Article

A quick look at prime numbers

David Alexander Lillis*
443 Muritai Road, Eastbourne, Lower Hutt, 5013

Our fascination with the primes

Since the time of the ancient Greeks, and probably long before then, we have been fascinated by prime numbers. One possible reason for our fascination with prime numbers is that the inter-relationships between them and the patterns that they appear to create are so difficult to understand. Another possible reason is that the primes appear to exhibit both deterministic behaviours (i.e. fully determined on the basis of naturally-occurring relationships) and random behaviours (i.e. occurring by chance). Nobody has succeeded in creating mathematical models that predict the magnitudes of prime numbers exactly or the numbers of prime numbers up to a given natural number. Other reasons for our fascination with primes may involve unexpected and quite exquisite relationships between prime numbers and certain mathematical functions and the seemingly disconnected islands of prime numbers, each surrounded by a sea of numbers that are not prime but which are related mysteriously to one another.

What are prime numbers?

Primes are positive natural numbers, greater than 1, that have two positive divisors (i.e. have no divisors other than themselves and 1). Numbers that are not prime (i.e. have divisors other than themselves and 1) are known as composite numbers. The so-called Fundamental Theorem of Arithmetic tells us that every whole number greater than 1 is either prime or is a unique product of primes, apart from the order of multiplication. For example, 30 can be written as the product of three primes: $2 \times 3 \times 5$. Figure 1 gives a table of the 25 prime numbers that lie between 1 and 100, in which the primes are shown against a shaded background.

Because 1 has only a single positive divisor (itself), it is not a prime. In fact, because a composite number has more than one divisor, 1 is neither prime nor composite.

We note that the gaps between these primes (which we define as the difference between a given prime and the next prime) vary considerably. We see from Figure 1 that, for numbers between 1 and 100, the gaps vary from 1 (i.e. the gap between 2 and 3) to 8 (that between 89 and 97). All prime gaps are even, apart from the first gap and, in fact, the first prime number (2) is the only even prime number.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

Figure 1: A table of the 25 prime numbers between 1 and 100.

The first 15 gaps are as follows (compare with Figure 1):

1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2, 4 and 6

However, though all prime gaps, except the first, are even natural numbers, many even natural numbers are not known prime gaps. Here are the fifteen smallest known gaps in ascending order:

1, 2, 4, 6, 8, 14, 18, 20, 22, 34, 36, 44, 52, 72 and 86.

We see that the even natural numbers 10, 12, 16 and others within the range of this list are not known gap sizes. Thus, the known set of prime gaps is a subset of the even natural numbers.

*Correspondence: david.lillis@nzis.ac.nz

David Lillis is the Senior Academic Manager (Research and Statistics) at the New Zealand Institute of Sport. Prior to joining NZIS, Dr Lillis was a senior statistician at the New Zealand Qualifications Authority.

A former secondary school teacher David has also worked as a Senior Research Evaluator for the Foundation of Research, Science and Technology and has spent over 25 years in the research field.

Dr Lillis is a past president of the New Zealand Association of Scientists.

How numerous are the primes?

There are 168 primes between 1 and 1000, 1229 primes between 1 and 10,000 and 78,498 of them between 1 and one million. We have tables of primes for much bigger natural numbers, but counting the primes becomes more and more difficult as we move along the number line to very large natural numbers.

Around 300 BC Euclid proved that there are infinitely many prime numbers. In more recent times, other famous names have been associated with prime number theory, notably Leonard Euler (1707–1783), Pafnuty Chebyshev (1821–1894), Bernard Riemann (1826–1866), G.H. Hardy (1877–1947), John Littlewood (1885–1977) and Srinivasa Ramanujan (1877–1920). Today, Yitang Zhang, Ben Green, Terence Tao and James Maynard are among the group that is advancing our knowledge in this very challenging field.

Figure 2 gives a graph of the prime-counting function, the actual number of primes up to and including a given natural number, usually denoted by the Greek symbol **π(n)**. Specifically, Figure 2 gives the number of primes up to 60.

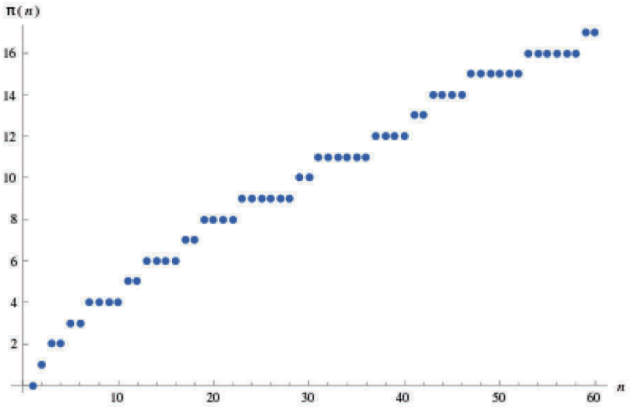


Figure 2: The prime-counting function for natural numbers up to 60.

We see that the number of primes increases in steps as we move along the natural number line and as we encounter each new prime number. Note that $\pi(1) = 0$ because 1 is not a prime, $\pi(2) = 1$, $\pi(3) = 2$, $\pi(4) = 2$ and $\pi(5) = 3$. Notice that there appears to be some curvature in this graph. In fact, as we move to greater and greater natural numbers, on average the primes do become sparser and the gaps between them tend to increase.

Some Mathematical Expressions

Over the last one hundred and fifty years, much work on primes has centered on the development of mathematical models that give:

1. Estimates of the number of primes up to a given natural number n
2. Estimates of the magnitude of the n'th prime number
3. Estimates of the size (average, minimum and maximum) of the gaps between adjacent primes at different scales of the natural numbers
4. Indications of the occurrence and recurrence of patterns within the set of primes (e.g. the occurrence of arithmetic series and even polynomial series).

We now start our beginner's review of prime number theory by examining some results that have been known for well over a

Bernhard Riemann (1826 – 1866)

century. One approximation for the magnitude of the n'th prime number is the following:

$$\mathbf{p(n) \sim n\log_e(n)}$$

. . . where the tilde means 'approximately equal to. In other words, we have a mathematical model of the size of the n'th prime, but not an exact model.

More of interest to researchers at present is the development of models of the number of primes up to a given natural number. Researchers have produced extensive tables of the numbers of primes that are updated almost every year to larger and larger natural numbers. One useful approximation (due to Carl Friedrich Gauss and Adrien Legendre) for the number of primes up to a given natural number n is:

$$\boldsymbol{\pi(n) \sim n / \log_e(n)}$$

. . . where, of course, **π(n)** represents the actual total number of primes up to **n**.

This expression is known as the prime number theorem and also as the asymptotic law of distribution of prime numbers. Though Chebyshev demonstrated that this expression is correct to within about ten percent of the true number of primes up to n, it appears to underestimate the true number persistently. For **n** = 10^8, the actual number of primes **π(n)** = 5,761,455, while the prime number theorem estimates **π(n)** at 5,438,681, which is 322,774 (about 5.6%) too few.

Aware of the limitations of the prime number theorem, Riemann provided a better estimate (Riemann 1859), as follows:

$$\mathbf{Li(n) \sim \int_0^N dt / \log_e(t)}$$

. . . where t is a dummy variable. This expression estimates the true number of primes up to a natural number n with greater precision. Thus, for **n** = 10^8, **Li(n)** = 5,762,209, which is only 754 (about 0.01%) too many. Thus, Riemann's expression represents a significant improvement.

For some years it was believed that **Li(n)** always overestimates the true number. However, Littlewood (1914) demonstrated that for very large natural numbers the expression underestimates the true number and, thereafter at greater and greater scales, successively underestimates and overestimates infinitely often (known as 'Littlewood Violations'). In other words, the difference **π(n) – Li(n)** changes sign infinitely often.

However, violations of the rule that **Li(n)** overestimates **π(n)** occur at numbers that are very large and it is not surprising that it took several decades from the original work of Riemann to identify these violations. We now know that Littlewood Viola-

John Littlewood (1885 – 1977)

tions occur at approximately 1.398×10^{316} (Bays and Hudson, 2000) where more than 10^{153} consecutive natural numbers exist for which the sign of $\pi(n) - Li(n)$ is positive. There are large zones of violation at much greater numbers again but Büthe (2015) has demonstrated that Littlewood Violations do not occur at natural numbers below 10^{19}.

Prime numbers and the zeta function

We now consider the zeta function whose relevance to prime numbers will be seen in the next section. One simple version of the zeta function is as follows:

$$Z(s) = 1 + 1/2^s + 1/3^s + 1/4^s + 1/n^s + \ldots = \sum_0^\infty (1 / n^s) = \sum_0^\infty n^{-s}$$

. . . where $\mathbf{s}$ is a real number.

In 1737 Leonard Euler demonstrated that the zeta function can be re-written as follows:

$$\sum_0^\infty (1 / n^s) = \prod_{primes} [1 / (1 - 1/P_s)]$$

. . .where P_s are the prime numbers and the symbol $\prod_{primes}$ is an instruction to multiply all terms in brackets involving all prime numbers. Thus, we already have a link between an infinite series of inverse powers of natural numbers and an infinite series of primes.

Figure 3 gives a graph of the zeta function.

For $\mathbf{s} > 1$ the zeta function always tends asymptotically to infinity as s tends towards 1 from above and tends asymptotically to 1 as $\mathbf{s}$ becomes larger (though the precise curve depends on the value of $\mathbf{s}$). For values of $\mathbf{s} < 1$ the graph shows peaks and troughs whose amplitudes depend on the value of $\mathbf{s}$. The graph is very complicated but we see roots (places where the graph crosses the horizontal axis) at every negative even whole number. These zeroes at the negative even whole numbers are known as the trivial roots of the zeta function. However, other non-trivial roots exist which are harder to find and harder to understand.

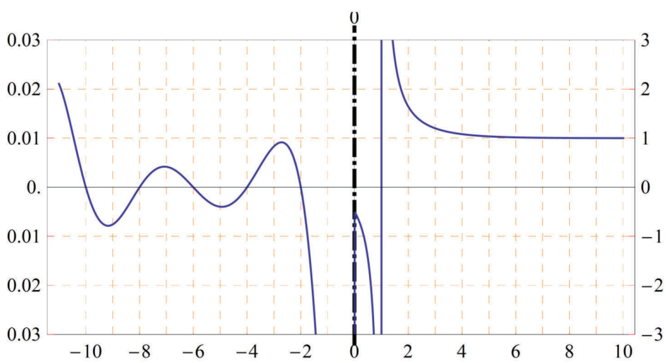


Figure 3: Graph of the zeta function.

For a complete discussion of these roots we must consider the zeta function applied on complex numbers. You may remember from high school that complex numbers are numbers of the form:

$$a + bi$$

where $\mathbf{i}$ is the square root of –1. Many texts discuss the zeta function very clearly, especially that by John Derbyshire (Derbyshire 2004). A discussion of the application of the zeta function on complex numbers is beyond the scope of this article. However, over the last century, a great deal of work has gone into understanding the non-trivial solutions to the zeta function applied on complex numbers.

The Riemann Hypothesis

The Riemann Hypothesis, one of the long-standing challenges of pure mathematics, has implications for prime number theory. The Riemann Hypothesis involves the following conjecture:

All non-trivial zeroes of the zeta function have real part equal to ½.

Hardy (1914) showed that infinitely many of the non-trivial roots have real part ½ but this result does not prove that absolutely all non-trivial roots have that value. If the Riemann Hypothesis can be proved, then certain conjectures about primes will be confirmed or proved untrue. However, it has not yet been proved or disproved and prime number theorists are divided in their opinions on whether the conjecture is true or not. If it is true, then we can write:

$$\pi(n) = Li(n) + O[n^{1/2}\log_e(n)]$$

. . . where the function $\mathbf{O}$ indicates an error term that remains bounded within the graph of the function $n^{1/2}\log_e(n)$. In other words, we can be more precise in our estimates of the numbers of primes. Figure 4 gives a graph of this error term up to 10,000 (which was created using R, an environment for statistical analysis and graphics).

The error term, the difference between the actual number of primes and the function $Li(n)$, is suppressed within the bounds of our function, enabling increased accuracy in our estimates of $\pi(n)$.

More mathematical expressions

Now we consider a particular function of the prime-counting function – the J function. Due to Riemann (Riemann 1859), it is written as follows:

$$J(n) = \pi(n) + ½\,\pi(n^{1/2}) + 1/3\,\pi(n^{1/3}) + 1/4\,\pi(n^{1/4}) + \ldots$$

This function can be inverted to give the prime-counting function on the left hand side:

$$\pi(n) = J(n) - ½\,J(n^{1/2}) - 1/3\,J(n^{1/3}) - 1/5\,J(n^{1/5}) + 1/6\,J(n^{1/6}) - 1/7\,J(n^{1/7}) + 1/10\,J(n^{1/10}) + \ldots$$

. . . where certain terms appear to be missing (in fact, terms involving 1/4 and 1/9, etc. have disappeared quite legitimately) and we now have both plus and minus signs. It can be demonstrated that:

$$\log_e[Z(s)] = s \int_0^\infty J(t).t^{-s-1}\,dt$$

. . . where $\mathbf{t}$ is a dummy variable. A proof of this expression is beyond the scope of this article, but it is demonstrated clearly

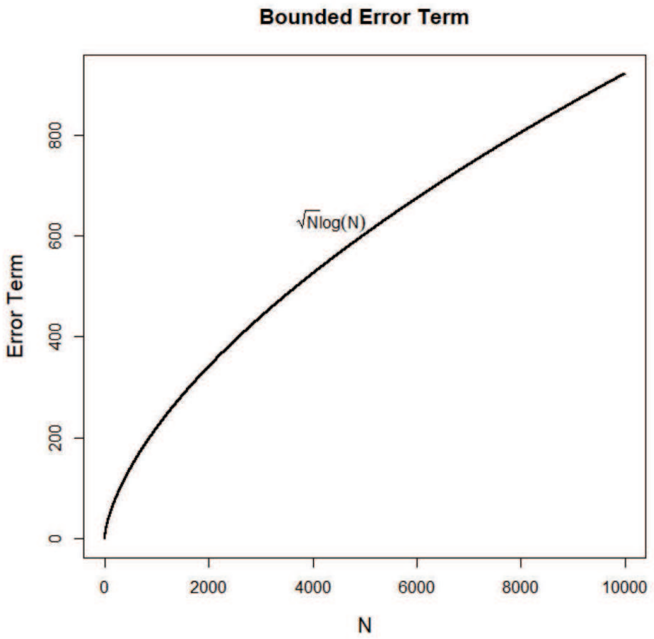


Figure 4: The bounded error term.

in John Derbyshire's book on pages 303–311. The result of interest for us is that, since **J(n)** is a function of the prime-counting function **π(n)**, we have a firm link between the prime-counting function and the zeta function. We have another link between the zeta function and the Riemann Hypothesis and therefore we have a link between the prime numbers and the Riemann Hypothesis.

The Green–Tao Theorem and arithmetic progressions within the primes

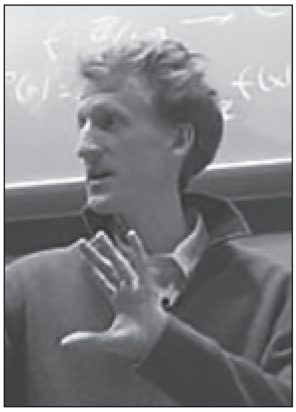
In 2004 Benjamin Green and Terence Tao (Green & Tao 2008) published the Green–Tao Theorem. They showed that the prime numbers include arithmetic progressions of arbitrary length. Thus, there are arithmetic progressions of prime numbers with every possible number of terms. Put another way, for every natural number n, the primes contain arithmetic progressions of length n.

As an example, consider the arithmetic progression of primes: 5, 11, 17,23, 29 in which the gaps are of magnitude 6. This progression is of length 5 and terminates at 29 because 35 is not a prime. The Green–Tao Theorem tells us that arithmetic progressions like this example exist within the primes but does not predict progression length or gap size, how to identify them or where they exist on the number line.

Small gaps between prime numbers

Over the last five years much progress has been made in our understanding of the gaps between the prime numbers. The lectures given by Professor Tao at the Department of Mathematics of the University of California Los Angeles on the gaps between the prime numbers, available on the Internet, provide a very helpful introductory synopsis (the URLs are given in the references for this article).

Much progress is being made on both how small and how large the gaps between adjacent primes can be at different scales of the natural numbers. In fact, gaps can be thought of in terms of the number of composite numbers lying between two adjacent primes (i.e. we could subtract 1 from the difference between the



Ben Green

Terry Tao

two adjacent primes). However, the gap **G[P(n)]** between two primes (the **n**'th and **n**+1'st primes) is usually defined as the difference between those adjacent primes, and that approach has been adopted in this article. Thus:

$$G[\ P(n)\ ] = P(n+1) - P(n)$$

We can relate the primes to the prime gaps as follows:

$$P(n + 1) = 2 + \sum_1^n G[\ P(t)\ ]$$

Apart from the first gap of size 1 (that between 2 and 3), the smallest possible gap is of size 2 (for example, that between the primes 11 and 13 and between 71 and 73). So, how can small gaps between primes at different scales of the natural numbers be characterised?

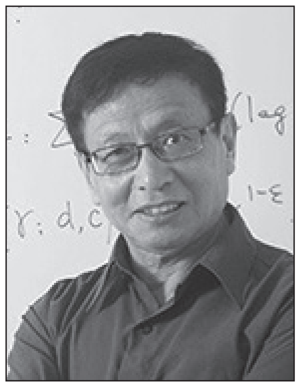
In May 2013 Yitang Zhang demonstrated that gaps less than or equal to 70,000,000 occur infinitely often. This number is not particularly special and simply emerges from the assumptions and approximations that Zhang adopted in his proof.

Within a few months other workers trimmed this number down to 4680 (a group within the Polymath Project – a project in which mathematicians collaborate in solving mathematical problems), then down to 600 (Maynard 2013) and then to 246 (the Polymath Project in April 2014).

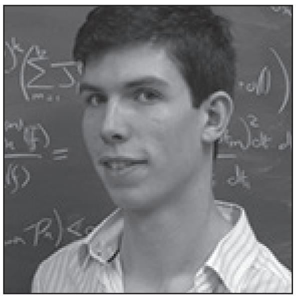
James Maynard's work in this area was undertaken as a very young mathematician, for the most part independently of other workers.

Under certain assumptions it can be demonstrated that gaps less than or equal to 6 occur infinitely often. However, because this result depends critically on those assumptions, we are not entirely sure that gaps of size 6 do occur infinitely often. Gaps less than or equal to 4 (the so-called Cousin Primes) may occur infinitely often but we do not yet have a definitive proof. Gaps equal to 2 may also occur infinitely often (the Twin Prime Con-

Yitang Zhang



James Maynard



jecture). At present, completely new approaches are required in order to verify the Twin Prime Conjecture, and Professor Tao believes that a proof could be more than a decade away.

Large gaps between the primes

Professor Tao's lectures provide an introduction to current work on large gaps between primes, particularly his lecture Small and Large Gaps between the Primes (see the first of the references to his lectures for the relevant URL). Professor Tao states that, across all of the primes, gaps can be of any size (i.e. arbitrarily large). He also states that our best estimate for the upper bound on the gap size between a particular pair of adjacent primes is as follows:

$$P(n+1) - P(n) < P(n)^{0.55}$$

Consider the adjacent primes 89 and 97. The gap between them has size 8 but the above expression predicts an upper bound of 11.8 (which we round up to 12). The expression works in the sense that it gives an estimate of the maximum size of a gap for a pair of adjacent primes, but it does not give us the actual gap size.

Thus, the prime gaps are at most just slightly larger than the square root of **P(n)**. However, if the Riemann Hypothesis is true, then we have a slight improvement:

$$P(n+1) - P(n) < P(n)^{0.5} \log_e[\, P(n) \,]$$

For the pair of adjacent primes 89 and 97, the predicted maximum gap size is now 42. Let's also use this expression for the adjacent primes 2393 and 2399. The improved expression predicts an upper bound on the gap size of 381, considerably larger than the actual gap of 6. However, we must remember that these expressions are intended to give upper bounds on gap sizes at different scales of the natural numbers, rather than precise gap sizes.

Even this expression is not expected to be the final, definitive result, and improvements will most probably be found. However, if the Riemann Hypothesis is verified, we will have more precise expressions for the upper bounds on gap sizes for adjacent primes at different scales of the natural numbers.

A voyage down the number line

On the basis of the Green–Tao theorem it appears that we can choose any natural number and there will be at least one arithmetic progression of this length. However, at any given scale of the natural numbers, gap sizes may be constrained to a subset of allowable gaps that are greater than or equal to 2 and less than or equal to certain maximum gap sizes that possibly are roughly consistent with the expressions of the previous section or with extensions of those expressions.

Here are some interesting questions that might be answered in the future:

1. A progression of primes at a certain scale may occur once or finitely often, but could it recur infinitely often and, if so, under what conditions could such a progression recur infinitely often? If progressions do recur, are progressions of small length and small gap size more likely to recur than longer progressions with large gap sizes?
2. If such a progression recurs finitely often or infinitely often then, just as the primes become sparser, on average, at greater and greater scales of the natural numbers, would repetitions of this progression tend to become sparser, on average, as we move to greater and greater scales of the natural numbers?
3. In many or all repetitions of this progression (and other similar progressions), all elements may occur purely deterministically but, in some repetitions, could certain elements appear by chance (i.e. from a purely random process) rather than deterministically?

Perhaps such questions cannot be answered directly through the Green–Tao Theorem (which states only that there are arithmetic progressions of arbitrary length but does not say anything about what those elements should be). It may appear counter-intuitive to have infinitely many recurring progressions of certain arithmetic progressions of particular lengths and gap sizes. However, the primes are infinite in extent and so the possibilities for such progressions to emerge may be limitless, except for possible inherent constraints on recurrences imposed by naturally-occurring relationships between the primes themselves. Nature provides different levels of infinity. Thus, even if we select only one whole number from every thousand trillion, we nevertheless select an infinite number of whole numbers. Applying this line of thought, perhaps it is possible for the primes to contain an infinite number of arithmetic progressions involving particular combinations of length and gap size.

Imagine that we are to travel down the number line at very high speed and stop to look at every occurrence of one particular progression. Our first stop may take only a few minutes to reach and perhaps that is the one and only stop. However, there may be others. Those other stops, further down the number line, if they exist, may take days, weeks or years to get to. As we get to very large scales, our next stops may take thousands of years and possibly much more, but our journey may last forever.

Future work on primes

The primes appear to be characterised by unexpected relationships with certain mathematical functions and those seemingly disconnected islands of prime numbers which are mysteriously related to one another, though they are surrounded by a sea of numbers that are not prime. Like other areas of mathematics, resolution of one question about primes seems to result in other questions.

Our understanding of prime numbers is increasing every year. Future work on primes may depend on a resolution of the Riemann Hypothesis, but could include attempts to prove the Cousin Prime Conjecture (that primes separated by 4 occur infinitely often) and the Twin Prime Conjecture (that primes separated by 2 occur infinitely often). Other valuable work may include refinements of our current models of the magnitudes of the n'th prime and the numbers of primes up to a given natural number.

References

Bays, C., Hudson, R. H. (2000). A new bound for the smallest **x** with **π(x) > i(x)**. *Mathematics of Computation 69 (231)*: 1285–1296.

Büthe, J. (2015). An analytic method for bounding **ψ(x)**. Retrieved from https://arxiv.org/pdf/1511.02032.pdf

Derbyshire, J. (2004). *Prime Obsession*. Plume.

Green, B., Tao, T. (2008). The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics 167(2)*: 481–547.

Hardy, G.H. (1914). On the zeros of Riemann's Zeta-function. *Proceedings of the London Mathematical Society.* (records of proceedings at meetings), ser. 2, vol. 13, 12, March 1914, p. 14.

Maynard, J. (2013). Small Gaps between Primes. Retrieved from https://arxiv.org/abs/1311.4600

Littlewood, J.E. (1914). Sur la distribution des nombres premiers. *Comptes Rendus 158*: 1869–1872, JFM 45.0305.01

Riemann, B. (1859). Über die Darstellbarkeit einer Function durch eine trigonometrische Reihe. Reprinted in *Gesammelte math. Abhandlungen*. New York: Dover, p. 227-264, 1957.

Professor Tao's Lectures on the Prime Numbers

Small and Large Gaps between the Primes. https://www.youtube.com/watch?v=pp06oGD4m00

Structure and Randomness in the Prime Numbers. https://www.youtube.com/watch?v=PtsrAw1LR3E