some specified sense (this concept seems to have been first considered by Rado [17]). For many different notions of universality, there are trivial lower bounds arising from counting arguments in a very similar way to Arratia's superpattern bound, and these bounds are very often known or suspected to be essentially best-possible. As a simple example, for any positive integers $k$ and $q$, there is known to exist a *de Brujin sequence* with these parameters, which is a string over a size-$q$ alphabet containing every possible length-$k$ string exactly once. A much less simple example is the case of universality for finite

graphs, a subject that has received a large amount of attention over recent years due in part to applications in computer science (see for example [3]). A graph is said to be *k-induced-universal* if it contains an induced copy of all possible graphs on $k$ vertices. There are at least $2^{\binom{k}{2}}/k!$ graphs on $k$ vertices, and an $n$-vertex graph has at most $\binom{n}{k}$ induced subgraphs with $k$ vertices, so a straightfoward computation shows that every $k$-induced-universal graph must have at least $(1 + o(1))2^{(k-1)/2}$ vertices (this was first observed by Moon [16]). In a recent breakthrough, Alon [1] proved that this trivial lower bound is in fact best-possible, by showing that a random graph with this many vertices typically contains *almost* all $k$-vertex graphs as induced subgraphs, and then proving that the remaining "exceptional" graphs can be handled separately without adding too many vertices.

It is very tempting to try to adapt Alon's proof strategy to permutations, to prove Arratia's conjecture. First, one might try to prove that a random permutation of length $n = (1/e^2 + o(1))k^2$ typically contains almost all permutations of length $k$ (one imagines that the $\binom{n}{k}$ subsequences of a random permutation tend to yield mostly distinct patterns, so the trivial counting bound should not be too far from the truth). Second, one hopes to be able to deal with the few remaining permutations in some more ad-hoc way, without substantially increasing the length of the permutation. In this paper we prove that not only is Arratia's conjecture false, but even the first of these two steps fails.

**Theorem 1.1.** *Suppose that* $n < (1.000076/e^2)k^2$. *Then every* $\sigma \in S_n$ *contains only* $o(k!)$ *different patterns* $\pi \in S_k$.

We have made some effort to optimize the constant in Theorem 1.1, where it would not negatively affect the readability of the proof. However, it would be quite complicated to fully squeeze the utmost out of our proof idea (see Section 4 for further discussion). In any case, obtaining a constant that is substantially larger than $1/e^2$ seems quite out of reach.

To very briefly describe the approach in our proof of Theorem 1.1, it is instructive to reinterpret Arratia's trivial bound in an "information-theoretic" way. Namely, for each pattern $\pi \in S_k$ that appears in $\sigma$, note that $\sigma$ gives a way to encode each $\pi \in S$ by a set of indices $t_1 < \cdots < t_k$. Since there are at most $\binom{n}{k}$ possible outcomes of this encoding, there are at most $\binom{n}{k}$ different patterns that we could have encoded, yielding Arratia's bound. The key observation towards improving this bound is that the aforementioned encoding is often slightly wasteful. For example, suppose that we are encoding a pattern $\pi$ that appears in $\sigma$ at indices $t_1 < \cdots < t_k$, and suppose that $t_{i+1} - t_{i-1} > k$ for some $i$. Now, having specified all the indices except $t_i$, there are more than $k$ possibilities for $t_i$. This means it would be "cheaper" simply to specify the relative position $\pi(i)$ itself, rather than to specify the index $t_i$ (note that there are only $k$ choices for this relative position, and it fully determines $\pi$). More generally, we can design an encoding scheme that decides for each $i$ whether to specify the index $t_i$ or the relative position $\pi(i)$ depending on the size of $t_{i+1} - t_{i-1}$. Note that the average value of $t_{i+1} - t_{i-1}$ is about $2k/e^2$ (which is substantially less than $k$), so for most $i$ we we will end up just specifying $t_i$, as in the trivial encoding scheme. Our small gain comes from the fact that for almost all choices of $t_1, \ldots, t_k$, there are a small number of $i$ for which the difference $t_{i+1} - t_{i-1}$ is much larger than average. The details of the proof are in Section 2.

At first glance the approach sketched above seems extremely general. For example, in the well-studied case of $k$-induced-universal graphs we can also view the trivial lower bound as coming from an encoding scheme (in fact, it is precisely this point of view that gives universal graphs their applications in computer science). We can put an ordering on the set of vertices and try to make similar local improvements to the trivial encoding scheme, taking advantage of large gaps between vertices according to our ordering. However, in the case of graphs this approach does not seem to give us even tiny (lower-order) improvements to the lower bound. We believe this warrants further investigation, and we discuss this in the concluding remarks (Section 4).

We were, however, able to find another problem closely related to Arratia's conjecture for which our approach is useful. To describe this, we generalize to the setting where $\sigma$ is a sequence of (not necessarily distinct) elements of $[r] := \{1, \ldots, r\}$. We define containment of a pattern identically: if $\sigma \in [r]^n$ is a sequence and $\pi \in S_k$ is a pattern, then $\sigma$ contains $\pi$ if there exist indices $t_1 < \cdots < t_k$ such that $\sigma(t_i) < \sigma(t_j)$ if and only if $\pi(i) < \pi(j)$. Again, $\sigma$ is said to be *k-universal* if it contains all $k$-patterns. Note that any $k$-universal sequence yields a $k$-superpattern of the same length, since it is possible to find a permutation with the same

relative ordering (breaking ties arbitrarily).

Engen and Vatter, in their survey paper [7], considered the question of finding the shortest possible $k$-universal sequence in the cases $r = k, k + 1$. In the case $r = k$, it is known that the answer must be $(1 + o(1))k^2$, as proved by Kleitman and Kwiatkowski [12]. The case $r = k + 1$ is of particular interest since Miller's construction of a $k$-superpattern of length $n = (k^2 + k)/2$ actually comes from a $k$-universal sequence in $[k + 1]^n$. Meanwhile, until now the best lower bound was the trivial one obtained as follows. Writing $a_m$ for the number of occurrences of each $m \in [r]$ in $\sigma$, the number of subsequences of $\sigma$ with elements $m_1, m_2, \ldots, m_k$ is equal to $a_{m_1} a_{m_2} \ldots a_{m_k}$. This expression is at most $(n/k)^k$ by convexity. Thus, the total number of subsequences that don't contain repeated elements is at most $\binom{r}{k}(n/k)^k$. This must be at least $k!$ for $\sigma$ to be a $k$-universal sequence, so as long as $r = (1 + o(1))k$ we must have $n \geq (1/e + o(1))k^2$. We improve upon this trivial bound as follows.

**Theorem 1.2.** *Suppose that $n < (1 + e^{-600})k^2/e$. Let $\sigma \in [(1 + e^{-1000})k]^n$ be a sequence of length $n$. Then, $\sigma$ can only contain $o(k!)$ patterns in $S_k$.*

The proof of Theorem 1.2 appears in Section 3. We made no effort to optimize the constants.

# 2 Lower-bounding the length of a superpattern

In this section we will prove Theorem 1.1. We assume, as we may, that $k$ is odd, and we will generally omit floor symbols in quantities that may not be integers.

Let $\sigma \in S_n$ be a permutation of length $n < 1.000076k^2/e^2$. Our objective is to prove that $\sigma$ contains $o(k!)$ patterns. As outlined in the introduction, the idea is to define an efficient encoding; namely, an injective function $\phi$ from the set of all patterns in $\sigma$ into a set of size $o(k!)$.

To describe our encoding, let $\pi$ be a pattern of length $k$ that is contained in $\sigma$. Fix a set of indices $T = \{t_1, \ldots, t_k\}$, where $t_1 < \cdots < t_k$, such that the values of $\sigma$ on $t_1, \ldots, t_k$ form the pattern $\pi$. For even $1 < i < k$, define $b_i = t_{i+1} - t_{i-1}$ to be the *width* around $t_i$ (so there are $(k - 1)/2$ widths). Note that the widths only depend on the indices $t_i$ for odd $i$. Also, note that if one knows $t_{i-1}$ and $t_{i+1}$, then the number of possibilities for $t_i$ is $b_i - 1$. That is to say, if the width around $t_i$ is large, then specifying $t_i$ is very "expensive." Now, to define our encoding $\phi$ we split into cases. With foresight, define $c = 0.00075$ and $d = 8.180$.

**Case 1:** If fewer than $ck$ of the widths are at least $dn/k$, then there is not much to be gained by a creative encoding scheme, and we simply encode $\pi$ by the indices $t_1, \ldots, t_k$ (formally, set $\phi(\pi) = (t_1, \ldots, t_k)$).

**Case 2:** Otherwise, at least $ck$ of the widths are at least $dn/k$. Let $I$ be a set of exactly $ck$ even integers such that $b_i \geq dn/k$ for each $i \in I$ (the specific choice of $I$ can be basically arbitrary, but we make sure it only depends on the indices $t_i$ where $i$ is odd). Then, encode $\pi$ by specifying $\pi(i)$ for each $i \in I$, and specifying $t_i$ for each $i \notin I$. Formally, we can define $\phi(\pi)$ to be the triple $(I, (t_i)_{i \notin I}, (\pi(i))_{i \in I})$.

It is important to note that our encoding $\phi$ is injective: in both cases, it is possible to recover $\pi$ by knowing $\sigma$ and $\phi(\pi)$. To see this in Case 2, note that if we know $I$ and $(\pi(i))_{i \in I}$, then we know the set of values of $\pi(i)$ for $i \notin I$ (but not their relative order). This relative order is obtainable from the relative order of the $\sigma(t_i)$, for $i \notin I$.

Now, in order to prove Theorem 1.1, we need to show that there are not too many possibilities (substantially fewer than $\binom{n}{k}$) for the value of $\phi(\pi)$. The main technical ingredient is the fact that Case 1 is needed only rarely, as follows. Let $f = 0.999924$.

**Lemma 2.1.** *For all but $O\left(f^k \binom{n}{k}\right)$ of the size-$k$ subsets $T \subseteq [n]$, at least $ck$ of the widths are at least $dn/k$.*

Before proving Lemma 2.1, we first show how to use it to deduce Theorem 1.1 via the encoding $\phi$. We return to Lemma 2.1 in section Section 2.1.

*Proof of Theorem 1.1.* First, using the bound $n < (1.000076/e^2)k^2$, we can use Stirling's approximation to

show that $O(f^k)\binom{n}{k} = o(k!)$ for large $k$. So, Lemma 2.1 shows there are only $o(k!)$ possible outcomes of $\phi(\pi)$ arising from Case 1.

Next, we consider Case 2. Let $\Phi$ be the set of all outcomes of $\phi(\pi)$ obtainable from Case 2, so our objective is to show that $|\Phi| = o(k!)$. We first want to show that there are not many possibilities for the pair $(I, (t_i)_{i \notin I})$ in the definition of $\phi$. Intuitively, this should be the case because we specifically decided to "forget" the indices $t_i$ surrounded by large widths, which are the most expensive indices to remember.

Note that our definition of the widths of a pattern $\pi$ really only depended on the subset $T$ of indices where $\pi$ appeared. So, let $\mathcal{F}$ be the set of all size-$k$ subsets $T \subseteq [n]$ with at least $ck$ widths of size at least $dn/k$. For every $T = \{t_1, \ldots, t_k\} \in \mathcal{F}$, we can choose a set $I$ of $ck$ even integers such that the widths $b_i$, for $i \in I$, are at least $dn/k$ (here $I$ only depends on the odd-numbered indices of $T$, just as in the definition of $\phi$). Let $\Psi$ be the set of all possibilities for the pair $(I, (t_i)_{i \notin I})$, among all $T \in \mathcal{F}$.

Now, note that we can encode the sets in $\mathcal{F}$ using the encoding $\psi(T) = (I, (t_i)_{i \notin I}, (t_i)_{i \in I})$. Given any $(I, (t_i)_{i \notin I}) \in \Psi$, the number of ways to extend this to a valid encoding $\psi(T)$ is at least

$$\left(\frac{dn}{k} - 1\right)^{ck},$$

because for each $i \in I$ there are at least $dn/k - 1$ possibilities for $t_i$, given $t_{i-1}$ and $t_{i+1}$ (here we are using the restriction that $I$ only consists of even-numbered indices). We deduce that $|\Psi|(dn/k - 1)^{ck} \le |\mathcal{F}| \le \binom{n}{k}$, so $|\Psi| \le \binom{n}{k}(dn/k - 1)^{-ck}$.

Given any pair $(I, (t_i)_{i \notin I}) \in \Psi$, the number of ways to choose $(\pi(i))_{i \in I}$ to extend our pair to an encoding $\phi(\pi)$ of some $\pi$ is at most $k!/(k - ck)!$. So,

$$|\Phi| \le |\Psi|\frac{k!}{(k - ck)!}.$$

Combining this with our upper bound for $|\Psi|$, we compute that $|\Phi|$ is bounded above by

$$\frac{k!}{(k - ck)!}\left(\frac{dn}{k} - 1\right)^{-ck}\binom{n}{k} \le k^{O(1)}k!\left(\frac{1}{(k(1-c)/e)^{(1-c)k}}\left(\frac{dn}{k}\right)^{-ck}\left(\frac{ne}{k}\right)^k\right)$$

$$= k^{O(1)}k!\left(\left(\frac{n}{(k^2/e^2)}\right)^{1-c}e^c d^{-c}(1 - c)^{-(1-c)}\right)^k = o(k!),$$

as desired (recall that $n < 1.000076 k^2/e^2$, $c = 0.00075$ and $d = 8.180$). All in all, we have proved that our encoding $\phi$ can output only $o(k!)$ different values, so the number of distinct patterns in $\sigma$ must be $o(k!)$.

$\square$

## 2.1  Large gaps in random sets

In this section we prove Lemma 2.1, showing that for all but $O(f^k)\binom{n}{k}$ choices of $T$, at least $ck$ of the widths are at least $dn/k$ (where $d = 8.180$, $c = 0.00075$ and $f = 0.999924$). The details are a bit technical, but the intuition is fairly simple, as follows. If we choose $T$ randomly, then its elements approximately correspond to a Poisson point process with intensity $k/n$ in the interval $[0, n]$. So, the lengths of the gaps between the elements of $T$ are approximately exponentially distributed with mean $n/k$. Each of the widths is a sum of a pair of these gap lengths, so its distribution is approximately Gamma$(2, k/n)$. The distribution function of Gamma$(2, k/n)$ is $x \mapsto 1 - e^{-xk/n}(xk/n + 1)$ (for $x \ge 0$), so we expect about an $e^{-d}(d + 1) > 2c$ fraction of the $(k - 1)/2$ widths to have size at least $dn/k$.

*Proof of Lemma 2.1.* Pick $T$ uniformly at random. We show that the probability that fewer than $ck$ widths are at least $dn/k$ is at most $O(f^k)$.

4

Define the *gaps* $a_0, a_1, \ldots, a_k$ by $a_i = t_{i+1} - t_i$ (where $a_0 = t_1$ and $a_k = n+1-t_k$). Then the widths can be represented as $b_i = a_i + a_{i-1}$. Note that $(a_0, \ldots, a_k)$ is uniformly distributed over all sequences of $k+1$ positive integers summing to $n+1$.

Let $(X_0, \ldots, X_k) \in \mathbb{R}^k$ be a point uniformly distributed in the simplex given by

$$\Gamma = \{(x_0, \ldots, x_k) \ : \ x_0, \ldots, x_k \geq 0, \ x_0 + \cdots + x_k = n+1\}.$$

Let $a'_i = \lfloor X_i \rfloor$ be obtained by rounding down $X_i$, for each $0 \leq i \leq k$. Then $(a'_1, \ldots, a'_k)$ is stochastically dominated by $(a_1, \ldots, a_k)$, in the sense that the two sequences can be coupled such that we always have $a_i \geq a'_i$ for each $i$. Let $b'_i = a'_i + a'_{i-1}$, and let $B_i = X_i + X_{i-1}$. It suffices to show that with probability $1 - O(f^k)$, we have $B_i \geq dn/k + 2$ (therefore $b'_i \geq dn/k$) for at least $ck$ different $i$.

Now, it is known that the distribution of $(X_0, \ldots, X_k)$ is identical to the distribution of

$$\frac{n+1}{\xi_0 + \cdots + \xi_k}(\xi_0, \ldots, \xi_k),$$

where $\xi_i$ are i.i.d. exponential random variables with rate 1 (see for example [6, Section 4.1]). Let $d' = 8.282 > d$. By a Chernoff bound for sums of exponential random variables (see for example [10]), we have

$$\Pr\left[\xi_0 + \cdots + \xi_k \geq \frac{d'}{d}(k+1)\right] \leq \exp\left(-(k+1)\left(\frac{d}{d'} - 1 - \ln\left(\frac{d}{d'}\right)\right)\right) \leq O(f^k).$$

Also, each of the $\xi_{i-1} + \xi_i$ (for even $i$) are i.i.d. with $\Pr(\xi_{i-1} + \xi_i \geq x) = e^{-x}(1+x)$. Take $d'' = 8.283$ slightly larger than $d'$, and let $p = e^{-d''}(1+d'') > 2c$, so that for each $i$ we have $\Pr(\xi_{i-1} + \xi_i \geq d'') \geq p$. Thus, by a Chernoff bound for the binomial distribution (see for example [2, Theorem A.1.13]), the probability that $\xi_{i-1} + \xi_i \geq d''$ for fewer than $ck$ different $i$ is at most

$$\Pr[\mathrm{Bin}((k-1)/2, p) \leq ck] \leq \exp\left(\frac{-(p(k-1)/2 - ck)^2}{2 \cdot p(k-1)/2}\right) \leq O(f^k).$$

Combining the above two bounds, we conclude that with probability $1 - O(f^k)$, we have

$$B_i = \frac{\xi_i + \xi_{i-1}}{\xi_1 + \cdots + \xi_k} \geq \frac{d''}{d'} \cdot \frac{d(n+1)}{k+1} \geq \frac{dn}{k} + 2,$$

for at least $ck$ different $i$, as desired. $\qquad\square$

## 3  Lower-bounding the length of a universal sequence

In this section we prove Theorem 1.2. It actually suffices to prove the following seemingly weaker result, where we do not allow the alphabet size to be greater than $k$.

**Proposition 3.1.** *Suppose $\sigma \in [k]^n$, where $n < (1 + e^{-600})k^2/e$. Then, for sufficiently large $k$, the number of $k$-patterns in $\sigma$ is at most $\exp(-e^{-600}k)\,k!$.*

Before proving Proposition 3.1, we deduce Theorem 1.2 from it.

*Proof of Theorem 1.2.* Suppose $k$ is sufficiently large, and let $t = e^{-1000}k$. Consider $\sigma \in [k+t]^n$. We call the elements of $[k+t]$ *symbols*. Every pattern $\pi \in S_k$ that appears in $\sigma$ uses some set $Y$ of $k$ symbols, and therefore it appears in the subsequence $\sigma_Y \in Y^{n'}$ of $\sigma$ obtained by keeping only the symbols in $Y$.

Now, by Proposition 3.1, for each of the $\binom{k+t}{k}$ choices of $Y$, there are at most $\exp(-e^{-600}k)k!$ patterns in $\sigma_Y$. Thus, the total number of patterns contained in $\sigma$ is at most

$$\binom{k+t}{k} \exp(-e^{-600}k)\,k! \leq \exp(-e^{-600}k)\frac{(k+t)^{k+t}}{k^k t^t}\,k!.$$

Recalling that $t/k = e^{-1000}$, one can verify that this expression is $o(k!)$, as desired. $\qquad\square$

5

Now we proceed to the proof of Proposition 3.1. Let $\sigma \in [k]^n$ be a sequence of length $n < (1+e^{-600})k^2/e$, and let $a_1, \ldots, a_k$ be the numbers of occurrences of the symbols $1, \ldots, k$ in $\sigma$. Recall from the introduction that we have a trivial upper bound of $a_1 a_2 \ldots a_k$ on the number of patterns $\pi \in S_k$ which appear in $\sigma$; we would like to improve on this by designing an efficient encoding scheme $\phi$ for the patterns which appear in $\sigma$.

Let $\pi \in S_k$ be a pattern that is contained in $\sigma$, and consider indices $t_1, \ldots, t_k$ which represent $\pi$ in $\sigma$, such that $\sigma(t_i) = i$ for each $i$. (This is different from the proof of Theorem 1.1 where $(t_1, \ldots, t_k)$ was an increasing sequence of indices, but note that $(t_1, \ldots, t_k)$ still uniquely determines $\pi$.) Let $c = e^{-290}$. Now, we encode $\pi$ as follows: first, we specify $t_1, \ldots, t_{k-ck}$. Then, instead of specifying $t_m$ for $k - ck < m \le k$, we will instead specify the relative position of $t_m$, with respect to all the other $t_i$. Formally, we let $\psi(m)$ be the binary vector $(\mathbf{1}_{t_m < t_i})_{i < m}$ indicating the position of $t_m$ relative to all $t_i$ with $i < m$, and define $\phi(\pi)$ to be the pair $((t_i)_{i \le k-ck}, (\psi(m))_{m > k-ck})$. This encoding is injective.

To see why this encoding should be more efficient than the trivial one, it is helpful to consider the extreme case where we first specify $t_1, \ldots, t_{k-1}$. Typically, the occurrences of the last symbol $k$ are not going to be perfectly distributed between the $t_i$, and there are multiple possible outcomes of $t_k$ which have the same position $\psi(k)$ relative to the other $t_i$. That is to say, specifying the relative position $\psi(k)$ of $t_k$ should be cheaper than specifying $t_k$ exactly.

To make precise the above intuition, we need to make some definitions. Fix $k - ck < m \le k$, and suppose that we have specified indices $t_1, \ldots, t_{k-ck}$, which divide the interval $[n]$ into $k - ck + 1$ disjoint subintervals. Suppose that in $\sigma$, the symbol $m$ appears in positions $s_1 < \ldots < s_{a_m}$. Then the number of possible positions to place $t_m$ relative to $t_1, \ldots, t_{k-ck}$ equals the number of subintervals that contain some $s_j$. Equivalently, this number of possible positions is one more than the number of adjacent pairs $s_j, s_{j+1}$ which are "split" by some $t_i$ between them. In this case, we say that the pair $s_j, s_{j+1}$ constitutes an *m-split with respect to* $t_i$. Our main technical ingredient is an upper bound on the number of $m$-splits for $k - ck < m \le k$. We also give ourselves the freedom to permute the symbols in $\sigma$, which allows us to choose the most convenient $ck$ symbols as the "last ones."

**Lemma 3.2.** *Consider $\sigma \in [k]^n$ with $n < (1 + e^{-600})k^2/e$. Either the conclusion of Proposition 3.1 holds for trivial reasons, or else it is possible to permute the symbols of $\sigma$ such that the following two conditions hold.*

(1) *For all $k - ck < m \le k$, we have $a_m \ge 0.1k$.*

(2) *For all but $k^{O(1)} \exp(-e^{-560}k) a_1 \ldots a_{k-ck}$ choices of $t_1, \ldots, t_{k-ck}$ (where each $\sigma(t_i) = i$), the following key property holds: For each $m > k - ck$, the total number of $m$-splits with respect to $t_1, \ldots, t_{k-ck}$ is at most $a_m(1 - e^{-280})$.*

Since the proof of Lemma 3.2 is rather technical, we defer it to the end of this section after deducing Proposition 3.1.

*Proof of Proposition 3.1.* We will assume that the symbols of $\sigma$ have been permuted to satisfy Lemma 3.2. Our main goal is to show that there are at most $k^{O(1)} \exp(-e^{-580}k) a_1 \ldots a_k$ possible outcomes of the encoding $\phi(\pi)$; we will then be able to deduce the desired bound by convexity.

Consider a pattern $\pi \in S_k$ appearing in $\sigma$ at indices $t_1, \ldots, t_k$. First, we deal with the case where $t_1, \ldots, t_{k-ck}$ do not satisfy the key property in Lemma 3.2. There are at most $\exp(-e^{-560}k) a_1 \ldots a_{k-ck}$ such possibilities for $t_1, \ldots, t_{k-ck}$. Then, we just trivially observe that there are at most $a_m$ ways to choose $t_m$, for all $m > k - ck$. So, the total number of possibilities for $\phi(\pi)$ among such $\pi$ is at most $k^{O(1)} \exp(-e^{-560}k) a_1 \ldots a_k$, which is substantially less than our target.

Next, we consider the case where $t_1, \ldots, t_{k-ck}$ do satisfy our key property. This implies that for each $m > k - ck$, there are at most $a_m(1 - e^{-280}) + 1$ distinct possibilities for the position of $t_m$ relative to $t_1, \ldots, t_{k-ck}$. If we additionally want to specify the position of $t_m$ relative to $t_{k-ck+1}, \ldots, t_{m-1}$, there are at most $m - (k - ck + 1)$ additional ways to make this choice. So, using condition (1) of Lemma 3.2, and

6

recalling that $c = e^{-290}$, the total number of possibilities for the relative position $\psi(m)$ of $t_m$ is at most

$$a_m(1 - e^{-280}) + m - (k - ck) \le a_m(1 - e^{-280}) + ck$$
$$\le a_m(1 - e^{-280} + 10c)$$
$$< a_m(1 - e^{-290}).$$

Thus, since there are at most $a_i$ ways to pick $t_i$ for each $i \le k - ck$, the number of possibilities for $\phi(\pi)$ among $\pi$ satisfying our key property is at most

$$\left(\prod_{i=1}^{k-ck} a_i\right)\left(\prod_{m>k-ck} a_m(1 - e^{-290})\right) = (1 - e^{-290})^{ck} a_1 \ldots a_k$$
$$\le \exp(-e^{-290}ck)a_1 \ldots a_k$$
$$\le \exp(-e^{-580}k)a_1 \ldots a_k.$$

All in all, accounting for both cases (when the key property is satisfied, and when it is not), the total number of possibilities for $\phi(\pi)$ is at most

$$\exp(-e^{-580}k)a_1 \ldots a_k + k^{O(1)}\exp(-e^{-560}k)a_1 \ldots a_k \le k^{O(1)}\exp(-e^{-580}k)a_1 \ldots a_k.$$

We have $a_1 + \cdots + a_k = n$, so by convexity, the number of possibilities for $\pi$ is at most

$$k^{O(1)}\exp(-e^{-580}k)\left(\frac{n}{k}\right)^k = k^{O(1)}\left(\exp(-e^{-580})\frac{n}{k^2/e}\right)^k k! \le \exp(-e^{-600}k)k!,$$

for sufficiently large $k$, as desired. This completes the proof of Proposition 3.1. $\qquad\square$

## 3.1 Few splits for random indices

In this section we prove Lemma 3.2. First, we want to be able to assume that most symbols occur fairly often. Call a symbol $m$ *common* if $a_m > 0.1k$.

**Lemma 3.3.** *Consider $\sigma \in [k]^n$ with $n < (1 + e^{-600})k^2/e$. Either the conclusion of Proposition 3.1 holds for trivial reasons, or else at least $0.99k$ symbols are common.*

*Proof.* Suppose without loss of generality that $1, \ldots, 0.01k$ are all not common. Then the number of $k$-patterns in $\sigma$ is at most

$$a_1 \ldots a_k \le (0.1k)^{0.01k} a_{0.01k+1} \ldots a_k \le (0.1k)^{0.01k}\left(\frac{n}{0.99k}\right)^{0.99k},$$

by convexity, since $a_{0.01k+1} + \cdots + a_k \le n$. We can rewrite this latter expression as

$$\left(\frac{0.1^{0.01}}{0.99^{0.99}}\right)^k k^k \left(\frac{n}{k^2}\right)^{0.99k} \le k^k \left(\frac{0.988n}{k^2}\right)^k \le \left(\frac{0.988n}{k^2/e}\right)^k k! \le \exp(-e^{-600}k)\, k!,$$

so the conclusion of Proposition 3.1 holds. $\qquad\square$

Next we will need a structural lemma. Suppose that in $\sigma$, the symbol $m$ appears in positions $s_1 < \ldots < s_{a_m}$. An *$m$-gap* is the (possibly empty) interval between a pair of adjacent indices $s_j, s_{j+1}$. We say that an $m$-gap is *full* if there exists some symbol $m' \ne m$ for which the gap contains at least $0.9a_{m'}$ occurrences of $m'$. In this case we say that the $m$-gap is *filled* by $m'$. See Fig. 1 for an example illustrating full gaps.

**Lemma 3.4.** *Consider $\sigma \in [k]^n$ with $n < (1 + e^{-600})k^2/e$, and suppose that at least $0.99k$ symbols are common. Then there are at least $0.03k$ common $m$ for which the number of full $m$-gaps is less than $0.9a_m$.*

$$
\underbrace{1 \quad \overbrace{2 \quad 3 \quad 3 \cdots 3 \quad 2}^{\text{2-gap filled by 3}} \quad 2 \quad \overbrace{1 \quad \underbrace{2 \quad 2}_{\text{not full}} \cdots 2 \quad \underbrace{2 \quad 3 \quad 2}_{\text{not full}}}^{\text{1-gap filled by 2}} \quad 3 \quad 1}_{\text{1-gap filled by 3}}
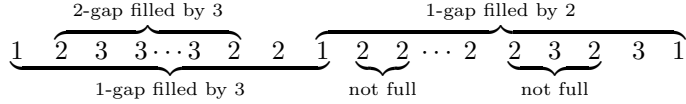$$

Figure 1: An example showing gaps that are full and not full

*Proof.* Consider some common $m$ for which the number of full $m$-gaps is at least $0.9a_m > 0.09k$ (if there is no such $m$ then we are immediately done). Note that each $m'$ can only fill at most one $m$-gap, so at least $0.9a_m - 0.01k > 0.08k$ of these full $m$-gaps are filled by different common symbols. Let $S$ be a set of $0.08k$ common symbols $m'$ which each fill a different $m$-gap.

Now, for any $m' \in S$, inside the $m$-gap filled by $m'$, there are at least $0.9a_{m'}$ instances of $m'$. These instances themselves form at least $0.9a_{m'} - 1$ different $m'$-gaps; denote the set of such $m'$-gaps by $G_{m'}$. Now, note that all the different gaps in $\bigcup_{m' \in S} G_{m'}$ are pairwise disjoint, so at most $k$ of them can be full (each of the $k$ symbols can fill at most one of these gaps). It follows that for at least $|S| - k/20 = 0.03k$ of the symbols $m' \in S$, at most 20 of the gaps in $G_{m'}$ are full. For each of these symbols $m'$, the total number of full $m'$-gaps is at most $0.1a_{m'} + 20$, because there are at most 20 full $m'$-gaps in $G_{m'}$, and $G_{m'}$ contains at least $0.9a_{m'} - 1$ of the $a_{m'} - 1$ different $m'$-gaps that exist. Since each such $m'$ is common, we have $a_{m'} > 0.1k$, so $0.1a_{m'} + 20 < 0.9a_{m'}$ for large $k$. The desired conclusion follows. $\square$

Now we are ready to prove Lemma 3.2.

*Proof of Lemma 3.2.* Consider $\sigma \in [k]^n$ with $n < (1 + e^{-600})k^2/e$. By Lemma 3.3 and Lemma 3.4, after a permutation of the symbols we may assume that the last $0.03k$ symbols $m > 0.97k$ are all common, and each have the property that the number of full $m$-gaps is less than $0.9a_m$. It is already clear that condition (1) of the lemma statement is satisfied (recall that $c = e^{-290} < 0.03$).

Now, for each $i \leq k - ck$, let $t_i$ be a uniformly random index satisfying $\sigma(t_i) = i$ (independently for each $i$). Fix some $m > k - ck$, and let $X$ be the random variable counting the number of $m$-splits with respect to $t_1, \ldots, t_{k-ck}$. It suffices to show that with probability at least $1 - k^{O(1)} \exp(-e^{-560}k)$, we have $X \leq a_m(1 - e^{-280})$. We will then be able to take the union bound over all $m > ck$.

As before, suppose that in $\sigma$, the symbol $m$ appears in positions $s_1, \ldots, s_{a_m}$. Let $J$ be the set of all $j$ such that $s_j$ and $s_{j+1}$ do not span a full $m$-gap. Then, $|J| \geq 0.1a_m$ (since we are assuming that the number of full $m$-gaps is less than $0.9a_m$). Also, for each $i \leq k - ck$ and $j \in J$, let $b_{ij}$ be the number of occurrences of $i$ in the $m$-gap between $s_j$ and $s_{j+1}$. Note that since the gap between $s_j$ and $s_{j+1}$ is not full, we always have $b_{ij}/a_i \leq 0.9$.

For any $j \in J$, there is an $m$-split at $s_j, s_{j+1}$ precisely when the $m$-gap between $s_j, s_{j+1}$ contains one of $t_1, \ldots, t_{k-ck}$. The probability that this event does not occur is

$$
\prod_{i=1}^{k-ck} \left(1 - \frac{b_{ij}}{a_i}\right) \geq \exp\left(-2.6 \sum_{i=1}^{k-ck} \frac{b_{ij}}{a_i}\right),
$$

where the inequality follows from the fact that each $b_{ij}/a_i \leq 0.9$. We deduce that

$$
\mathbb{E}\,X \leq a_m - 1 - \sum_{j \in J} \exp\left(-2.6 \sum_{i=1}^{k-ck} \frac{b_{ij}}{a_i}\right).
$$

Recalling that $\sum_j b_{ij} \leq a_i$ for each $i$, and using convexity, we further deduce that

$$\mathbb{E}\,X \leq a_m - 1 - |J|\exp\left(-2.6\frac{1}{|J|}\sum_{j \in J}\sum_{i=1}^{k-ck}\frac{b_{ij}}{a_i}\right) \leq a_m - |J|\exp\left(-2.6\frac{1}{|J|}\sum_{i=1}^{k-ck}\sum_{j \in J}\frac{b_{ij}}{a_i}\right)$$
$$\leq a_m - |J|\exp\left(\frac{-2.6(k-ck)}{|J|}\right).$$

Now, recall that $|J| \geq 0.1a_m \geq 0.01k$ (since $m$ is common), so we can compute that $\mathbb{E}\,X < a_m\left(1 - e^{-270}\right)$. Finally, note that if any $t_i$ (for $1 \leq i \leq k-ck$) is changed, then this can cause the number of $m$-splits $X$ to increase or decrease by at most one. By the Azuma–Hoeffding inequality (see for example [2, Theorem 7.2.1]), we conclude that

$$\Pr(X \geq a_m(1 - e^{-280})) \leq \exp\left(-\frac{(a_m(e^{-270} - e^{-280}))^2}{2(k-ck)}\right) \leq \exp(-e^{-560}k),$$

as desired (we have used that $a_m \geq 0.1k$, since $m$ is common). $\qquad\square$

# 4   Concluding remarks

In this paper we have introduced a new method to prove lower bounds for universality-type problems, by identifying local inefficiencies in trivial encoding schemes. We have used this method to improve bounds on two different problems. It would be very interesting to find further applications of this idea. For example, Alon showed that the minimum number of vertices in a $k$-induced-universal graph is asymptotic to the trivial lower bound, but he observed that the trivial lower bound is not exactly tight, and raised the question of better understanding lower-order terms (see [1, Section 5]). There are also many other problems about universality in graphs (for example, universality with respect to containment of trees or bounded-degree graphs; see for example [3]), where it is not yet known whether trivial lower bounds are asymptotically tight.

However, there seem to be some difficulties in applying our methods to graph problems. Roughly speaking, the reason that we were able to obtain improvements in the setting of permutation patterns is as follows. For a pattern $\pi \in S_k$, the amount of information (entropy) carried by a single value $\pi(t)$ is about $\log(k/e) = \log k - 1$ "on average" (because the information carried by $\pi$ itself is $\log(k!) \approx k\log(k/e)$). Taking a different point of view, there are $k$ possibilities for $\pi(k)$, so when viewed in isolation, the amount of information carried by $\pi(k)$ is $\log k$. The first point of view is relevant for computing trivial lower bounds, and the second point of view is relevant for improving local inefficiencies. It was important for our proof strategy that these two points of view gave very similar answers. However, in most graph problems, these two points of view tend to give quite different answers: specifying the adjacencies of a single vertex requires much more information than the "average information per vertex," and therefore our methods do not seem to be directly applicable to graphs. It would be interesting to investigate this further.

On the subject of superpatterns, obviously there is still a large gap between our new bound and the upper bound $(1/2 + o(1))k^2$ obtained by Miller. It should be clear from the proof of Theorem 1.1 that it is possible to make various small improvements to our lower bound: for example, it was convenient to restrict our attention to widths $t_{i+1} - t_{i-1}$ only for even $i$, but with a more sophisticated argument one could take both even and odd $i$ into account. Also, the bounds in Lemma 2.1 were rather crude, and presumably one could prove exact large deviation bounds for the number of widths above a given threshold. However, it would be very complicated to fully optimize all aspects of our argument, and it seems unlikely that one could prove a lower bound much larger than $k^2/e^2$ without substantial new ideas. At present, we do not have a strong conjecture for the true minimal length of a $k$-superpattern.

It is worth mentioning some related problems which may be more tractable, and may shed light on the true minimal length of a $k$-superpattern. For example, instead of demanding that our permutation $\sigma$ contains every pattern of length $k$, we can ask for permutations that contain *almost* every pattern of length $k$. Stronger upper bounds are known in this case: a construction by Eriksson, Eriksson, Linusson and

Wastlund [8] captures all but an exponentially small proportion of $k$-patterns in a permutation of length $(1 + o(1))k^2/4$. We remark that this construction represents an obstruction to counting-based arguments: even if Miller's upper bound of $(1 + o(1))k^2/2$ turns out to be best-possible (as conjectured in [8]), a proof of this would have to be sensitive to the difference between containing *all* patterns, and containing *almost all* of them. As suggested by He and Kwan [9], it would also be very interesting to explore for which $n$ a *random* permutation of length $n$ contains almost all $k$-patterns. It could be that this holds for $n$ quite close to $k^2/e^2$, and by analogy to Alon's study of universal graphs, this would suggest that there is also a $k$-superpattern of approximately the same length.

The study of pattern containment in random permutations is also of independent interest. It is a celebrated result in probability theory that the longest increasing subsequence in a random permutation of length $n$ is typically about $2\sqrt{n}$ (see for example [18]). This tells us that the "threshold" value of $n$, above which a random $n$-permutation is likely to contain the increasing pattern $1\,2\ldots k \in S_k$, is approximately $k^2/4$. It would be interesting to understand how the threshold for containment of a pattern $\pi \in S_k$ depends on the structure of $\pi$. Alon (see [4]) conjectured that this threshold is never more than $(1/4 + o(1))k^2$; in fact he conjectured that $(1/4 + o(1))k^2$ is the threshold for being a $k$-superpattern. The best known bounds are still quite far from this conjecture: recently He and Kwan [9] proved that a random permutation of length $n = 2000k \log \log k$ is typically a $k$-superpattern.

Similar considerations are also relevant for the study of $k$-universal $r$-ary sequences. Though Kleitman and Kwiatkowski [12] found the asymptotics of the minimal length of a $k$-universal $k$-ary sequence, it is not obvious how the situation changes if $r > k$ or if one only requires containment of almost all patterns. There may also be interesting related problems about random sequences.

**Acknowledgments:** we would like to thank Noga Alon and Xiaoyu He for insightful discussions.

# References

[1]  Noga Alon. "Asymptotically optimal induced universal graphs". In: *Geom. Funct. Anal.* 27.1 (2017), pp. 1–32. ISSN: 1016-443X. DOI: 10.1007/s00039-017-0396-9.

[2]  Noga Alon and Joel H. Spencer. *The probabilistic method*. Fourth. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 2016, pp. xiv+375. ISBN: 978-1-119-06195-3.

[3]  Stephen Alstrup et al. "Adjacency labeling schemes and induced-universal graphs". In: *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*. ACM, New York, 2015, pp. 625–634.

[4]  Richard Arratia. "On the Stanley-Wilf conjecture for the number of permutations avoiding a given pattern". In: *Electron. J. Combin.* 6 (1999), Note, N1, 4.

[5]  Miklós Bóna. *Combinatorics of permutations*. Discrete Mathematics and its Applications (Boca Raton). With a foreword by Richard Stanley. Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. xiv+383. ISBN: 1-58488-434-7. DOI: 10.1201/9780203494370.

[6]  Luc Devroye. *Nonuniform random variate generation*. Springer-Verlag, New York, 1986, pp. xvi+843. ISBN: 0-387-96305-7. DOI: 10.1007/978-1-4613-8643-8.

[7]  Michael Engen and Vincent Vatter. *Containing all permutations*. 2018. arXiv: 1810.08252 [math.CO].

[8]  Henrik Eriksson et al. "Dense packing of patterns in a permutation". In: *Ann. Comb.* 11.3-4 (2007), pp. 459–470. ISSN: 0218-0006. DOI: 10.1007/s00026-007-0329-7.

[9]  Xiaoyu He and Matthew Kwan. *Universality of random permutations*. 2019. arXiv: 1911.12878 [math.CO].

[10]  Svante Janson. "Tail bounds for sums of geometric and exponential variables". In: *Statist. Probab. Lett.* 135 (2018), pp. 1–6. ISSN: 0167-7152. DOI: 10.1016/j.spl.2017.11.017.

[11]  Sergey Kitaev. *Patterns in permutations and words*. Monographs in Theoretical Computer Science. An EATCS Series. With a foreword by Jeffrey B. Remmel. Springer, Heidelberg, 2011, pp. xxii+494. ISBN: 978-3-642-17332-5; 978-3-642-17333-2. DOI: 10.1007/978-3-642-17333-2.

[12] D. J. Kleitman and D. J. Kwiatkowski. "A lower bound on the length of a sequence containing all permutations as subsequences". In: *J. Combinatorial Theory Ser. A* 21.2 (1976), pp. 129–136. ISSN: 0097-3165. DOI: 10.1016/0097-3165(76)90057-1.

[13] Donald E. Knuth. *The art of computer programming. Vol. 1.* Fundamental algorithms, Third edition [of MR0286317]. Addison-Wesley, Reading, MA, 1997, pp. xx+650. ISBN: 0-201-89683-4.

[14] Percy A. MacMahon. *Combinatory analysis. Vol. I, II (bound in one volume).* Dover Phoenix Editions. Reprint of ıt An introduction to combinatory analysis (1920) and ıt Combinatory analysis. Vol. I, II (1915, 1916). Dover Publications, Inc., Mineola, NY, 2004, pp. ii+761. ISBN: 0-486-49586-8.

[15] Alison Miller. "Asymptotic bounds for permutations containing many different patterns". In: *J. Combin. Theory Ser. A* 116.1 (2009), pp. 92–108. ISSN: 0097-3165. DOI: 10.1016/j.jcta.2008.04.007.

[16] J. W. Moon. "On minimal $n$-universal graphs". In: *Proc. Glasgow Math. Assoc.* 7 (1965), 32–33 (1965). ISSN: 2040-6185.

[17] R. Rado. "Universal graphs and universal functions". In: *Acta Arith.* 9 (1964), pp. 331–340. ISSN: 0065-1036. DOI: 10.4064/aa-9-4-331-340.

[18] Dan Romik. *The surprising mathematics of longest increasing subsequences.* Vol. 4. Institute of Mathematical Statistics Textbooks. Cambridge University Press, New York, 2015, pp. xi+353. ISBN: 978-1-107-42882-9; 978-1-107-07583-2.

[19] Rodica Simion and Frank W. Schmidt. "Restricted permutations". In: *European J. Combin.* 6.4 (1985), pp. 383–406. ISSN: 0195-6698. DOI: 10.1016/S0195-6698(85)80052-4.

[20] Richard P. Stanley. "Increasing and decreasing subsequences and their variants". In: *International Congress of Mathematicians. Vol. I.* Eur. Math. Soc., Zürich, 2007, pp. 545–579. DOI: 10.4171/022-1/21.

[21] Einar Steingrímsson. "Some open problems on permutation patterns". In: *Surveys in combinatorics 2013.* Vol. 409. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 2013, pp. 239–263.

[22] Bridget Tenner. *Database of Permutation Pattern Avoidance.* https://math.depaul.edu/~bridget/patterns.html.