Prime numbers, Graphs and Cryptography

Gabor Sarkozy

- Prime numbers and some of their basic properties
- Relatively prime numbers and graphs
- An application: the RSA cryptosystem

Prime numbers

Natural numbers which are not multiples of any smaller natural number (except 1). E.g. 2, 3, 5, 7, 11, ...

The "oddest" prime number is 2, because it is not odd.

If a natural number is neither 1 nor a prime, it is called *composite*.

How many prime numbers are there?

Theorem 1 (Euclid, 3rd century BC) *There exist infinitely many primes.*

There are many-many proofs (altough not infinitely many) of this theorem. Here we give two.

Euclid's proof: Suppose that $p_1 = 2 < p_2 = 3 < \ldots < p_r$ are all the primes. Let $P = p_1 p_2 \ldots p_r + 1$ and let p be a prime dividing P; then p cannot be any of $p_1, p_2, \ldots, p_r$ otherwise p would divide the difference $P - p_1 p_2 \ldots p_r = 1$, which is impossible. So this prime p is still another prime, and $p_1, p_2, \ldots, p_r$ would not be all the primes, a contradiction.$\Box$

Pólya's proof: This proof uses the following idea: It is enough to find an infinite sequence of natural numbers $1 < a_1 < a_2 < \ldots$ that are pairwise relatively prime (i.e. without a common prime factor). So if p_1 is a prime dividing a_1, p_2 is a prime dividing a_2, etc. then $p_1, p_2, \ldots$ are all different.

For this proof, the numbers a_n are chosen to be the Fermat numbers $F_n = 2^{2^n}+1, (n \geq 0)$. Indeed, it is easy to see, by induction on m, that

$$F_m - 2 = F_0 F_1 \ldots F_{m-1}$$

hence, if $n < m$, then F_n divides $F_m - 2$. If a prime would divide both F_n and F_m, then it would divide $F_m - 2$ and F_m, hence also 2, so $p = 2$, contradiction.$\square$

$F_0 = 3, F_1 = 5, F_2 = 17, F_3 = 257, F_4 = 65537$ are all primes.

Are all Fermat numbers primes?

No. $F_5 = 641 \times 6700417$. (Euler)

Open problems: Are there infinitely many prime Fermat numbers? Are there infinitely many composite Fermat numbers?

Largest known Fermat prime is F_4. Largest known composite Fermat number is F_{23471}.

Fermat is famous for two other things.

Theorem 2 (Fermat's Little Theorem) *If p is a prime and if a is a natural number, then $a^p \equiv a \pmod{p}$. In particular, if p does not divide a, then $a^{p-1} \equiv 1 \pmod{p}$.*

Proof: It is true for $a = 1$. Assuming that it is true for a, then, by induction,

$$(a+1)^p \equiv a^p + 1 \equiv a + 1 \pmod{p}.$$

So the theorem is true for every natural number a.$\square$

Theorem 3 (Fermat's Last Theorem) *If n is an integer greater than 2, then $x^n+y^n = z^n$ has no solution in positive integers.*

Fermat wrote in 1637: "I discovered a marvelous proof but the margins were too small to hold it."

Finally it was proved by Taylor and Wiles in 1995.

Euler generalized Fermat's Little Theorem by introducing $\phi(n)$, or Euler's function. This is the main idea in RSA.

(n, m) denotes the largest common divisor of n and m. For every $n \geq 1$, $\phi(n)$ denotes the number of integers a, $1 \leq a \leq n$, such that $(a, n) = 1$. Thus if $n = p$ is a prime, then $\phi(p) = p - 1$; also

$$\phi(p^k) = p^{k-1}(p-1) = p^k \left(1 - \frac{1}{p}\right).$$

Moreover, if $m, n \geq 1$ and $(m, n) = 1$, then $\phi(mn) = \phi(m)\phi(n)$, that is ϕ is a multiplicative function. So if p, q are two primes then (RSA)

$$\phi(pq) = (p-1)(q-1).$$

Theorem 4 (Euler-Fermat Theorem) *If* $(a, n) = 1$, *then* $a^{\phi(n)} \equiv 1 \pmod{n}$.

Proof: Let $r = \phi(n)$ and let $b_1, b_2, \ldots, b_r$ be integers, pairwise incongruent modulo n, such that $(b_i, n) = 1$. Then $ab_1, ab_2, \ldots, ab_r$ have the same properties, thus the sets

$$\{b_1 \bmod n, \ldots, b_r \bmod n\}$$

and

$$\{ab_1 \bmod n, \ldots, ab_r \bmod n\}$$

are equal. Now,

$$a^r \prod_{i=1}^{r} b_i \equiv \prod_{i=1}^{r} ab_i \equiv \prod_{i=1}^{r} b_i \pmod{n}.$$

Hence,

$$(a^r - 1) \prod_{i=1}^{r} b_i \equiv 0 \pmod{n}$$

and so $a^r \equiv 1 \pmod{n}$.$\square$

Can we reverse Fermat's Little Theorem? For example is it true that if $2^{n-1} \equiv 1 \pmod{n}$, then n must be a prime?

No, $2^{340} \equiv 1 \pmod{341}$ yet $341 = 11 \times 31$ (Straus, 1819). These numbers are called pseudoprimes and there are infinitely many pseudoprimes.

Even worse, there are composite numbers for which $a^{n-1} \equiv 1 \pmod{n}$ for every integer a, $1 < a < n$, $(a, n) = 1$. These are called Carmichael numbers. The smallest Carmichael number is $561 = 3 \times 11 \times 17$. Open problem: Are there infinitely many Carmichael numbers?

Mersenne numbers: If a number of the form $2^m - 1$ is a prime, then $m = q$ is a prime. Indeed, if p divides m, then $2^p - 1$ divides $2^m - 1$. The numbers $M_q = 2^q - 1$ (with q prime) are called Mersenne numbers. Some of them prime, others composite. For example $M_2 = 3, M_3 = 7, M_5 = 31, M_7 = 127$ are primes, but $M_{11} = 23 \times 89$ is composite.

Open problems: Are there infinitely many Mersenne primes? Are there infinitely many composite Mersenne numbers?

The known largest primes are Mersenne primes.

Mersenne primes are related to perfect numbers. A natural number $n > 1$ is perfect if it is equal to the sum of its proper divisors, i.e. divisors d with $d < n$. E.g. 6, 28, 496, 8128 are perfect. The following is known (Euclid, Euler): If q is a prime and $M_q = 2^q - 1$ is a Mersenne prime, then $n = 2^{q-1}(2^q - 1)$ is a perfect number and these are the only even perfect numbers. Open problem: Are there odd perfect numbers?

How many primes are there? Let us denote the number of primes up to x by $\pi(x)$. We know this number quite precisely. The notation $f(x) \sim g(x)$ means that

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1,$$

$f(x)$ and $g(x)$ are then said to be asymptotically equal as x tends to infinity.

Here is the most important theorem about primes:

Theorem 5 (Prime Number Theorem)

$$\pi(x) \sim \frac{x}{\log x}.$$

History: Euler, Legendre, Gauss, Tschebycheff, Riemann,...

First proof using analysis (properties of $\zeta(s)$): de la Vallee Poussin, Hadamard (1896).

First elementary proof: Erdős, Selberg 1949.

How do we find the prime numbers?

Sieve of Eratosthenes: We determine all the prime numbers up to 101. We write all the numbers up to 101; cross out all the multiples of 2, bigger than 2; in each subsequent step, cross out all the multiples of the smallest remaining number p, which are bigger than p. It suffices to do it for $p^2 < 101$. Primes remaining: 2, 3, 5, 7, 11, 12, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101.

This procedure is the basis of sieve theory.

How are the prime numbers distributed?

There can be arbitrarily large gaps between primes; for every $N > 1$ there exists a string of at least N consecutive composite integers, e.g.

$$(N+1)!+2, (N+1)!+3, \ldots, (N+!)!+(N+1)$$

Small gaps: If p and $p + 2$ are primes, they are called twin primes. Smallest pairs of twin primes are $(3, 5), (5, 7), (11, 13)$ and $(17, 19)$. Open problem: Are there infinitely many twin primes?

Relatively prime numbers and graphs

We consider the coprime graph on the integers. This is the graph whose vertex set is the set of integers and two integers a, b are connected by an edge if and only if $(a, b) = 1$. Let $A \subset \{1, 2, \ldots, n\}$ be a set of positive integers. The *coprime graph* of A, denoted by $G(A)$, is the induced coprime graph on A.

What kind of properties does this graph have?

What can we say about complete subgraphs in $G(A)$?

Erdős (1962) (K_k is the complete graph on k vertices: What is largest set

$$A \subset \{1, 2, \ldots, n\}$$

such that $K_k \not\subset G(A)$? Of course, the set of $m \leq n$ which have a prime factor among the first $k - 1$ primes is such a set, and Erdős conjectured that this set gives the maximum.

$k = 2, 3$ trivial,

$k = 4$ Szabó, Tóth (1985)

$k = 5$ Mócsy (1989).

However, Ahlswede and Khachatrian disproved the conjecture in 1995 for $k = 212$. They also gave some positive results, for example the conjecture is true if $n \geq n_0(k)$.

How about perfect matchings?

Newman conjectured more than 25 years ago, that if $I_1 = \{1, 2, \ldots, n\}$ and I_2 is any interval of n consecutive integers, then there is a perfect coprime matching from I_1 to I_2. This conjecture was proved by Pomerance and Selfridge (1985). Note that the statement is not true if I_1 is also an arbitrary interval of n consecutive integers. Example: $I_1 = \{2, 3, 4\}$ and $I_2 = \{8, 9, 10\}$, any one-to-one correspondance between I_1 and I_2 must have at least one pair of even numbers in the correspondance.

Finally what about cycles?

Let C_l denote the cycle of length l.

Even cycles: For the largest set $A \subset \{1, 2, \ldots, n\}$ with $C_{2l} \not\subset G(A)$, we have $|A| = \frac{1}{2}n + (l-1)$ (Erdős). This cardinality can be obtained by taking all the even numbers and the first $l-1$ odd primes, then obviously $C_{2l} \not\subset G(A)$. Thus the bound increases as the cycle length increases.

Odd cycles: C_3 is just K_3, so assuming 6 divides n, to get a C_3 we need more than $\frac{2}{3}n$ numbers. Somewhat surprisingly this cardinality already guarantees the existence of odd cycles of "almost" every length. More precisely:

Theorem 6 (Erdős, G.S. 1996)) *There exist contants c, n_0 such that if $n \geq n_0$, $A \subset \{1, 2, \ldots, n\}$ and $|A| > \frac{2}{3}n$, then $C_{2l+1} \subset G(A)$ for every positive integer $l \leq cn$.*

Open problem: What is the best possible value of the constant c? Conjecture: $c = 1/6$ (a trivial upper bound).

The RSA cryptosystem

This is a public key cryptosystem. In such a system, every person has a publicly known encryption key. Only the decryption key is kept secret, and the idea is that the encryption key does not let someone find the decryption key without an extraordinary amount of work.

It was invented at MIT by Ron Rivest, Adi Shamir and Len Adleman in 1976. Here is how it works. Each individual has an encryption key (n, s) where $n = pq$ is a product of two large primes, say with 200 digits each. s is a number that is relatively prime to $(p-1)(q-1)$.

Remember that:

$$\phi(pq) = (p-1)(q-1)$$

and, if $(a, n) = 1$, then

$$a^{\phi(pq)} \equiv a^{(p-1)(q-1)} \equiv 1 \pmod{n}.$$

Say A wants to send message M to B. First A converts the message into a number. Each letter corresponds to a two-digit number, a blank to the number 00, and $a = 01, b = 02, c = 03, \ldots$

Then A sends to B the encoded message

$$M_1 \equiv M^s \pmod{n}.$$

In order to decode M_1, B calculates a t such that

$$1 \leq t \leq \phi(pq) = (p-1)(q-1)$$

and

$$ts \equiv 1 \pmod{(p-1)(q-1)}.$$

This is not hard by using the Euclidean algorithm and it is done once and for all. Then to decode (we may assume that $(M, n) = 1$)

$$M_1^t \equiv M^{ts} \equiv M \pmod{n}$$

so B may read the message M.

It is also possible for A to send a "signed message" to B, so that B is able to know that unmistakably the message does come from A and could not have been forged.

For this purpose every user A has a different s_A (public). A also has the secret t_A for which

$$1 \leq t_A \leq \phi(pq) = (p-1)(q-1)$$

and

$$t_A s_A \equiv 1 \pmod{(p-1)(q-1)}.$$

Then A sends to B the encoded message

$$M_1 \equiv M^{s_B t_A} \pmod{n}.$$

B decodes this to get

$$M_1^{t_B} \equiv M^{t_B s_B t_A} \equiv M^{t_A} \pmod{n}.$$

Then using the public s_A

$$(M^{t_A})^{s_A} \equiv M^{s_A t_A} \equiv M \pmod{n},$$

so the message M must indeed come from A.

Example: We receive the encrypted message 0981 0461. What is the decrypted message if $n = 43 \times 59 = 2537$ and $s = 13$. Computation shows that $t = 937$, so $13 \times 937 \equiv 1$ modulo $42 \times 58 = 2436$. We do it usually in blocks of size four:

$$0981^{937} \equiv 0704 \pmod{2537}$$

$$0461^{937} \equiv 1115 \pmod{2537}$$

So the decrypted message is 0704 1115 and the message is HELP.

The original RSA 129 contained two primes with 129 digits and the inventors predicted that it would take others at least 40 quadrillion years to find them. It took a little bit less than that: 19 years. By using the Internet and volunteers a group directed by Lenstra found the original primes in 1995. Of course now they are using bigger primes around 200 digits, which are safe for now.