THE DISTANCE TO AN IRREDUCIBLE POLYNOMIAL, II

MICHAEL FILASETA AND MICHAEL J. MOSSINGHOFF

Abstract. P. Turán asked if there exists an absolute constant C such that for every polynomial $f \in \mathbb{Z}[x]$ there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) \le \deg(f)$ and $L(f-g) \le C$, where $L(\cdot)$ denotes the sum of the absolute values of the coefficients. We show that $C = 5$ suffices for all integer polynomials of degree at most 40 by investigating analogous questions in $\mathbb{F}_p[x]$ for small primes p. We also prove that a positive proportion of the polynomials in $\mathbb{F}_2[x]$ have distance at least 4 to an arbitrary irreducible polynomial.

1. Introduction

For a polynomial $f(x) = \sum_{k=0}^{n} a_k x^k$, let $L(f)$ denote its *length*, defined by

$$L(f) = \sum_{k=0}^{n} |a_k| . \tag{1}$$

More than 40 years ago, P. Turán [7] asked if every polynomial with integer coefficients lies near an irreducible polynomial with the same degree or smaller, where distance is measured by using the length. More precisely, he asked if there exists an absolute constant C such that for every polynomial $f \in \mathbb{Z}[x]$ there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) \le \deg(f)$ and $L(f-g) \le C$. Note that if such a constant C exists, then certainly $C \ge 2$, as this value is required for $f(x) = x^n$ when n is odd and $n \ge 3$, or for $f(x) = x^{n-2}(x^2 + x - 1)$ when n is even and $n \ge 4$.

While Turán's problem remains open, a number of partial results are known. In 1970, Schinzel [8] proved that $C = 3$ suffices if one removes the restriction on the degree of $g(x)$. He showed in fact that if f has degree n, then an irreducible polynomial g exists with $L(f-g) \le 3$ and

$$\deg(g) \le \exp((5n+7)(\|f\|^2 + 3)),$$

where $\|f\|^2$ denotes the sum of the squares of the coefficients of f. (Furthermore, $C = 2$ suffices if $f(0) \ne 0$.) Recently, Banerjee and the first author [1] improved this by showing that an irreducible polynomial g with $L(f-g) \le 3$ must exist with the bound on the degree of the irreducible polynomial g depending only linearly on that of f (though exponentially on $\|f\|^2$). More precisely, they showed that the degree of g satisfies

$$\deg(g) \le 8 \max\{n+3, n_0\} 5^{8\|f\|^2+9},$$

where n_0 is an effectively computable constant.

Date: March 11, 2011.

2000 *Mathematics Subject Classification.* Primary: 11C08; Secondary: 11R09, 11Y40.

Key words and phrases. Turán's problem, irreducible polynomial, distance.

Research of the second author supported in part by NSA grant number H98230-08-1-0052.

Turán's original problem has been verified for polynomials of small degree n by using explicit computations. In 1997, Bérczes and Hajdu [2] showed in effect that $C = 5$ suffices for polynomials of degree $n \leq 22$, and extended this to $n \leq 24$ in a subsequent article in 1998 [3]. In 2008, Lee, Ruskey, and Williams [5] established that $C = 5$ is sufficient for $n \leq 32$. More recently, the second author [6] demonstrated that this bound suffices for $n \leq 34$.

In this article, we investigate Turán's problem further. First, using a computational strategy we answer this question for all integer polynomials of degree at most 40. We prove the following theorem.

Theorem 1. *If $f \in \mathbb{Z}[x]$ has degree $n \leq 40$, then there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) = n$ and $L(f - g) \leq 5$.*

Section 2 briefly reviews prior investigations of Turán's problem, and Section 3 describes our current method and its results. Section 4 then discusses some heuristic models for the distribution of distances in Turán's problem, as additional evidence toward a favorable resolution of this question. Throughout the article, Turán's problem is investigated in various local settings, especially in $\mathbb{F}_2[x]$, and the main results over $\mathbb{Z}[x]$ are derived from such settings.

Section 5 then proves our second principal result, that one cannot expect improved results for larger degrees by working modulo 2. We show in this last section that the set of polynomials in $\mathbb{F}_2[x]$ having distance at least 4 to every irreducible polynomial of *any* degree in this ring has positive density. More precisely, we establish the following theorem.

Theorem 2. *A positive proportion of polynomials in $\mathbb{F}_2[x]$ has distance ≥ 4 to every irreducible polynomial. More precisely, if $D_4(n)$ denotes the number of polynomials of degree n that are a distance ≥ 4 to each irreducible polynomial in $\mathbb{F}_2[x]$ and $n \geq 246$, then $D_4(n) \geq 2^{n-246}$.*

2. Prior computations

In order to prove Theorem 1, it suffices to show that distance $C = 3$ suffices for polynomials whose leading and constant terms are both odd. For any such polynomial f with degree n, by Eisenstein's criterion with prime $p = 2$, there certainly exists an irreducible polynomial $g(x)$ with $\deg(g) = n$ and $L(f - g) \leq n$. For any positive integer n, let c_n denote the smallest positive integer having the property that for every $f \in \mathbb{Z}[x]$ with degree n and odd leading and trailing terms, there exists an irreducible polynomial $g \in \mathbb{Z}[x]$ with $\deg(g) \leq \deg(f)$ and $L(f - g) \leq c_n$.

Next, consider a local version of Turán's problem. For a polynomial $f \in \mathbb{F}_2[x]$, define its length $L_2(f)$ as its number of monomials. Let $c_n(2)$ be the smallest positive integer with the property that for every $f \in \mathbb{F}_2[x]$ with degree n and constant term 1, there exists an irreducible polynomial $g \in \mathbb{F}_2[x]$ with the same degree and $L_2(f - g) \leq c_n(2)$. Since any polynomial $g \in \mathbb{Z}[x]$ with odd leading coefficient is necessarily irreducible in $\mathbb{Z}[x]$ if its reduction modulo 2 is irreducible in $\mathbb{F}_2[x]$, it follows that $c_n \leq c_n(2)$. Thus, to establish Theorem 1, it suffices to prove that $c_n(2) \leq 3$ for $n \leq 40$.

Bérczes and Hajdu [2, 3] used this strategy to establish their result for $n \leq 24$. They used Maple for testing irreducibility mod 2, and kept a large table recording the result of each irreducibility test performed in order to avoid testing the same

polynomial more than once. They also employed a pair of tables to facilitate the computation of the parity of $L_2(f)$, and used this information to simplify the search for nearby irreducible polynomials, since one need only test polynomials with an odd number of terms. The space requirement in their method was substantial, at $O(2^n)$ for degree n.

Lee, Ruskey, and Williams [5] also used a local strategy to verify that $c_n(2) \leq 3$ for $n \leq 32$, again working modulo 2. More recently, the second author [6] proved that $c_n(2) \leq 3$ for $n \leq 34$. We briefly describe method of the latter paper, as a similar strategy is employed in the present research. This method has two principal phases. First, one determines all the irreducible polynomials of a given degree n. Second, for each $f \in \mathbb{F}_2[x]$ of degree n, and using the list from the first part, one computes the distance from f to an irreducible polynomial. A polynomial $f \in \mathbb{F}_2[x]$ was represented in this method by using a single 32-bit word that recorded its sequence of coefficients. (For degrees 32 and 33, the leading or constant bit or both were omitted; degree 34 required a further adjustment.) This way, many operations on coefficients could be performed in parallel. For example, two polynomials could be added very quickly by simply computing the exclusive or (xor) of their respective binary representations. Other operations on polynomials were likewise simplified, including division and testing for equality.

In the first phase of the algorithm, approximately half of the polynomials $f \in \mathbb{F}_2[x]$ with $f(0) = 1$ and $\deg(f) = n$ were tested for irreducibility. This test was expedited by using a Gray code ordering on the set of polynomials of fixed degree, and packing the values of remainders modulo irreducible polynomials of small degree into long bit vectors. This way, many remainders for one polynomial could be computed from those of the prior polynomial with just a few xor operations. Larger factors were tested by using trial division, although half the tests were avoided by maintaining along with $f(x) = \sum_{k=0}^{n} a_k x^k$ its *reciprocal polynomial* $f^*(x) = \sum_{k=0}^{n} a_{n-k} x^k$. Since $f(0) = 1$, clearly f^* is irreducible if and only if f is irreducible, so the method avoided trial division on f if it exceeded f^* in the lexicographic ordering.

For the second phase of the algorithm, the irreducible polynomials of degree n were stored in a hash table (in fact, just one of f and f^* was stored), using a double hashing strategy to resolve collisions. Define the *load factor* α of a hash table as the proportion of filled entries in the table. It is well known that the expected number of probes in such a table on a successful search is $\frac{1}{\alpha} \log(\frac{1}{1-\alpha})$, and $1/(1-\alpha)$ for an unsuccessful search. Using $\alpha = 2/3$ allowed the second phase of the algorithm to run on a single computer with two gigabytes of memory up to degree $n = 34$, while maintaining very good search times.

By using a Gray code again to iterate over the set of polynomials with degree n and constant term 1, the parity of the number of terms is simple to maintain, as it flips at each iteration. One could then test polynomials having an odd number of terms for distance 0 or 2 from an irreducible polynomial, and those with even length for distance 1 or 3. The revolving door algorithm was used to enumerate the subsets of size 2 or 3 in an efficient way. In this method, each subsequent subset considered differs from the prior one in a minimal way—one element is removed from the last set, and another one replaces it.

This algorithm verified Theorem 1 for polynomials of degree $n \leq 34$, and in addition it determined the exact number of polynomials in $\mathbb{F}_2[x]$ of each degree

in this range having distance k from an irreducible polynomial, for each k. These results are summarized in Figure 1 in Section 4, which shows the proportion of the 2^n polynomials of degree n in $\mathbb{F}_2[x]$ at distance k, for $0 \le k \le 4$. (This figure also incorporates data on polynomials with constant term 0.)

3. New computations

We follow a similar strategy in the present research: first generate irreducible polynomials in $\mathbb{F}_2[x]$, and then compute distances to these irreducible polynomials. However, the new method incorporates three principal changes. First, we reduce the space requirement for the table of irreducible polynomials, since memory was the most important constraining factor in the prior computations. Second, we investigate some different methods for generating the irreducible polynomials in $\mathbb{F}_2[x]$. Third, we incorporate a parallel strategy in the distance computation.

For the first improvement, we observe that prior computations indicated that nearly all polynomials in $\mathbb{F}_2[x]$ have many irreducible neighbors of the same degree at minimal distance. It seems then that one might store just a sample of the irreducible polynomials of a particular degree in a hash table, without much loss in the distance computations. In fact, we can estimate the number of polynomials expected to require distance $C > 3$ if only a proportion q of the irreducible polynomials are used, for a parameter $q \in [0, 1]$.

Since the number of irreducible polynomials of degree n in $\mathbb{F}_2[x]$ is precisely $\frac{1}{n}\sum_{d|n} \mu(d)2^{n/d}$, where $\mu(\cdot)$ is the Möbius function, it follows that the probability that a polynomial $g \in \mathbb{F}_2[x]$ is irreducible, given that $\deg(g) = n$, $g(0) = 1$, and g has odd length, is approximately $\frac{2^n}{n} \cdot \frac{1}{2^{n-2}} = \frac{4}{n}$. Suppose that $f \in \mathbb{F}_2[x]$ has degree n, $f(0) = 1$, and $L_2(f)$ is odd. If we assume that the irreducible polynomials of degree n are uniformly distributed among the polynomials with constant term 1 and odd length, and we have selected a random proportion q of these polynomials, then the probability that neither $f(x)$ nor any polynomials of the form $f(x)+x^i+x^j$ with $0 < i < j < n$ lies in our selected set of irreducible polynomials is

$$\left(1 - \frac{4q}{n}\right)^{1+\binom{n-1}{2}} = \exp\left(6q - 4q^2 - 2qn\right)\left(1 + O\left(\frac{1}{n}\right)\right).$$

Thus, we expect

$$2^{n-2}\left(1 - \frac{4q}{n}\right)^{1+\binom{n-1}{2}} = \exp\left((n-2)\log 2 + 6q - 4q^2 - 2qn\right)\left(1 + O\left(\frac{1}{n}\right)\right) \tag{2}$$

polynomials with odd length to require $C > 3$ when a random proportion q of the polynomials are stored. This expression has constant limiting value $2^{1-\log 2}$ when $q = (\log 2)/2 = 0.3465\ldots$, so it follows that we should be able to discard nearly $2/3$ of the irreducible polynomials with only minor effect when checking if distance 3 suffices for polynomials of odd length and constant term 1.

A similar analysis handles the case where the length of f is even. In this case, the number of polynomials that we expect to fail all tests at both distance 1 and

distance 3 is

$$2^{n-2}\left(1-\frac{4q}{n}\right)^{n-1+\binom{n-1}{3}} = \exp\Bigg((n-2)\log 2 - \frac{2q}{3}\left(n^2-6n+17\right) - \frac{4q^2}{3}(n-6) - \frac{32q^3}{9}\Bigg)\left(1+O\left(\frac{1}{n}\right)\right), \tag{3}$$

and this expression tends to 0 as n grows large if $q > (3\log 2)/(2n)$.

We set $q = 1/4$ at degrees $n = 35$ and 36, so that only about an eighth of the irreducible polynomials need to be stored in these tests, since we only store one of f and its reciprocal f^*. We reduce this to $q = .175$ at $n = 37$, then to $q = 1/8$ at $n = 38$ and 39, and finally to $q = 1/9$ at $n = 40$. Table 1 shows the number of irreducible polynomials selected at each degree using these parameters. We use (2) to estimate the number of polynomials that we expect will fail the test for distance $C \le 3$. (We expect no exceptional polynomials from (3) since q remains substantially larger than the critical value here.) Table 4 lists both this estimated number of exceptions (to two significant digits), and the actual number produced in our computations at each degree. Each estimate shown in this table is in fact half the value produced by (2), since our implementation tests only about half of the polynomials of each degree, owing to the symmetry of f with f^*.

TABLE 1. Generating irreducible polynomials: q = sampling rate, N = irreducible polynomials selected.

n	q	N
35	.250	122 698 486
36	.250	238 588 443
37	.175	325 039 433
38	.125	452 116 480
39	.125	880 977 087
40	.111	1 525 544 625

The second change in the algorithm involved the method for generating irreducible polynomials in $\mathbb{F}_2[x]$. We used a variation on the procedure employed in [6], reducing the time required, at the expense of a larger space requirement. For example, at degree $n = 40$ we used the bit-packing strategy to scan quickly for divisors of degree up to 16, and then checked if the polynomial occurred in a precomputed hash table containing the reducible polynomials of degree 40 having no factor with degree less than 17. However, after these computations were completed, we found that the algorithm of [4] is significantly more efficient in terms of both time and space. This method, which was employed in [5], computes the set of irreducible polynomials in $\mathbb{F}_2[x]$ of fixed degree n in a more efficient way by exploiting a correspondence with certain equivalence classes of binary Lyndon words of length $n+1$. (Recall that a binary Lyndon word is a finite sequence of 0s and 1s that exhibits no periodic structure.)

Our third main improvement lies in the implementation of the distance computation for large degrees. For $n = 35$ through 37, we used a single iMac computer with 4 GB of memory and a dual-core 2.93 GHz Intel processor for this check. However, due to a restriction in the operating system, we found that one process could not allocate more than about 3 GB of memory, and as a result we needed to increase

α slightly at $n = 37$. Starting with degree $n = 38$, we used OpenMP in C++ to implement shared memory on multicore machines to speed up the distance check. These computations were performed on a cluster of dual quad-core Intel 2.33 GHz Xeon processors, housed at the Centre for Interdisciplinary Research in the Mathematical and Computational Sciences (IRMACS) at Simon Fraser University. Using all eight cores on one blade of the cluster, and with 16 GB of memory available to the group, we could store just one copy of the very large hash table in memory, with eight processors querying it simultaneously.

In shared memory computations, one should of course distribute the computation among the cores as evenly as possible for maximum benefit of the parallelization. Splitting the work among the cores based on prefix (or suffix) bits of the polynomials however produces uneven times, since our program discards any polynomial that is lexicographically larger than its reciprocal. After some experimentation, we found a good strategy was to pair two complementary segments in each process, such as prefix bit sequences 10011 and 01100.

Table 2 summarizes the distance computations. Each polynomial used exactly 40 bits of storage in our C++ implementation. (This required discarding the high bit at $n = 40$.) In the table, c denotes the number of cores used per process in the shared memory calculations, J is the total number of jobs, α is the load factor for the hash table, S_g is the space needed in gigabytes for the hash table, and T_h is the total time in hours for the jobs to complete. In addition, the value b indicates the load balance in the shared memory calculations, defined as the total CPU time used by all c cores, divided by the time required by the longest-running core; its optimum value is therefore 8. Our load-balancing strategy thus improved with each successive degree.

TABLE 2. Distance computation: c = cores, J = jobs, α = load factor, S_g = space (GB), b = load factor, T_h = time (hrs).

n	c	J	$1/\alpha$	S_g	b	T_h
35	1	1	2	1.1	—	7.4
36	1	1	2	2.2	—	14.6
37	1	1	1.85	2.8	—	45.7
38	8	1	3	6.3	5.1	32.6
39	8	1	3	12.3	6.0	58.1
40	8	2	2	14.2	7.2	146.2

Table 3 exhibits the results of the distance computations. It indicates the number of polynomials found at each distance $k \le 3$, and the number that fail the test. These numbers account for both f and f^*, so the sum across the row at degree n is exactly 2^{n-1}. Of course, due to our sampling of the irreducible polynomials, these distances are upper bounds—for example, some of the polynomials counted in the $k = 2$ column are in fact irreducible. This leaves a relatively small number of polynomials to check in a third stage in order to verify Theorem 1. A separate program verifies that each of these exceptional polynomials has distance 0 or 2. This program uses the `IterIrredTest` function in the NTL library [9] called from a C++ program to test irreducibility. Table 4 summarizes these computations, and shows the modest time in minutes T_m required to complete the test.

TABLE 3. Number of polynomials detected at distance k.

n	$k=0$	$k=1$	$k=2$	$k=3$	$k\geq 4$
35	245 396 972	5 347 650 573	8 344 536 124	3 242 284 019	1 496
36	477 175 020	10 695 545 232	16 702 692 131	6 484 323 952	2 033
37	650 078 866	16 980 159 922	33 709 365 412	17 379 578 446	294 090
38	904 231 253	26 485 132 540	67 802 867 444	42 234 344 196	12 378 039
39	1 761 954 174	52 979 923 472	135 657 311 536	84 459 030 000	19 687 762
40	3 051 086 345	96 526 950 648	271 742 478 661	178 350 956 296	84 341 938

TABLE 4. Checking exceptional polynomials: estimated and actual number of exceptions, true distance results, time in minutes.

n	Est.	Actual	$k=0$	$k=2$	T_m
35	360	748	18	730	.002
36	440	1 043	23	1 020	.003
37	$1.0\cdot 10^5$	147 045	6 653	140 392	0.38
38	$5.0\cdot 10^6$	6 190 937	371 487	5 819 450	16.1
39	$7.8\cdot 10^6$	9 843 881	558 390	9 285 491	26.9
40	$3.5\cdot 10^7$	42 176 124	2 495 792	39 680 332	119.0

4. Distance distributions

Suppose that $f\in\mathbb{F}_2[x]$ has constant term 1 and degree n. Estimating the conditional probability that f is irreducible given that it has odd length as $4/n$, we may determine a conjectural distribution for the distance from f to an irreducible polynomial of the same degree in $\mathbb{F}_2[x]$. If f has odd length, then the probability that f has distance at least 4 is negligible at

$$\left(1-\frac{4}{n}\right)^{1+\binom{n-1}{2}}=e^{2-2n}\left(1-\frac{20}{3n}+O\left(\frac{1}{n^2}\right)\right),$$

so the probability that f has distance 2 is approximately $1-4/n$. If f has even length, then it is not adjacent to an irreducible polynomial with approximate probability

$$\left(1-\frac{4}{n}\right)^{n-1}=e^{-4}\left(1-\frac{4}{n}+O\left(\frac{1}{n^2}\right)\right).$$

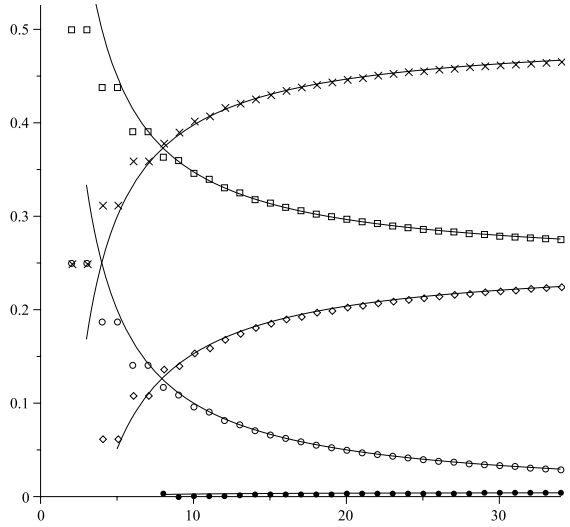
Since the probability that f has distance at least 5 is again negligible, decaying like $\exp(-2n^2/3)$, we estimate the probability that f has distance 3 as $e^{-4}(1-4/n)$, and distance 1 as $1-e^{-4}(1-4/n)$.

If $f(0)=0$, then clearly the probability that f has distance k from an irreducible polynomial is the same as the probability that $f+1$ has distance $k-1$. As shown in [5,6], one can then determine the heuristic distribution for the probability $r_2(k,n)$ that a randomly selected polynomial in $\mathbb{F}_2[x]$ of degree n has distance k from an irreducible polynomial:

$$r_2(n,0)\approx\frac{1}{n},\quad r_2(n,1)\approx\frac{1-e^{-4}}{4}+\frac{1+e^{-4}}{n},\quad r_2(n,2)\approx\frac{2-e^{-4}}{4}-\frac{1-e^{-4}}{n},$$
$$r_2(n,3)\approx\frac{1+e^{-4}}{4}\left(1-\frac{4}{n}\right),\quad r_2(n,4)\approx\frac{e^{-4}}{4}\left(1-\frac{4}{n}\right).$$

Figure 1 (from [6]) shows that these estimates fit the experimental data very well.

FIGURE 1. Predicted proportions versus experimental data for distances in $\mathbb{F}_2[x]$ ($k = 0$: open circles; $k = 1$: boxes; $k = 2$: crosses; $k = 3$: diamonds; $k = 4$: filled circles).



Thus, for large n, one expects that approximately 24.54% of the polynomials of degree n will have distance 1, about 49.54% distance 2, another 25.46% distance 3, and the remaining .46% distance 4. We may also use this model to estimate the number of polynomials $f \in \mathbb{F}_2[x]$ with constant term 1 of any degree with distance at least 4 to an irreducible polynomial. For polynomials with odd length, the heuristic predicts this total number to be

$$\sum_{n \geq 41} 2^{n-2} \left(1 - \frac{4}{n}\right)^{1+\binom{n-1}{2}} < 1.2 \cdot 10^{-23},$$

since we have established Theorem 1. Similarly, for polynomials with even length, the expected total is

$$\sum_{n \geq 41} 2^{n-2} \left(1 - \frac{4}{n}\right)^{n-1+\binom{n-1}{3}} < 3.1 \cdot 10^{-431}.$$

This then presents some additional heuristic evidence for a favorable resolution of Turán's question with constant $C = 5$.

Finally, Turán's problem has been investigated using other primes besides $p = 2$. For a polynomial $f \in \mathbb{F}_p[x]$, define its length $L_p(f)$ by choosing each of its coefficients in the interval $(-p/2, p/2]$, and then summing their absolute values (in $\mathbb{Z}$) as in (1). Define $c_n(p)$ as the smallest positive integer with the property that for every monic polynomial $f \in \mathbb{F}_p[x]$ with degree n and $f(0) \neq 0$, there exists an irreducible polynomial $g \in \mathbb{F}_p[x]$ with the same degree and $L_p(f - g) \leq c_n(p)$. Bérczes and Hajdu [2] showed in effect that $c_n(3) \leq 2$ for $n \leq 12$, so that $C = 4$

suffices in Turán's question for polynomials in $\mathbb{Z}[x]$ up to degree 12. The second author extended this result to $n \leq 18$ in [6], and also investigated the problem in $\mathbb{F}_p[x]$ for other primes $p \leq 31$ across a range of degrees.

We may apply a similar analysis for the prime $p = 3$ to obtain heuristic approximations for the probability $r_3(n,k)$ that a polynomial $f \in \mathbb{F}_3[x]$ with degree n has distance k from an irreducible polynomial of the same degree. We assume that irreducible polynomials of fixed degree in $\mathbb{F}_3[x]$ are uniformly distributed provided that no linear factors appear. We estimate the probability that a polynomial f with degree n and $f(0)f(1)f(-1) \neq 0$ is irreducible as $27/8n$, and a lengthy calculation produces the following estimates for $r_3(k,n)$, where we use t for $\exp(-27/16) = 0.18498\ldots$:

$$\begin{aligned}
r_3(0,n) &= \frac{1}{n} + O\left(\frac{1}{n^2}\right), \\
r_3(1,n) &= \frac{2}{3} - \frac{2t}{27}\left(t^3+4t+4\right) + \frac{t}{32n}\left(19t^3+38t+51\right) + O\left(\frac{1}{n^2}\right) \\
&\approx 0.60163 + \frac{0.33614}{n}, \\
r_3(2,n) &= \frac{1}{3} - \frac{t}{27}\left(2t^4+4t^2-7t-8\right) \\
&\qquad + \frac{1}{128n}\left(87t^5-32t^4+130t^3-109t^2-204t-128\right) + O\left(\frac{1}{n^2}\right) \\
&\approx 0.39606 - \frac{1.3177}{n}, \\
r_3(3,n) &= \frac{t^2}{27}\left(2t^3+2t^2+4t+1\right) - \frac{t^2}{128n}\left(87t^3+44t^2+130t+43\right) + O\left(\frac{1}{n^2}\right) \\
&\approx 0.0023078 - \frac{0.018473}{n},
\end{aligned}$$

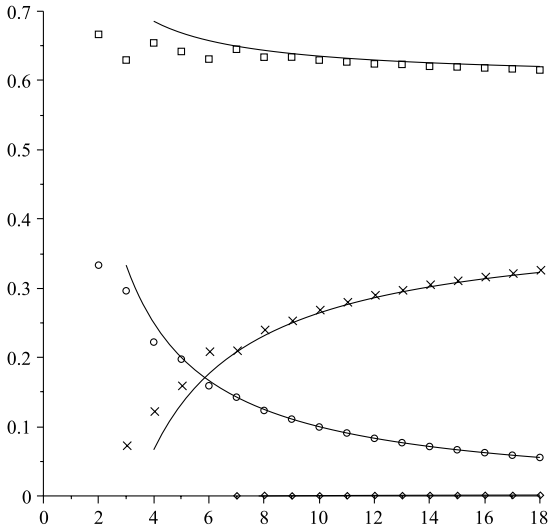
and $r_3(k,n)$ is minuscule for $k \geq 4$. Figure 2 shows that these estimates fit the data for $n \leq 18$ from [6] very well. It seems plausible then that $C = 4$ may in fact hold in Turán's problem in $\mathbb{Z}[x]$.

It is argued in [6] that larger primes are unlikely to produce better bounds in Turán's problem when working over $\mathbb{F}_p[x]$. However, in $\mathbb{Z}[x]$ it is quite possible that even $C = 2$ suffices, as it appears that no integer polynomial requiring distance 3 is even known.

5. An explicit density result

In Section 4, heuristic arguments indicated that it is plausible that every polynomial in $\mathbb{F}_2[x]$ has distance at most 4 from an irreducible polynomial of the same degree. In this section, we show on the other hand that distance 4 is certainly required in $\mathbb{F}_2[x]$, in a particularly strong sense. Throughout this section, we work in the field $\mathbb{F}_2$ of arithmetic modulo 2. We show that a positive proportion of polynomials $f(x)$ in $\mathbb{F}_2[x]$ have minimal distance at least 4 to an irreducible polynomial in this ring. More precisely, we show that a positive proportion of polynomials $f(x)$ in $\mathbb{F}_2[x]$ have the property that if $g(x)$ is an irreducible polynomial in $\mathbb{F}_2[x]$ of arbitrary degree, then the number of terms in $f(x) + g(x)$ is at least 4. To clarify, the trinomial $x^{40} + x^{20} + x$ arose in our computations earlier in the paper as an

FIGURE 2. Predicted proportions versus experimental data for distances in $\mathbb{F}_3[x]$ ($k = 0$: open circles; $k = 1$: boxes; $k = 2$: crosses; $k = 3$: diamonds).



example of a polynomial of degree 40 that has distance 4 from every irreducible polynomial in $\mathbb{F}_2[x]$ of degree ≤ 40. However, $x^{57} + x^{40} + x^{20} + x + 1$ is irreducible, so $x^{40} + x^{20} + x$ has distance 2 from some irreducible polynomial. We show in this section that a positive proportion of polynomials $f(x)$ in $\mathbb{F}_2[x]$ have distance at least 4 from *every* irreducible polynomial in $\mathbb{F}_2[x]$. We begin by explaining how to construct one such $f(x)$.

We first state some essential motivating facts.

(i) For each odd positive integer n, the nth cyclotomic polynomial $\Phi_n(x)$ factors in $\mathbb{F}_2[x]$ as

$$\Phi_n(x) = \rho_n^{(1)}(x)\rho_n^{(2)}(x)\cdots\rho_n^{(\kappa)}(x),$$

where the $\rho_n^{(j)}(x)$ are mutually incongruent irreducible polynomials modulo 2, each having degree $\mathrm{ord}_n(2)$, and $\kappa(n) = \varphi(n)/\mathrm{ord}_n(2)$. Here, $\varphi(\cdot)$ is Euler's totient function, and $\mathrm{ord}_n(2)$ is the smallest positive integer t such that $2^t \equiv 1 \pmod{n}$.

(ii) The infinite collection of polynomials

$$\rho_n^{(j)}(x),\ n \geq 1,\ n \text{ odd},\ 1 \leq j \leq \kappa(n),$$

are distinct, and different from x and $x + 1$.

The arguments establishing these statements are not difficult, given classical literature on the subject. In this section, we will make use of a polynomial $f(x)$ constructed explicitly from the Chinese Remainder Theorem, where the moduli are among factors of $\Phi_n(x)$ described in (i), together with the polynomials x and $x+1$. This construction will enable us to justify our main result, without making explicit use of any of the motivating facts above.

For our argument, as in (i), we define $\kappa(n)$ to be the number of distinct irreducible factors of $\Phi_n(x)$ modulo 2. We will be interested in the value of $\kappa(n)$ for each nontrivial positive divisor of 315. A direct computation (or (i)) can be used to obtain Table 5, which exhibits these values.

TABLE 5. Number of irreducible factors $\kappa(n)$ of $\Phi_n(x)$ modulo 2.

n	3	5	7	9	15	21	35	45	63	105	315
$\kappa(n)$	1	1	2	1	2	2	2	2	6	4	12

We now take advantage of the polynomials $\rho_n^{(j)}(x)$ described in (i). The following lemma is easily established, and we omit its proof.

Lemma 1. *Let m be a positive integer. Let $\rho(x)$ denote a factor (not necessarily irreducible) of $\Phi_m(x)$ in $\mathbb{F}_2[x]$. Let $h(x) \in \mathbb{F}_2[x]$ be a polynomial divisible by $\rho(x)$. Then $h(x) + x^u + x^v$ is divisible by $\rho(x)$ for all positive integers u and v with $u - v$ divisible by m.*

We next require a *covering system* of the integers: a system of congruences with the property that every integer satisfies at least one congruence from the system. While there is some flexibility in the system we construct, the left column in Table 6 shows the specific covering system that we employ here. The first column in this table exhibits expressions of the form $a_j \pmod{m_j}$, with $1 \le j \le 29$. We therefore assert that for each integer x, there is at least one $j \in \{1, 2, \dots, 29\}$ such that $x \equiv a_j \pmod{m_j}$. To verify that this is in fact the case, we may proceed as follows. Observe that each m_j divides 315. Given $x \in \mathbb{Z}$, we can select $y \in \{1, 2, \dots, 315\}$ such that $x \equiv y \pmod{315}$. This implies then that $x \equiv y \pmod{m_j}$ for each $j \in \{1, 2, \dots, 29\}$. Therefore, to see that every integer x satisfies some congruence $x \equiv a_j \pmod{m_j}$, it suffices to show that every integer $y \in \{1, 2, \dots, 315\}$ satisfies some congruence $y \equiv a_j \pmod{m_j}$. As $y \in \{1, 2, \dots, 315\}$ and $j \in \{1, 2, \dots, 29\}$, this is a simple computational check.

While the moduli in our covering system are not distinct, it is important that each modulus m is repeated at most $\kappa(m)$ times. With $a_j \pmod{m_j}$ as in Table 6, we consider congruences

$$f(x) \equiv x^{a_j} + 1 \pmod{g_j(x)}, \quad \text{for } 1 \le j \le 29, \tag{4}$$

where each $g_j(x)$ is chosen to be an irreducible factor of $\Phi_{m_j}(x)$ in $\mathbb{F}_2[x]$, and where the collection of $g_j(x)$ are distinct. The particular $g_j(x)$ corresponding to a given $a_j \pmod{m_j}$ that we use appears in the right column of Table 6. In addition, we combine the 29 congruences in (4) with the two additional requirements

$$f(x) \equiv 0 \pmod{x} \quad \text{and} \quad f(x) \equiv 1 \pmod{x+1}. \tag{5}$$

In particular, if $f(x)$ satisfies (5), then $f(0) = 0$ and $f(1) = 1$ in $\mathbb{F}_2$.

Suppose now that $f(x)$ satisfies the congruences in (4) and (5), and suppose further that $f(x)$ has at least three terms with degree > 12. Let $g(x)$ be an irreducible polynomial that has minimal distance τ from $f(x)$, so that $f(x) + g(x)$ has as few terms as possible in $\mathbb{F}_2[x]$. We do not require here that $\deg g \le \deg f$. Our immediate goal is to show $\tau \ge 4$.

We show first that if $\tau \le 3$, then $g(0) = 1$. Assume $\tau \le 3$ and $g(0) = 0$. Since $g(x)$ is irreducible, we deduce $g(x) = x$. We use that $f(x)$ has at least three terms

TABLE 6. Covering system with polynomial selection.

1 (mod 3)	x^2+x+1
1 (mod 5)	$x^4+x^3+x^2+x+1$
1 (mod 7)	x^3+x+1
2 (mod 7)	x^3+x^2+1
3 (mod 9)	x^6+x^3+1
2 (mod 15)	x^4+x^3+1
8 (mod 15)	x^4+x+1
11 (mod 21)	$x^6+x^5+x^4+x^2+1$
17 (mod 21)	$x^6+x^4+x^2+x+1$
19 (mod 35)	$x^{12}+x^{11}+x^{10}+x^8+x^5+x^4+x^3+x^2+1$
34 (mod 35)	$x^{12}+x^{10}+x^9+x^8+x^7+x^4+x^2+x+1$
33 (mod 45)	$x^{12}+x^3+1$
42 (mod 45)	$x^{12}+x^9+1$
0 (mod 63)	x^6+x^5+1
18 (mod 63)	$x^6+x^4+x^3+x+1$
27 (mod 63)	$x^6+x^5+x^2+x+1$
42 (mod 63)	$x^6+x^5+x^3+x^2+1$
45 (mod 63)	$x^6+x^5+x^4+x+1$
54 (mod 63)	x^6+x+1
5 (mod 105)	$x^{12}+x^{10}+x^9+x^7+x^6+x^4+1$
14 (mod 105)	$x^{12}+x^{11}+x^9+x^8+x^7+x^3+1$
20 (mod 105)	$x^{12}+x^8+x^6+x^5+x^3+x^2+1$
35 (mod 105)	$x^{12}+x^9+x^5+x^4+x^3+x+1$
24 (mod 315)	$x^{12}+x^9+x^6+x^2+1$
60 (mod 315)	$x^{12}+x^{10}+x^8+x^6+x^3+x+1$
150 (mod 315)	$x^{12}+x^{11}+x^9+x^6+x^4+x^2+1$
195 (mod 315)	$x^{12}+x^{11}+x^9+x^8+x^7+x^6+x^5+x+1$
249 (mod 315)	$x^{12}+x^{11}+x^{10}+x^8+1$
285 (mod 315)	$x^{12}+x^{11}+x^{10}+x^8+x^6+x^5+x^2+x+1$

with degree >12. Since $f(x)+g(x)$ has $\tau \le 3$ terms, we must have that $f(x)$ has exactly four terms, three of degree >12 together with the term x. However, this contradicts that $f(1)=1$. Hence, if $\tau \le 3$, then $g(0)=1$.

We consider the four cases $\tau=0$, $\tau=1$, $\tau=2$, and $\tau=3$, and show that in each case we obtain a contradiction.

- Suppose $\tau=0$. In this case, $f(x)$ is irreducible in $\mathbb{F}_2[x]$. Since $f(0)=0$, we deduce $f(x)=x$. Since $f(x)$ has degree >12, we obtain a contradiction.
- Suppose $\tau=1$. Since $f(0)=0$ and $g(0)=1$, we deduce $g(x)=f(x)+1$. But in this case, $g(1)=f(1)+1=1+1=0$. Since $g(x)$ is irreducible, we deduce $g(x)=x+1$ and so $f(x)=x$. This contradicts that $f(x)$ has degree >12.

- Suppose $\tau = 2$. Since $f(0) = 0$ and $g(0) = 1$, we deduce $g(x) = f(x)+x^a+1$ for some positive integer a. Since the congruences $x \equiv a_j \pmod{m_j}$ form a covering of the integers, there is a $k \in \{1, 2, \dots, 29\}$ such that $a \equiv a_k \pmod{m_k}$. From (4), we obtain
$$f(x) \equiv x^{a_k} + 1 \pmod{g_k(x)}.$$
We take $m = m_k$, $\rho(x) = g_k(x)$, $h(x) = f(x)+x^{a_k}+1$, $u = a_k$ and $v = a$ in Lemma 1 to deduce that $g(x) = f(x) + x^a + 1$ is divisible by $g_k(x)$. Since $g(x)$ is irreducible, we deduce $g(x) = g_k(x)$ so that $f(x) = x^a + g_k(x) + 1$. Recall that m_k divides 315 and $g_k(x)$ is an irreducible factor of $\Phi_{m_k}(x)$ in $\mathbb{F}_2[x]$. It is easy to check that the degree of $g_k(x)$ is at most 12. In fact, $g_k(x)$ is one of the polynomials in the right column of Table 6 of degree ≤ 12. Since $f(x) = x^a + g_k(x) + 1$, we obtain a contradiction since $f(x)$ has at least three terms with degree > 12.
- Suppose $\tau = 3$. Here, $f(0) = 0$ and $g(0) = 1$ implies $g(x) = f(x) + x^a + x^b + 1$, for some positive integers a and b. Then $g(1) = f(1)+1+1+1 = 0$, so $g(x) = x+1$ and $f(x) = x^a+x^b+x$, contradicting that $f(x)$ has at least three terms with degree > 12.

We deduce then that $\tau \ge 4$.

Before proceeding, we clarify for later purposes what we have just shown: Every polynomial $f(x)$ having at least three terms of degree > 12 and satisfying the congruences in (4) and (5) has the property that it has distance ≥ 4 to every irreducible polynomial in $\mathbb{F}_2[x]$.

We now construct a polynomial $f(x)$ satisfying the congruences in (4) and (5), noting that we require $f(x)$ to have at least three terms with degree > 12. We apply the Chinese Remainder Theorem on the system of congruences given by (4) and (5). The polynomial $f(x)$ we obtain has degree 243, and is displayed below.

$$\begin{aligned} f(x) = x^{243} &+ x^{238} + x^{233} + x^{232} + x^{231} + x^{227} + x^{225} + x^{223} + x^{222} + x^{221} \\ &+ x^{217} + x^{216} + x^{214} + x^{208} + x^{206} + x^{203} + x^{202} + x^{201} + x^{199} \\ &+ x^{197} + x^{196} + x^{192} + x^{186} + x^{184} + x^{180} + x^{175} + x^{174} + x^{171} \\ &+ x^{169} + x^{167} + x^{164} + x^{163} + x^{162} + x^{160} + x^{157} + x^{155} + x^{149} \\ &+ x^{147} + x^{146} + x^{145} + x^{143} + x^{141} + x^{136} + x^{133} + x^{130} + x^{129} \\ &+ x^{125} + x^{124} + x^{116} + x^{115} + x^{114} + x^{108} + x^{103} + x^{100} + x^{99} \\ &+ x^{98} + x^{95} + x^{94} + x^{92} + x^{88} + x^{83} + x^{81} + x^{72} + x^{68} + x^{63} \\ &+ x^{61} + x^{55} + x^{52} + x^{50} + x^{49} + x^{47} + x^{46} + x^{43} + x^{36} + x^{35} \\ &+ x^{29} + x^{26} + x^{23} + x^{22} + x^{20} + x^{18} + x^{14} + x^{10} + x^7 + x^6. \end{aligned} \tag{6}$$

Although we have already justified this polynomial has the properties we require, we note that this is straightforward to verify directly. First, since $f(0) = 0$, the polynomial $f(x)$ is itself reducible. Second, since the number of terms in $f(x)$ is 85, we have $f(x) + 1$ and $f(x) + x^a + x^b + 1$ are divisible by $x + 1$ and, hence, reducible for all nonnegative integers a and b. Third, a direct computation shows that $\gcd\big(f(x) + x^a + 1, x^{315} + 1\big)$, computed in $\mathbb{F}_2[x]$, has positive degree for all integers a satisfying $0 \le a < 315$. Last, since

$$\gcd\big(f(x) + x^a + 1, x^{315} + 1\big) = \gcd\big(f(x) + x^{a+315} + 1, x^{315} + 1\big),$$

it follows that $f(x) + x^a + 1$ is reducible for all nonnegative integers a.

We have therefore constructed a single polynomial $f(x)$ that has distance ≥ 4 to an irreducible polynomial. We next prove Theorem 2, which shows that there are in fact many such polynomials. Recall that $D_4(n)$ denotes the number of polynomials of degree n that have minimal distance ≥ 4 to an irreducible polynomial in $\mathbb{F}_2[x]$. We also clarify here that again we do not restrict to irreducible polynomials of degree $\leq n$ in our definition of $D_4(n)$. Instead, we count only polynomials of degree n that have distance ≥ 4 from irreducible polynomials in $\mathbb{F}_2[x]$ of *any* degree.

Proof of Theorem 2. Let $n \geq 246$, and set $\varepsilon_n = 1$. We want to show that there are 2^{n-246} different choices of the n-tuple $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{n-1})$, with each ε_j in $\{0, 1\}$, for which

$$F(x) = \varepsilon_0 + \varepsilon_1 x + \cdots + \varepsilon_n x^n$$

has distance ≥ 4 to an irreducible polynomial. Let $f(x)$ be the polynomial from (6). Let $w(x)$ denote the product of x, $x+1$, and all of the polynomials in the right column of Table 6, so that $\deg w = 244$, and $w(0) = w(1) = 0$.

Select ε'_{244}, ε'_{245}, $\ldots$, ε'_{n-1}, each in $\{0, 1\}$, ensuring that a positive even number of these ε'_j are equal to 1. There are therefore $2^{n-245} - 1 \geq 2^{n-246}$ different choices for the ε'_j. Given such a selection, we next show that there exist ε'_0, ε'_1, $\ldots$, ε'_{243}, for which $x^n + \sum_{j=0}^{n-1} \varepsilon'_j x^j$ has distance ≥ 4 from every irreducible polynomial in $\mathbb{F}_2[x]$.

Let $\varepsilon'_n = 1$ and let $g(x) = \sum_{j=244}^{n} \varepsilon'_j x^j$. Observe that $g(1) = 1$ in $\mathbb{F}_2$, and that $g(x)$ has at least three terms of degree ≥ 244 (since $\varepsilon'_n = 1$ and at least two $j \in \{244, 245, \ldots, n-1\}$ satisfy $\varepsilon'_j = 1$). Let $h(x)$ be the unique polynomial in $\mathbb{F}_2[x]$ of degree < 244 satisfying

$$h(x) \equiv g(x) + f(x) \pmod{w(x)}. \tag{7}$$

For $0 \leq j \leq 243$, we define ε'_j in $\{0, 1\}$ so that $h(x) = \sum_{j=0}^{243} \varepsilon'_j x^j$.

We claim that $F(x) = g(x) + h(x)$ has distance ≥ 4 from every irreducible polynomial in $\mathbb{F}_2[x]$. For each factor $\rho(x)$ of $w(x)$, we obtain from (7) that

$$F(x) \equiv g(x) + h(x) \equiv f(x) \pmod{\rho(x)}.$$

Hence, $F(x)$ satisfies the same congruences as those given for $f(x)$ in (4) and (5). More precisely, we have

$$F(x) \equiv x^{a_j} + 1 \pmod{g_j(x)}, \quad \text{for } 1 \leq j \leq 29,$$
$$F(x) \equiv 0 \pmod{x} \quad \text{and} \quad F(x) \equiv 1 \pmod{x+1}.$$

From the definition of $F(x)$, we see that $F(x)$ has at least three terms of degree ≥ 244. As noted earlier, these congruence conditions together with $F(x)$ having at least three terms of degree > 12 are sufficient to ensure that every irreducible polynomial in $\mathbb{F}_2[x]$ has distance ≥ 4 to $F(x)$, completing the proof. □

We remark that the lower bound in Theorem 2 can be improved easily by accounting for a multitude of similar constructions for the polynomial $f(x)$. We do not attempt to make this more precise. We also note that for $n \geq 244$, the polynomial $x^{n-244}w(x) + f(x)$ provides an explicit example of a polynomial of degree n with distance ≥ 4 to every irreducible polynomial in $\mathbb{F}_2[x]$.

Acknowledgements

We thank the Centre for Interdisciplinary Research in the Mathematical and Computational Sciences (IRMACS) at Simon Fraser University for computational resources. We also thank Frank Ruskey for kindly bringing the articles [4, 5] to our attention.

References

[1] P. Banerjee and M. Filaseta, *On a polynomial conjecture of Pál Turán*, Acta Arith. **143** (2010), no. 3, 239–255. MR2652578

[2] A. Bérczes and L. Hajdu, *Computational experiences on the distances of polynomials to irreducible polynomials*, Math. Comp. **66** (1997), no. 217, 391–398. MR1377660 (97c:11035)

[3] ———, *On a problem of P. Turán concerning irreducible polynomials*, Number Theory: Diophantine, Computational and Algebraic Aspects (Eger, Hungary, 1996) (K. Győry, A. Pethő, and V. T. Sós, eds.), de Gruyter, Berlin, 1998, pp. 95–100. MR1628834 (99f:11032)

[4] K. Cattell, F. Ruskey, J. Sawada, M. Serra, and C. R. Miers, *Fast algorithms to generate necklaces, unlabeled necklaces, and irreducible polynomials over GF(2)*, J. Algorithms **37** (2000), no. 2, 267–282.

[5] G. Lee, F. Ruskey, and A. Williams, *Hamming distance from irreducible polynomials over* $\mathbb{F}_2$, Discrete Math. Theor. Comput. Sci. Proc. **AH** (2008), 169–180.

[6] M. J. Mossinghoff, *The distance to an irreducible polynomial*, Gems in Experimental Mathematics (T. Amdeberhan, L. A. Medina, and V. H. Moll, eds.), Contemp. Math., vol. 517, Amer. Math. Soc., Providence, RI, 2010, pp. 275–288.

[7] A. Schinzel, *Reducibility of polynomials and covering systems of congruences*, Acta Arith. **13** (1967), 91–101. MR0219515 (36 #2596)

[8] ———, *Reducibility of lacunary polynomials, II*, Acta Arith. **16** (1970), 371–392. MR0265323 (42 #233)

[9] V. Shoup, *NTL: A library for doing number theory.* www.shoup.net/ntl.

Department of Mathematics, University of South Carolina, Columbia, South Carolina 29208

E-mail address: filaseta@math.sc.edu

Department of Mathematics, Davidson College, Davidson, North Carolina 28035-6996

E-mail address: mimossinghoff@davidson.edu