New Bounds for the Optimal Density of Covering Single-Insertion Codes via the Turán Density

Oleg Pikhurko, Oleg Verbitsky, and Maksim Zhukovskii

***Abstract*—We prove that the density of any covering single-insertion code $C \subseteq X^r$ over the n-symbol alphabet X cannot be smaller than $1/r + \delta_r$ for some positive real δ_r not depending on n. This improves the volume lower bound of $1/(r+1)$. On the other hand, we observe that, for all sufficiently large r, if n tends to infinity then the asymptotic upper bound of $7/(r+1)$ due to Lenz et al. (2021) can be improved to $4.911/(r+1)$. Both the lower and the upper bounds are achieved by relating the code density to the Turán density from extremal combinatorics. For the last task, we use the analytic framework of measurable subsets of the real cube $[0,1]^r$.**

***Index Terms*—Covering insertion codes, Turán systems.**

I. Introduction

Let $r < k$ be positive integers and X be a (not necessarily finite) set. We say that a sequence $x \in X^r$ *covers* a sequence $a \in X^k$ if x is a subsequence of a, i.e., if x is obtainable by removing $k-r$ elements from a (while keeping the ordering of the remaining elements). We say that a set $C \subseteq X^r$ *covers* a set $A \subseteq X^k$ if every sequence in A is covered by at least one sequence in C.

Definition 1: A set $C \subseteq X^r$ covering X^k is called a *covering $(k-r)$-insertion code* over X. If $X = [n]$, where we denote $[n] = \{0, 1, \dots, n-1\}$, we speak of a covering code over the n-symbol alphabet. The minimum possible cardinality of such a code will be denoted by $S(n,k,r)$.

Example 1 (Grozea [9]:) $S(3,4,3) = 12$ and the unique, up to renaming the symbols, optimal code consists of the sequences $(0,0,0)$, $(1,1,1)$, $(2,2,2)$, $(0,0,1)$, $(0,1,0)$, $(1,0,0)$, $(1,1,2)$, $(1,2,1)$, $(2,1,1)$, $(2,2,0)$, $(2,0,2)$, $(0,2,2)$.

It is not hard to show (see Section II) that for each k and r, the optimal density $S(n,k,r)/n^r$ of a code converges to a limit $s(k,r)$ as n increases and that

$$S(n,k,r)/n^r \geq s(k,r) \tag{1}$$

for all n. This motivates estimating the limit value $s(k,r)$, especially because determining the exact values of $S(n,k,r)$ is computationally infeasible even for relatively small parameters n, k, and r (cf. [9] where the exact values of $S(n,4,3)$ are determined for $n \leq 5$).

We are especially interested in single-insertion codes, that is, in the case of $k = r+1$. As observed by various researchers (e.g. [13, Eq. (2)] for $n = 2$ and [22, Lemma 4.1] for general n), every sequence $x \in [n]^r$ covers exactly $(r+1)(n-1)+1$ sequences in $[n]^{r+1}$, which immediately yields

$$S(n, r+1, r) \geq \frac{n^{r+1}}{(r+1)(n-1)+1}.$$

On the other hand, Lenz et al. [12] proved that

$$S(n, r+1, r) \leq \frac{7n^{r+1}}{(r+1)(n-1)+1}.$$

These estimates readily imply that

$$\frac{1}{r+1} \leq s(r+1, r) \leq \frac{7}{r+1}. \tag{2}$$

In the present paper, we aim at improving the lower and the upper bound in (2).

We begin with addressing the question whether or not the lower bound in (2) is sharp. The equality $s(r+1,r) = 1/(r+1)$ would mean the existence of asymptotically perfect covering single-insertion codes. Our first estimate rules out this possibility by showing that

$$s(r+1, r) \geq \frac{1}{r}. \tag{3}$$

We prove the lower bound (3) in a natural analytic framework of measurable covering single-insertion codes over the real segment $[0,1]$. This framework is useful for establishing a relationship between covering codes and Turán systems, the classical and actively studied subject in combinatorics [10], [18], [20]. Of crucial importance for us is the concept of the extremal Turán density $t(k,r)$ (see Section V for the definition). We notice that

$$s(k,r) \leq t(k,r) \tag{4}$$

and, therefore, any upper bound for $t(k,r)$ yields also an upper bound for $s(k,r)$. The currently best upper bounds for the Turán density $t(r+1,r)$ have recently been obtained by Pikhurko [16] who proved that $t(r+1,r) \leq 6.239/(r+1)$ for all r and $t(r+1,r) \leq 4.911/(r+1)$ for all sufficiently large r. Both of these bounds imply an improvement of the upper bound in (2). These improvements can also be derived

Received 17 September 2024; revised 21 January 2025; accepted 23 March 2025. Date of publication 3 April 2025; date of current version 21 May 2025. The work of Oleg Pikhurko was supported by the European Research Council (ERC) Advanced Grant 101020255. The work of Oleg Verbitsky was supported by the German Research Foundation (DFG) under Grant KO 1053/8–2. *(Corresponding author: Oleg Pikhurko.)*

Oleg Pikhurko is with the Mathematics Institute and DIMAP, University of Warwick, CV4 7AL Coventry, U.K. (e-mail: pikhurko@gmail.com).

Oleg Verbitsky is with the Institut für Informatik, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, on leave from IAPMM, Lviv, Ukraine.

Maksim Zhukovskii is with the School of Computer Science, The University of Sheffield, S1 4DP Sheffield, U.K.

Communicated by X. Zhang, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2025.3557393

by a more careful analysis of the construction in [12].[1] In any case, it is remarkable that the state-of-the-art upper bounds for $s(r+1,r)$ are actually provided by the available upper bounds for $t(r+1,r)$.

Somewhat surprisingly, we obtain a relation between $s(r+1,r)$ and $t(r+1,r)$ also in the other direction: Any lower bound for $t(r+1,r)$ better than $1/r$ implies a lower bound for $s(r+1,r)$ also better than $1/r$. Lower bounds $t(r+1,r) \geq 1/r + \epsilon_r$ for $\epsilon_r > 0$ are obtained by Chung and Lu [4] and Lu and Zhao [15] and, therefore, our initial lower bound (3) can be further improved to $s(r+1,r) \geq 1/r + \delta_r$ for some $\delta_r > 0$. Though we provide explicit values of δ_r in the main body of the paper, right now we prefer to summarize the new bounds for the optimal density of covering single-insertion codes, improving the current bounds (2), in a somewhat simplified form.

Theorem 1: For $s(k,r) = \lim_{n\to\infty} S(n,k,r)/n^r$,

$$\frac{1}{r} < s(r+1,r) \leq \frac{4.911}{r+1}$$

where the former inequality is true for all r and the latter inequality is true for all sufficiently large r.

Taking into account the inequality (1), note that the lower bound stated in Theorem 1 is not just asymptotic, as it yields a lower bound $S(n,k,r)/n^r \geq 1/r + \delta_r$ for some real δ_r not depending on n (an explicit value of δ_r will be specified in the sequel). Covering codes over large alphabets naturally arise in research driven by applications in computational biology and genomics. Notably, DNA and RNA sequences are constructed from five canonical nucleobases: A, C, G, T, and U. Furthermore, the genetic code of life involves 22 proteinogenic amino acids, while over 500 amino acids are known to occur in nature. For results concerning the Hamming metric, we refer to [11], while results for the Levenshtein metric (which can be described using insertion/deletion codes) can be found in [3]. Covering insertion/deletion codes over arbitrarily large alphabets have also been studied in [1] and [2] in the context of the MapReduce framework for data analytics.

The paper is organized as follows. The convergence of the optimal code density $S(n,k,r)/n^r$ to a limit $s(k,r)$ is showed in Section II. An analytic framework for estimation of $s(k,r)$ is suggested in Section III. Our first lower bound (3) is established in Section IV. In Section V we introduce Turán systems and prove the relation (4), thereby obtaining the upper bound in Theorem 1 (restated as Corollary 2). A reverse relation between $s(r+1,r)$ and $t(r+1,r)$ is proved in Section VI as Theorem 5 and Corollary 3, which allows us to improve (3) to a strict inequality stated in Theorem 1 in a simplified form and made more precise in Corollaries 4 and 5. Note that the proof of Theorem 5 is heavily based on the argument used in Section IV for obtaining the bound in (3).

II. Preliminary Lemmas

Given a function $f : Y \to X$, we define a function $f^r : Y^r \to X^r$ by $f^r(y_1,\dots,y_r) = (f(y_1),\dots,f(y_r))$. The preimage of a set $C \subseteq X^r$ under f^r will be denoted by $f^{-r}(C)$.

[1]Using the same argument as in [16, Lemma 2.3], it can be showed that [12, Lemma 8] in fact holds if $\mu_I \leq 4.911$ and q is large enough.

Lemma 1: If $C \subseteq X^r$ covers X^k and f is an arbitrary function from Y to X, then $f^{-r}(C)$ covers Y^k.

Proof: Consider an arbitrary $(y_1,\dots,y_k) \in Y^k$ and denote $(x_1,\dots,x_k) = f^k(y_1,\dots,y_k)$. Since C covers X^k, some r-dimensional projection of $(x_1,\dots,x_k)$ belongs to C. Let, say, $(x_1,\dots,x_r) \in C$. It remains to note that $(y_1,\dots,y_k)$ is covered by the vector $(y_1,\dots,y_r)$ in $f^{-r}(C)$.■

Lemma 2: For all positive integers $k > r$, the optimal code density $S(n,k,r)/n^r$ converges to a limit $s(k,r)$ as n grows, and $S(n,k,r)/n^r \geq s(k,r)$ for all n.

Proof: Let $C \subseteq [n]^r$ be an optimal covering $(k-r)$-insertion code, that is, $|C| = S(n,k,r)$. Let $m > n$ and define $f : [m] \to [n]$ by $f(y) = y \bmod n$ for all $y \in [m]$. By Lemma 1, the preimage $f^{-r}(C)$ is a covering $(k-r)$-insertion code over $[m]$. Let $q = \lfloor m/n \rfloor$. Thus,

$$S(m,k,r) \leq |f^{-r}(C)| \leq (q+1)^r\,|C|.$$

It follows that

$$\limsup_{m\to\infty} \frac{S(m,k,r)}{m^r} \leq \frac{S(n,k,r)}{n^r},$$

implying both statements in the lemma.■

III. Analytic Reformulation

Definition 1 admits consideration of an infinitary setting. In order to be able to speak about the size of a code, we suppose that X is a measurable space endowed with a probability measure λ. The Cartesian power X^r is endowed with the product measure, which for brevity will be denoted also by λ. We define $s(X,k,r) = \inf_C \lambda(C)$ where the infimum is taken over all measurable $(k-r)$-insertion covering codes $C \subseteq X^r$. In the discrete case, we endow $X = [n]$ with the uniform probability measure, getting $s([n],k,r) = S(n,k,r)/n^r$, which is just another notation for the optimal code density over a finite alphabet. For the unit segment of reals $X = [0,1]$, let λ be the Lebesgue measure. In this case,

$$s([0,1],k,r) = \inf_C \lambda(C) \tag{5}$$

where the infimum is taken over Lebesgue measurable $(k-r)$-insertion covering codes $C \subseteq [0,1]^r$.

Theorem 2: $s(k,r) = s([0,1],k,r)$.

Proof: We first prove that $s([0,1],k,r) \leq s(k,r)$. Let $C \subseteq [n]^r$ be an optimal covering $(k-r)$-insertion code, that is, $|C| = S(n,k,r)$. Define $f_n : [0,1] \to [n]$ by $f_n(0) = 0$ and $f_n(x) = i$ for all $x \in \left(\frac{i}{n}, \frac{i+1}{n}\right]$. By Lemma 1, the preimage $f_n^{-r}(C) \subseteq [0,1]^r$ is a covering $(k-r)$-insertion code over $[0,1]$. Note that f_n is a measurable function, and $\lambda(f_n^{-r}(C)) = |C|/n^r$. This shows that

$$s([0,1],k,r) \leq \lambda\left(f_n^{-r}(C)\right) = \frac{|C|}{n^r} = s([n],k,r)$$

for all $k > r$, implying the required inequality.

In order to prove that $s(k,r) \leq s([0,1],k,r)$, we use the following convention. A mapping $\tau : [t] \to Z$ can be identified with the sequence $(\tau(0),\tau(1),\dots,\tau(t-1)) \in Z^t$. After this, it makes sense to say, for example, that $\rho : [r] \to Z$ covers $\kappa : [k] \to Z$.

Let $C \subseteq [0,1]^r$ be a covering $(k-r)$-insertion code over $[0,1]$. Given a sequence $y = (y_0,y_1,\dots,y_{n-1})$ in $[0,1]^n$, we define $C_n = C_n(y)$ as the set of all mappings $\rho : [r] \to [n]$ such that $(y_{\rho(0)}, y_{\rho(1)}, \ldots, y_{\rho(r-1)}) \in C$. According to our convention, we view C_n as a subset of $[n]^r$ and claim that C_n covers $[n]^k$, i.e., that it is a covering $(k-r)$-insertion code over $[n]$. Indeed, take an arbitrary mapping $\kappa : [k] \to [n]$. Since C covers $[0,1]^k$, the sequence $(y_{\kappa(0)}, y_{\kappa(1)}, \ldots, y_{\kappa(k-1)})$ is covered by some subsequence $(y_{\kappa(i_0)}, \ldots, y_{\kappa(i_{r-1})}) \in C$ where $0 \le i_0 < i_1 < \ldots < i_{r-1} < k$. Define $\rho : [r] \to [n]$ by setting $\rho(0) = \kappa(i_0), \ldots, \rho(r-1) = \kappa(i_{r-1})$ and note that $\rho \in C_n$ covers κ.

It follows that $S(n,k,r) \le |C_n(y)|$ for every $y \in [0,1]^n$. This implies that if we take y uniformly at random in $[0,1]^n$ or, equivalently, if we take independent random variables $y_0, \ldots, y_{n-1}$ uniformly distributed in $[0,1]$, then $S(n,k,r)$ does not exceed the expectation $\mathsf{E}|C_n(y)|$, which by linearity is equal to the sum $\sum_\rho \mathsf{P}[(y_{\rho(0)}, \ldots, y_{\rho(r-1)}) \in C]$ of probabilities over all maps $\rho : [r] \to [n]$. If ρ is injective, then $(y_{\rho(0)}, \ldots, y_{\rho(r-1)}) \in C$ with probability $\lambda(C)$ and, therefore,

$$S(n,k,r) \le n(n-1)\cdots(n-r+1)\lambda(C) + (n^r - n(n-1)\cdots(n-r+1)).$$

Using Lemma 2, we derive from here

$$s(k,r) \le \frac{S(n,k,r)}{n^r} \le \lambda(C) + o(1).$$

As C can be chosen with $\lambda(C)$ arbitrarily close to $s([0,1],k,r)$, this proves the inequality $s(k,r) \le s([0,1],k,r)$.■

While some of the forthcoming proofs use the analytic setting of a measurable subset $X \subseteq [0,1]^r$, they can also be easily re-written to work directly with subsets of $[n]^r$. The choice of which language to use is just the matter of convenience.

IV. A Better Lower Bound for $s(r+1,r)$

We now improve the lower bound in (2).

Theorem 3: $s(r+1,r) \ge 1/r$.

Recall that Theorem 2 allows us to switch to the analytic setting, where we have to prove that $s([0,1], r+1, r) \ge 1/r$. The proof is based on two lemmas below.

Let $C_1, \ldots, C_k$ be measurable sets in a space Ω with probability measure λ. For two indices i and j such that $i<j$, let $C_{i,j} = C_i \cap C_j$. By Bonferroni's inequality,

$$\lambda\left(\bigcup_{i=1}^k C_i\right) \ge \sum_{i=1}^k \lambda(C_i) - \sum_{1\le i<j\le k} \lambda(C_{i,j}).$$

We, however, need an inequality in the opposite direction.

Lemma 3 (An inverse Bonferroni's inequality) Let T be a tree with vertex set $V(T) = \{1,2,\ldots,k\}$ and edge set $E(T)$. Then

$$\lambda\left(\bigcup_{i=1}^k C_i\right) \le \sum_{i=1}^k \lambda(C_i) - \sum_{e\in E(T)} \lambda(C_e). \tag{6}$$

Proof: Note that $\lambda\left(\bigcup_{i=1}^k C_i\right) = \sum_A \lambda(A)$, where the sum is over all atomic sets A in the Boolean algebra generated by the subsets $C_1, \ldots, C_k$ of Ω, apart $A = \Omega \setminus \bigcup_{i=1}^k C_i$. Consider a particular atomic set A and let $t(A)$ denote the number of sets C_i including A as a subset. If $t(A) > 0$, then A is included in at most $t(A)-1$ of the sets C_e with $e \in E(T)$. This is true because a set of t vertices spans the subgraph of the tree T with at most $t-1$ edges. It follows that

$$\sum_{e\in E(T)} \lambda(C_e) \le \sum_{A:\, t(A)>0} (t(A)-1)\cdot\lambda(A).$$

We conclude that

$$\sum_{e\in E(T)} \lambda(C_e) + \lambda\left(\bigcup_{i=1}^k C_i\right) \le \sum_A t(A)\cdot\lambda(A) = \sum_{i=1}^k \lambda(C_i),$$

completing the proof.■

Let $X = [0,1]$ and λ be the Lebesgue measure on X. We write λ also to denote the corresponding product measure on X^{r+1}.

Let $C \subseteq X^r$. For $i \le r+1$, we define $C_i \subseteq X^{r+1}$ as the set of all sequences x in X^{r+1} such that the subsequence obtained by removing the i-th element of x belongs to C. Theorem 3 immediately follows from the next lemma.

Lemma 4: If $C \subseteq X^r$ covers X^{r+1}, i.e., $\bigcup_{i=1}^{r+1} C_i = X^{r+1}$, then $\lambda(C) \ge 1/r$.

Proof: Applying Lemma 3 to $C_1, \ldots, C_{r+1} \subseteq [0,1]^{r+1}$ for any fixed tree T, we readily obtain

$$\sum_{e\in E(T)} \lambda(C_e) \le \sum_{i=1}^{r+1} \lambda(C_i) - \lambda\left(\bigcup_{i=1}^{r+1} C_i\right) = (r+1)\lambda(C) - 1.$$

We now estimate the left-hand side from below. Denote the characteristic function of C by χ_C. Using Fubini's theorem along with the Cauchy-Bunyakovsky-Schwarz inequality, we get

$$\begin{aligned}
\lambda(C_{1,2}) &= \int_{X^{r+1}} \chi_C(x_1, x_3, \ldots, x_{r+1}) \\
&\qquad \times \chi_C(x_2, x_3, \ldots, x_{r+1})\,\mathrm{d}x_1 \cdots \mathrm{d}x_{r+1} \\
&= \int_{X^{r-1}} \left(\int_X \chi_C(x_1, x_3, \ldots, x_{r+1})\,\mathrm{d}x_1 \right.\\
&\qquad \left. \times \int_X \chi_C(x_2, x_3, \ldots, x_{r+1})\,\mathrm{d}x_2\right)\mathrm{d}x_3 \cdots \mathrm{d}x_{r+1} \\
&= \int_{X^{r-1}} \left(\int_X \chi_C(x, x_3, \ldots, x_{r+1})\,\mathrm{d}x\right)^2 \mathrm{d}x_3 \cdots \mathrm{d}x_{r+1} \\
&\ge \left(\int_{X^{r-1}} \int_X \chi_C(x, x_3, \ldots, x_{r+1})\,\mathrm{d}x\,\mathrm{d}x_3 \cdots \mathrm{d}x_{r+1}\right)^2 \\
&= \left(\int_{X^r} \chi_C(x, x_3, \ldots, x_{r+1})\,\mathrm{d}x\,\mathrm{d}x_3 \cdots \mathrm{d}x_{r+1}\right)^2 = \lambda(C)^2.
\end{aligned}$$

Each of the r values $\lambda(C_e)$ for $e \in E(T)$ is estimated similarly. It follows that

$$r\lambda(C)^2 \le (r+1)\lambda(C) - 1.$$

Rewriting this as

$$(1-\lambda(C))(r\lambda(C)-1) \ge 0, \tag{7}$$

we conclude that $\lambda(C) \ge 1/r$.■

We conclude this section with a discussion of a consequence of Theorem 3. We call $X \subseteq [n]^{r+1}$ a *1-packing* if no two sequences in X have a common subsequence of length r (or, equivalently, if the minimum Levenshtein distance between two elements of X is larger than 2). Denote the maximum size $|X|$ of a 1-packing $X \subseteq [n]^{r+1}$ by $P(n, r+1, r)$. The packing and covering numbers are related by the inequality

$$P(n, r+1, r) \le S(n, r+1, r); \tag{8}$$

indeed, every element of a covering 1-insertion code $C \subseteq [n]^r$ can cover at most one element of a maximum packing $X \subseteq [n]^{r+1}$. It is known [14, Cor. 5.1] that $P(n, r+1, r)/n^r \sim 1/(r+1)$ if $n/r \to \infty$. Taking this result into account, our Theorem 3 separates the two values in (8) by showing an additive gap at least $1/(r(r+1))$ between their density versions in the setting when r is fixed and n grows.

V. Turán Systems

If $X \subseteq Y$ are any sets, then one can say that Y covers X, but by a kind of duality we will also say that X *covers* Y. Let $1 < r < k < n$. A family C of r-element subsets of $[n]$ is called a *Turán* (n, k, r)*-system* if every k-element subset of $[n]$ is covered by at least one member of C. The minimum possible cardinality of C is denoted by $T(n, k, r)$. A well-known argument [10], [20] shows that $(n-r)\,T(n, k, r) \ge n\,T(n-1, k, r)$, which implies that the densities $T(n, k, r)/\binom{n}{r}$ form a non-decreasing sequence for each k and r. The limit is called the *Turán density* and denoted by $t(k, r)$. For surveys including Turán systems, see [10], [18], [20].

We connect Turán systems and covering insertion codes by showing that the former concept can in limit be seen as a symmetric version of the latter concept.

Call a set $C \subseteq X^r$ *symmetric* if C is closed with respect to all permutations of the r coordinates. Let us start with a simple observation.

Lemma 5: If $C \subseteq X^r$ is symmetric, then $f^{-r}(C)$ is also symmetric for any function $f : Y \to X$.

The Turán density $t(k, r)$ can be characterized in terms of an analytic object similarly to Theorem 2. Specifically, we define $s^\star([0, 1], k, r)$ similarly to (5) with the additional condition that the infimum is taken over symmetric codes.

Theorem 4: For all positive integers $k > r$, we have $t(k, r) = s^\star([0, 1], k, r)$.

Proof: We first prove the inequality $s^\star([0, 1], k, r) \le t(k, r)$. For a set $A \subseteq [n]^r$, let $A^\dagger$ denote the set of all sequences in A with pairwise distinct elements and $A^\ddagger = A \setminus A^\dagger$ be the remaining part of A.

Let T_n be an optimal Turán (n, k, r)-system, that is, $|T_n| = T(n, k, r)$. Convert T_n into a covering $(k-r)$-insertion code $C_n \subseteq [n]^r$ as follows. For each r-element set in T_n, place all its $r!$ orderings in C_n, thereby covering all sequences in $([n]^k)^\dagger$. In order to cover the remaining sequences, we just add $([n]^r)^\ddagger$ to C_n. Note that C_n is symmetric and that

$$\begin{aligned} |C_n| \le |T_n|\, r! + |([n]^r)^\ddagger| = |T_n|\, r! + \\ (n^r - n(n-1)\cdots(n-r+1)). \end{aligned} \tag{9}$$

Consequently,

$$\frac{|C_n|}{n^r} \le \frac{|T_n|}{\binom{n}{r}} + o(1)$$

where the little-o term approaches 0 as n increases. Consider $D_n = f_n^{-r}(C_n)$ for the function $f_n : [0, 1] \to [n]$ defined in the proof of Theorem 2. Note that f_n has preimages of measure $1/n$ each. By Lemma 1, $D_n \subseteq [0, 1]^r$ is a covering $(k-r)$-insertion code over $[0, 1]$. Since C_n is symmetric, D_n is also symmetric by Lemma 5. It follows that

$$\begin{aligned} s^\star([0, 1], k, r) \le \lambda(D_n) &= \frac{|C_n|}{n^r} \\ &\le \frac{|T(n, k, r)|}{\binom{n}{r}} + o(1) \le t(k, r) + o(1), \end{aligned}$$

yielding the required inequality.

We now prove that, conversely, $t(k, r) \le s^\star([0, 1], k, r)$. Let $C \subseteq [0, 1]^r$ be an arbitrary measurable symmetric covering $(k-r)$-insertion code over $[0, 1]$. Given an integer $n \ge k$, consider a sequence $y = (y_0, \ldots, y_{n-1})$ in $[0, 1]^n$ with pairwise different elements. Define a family $G = G(y)$ of r-element subsets of $[n]$ by putting $\{i_1, \ldots, i_r\} \subseteq [n]$ in G if and only if $(y_{i_1}, \ldots, y_{i_r}) \in C$. The last condition does not depend on the order of indices by the symmetry of C.

Let us show that G is a Turán (n, k, r)-system. Take any k-element set $K \subseteq [n]$. Since the subsequence $(y_i)_{i \in K}$ of y is covered by some sequence in C, there is an r-element set $\{i_1, \ldots, i_r\} \subseteq K$ such that $(y_{i_1}, \ldots, y_{i_r}) \in C$. By definition, this means that $\{i_1, \ldots, i_r\} \in G$. Since K was an arbitrary k-element subset of $[n]$, G is indeed a Turán (n, k, r)-system.

Now, take a uniformly random $y = (y_0, \ldots, y_{n-1})$ in $[0, 1]^n$; equivalently, we take independent uniform $y_0, \ldots, y_{n-1} \in [0, 1]$. With probability 1, all y_i are different. To compute the expected number of r-element sets belonging to G, we sum the probability that $R \in G$ over all $R = \{i_1, \ldots, i_r\} \subseteq [n]$. By the uniformity of $(y_0, \ldots, y_{n-1}) \in [0, 1]^n$, we have that $(y_{i_1}, \ldots, y_{i_r})$ is a uniform element of $[0, 1]^r$. Thus, the probability that $(y_{i_1}, \ldots, y_{i_r}) \in C$ (which is exactly the probability that $R \in G$) is equal to the measure $\lambda(C)$ of C. We conclude that $\mathsf{E}|G| = \binom{n}{r}\lambda(C)$.

Of course, if we remove from $[0, 1]^n$ the null-set D of points y where some two coordinates y_i's coincide, then the expectation does not change. Take $(y_0, \ldots, y_{n-1}) \in [0, 1]^n \setminus D$ such that $|G|$ is at most its expected value $\binom{n}{r}\lambda(C)$. Then the density of G is most $\lambda(C)$. Since n and C were arbitrary, with $\lambda(C)$ arbitrarily close to $s^\star([0, 1], k, r)$, the required inequality follows.■

One can show that the appropriately defined parameter $s^\star(X, k, r)$ (resp. $s(X, k, r)$) is the same for all atomless probability spaces X, since each such space admits, for every n, a measurable partition into parts of measure $1/n$ each.

For fixed k and r, one can alternatively define the function $s^\star(k, r)$ using the *r-hypergraphon* limit object introduced by Elek and Szegedy [8]. While the advantage of this approach is that the infimum in the definition would be in fact the minimum (that is, would be attained) potentially allowing for further methods like variational calculus, the limit object is rather complicated and requires a lot of technical preliminaries. So we stay with our simple setting of measurable subsets of $[0, 1]^r$.

We state an immediate consequence of (9) (which also follows from Theorems 2 and 4).

Corollary 1: $s(k, r) \le t(k, r)$.

Theorem 3, therefore, implies that

$$t(r+1,r) \geq s(r+1,r) \geq 1/r.$$

This lower bound $t(r+1,r) \geq 1/r$ was shown independently by de Caen [5], Sidorenko [19], and Tazawa and Shirakura [21] and generalized by de Caen [6]. Thus, Theorem 3 is an extension of this classical result to the realm of covering insertion codes.

On the other hand, no analogue of the upper bound $s(r+1,r) = O(1/r)$ (see (2)) was known for $t(r+1,r)$. Quite the contrary, de Caen [7] conjectured that $r \cdot t(r+1,r) \to \infty$ as r grows. Inspired by the relationship between Turán systems and covering insertion codes, which we pinpoint here, Pikhurko [16] disproved this conjecture by showing that $t(r+1,r) \leq 6.239/(r+1)$ for all r and $t(r+1,r) \leq 4.911/(r+1)$ for all sufficiently large r. By Corollary 1, the same upper bounds apply to $s(r+1,r)$. Alternatively, Corollary 2 below follows from the recurrence in [12], via the same analysis as that in the proof of [16, Lemma 2.3].

Corollary 2:

1. $s(r+1,r) \leq 6.239/(r+1)$ for all r.
2. $s(r+1,r) \leq 4.911/(r+1)$ for all sufficiently large r.

In the particular case of $r = 2$, we have

$$s([0,1],3,2) = s^\star([0,1],3,2) = 1/2.$$

Indeed, $s([0,1],3,2) \geq 1/2$ by Theorem 3. The upper bound $s^\star([0,1],3,2) \leq 1/2$ is provided by the symmetric single-insertion code $\left[0,\frac{1}{2}\right]^2 \cup \left(\frac{1}{2},1\right]^2$ covering the cube $[0,1]^3$, which is an analog of the single-insertion code $\{(0,0),(1,1)\}$ covering the Boolean cube $\{0,1\}^3$.

Along with Theorem 4, this implies that $t(3,2) = \frac{1}{2}$, which is a well-known fact belonging to the basics of graph theory. The lower bound $t(3,2) \geq \frac{1}{2}$ is known as Mantel's theorem. The upper bound $t(3,2) \leq \frac{1}{2}$ follows by considering the disjoint union of complete graphs $K_{\lfloor n/2 \rfloor}$ and $K_{\lceil n/2 \rceil}$.

We conclude this section with an overview of the known bounds on $t(r+1,r)$ for $r \geq 3$.

A. Bounds for Small r

For $r = 3$ it is known that

$$0.438334 \leq t(4,3) \leq \frac{4}{9} = 0.444\ldots. \tag{10}$$

The lower bound is due to Razborov [17]. The upper bound, conjectured to be optimal, is given by many different constructions, one of which is the following. Split $[n]$ into three parts V_0, V_1, V_2 as evenly as possible and put a 3-element set in C if it either lies entirely inside some V_i or, for some residue i modulo 3, has two elements in V_i and one element in V_{i+1}.

An account of the known bounds on $t(r+1,r)$ for other small values of r can be found in the survey [20].

B. General Bounds

The bound $t(r+1,r) \geq 1/r$ is improved in [4] for odd r and in [15] for even r. For all odd $r \geq 3$, it is shown in [4] that

$$t(r+1,r) \geq \frac{5r - \sqrt{9r^2+24r}+12}{2r(r+3)} = \frac{1}{r} + \frac{1}{r^2} + O(r^{-3}). \tag{11}$$

For all even $r \geq 4$, it is shown in [15] that

$$t(r+1,r) \geq \frac{1}{r} + \frac{(1-1/r^{p-1})(r-1)^2}{2r^p\left(\binom{r+p}{p-1} + \binom{r+1}{2}\right)}, \tag{12}$$

where p is the least prime factor of $r-1$. This bound is the strongest if $p = 3$, that is, $r = 4 \pmod 6$. In this case, it reads

$$t(r+1,r) \geq \frac{1}{r} + \frac{1}{2r^3} + O(r^{-4}).$$

In the worst case, which happens when $p = r-1$, Bound (12) yields

$$t(r+1,r) \geq \frac{1}{r} + \frac{1-o(1)}{4r^{r-3}\binom{2r}{r}}.$$

VI. A Further Improvement of the Lower Bound for $s(r+1,r)$

We now improve Theorem 3 by showing that any lower bound for $t(r+1,r)$ better than $1/r$ implies a lower bound for $s(r+1,r)$ better than $1/r$.

We have $s(r+1,r) \leq t(r+1,r)$ by Corollary 1. Let us prove a relation in the opposite direction.

Theorem 5: For every $r \geq 3$, it holds that

$$t(r+1,r) \leq s(r+1,r) + 2r!\sqrt{r(r+1)(1-s(r+1,r))\left(s(r+1,r)-\frac{1}{r}\right)}.$$

Proof: Let $X = [0,1]$ and let λ denote the Lebesgue measure on X. Moreover, we write λ to denote also the corresponding product measure on any k-dimensional cube X^k. Consider a measurable set $C \subseteq X^r$. As in Section IV, for each $i \leq r+1$ we define $C_i \subseteq X^{r+1}$ to be the set of all sequences x in X^{r+1} such that the subsequence obtained by removing the i-th element of x belongs to C. Suppose that C covers X^{r+1}, that is, $X^{r+1} = \bigcup_{i=1}^{r+1} C_i$. We define

$$K = \bigcap_{i=1}^{r+1} C_i,$$

$$P_i = C_i \setminus \bigcup_{j \neq i} C_j,$$

$$R = X^{r+1} \setminus \left(K \cup \bigcup_{i=1}^{r+1} P_i\right),$$

$$R_i = R \setminus C_i.$$

In other words, K is the atomic set of the Boolean algebra generated by $C_1,\ldots,C_{r+1}$ occurring in these sets with the maximum possible multiplicity $t(K) = r+1$. For each $i \leq r+1$, P_i is an atomic set of minimum possible multiplicity $t(P_i) = 1$ (as 0 is impossible by the covering property). The remaining part R is the union of all atomic sets A of intermediate multiplicity $1 < t(A) \leq r$. Finally, R_i is the part of R formed by the atomic sets outside C_i.

Let $k = r+1$. The inverse Bonferroni's inequality given by Lemma 3 can be somewhat improved. While T was in this lemma an arbitrary tree on vertices $1,\ldots,k$, let T_j be now the star with centre at j, that is, $E(T_j) =$

$\{\{j, i\} : 1 \le i \le r+1,\ i \ne j\}$. In the case that $T = T_j$, Inequality (6) can be improved to

$$\lambda\left(\bigcup_{i=1}^{k} C_i\right) \le \sum_{i=1}^{k} \lambda(C_i) - \sum_{e \in E(T_j)} \lambda(C_e) - \lambda(R_j),$$

which can be routinely verified by looking at the contribution of each atomic set. Since C covers X^{r+1}, the left hand side is equal to 1. Arguing as in the proof of Lemma 4, in place of Inequality (7) we obtain

$$\lambda(R_j) \le (1 - \lambda(C))(r\lambda(C) - 1)$$

for each $j \le r+1$. It follows that

$$\lambda(R) \le \sum_{j=1}^{r+1} \lambda(R_j) \le (r+1)(1-\lambda(C))(r\lambda(C) - 1)$$
$$= r(r+1)(1-\lambda(C))\left(\lambda(C) - \frac{1}{r}\right). \qquad (13)$$

This shows that if the covering code C has density sufficiently close to $1/r$, then up to a small set R, the Boolean algebra generated by $C_1, \ldots, C_{r+1}$ is the sunflower with kernel K and petals $P_1, \ldots, P_{r+1}$.

Given a set $M \subseteq X^k$, we define its *symmetric closure* $\overline{M}$ to be the inclusion-minimal symmetric superset of M:

$$\overline{M} = \big\{(x_1, \ldots, x_k) \in X^k : \exists \text{ permutation } \sigma \text{ of } \{1, \ldots, k\} \text{ with } (x_{\sigma(1)}, \ldots, x_{\sigma(k)}) \in M\big\}.$$

Claim A. $K \setminus \overline{R}$ is symmetric.

Proof of Claim A. Let $(x_1, x_2, x_3, \ldots, x_{r+1}) \in K \setminus \overline{R}$. Since $(x_1, x_2, x_3, \ldots, x_{r+1}) \in K \subseteq C_1 \cap C_2$, we have $(x_2, x_3, \ldots, x_{r+1}) \in C$ and $(x_1, x_3, \ldots, x_{r+1}) \in C$. By the definition of C_i, this implies that $(x_2, x_1, x_3, \ldots, x_{r+1}) \in C_1 \cap C_2$. Since $(x_1, x_2, x_3, \ldots, x_{r+1}) \notin \overline{R}$, the vector $(x_2, x_1, x_3, \ldots, x_{r+1})$ does not belong to R and, therefore, belongs to K. This argument actually shows that $(x_1, x_2, x_3, \ldots, x_{r+1})$ still belongs to $K \setminus \overline{R}$ after transposing any two coordinates. It follows that every permutation of $(x_1, x_2, x_3, \ldots, x_{r+1})$ stays in $K \setminus \overline{R}$.$\lhd$

Claim A shows that if C is a covering code of density $\lambda(C) \approx 1/r$ and, therefore, $\lambda(R)$ is small, then the kernel K is almost symmetric.

Given $(x_1, \ldots, x_{r-1}, x_r) \in C$, define the *splinter* of C at $(x_1, \ldots, x_{r-1}, x_r)$ with respect to the last coordinate as the set

$$S\,(x_1, \ldots, x_{r-1}, x_r) = \{x \in X : (x_1, \ldots, x_{r-1}, x) \in C\}.$$

Let $\delta \in (0, 1)$ be the parameter whose value will be chosen later. Consider the part C' of C consisting of the vectors with small splinters. Specifically,

$$C' = \{(x_1, \ldots, x_r) \in C : \lambda\,(S\,(x_1, \ldots, x_r)) \le \delta\}.$$

Note that

$$\lambda(C') \le \delta. \qquad (14)$$

Given $(x_1, \ldots, x_{r-1}, x_r) \in C$, we also define its *extension-deletion set* $E(x_1, \ldots, x_{r-1}, x_r) \subseteq X^{r+1}$ by

$$E(x_1, \ldots, x_{r-1}, x_r) = \big\{(x_1, \ldots, x_{r-1}, x_r, x_{r+1}) \in X^{r+1} : (x_1, \ldots, x_{r-1}, x_{r+1}) \in C\big\}.$$

Finally, let

$$W = \big\{(x_1, \ldots, x_r) \in C \setminus C' : E\,(x_1, \ldots, x_r) \subseteq \bar{R}\big\}.$$

Claim B. $\lambda(\overline{W}) \le \lambda(\overline{R})/\delta$.

Proof of Claim B. Consider the set

$$W^+ = \bigcup_{(x_1, \ldots, x_r) \in W} E(x_1, \ldots, x_r).$$

Since $W^+ \subseteq \overline{R}$, we have $\overline{W^+} \subseteq \overline{R}$ and, therefore,

$$\lambda(\overline{W^+}) \le \lambda(\overline{R}). \qquad (15)$$

On the other hand,

$$\lambda(\overline{W^+}) \ge \lambda(\overline{W}) \cdot \delta. \qquad (16)$$

Indeed, for $(x_1, \ldots, x_r) \in \overline{W}$ let σ be the lexicographically smallest permutation of $\{1, \ldots, r\}$ such that $(x_{\sigma(1)}, \ldots, x_{\sigma(r)}) \in W$. For every $x \in S(x_{\sigma(1)}, \ldots, x_{\sigma(r)})$, we have $(x_{\sigma(1)}, \ldots, \underline{x_{\sigma(r)}}, x) \in E(x_{\sigma(1)}, \ldots, x_{\sigma(r)}) \subseteq W^+$ and, hence, $(x_1, \ldots, x_r, x) \in \overline{W^+}$. To obtain Inequality (16), it suffices to note that $\lambda(S(x_{\sigma(1)}, \ldots, x_{\sigma(r)})) > \delta$ because $(x_{\sigma(1)}, \ldots, x_{\sigma(r)}) \notin C'$.

The claim readily follows from Inequalities (15) and (16).$\lhd$

We now show that if δ is chosen so that $\lambda(C')$ and $\lambda(W)$ are small, then C is almost symmetric.

Claim C. $\overline{C \setminus (C' \cup W)} \subseteq C$.

Proof of Claim C. For any $(x_1, \ldots, x_r) \in C$, note that its extension $(x_1, \ldots, x_r, x_{r+1})$ belongs to $E(x_1, \ldots, x_r)$ if and only if it belongs to $C_r \cap C_{r+1}$. Suppose that $(x_1, \ldots, x_r) \in C \setminus (C' \cup W)$. By the definition of W, we have $E(x_1, \ldots, x_r) \not\subseteq \overline{R}$. This means that there exists $x_{r+1} \in X$ such that $(x_1, \ldots, x_r, x_{r+1})$ belongs to $C_r \cap C_{r+1}$ but not to $\overline{R}$. It follows that

$$(x_1, \ldots, x_r, x_{r+1}) \in (C_r \cap C_{r+1}) \setminus \overline{R}$$
$$\subseteq (C_r \cap C_{r+1}) \setminus R \subseteq K.$$

We conclude that $(x_1, \ldots, x_r, x_{r+1}) \in K \setminus \overline{R}$. Let σ be an arbitrary permutation of $\{1, \ldots, r\}$. By Claim A, we have $(x_{\sigma(1)}, \ldots, x_{\sigma(r)}, x_{r+1}) \in K$. Since $K \subseteq C_{r+1}$, this implies that $(x_{\sigma(1)}, \ldots, x_{\sigma(r)}) \in C$.$\lhd$

Since C covers X^{r+1}, its symmetrization $\overline{C}$ covers X^{r+1} as well and, by Claim C, we have

$$s^\star([0,1], r+1, r) \le \lambda(\overline{C}) \le \lambda(C) + \lambda(\overline{C'}) + \lambda(\overline{W}).$$

Taking into account Bound (14) and Claim B, we obtain

$$s^\star([0,1], r+1, r) \le \lambda(C) + r!\,\lambda(C') + \lambda(\overline{R})/\delta$$
$$\le \lambda(C) + r!\,\delta + r!\,\lambda(R)/\delta.$$

Setting $\delta = \sqrt{\lambda(R)}$, we conclude that

$$s^\star([0,1], r+1, r) \le \lambda(C) + 2r!\,\sqrt{\lambda(R)}.$$

Along with Bound (13), this implies that

$$s^\star([0,1], r+1, r) \le \lambda(C) + 2r!\sqrt{r(r+1)(1-\lambda(C))\left(\lambda(C) - \frac{1}{r}\right)}.$$

Since $\lambda(C)$ can be taken arbitrarily close to $s([0,1], r+1, r)$, the proof is completed by applying Theorems 2 and 4. ■

Corollary 3: $s(r+1,r) \geq \frac{1}{r} + (1-o(1))\left(\frac{t(r+1,r)-1/r}{2r\cdot r!}\right)^2$.

Proof: Let $s_r = s(r+1,r) - 1/r$ and $t_r = t(r+1,r) - 1/r$. Set $R = r \cdot r!$. Using the lower bound $s(r+1,r) \geq 1/r$ of Theorem 3, from Theorem 5 we derive

$$t_r \leq s_r + 2r!\,\sqrt{r(r+1)(1-1/r)s_r} < s_r + 2R\sqrt{s_r}.$$

This readily implies

$$\sqrt{s_r} > \sqrt{t_r + R^2} - R = \frac{t_r}{R + \sqrt{R^2 + t_r}} > \frac{t_r}{R + \sqrt{R^2 + 1}},$$

yielding the desired bound. ■

Plugging in Bound (11), we obtain the following.

Corollary 4: For odd r,

$$s(r+1,r) \geq \frac{1}{r} + \frac{1-o(1)}{4r^6(r!)^2}.$$

An analog of Corollary 4 for even r follows from Corollary 3 by using Bound (12).

Combining Theorem 5 for $r = 3$ with the lower bound in (10), we get

$$0.438334 \leq s(4,3) + 24\sqrt{(1-s(4,3))(3s(4,3)-1)}.$$

This allows us to slightly improve the lower bound $s(4,3) \geq 1/3$ given by Theorem 3.

Corollary 5: $s(4,3) \geq 0.3333429$.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for careful reading the manuscript and useful comments.

REFERENCES

[1] F. N. Afrati, A. D. Sarma, D. Menestrina, A. Parameswaran, and J. D. Ullman, "Fuzzy joins using MapReduce," in *Proc. IEEE 28th Int. Conf. Data Eng.*, Apr. 2012, pp. 498–509.

[2] F. Afrati, A. D. Sarma, A. Rajaraman, P. Rule, S. Salihoğlu, and J. D. Ullman, "Anchor-points algorithms for Hamming and edit distances using MapReduce," in *Proc. 17th Int. Conf. Database Theory (ICDT)*, Jan. 2014, pp. 4–14.

[3] V. Bhardwaj, P. A. Pevzner, C. Rashtchian, and Y. Safonova, "Trace reconstruction problems in computational biology," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3295–3314, Jun. 2021.

[4] F. R. K. Chung and L. Lu, "An upper bound for the Turán number *t_{3} (n, 4)*," *J. Combinat. Theory, A*, vol. 87, no. 2, pp. 381–389, 1999.

[5] D. de Caen, "Extension of a theorem of moon and Moser on complete subgraphs," *Ars Combinatoria*, vol. 16, pp. 5–10, Aug. 1983.

[6] D. de Caen, "A note on the probabilistic approach to Turán's problem," *J. Combinat. Theory, B*, vol. 34, no. 3, pp. 340–349, 1983.

[7] D. de Caen, "The current status of Turán's problem on hypergraphs," in *Extremal Problems for Finite Sets* (Bolyai Society Mathematical Studies), vol. 3. Budapest, Hungary: János Bolyai Math. Soc., 1991, pp. 187–197.

[8] G. Elek and B. Szegedy, "A measure-theoretic approach to the theory of dense hypergraphs," *Adv. Math.*, vol. 231, nos. 3–4, pp. 1731–1772, Oct. 2012.

[9] C. Grozea. (2024). *The Optimal Densities of Covering Single-Insertion Codes*. [Online]. Available: https://www.brainsignals.de/users/cristian.grozea/combinatorial.html

[10] P. Keevash, "Hypergraph Turán problems," in *Surveys in Combinatorics 2011*. Cambridge, U.K.: Cambridge Univ. Press, 2011, pp. 83–139.

[11] G. Kéri and P. R. J. Östergård, "Bounds for covering codes over large alphabets," *Designs, Codes Cryptography*, vol. 37, no. 1, pp. 45–60, Oct. 2005.

[12] A. Lenz, C. Rashtchian, P. H. Siegel, and E. Yaakobi, "Covering codes using insertions or deletions," *IEEE Trans. Inf. Theory*, vol. 67, no. 6, pp. 3376–3388, Jun. 2021.

[13] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Sov. Phys. Doklady*, vol. 10, pp. 707–710, Jul. 1966.

[14] V. I. Levenshtein, "Perfect codes in the metric of deletions and insertions," *Discrete Math. Appl.*, vol. 3, no. 2, pp. 241–258, 1992.

[15] L. Lu and Y. Zhao, "An exact result for hypergraphs and upper bounds for the Turán density of K^{r}_{r+1}," *SIAM J. Discrete Math.*, vol. 23, no. 3, pp. 1324–1334, Jan. 2009.

[16] O. Pikhurko, "Constructions of Turán systems that are tight up to a multiplicative constant," *Adv. Math.*, vol. 464, Mar. 2025, Art. no. 110148.

[17] A. Razborov, "On 3-hypergraphs with forbidden 4-vertex configurations," *SIAM J. Discret. Math.*, vol. 24, no. 3, pp. 946–963, Jan. 2010.

[18] M. Ruszinkó, "Turán systems," in *CRC Handbook of Combinatorial Designs*, 2nd ed., Boca Raton, FL, USA: CRC Press, 2007, pp. 649–651.

[19] A. F. Sidorenko, "The quadratic form method in the combinatorial Turan problem," *Vestn. Mosk. Univ., I*, vol. 1982, no. 1, pp. 3–6, 1982.

[20] A. F. Sidorenko, "What we know and what we do not know about Turán numbers," *Graphs Combinatorics*, vol. 11, no. 2, pp. 179–199, 1995.

[21] S. Tazawa and T. Shirakura, "Bounds on the cardinality of clique-free family in hypergraphs," *Math. Sem. Notes Kobe Univ.*, vol. 11, pp. 277–281, Jul. 1983.

[22] G. Viennot, "Maximal chains of subwords and up-down sequences of permutations," *J. Combinat. Theory A*, vol. 34, no. 1, pp. 1–14, Jan. 1983.

Oleg Pikhurko received the Diploma degree in mathematics from Lviv National University, Ukraine, in 1995, the Master of Advanced Study degree in mathematics from Cambridge University in 1996, and the Ph.D. degree from Cambridge University in 2000. He was a Junior Research Fellow with the St. John's College, Cambridge, from 2000 to 2003, and an Assistant and Associate Professor with Carnegie Mellon University from 2003 to 2012. He is currently a Professor of mathematics with the University of Warwick, which he joined in 2011. His research interests include extremal combinatorics and its connections to other areas (such as analysis, probability, and theoretical computer science).

Oleg Verbitsky has been with the Institute of Computer Science, Humboldt University of Berlin, Germany, since 2011, working on several research projects funded by German Research Foundation (DFG). From 1994 to 2009, he taught with Lviv National University, Ukraine, and since 2006, he has been affiliated with the Institute of Applied Problems of Mechanics and Mathematics, Lviv, Ukraine. In 1999 and 2000, he was a Lise Meitner Fellow with Vienna University of Technology, Austria, and in 2005 and 2006, he was an Alexander von Humboldt Fellow with the Humboldt University of Berlin. His research interests lie at the intersection of discrete mathematics and theoretical computer science.

Maksim Zhukovskii received the Diploma and Ph.D. degrees in mathematics from Moscow State University, Russia, in 2009 and 2012, respectively. He was an Assistant and then Associate Professor with Moscow Institute of Physics and Technology from 2010 to 2022. In 2022, he held visiting positions with the Weizmann Institute of Science and Tel Aviv University. Since December 2022, he has been a Senior Lecturer with the University of Sheffield. His research interests include extremal and probabilistic combinatorics as well as several areas of theoretical computer science.