



HAL
open science

Number Theory of Real Numbers: A Canonical Representation Approach

Luciano da Fontoura Costa

► **To cite this version:**

Luciano da Fontoura Costa. Number Theory of Real Numbers: A Canonical Representation Approach. 2024. <hal-04823322>

HAL Id: hal-04823322

<https://hal.science/hal-04823322v1>

Preprint submitted on 6 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

Number Theory of Real Numbers: A Canonical Representation Approach

Luciano da F. Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

18th Oct. 2024

Abstract

Number theory is an interesting mathematical area focusing on integer values and arithmetic functions, especially on topics including prime factor decompositions and properties. Relatively fewer developments have been developed respectively to real numbers, including their rational and irrational subsets. In the present work, real numbers are approached from a particular number-theoretic perspective focusing on canonical representations. In particular, it is shown that canonical representations can be generalized to rational numbers by using np-sets, namely multisets involving non-negative and non-positive multiplicities. This np-set based approach also allows not only effective representation and performance of products and divisions between rational numbers, but can also be directly related to the identification of the least common denominator and greatest common divider of two rational numbers. In addition, it is also argued that the comparison between the canonical representations of two rational numbers can be performed in terms of the np-set Jaccard similarity index between those representations, actually corresponding to the quotient between the greatest common divider and least common denominator. Also developed in this work is the possibility to characterize the complexity of both rational and irrational numbers in terms of the Jaccard similarity between the number of primes factors and total multiplicity signatures obtained by successive respective decimal expansion of this type of real numbers.

1 Introduction

Through the long history of science, great interest has been continuously placed on the area known as *number theory* (e.g. [1]), an area involving both theoretical (e.g. [2, 3, 4, 5, 6, 7, 8]) and computational aspects (e.g. [9, 10, 7, 11, 12]) which addresses several aspects and properties related to the basic structure of numbers, with emphasis on their prime factor structure and arithmetic operations. Introductory texts to number theory include (e.g. [8, 13, 14]). Number theory has not only been approached from theoretical and computational perspectives, but also motivated a wide spectrum of interest ranging from scientific dissemination works involving accessible mathematical concepts (e.g. [15, 16, 17]) to the most intricate mathematical approaches including the proof of Fermat's last theorem (e.g. [18, 19]).

Yet, despite the continuous and intense attention the area of number theory has received, most developments in this interesting area have been focused on positive integer values, while real numbers have received relatively less attention (e.g. [20, 21]).

In the present work, some possible developments related to the extension of number theory to real numbers are described, with emphasis on the adoption of concepts and methods related to np-sets [22], namely a generalization of multisets (e.g. [23, 24, 24, 25, 26]), involving two respective multiplicities, one non-negative and another non-positive. These two multiplicities are needed in order to accommodate the results from the union of two np-set, which can possibly include both types of multiplicities.

The adoption of np-set concepts [22], properties and methods allowed several interesting results, including the effective expression of canonical representation of positive rational numbers, as well as the performance of their pairwise products and divisions, which are directly associated to the sum and subtraction between the np-set representations of the canonical forms. It has also been shown that the least common denominator and greatest common divider are directly related to the np-set operations of union and intersection. The np-set Jaccard similarity index was also verified to correspond to the division between these two quantities.

Emphasis has also been placed on an attempt to quantify the complexity of real numbers from the specific perspective of their prime factors structures. The approach respectively developed in the current work to characterize the prime factors structure of real numbers, including irrational numbers, corresponds to obtaining successive respective decimal expansions which, as integer values, can be decomposed into prime factors.

This work starts by revising some basic concepts from rational and irrational numbers, as well as the subjects of prime factor factorization and canonical representations. Then, some basic concepts of np-sets are summarized, which is followed by applying them to canonical representations, as well as integer and rational arithmetics. The subjects of least common denominator and greatest common divider are subsequently addressed in terms of the np-set extension of the Jaccard similarity index. The possibility to employ the described concepts and methods to estimate the complexity, from the perspective of prime factor structure, of rational and real numbers is then described and illustrated.

2 Rational and Irrational Numbers

Fractional numbers are characterized by possibly presenting a fractional part, placed at the right-hand side of the decimal point. Though integer values can be understood as particular cases of fractional numbers (with null fractional parts), the term fractional number is typically applied to non-null decimal portion. Generally speaking, real numbers are fractional numbers including integer, rational and irrational numbers (e.g. [27]) as subsets.

Given a fractional number its i -th *decimal expansion* corresponds to the truncation of its product by 10^i , i.e.:

$$a_i = \lfloor a 10^i \rfloor \quad (1)$$

As an example, in the case of $a = 0.2468$, we would have:

$$\begin{aligned} a_0 &= 0 \\ a_1 &= 2 \\ a_2 &= 24 \\ a_3 &= 246 \\ a_4 &= 2468 \end{aligned}$$

Two main types of fractional numbers are contained in the set of real numbers \mathbb{R} : *rational* and *irrational*. *Rational numbers* are those which can be expressed as the division between two integer values, which is not the case for *irrational numbers*. Examples of irrational numbers

include π , $\sqrt{2}$, e , as well as the golden ratio $(1 + \sqrt{5})/2$. In addition, every square root of a non-perfect square number can be shown to correspond to an irrational number. There is, however, no definite procedural specification of all possible irrational numbers. For instance, it is now known if πe or $\pi + e$ are irrational numbers. At the same time, almost all real numbers are irrational. The study of irrational numbers, including additional such properties, constitutes an interesting research area (e.g. [27])

It follows from the own nature of rational numbers of being expressible as the division between two integer numbers that the the inverse of a rational number is necessarily a rational number. At the same time, the sum or product of two irrational numbers may or not be an irrational number, but a sum of a rational and an irrational number is necessarily an irrational number.

The *fractional part* of a real number corresponds to the digits at the right of the *decimal point*. A *terminating* real number has only zeros following one of its finite fractional positions. For instance, the number $a = 1.234$ is a finite real number with fractional part 0.234. This type of numbers are necessarily rational, because they can be readily expressed as a quotient between two integer values. For instance, the number a in the previous example can be represented as:

$$a = \frac{1234}{1000} = 0.1234 = 0.1234000 \dots$$

There is another type of rational number, called *repeating decimals*, which are characterized by unending but periodical fractional parts. As an example, we have $b = 1/3 = 0.333 \dots$, involving the repetition of only a single digit. and therefore having a period of 1 However, it is also possible to have larger repetition periods, such as:

$$c = 19.3582828282 \dots$$

The sequence of digits which is repeated in the fractional part of a rational number is called its *repetend*, which therefore corresponds to the period of the decimal sequence. For instance, in the case of the above examples, the respective repetends are:

$$\begin{aligned} \text{rep}(a) &= 0 \dots = \bar{0} \\ \text{rep}(b) &= 3 \dots = \bar{3} \\ \text{rep}(c) &= 8282 \dots = \overline{82} \end{aligned}$$

The repetend (or period) is often identified under a line. For instance, in the case of the number c above, we could also write:

$$c = 19.35\overline{82}$$

Given a generic repeated decimal number a , it is always possible to express it as a division between two integer

values. For instance, in the case of the number c above, we could write:

$$\begin{aligned} c = 19.35\overline{82} &\implies 100c = 1935.\overline{82} \implies \\ &\implies 10000c = 193582.\overline{82} \implies \\ &\implies 10000c - 100c = 193582 - 1935 \implies \\ &\implies 9900c = 191647 \implies c = \frac{191647}{9900} \end{aligned}$$

which took into account that $82.\overline{82} - 0.\overline{82} = 82$.

The basic procedure consists of obtaining the respective decimal expansion so that the repetend is placed just after the decimal point, then obtaining a second decimal expansion by the length of the repetend, and subtracting these expansions. The fact that this procedure always apply is compatible with the fact that both repeating or terminating rational numbers can be expressed as the quotient between two positive integer values, which is not the case for irrational decimals.

A particular interesting perspective to approach real numbers is in terms of the information content, which can be gauged in terms of *minimum length descriptions* (e.g. [28, 29]) in terms of digits. It follows from the discussion above that repeating decimals tend to be less complex than irrational numbers, because the former type of number can be expressed in terms of two finite positive integer values, while irrational numbers require an infinite numeric description. It is interesting to observe that minimum description quantifications of complexity are relative to specific types of description. For instance, the irrational number $1.414214\dots$ can be simply represented as $\sqrt{2}$, while its fractional form requires an infinite number of digits. Henceforth, the complexity of real numbers is approached from the perspective of the minimum length of their respective canonical forms and prime factor structure.

Though the complexity of methods and algorithms for generating and recognizing prime numbers has motivated related approaches (e.g. [30, 31, 32, 33, 11]), relatively fewer works have addressed possible means to characterize the patterns of structural complexity of integer and real numbers themselves (e.g. [34, 35, 36]).

The Venn diagram in Figure 1 illustrates the sets of numbers considered in the present work, as well as their interrelationships. In particular, observe that the positive integers have been understood as a special case of rational numbers with denominator equal to 1.

3 Primes and Canonical Representations

Prime numbers are positive integer numbers which can only be divided (integer division) by the numbers 1 or

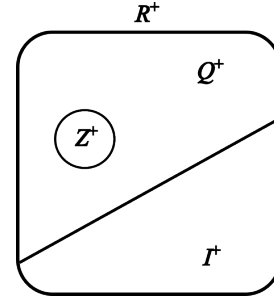


Figure 1: Venn diagram illustrating the main sets of numbers considered in this work: (a) positive integers \mathbb{Z}^+ ; (b) positive rationals \mathbb{Q}^+ , and (c) positive irrationals \mathbb{I}^+ , all of which being subsets of \mathbb{R}^+ and \mathbb{R} .

themselves. Examples of prime numbers include 2, 3, 5, 7, 11, 13. Observe that 1 is not typically considered a prime number. Numbers which are not prime are said to be *composite numbers*. According to *Euclid's theorem* [37], there is an infinite number of prime numbers.

Prime factors are prime numbers used to express generic positive integer numbers. A prime factor can be added to a given number n (in case it is not contained yet) by taking the *product* of that prime factor and n . Similarly, prime factors can be removed from a positive integer by *dividing* it by that prime factor. Therefore, if prime factors are understood as the ‘ingredients’ from which all integer numbers are made, the operations of product and division constitute a primary way of specifying ‘recipes’ for manipulating integer numbers.

To a good extent, the interest of integer prime numbers from the perspective of integer arithmetics stems from the fact that operations between this type of numbers is not *closed*, indicating that integer numbers are respectively *incomplete*. More specifically, in integer arithmetics, prime numbers cannot be divided by numbers other than 1 and themselves, while real a division always yields a real number as result. That is why prime numbers have often been approached from the integer perspective.

Given a positive integer value n , it can be expressed in terms of its prime factors, i.e. the set of P prime numbers p_i , $i = 1, 2, \dots, P$, so that:

$$n = (p_1)^{m_1} (p_2)^{m_2} \dots (p_P)^{m_P} = \prod_{i=1}^P (p_i)^{m_i} \quad (2)$$

where m_i are the respective *multiplicities* associated to each prime factor p_i . Equation 2 is called the *canonical representation* of the positive integer number n . In addition, The sequence of exponents of the prime factors (in increasing order) of a given positive integer number n has been called the *prime signature* of n (e.g. [36]).

Thus, the *prime factor decomposition* of a given positive integer value n consists of finding the respective

prime factors p_i as in Equation 2, which can be represented as the elements of the vector $\vec{P} = [p_1, p_2, \dots, p_P]$, with $p_1 < p_2 < \dots < p_P$, which is called the respective *canonical representation*. The set of multiplicities of n can be similar organized into a respective vector $\vec{M} = [m_1, m_2, \dots, m_P]$. The number 1 is typically not considered in prime factor decompositions.

As examples of prime factor decomposition and canonical representations, we have:

$$\begin{aligned} n &= 432 = (2)^2 (3)^3 (4) \\ m &= 720 = (2)^2 (3)^2 (4) (5) \end{aligned}$$

The quotient representation of rational numbers where both the numerator and denominator do not share common factors are said to be *irreducible*.

The prime factors of a number n are of particular importance because they can be understood as the unique ‘ingredients’ from which the positive integer value n is composed. In addition, the quantity of prime factors in a given number n , taken jointly with their respective multiplicity, can be related to the amount of information of that number, in the sense of minimum respective description in terms of the Equation 2.

For instance, from this perspective, the number $65.536 = 2^1 6$ could be described by the numbers 2 and 16 (three digits), while $40320 = (2)(3)(4)(5)(6)(7)(8)$ would require seven digits, meaning that the former number has a shorter description, being therefore associated to a smaller amount of information (or complexity). Henceforth, the total number of unique prime factors in a given number n is expressed as $P(n)$.

The amount of information (and complexity) of a positive integer value n can also be related to the multiplicities of its respective prime factors, in the sense that the larger the multiplicities the less complex the number tends to be. In this work, we shall consider the sum of all multiplicities of a number n , namely $M(n) = \sum_{i=1}^P m_i$, as a measurement of interest.

While larger numbers can have more prime factors, it does not necessarily follow that a large number will be more complex than a smaller one, as is the case in the example above. However, it has been shown that prime numbers become less frequent as the numbers increase. In particular, let $\pi(x)$ be the *prime counting function*, providing the number of primes which are smaller or equal to x . This function can be asymptotically approximated as:

$$\pi(x) \approx c(x) = \frac{x}{\log(x)} \quad (3)$$

which has derivative as:

$$\frac{dc(x)}{dx} = \frac{\log(x) - 1}{\log^2(x)} \quad (4)$$

Figure 2 illustrates the above derivative for a long range of x . The fact that the derivative values decrease with x are related to the smaller density of large prime numbers. However, this is only an informal indication of this important result, which can be more formally approached.

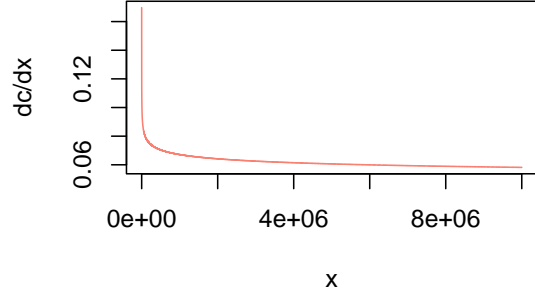


Figure 2: The derivative of the approximated number of prime numbers smaller or equal to x decreases with x , therefore indicating the smaller density of large prime numbers.

In addition, prime factor decomposition can also help the study and computation of several operations between numbers, especially the product and division. For instance, if two numbers n and m have respective prime factor decompositions specified by $\vec{P}_n, \vec{M}_n, \vec{P}_m$, and \vec{M}_m , the number r corresponding to their product will necessarily have its prime factors corresponding to the elements of the union of the sets of original prime factors, while the multiplicities would correspond to the respective sums. For instance, if $n = 432$ and $m = 720$, we would have (as from our previous examples):

$$r = nm = 311040 = (2)^4 (3)^5 (4)^2 (5)$$

This interesting result follows directly from the property that $(a)^b (a)^c = (a)^{b+c}$.

In the case of the division of two numbers n and m , the result $s = n/m$ will correspond to a fraction between two integer values with respective prime factors which are not common to both numerator and denominator. As an example:

$$s = \frac{n}{m} = \frac{432}{720} = \frac{(2)^2 (3)^3 (4)}{(2)^2 (3)^2 (4) (5)} = \frac{3}{5} \quad (5)$$

Prime factor decompositions and canonical representations are interesting not only from the theoretic-analytical perspective, but also have several interesting applications, including in signal processing (e.g. [38, 39, 40, 32, 41]), cryptography (e.g. [42, 43, 44, 45]), computing (e.g. [46, 47, 48, 49]), and even biology (e.g. [50, 51, 52, 53]).

4 NP-Sets

Multisets (e.g. [23, 24, 24, 26]) are generalizations of traditional sets allowing for the multiplicity of each element. Though the multiplicity of multisets is often limited to non-negative values, here we adopt the extension of multisets as np-sets [22] as a means to integrate properties of sets and real vector spaces.

An np-set [22] \vec{a} is a set of tuples:

$$\vec{a} = \{[a_1, m_{a_1}^p, m_{a_1}^n], [a_2, m_{a_2}^p, m_{a_2}^n], \dots, [a_N, m_{a_N}^p, m_{a_N}^n]\} \quad (6)$$

where a_i is an identifier or label of each element, while m_a^p and m_a^n are its non-negative and non-positive real-valued multiplicities, respectively. The *support* of a multiset \vec{a} is the set of its elements, and the *cardinality* of \vec{a} corresponds to the sum of the absolute values of all its multiplicities. The *net value* of an np-set corresponds to the sum of its signed multiplicities. The *null* scalar np-set has the general form $\vec{0} = [0, x, -x]$, where x is any non-negative real value. Therefore, null np-sets have all possible elements in the respective support with net values equal to zero. Every np-set is henceforth understood to contain the null scalar tuple of any possible element.

As an example, we have the following np-sets:

$$\vec{a} = \{[2, 1, 0], [5, 3, -2], [7, 3, -1]\} \quad (7)$$

$$\vec{b} = \{[2, 0, -2], [5, 2, -1], [8, 2, 0]\} \quad (8)$$

$$\vec{c} = \{[a, 2, 0], [b, 3, -1], [c, 1, -2]\} \quad (9)$$

which have supports $S_a = \{a, b, c\}$, $S_b = \{2, 5, 7\}$, $S_c = \{3, 7, 8\}$ and cardinalities $|\vec{a}| = 9$, $|\vec{b}| = 8$, and $|\vec{c}| = 6$.

Given two np-sets \vec{a} and \vec{b} , their respective *union* $\vec{a} \cup \vec{b}$ has support corresponding to the union of the original supports, and can be expressed as:

$$\vec{a} \cup \vec{b} = \{[c_i, \max(m_{a_i}^p, m_{b_i}^p), \min(m_{a_i}^n, m_{b_i}^n)]\} \quad (10)$$

The *intersection* between two np-sets has as its support the intersection between the two original supports, and can be calculated as:

$$\vec{a} \cap \vec{b} = \{[c_i, \min(m_{a_i}^p, m_{b_i}^p), \max(m_{a_i}^n, m_{b_i}^n)]\} \quad (11)$$

The *sum* between two np-sets has as its support the union between the two original supports, and can be obtained as:

$$\vec{a} + \vec{b} = \{[c_i, m_{a_i}^p + m_{b_i}^p, m_{a_i}^n + m_{b_i}^n]\} \quad (12)$$

The *difference* between two np-sets has as its support the union between the two original supports, and can be obtained as:

$$\vec{c} = \vec{a} - \vec{b} = \{[c_i, m_{a_i}^p - m_{b_i}^n, m_{a_i}^n - m_{b_i}^p]\} \quad (13)$$

It is also possible to consider the following alternative form of the difference operation, which is henceforth called *paired difference*:

$$\vec{c} = \vec{b} \ominus \vec{a} \leftrightarrow \{[c_i, m_{c_i}^p, m_{c_i}^n]\} \quad (14)$$

where:

$$m_{c_i}^p = \max(m_{b_i}^p - m_{a_i}^p, 0) + \max(m_{b_i}^n - m_{a_i}^n, 0)$$

$$m_{c_i}^n = \min(m_{b_i}^p - m_{a_i}^p, 0) + \min(m_{b_i}^n - m_{a_i}^n, 0)$$

Observe that both the definitions of the difference above, for a same set of operands, yield results with the same net value.

As an example, in the case of the np-sets in Equations 7 and 8, we would have:

$$\vec{a} \cup \vec{b} = \{[2, 1, -2], [5, 3, -2], [7, 3, -1], [8, 2, 0]\}$$

$$\vec{a} \cap \vec{b} = \{[2, 0, 0], [5, 2, -1], [7, 0, 0], [8, 0, 0]\}$$

$$\vec{a} + \vec{b} = \{[2, 1, -2], [5, 5, -3], [7, 3, -1], [8, 2, 0]\}$$

$$\vec{a} - \vec{b} = \{[2, 3, 0], [5, 4, -4], [7, 3, -1], [8, 0, -2]\} \quad \vec{a} - \ominus \vec{b} = \{[2, 3, 0], [5, 4, -4], [7, 3, -1], [8, 0, -2]\}$$

Observe that all the four above np-set operations are *associative*, while only the first three are *commutative*.

5 NP-Sets and Canonical Representations

The prime factor decomposition of any *positive integer number* n can be readily represented in terms of the following np-set:

$$\vec{c}(n) = \{[p_1, m_1, 0], [p_2, m_2, 0], \dots, [p_P, m_P, 0]\} \quad (15)$$

For instance, if $n = 4725$, it follows that:

$$\vec{c}(n) = \{[3, 3, 0], [5, 2, 0], [7, 1, 0]\}$$

Real numbers (and real vectors) can be associated to np-sets with net value equal to the represented number [22]. Interestingly, a same positive integer value n can have an infinite number of respective np-set representations, with each element having net value equal to the original multiplicity of each prime factor. For instance, the above number n can also be represented by np-sets so that each respective tuple has the same net value:

$$\vec{c}(n) = \{[3, 4, -1], [5, 4, -2], [7, 3, -2]\}$$

In case every tuple of an np-set has only one of its two multiplicities non-zero, the np-set is said to be in *standard representation*, of which the np-set in Equation 15 is an example. Observe that there is only one standard representation associated to each possible real number.

The prime signatures of positive integer values (e.g. [36]) are intrinsically related to multisets (e.g. [54, 55]). Np-set representations can be used to express the canonical form of terminating or repeating positive real numbers, or even the *real-valued division* between two integers. For instance, the number s in Equation 5 can be associated to the following np-set representation:

$$\vec{c}(s) = \{[2, 2, -2], [3, 3, -2], [4, 1, -1], [5, 0, -1]\}$$

where the multiplicities of each prime factor are represented in terms of respective non-negative or non-positive multiplicities depending on whether the prime factor appears in the numerator or denominator of the quotient.

The simplification of the above representation can be obtained by consider the respective standard representation of the respective multiset, i.e.:

$$\vec{c}(s) = \{[2, 0, 0], [3, 1, 0], [4, 0, 0], [5, 0, -1]\} \longleftrightarrow s = \frac{3}{5}$$

so that the irreducible fraction of a given rational number is associated to the standard form of that same number in an one-to-one manner.

The product of two positive integers can also be readily represented in terms of sum of the respective np-set representations. For instance, in the case $n = 432$ and $m = 720$, we have:

$$n m \longleftrightarrow \{[2, 4, 0], [3, 5, 0], [4, 2, 0], [5, 1, 0]\}$$

The above described np-set representation of prime factor decompositions can be readily employed to obtain the product and division between two terminating or repeating real decimal values. For instance, if $a = \frac{20}{140}$ and $b = \frac{54}{60}$, it follows that:

$$\begin{aligned} \frac{a}{b} &= \frac{20}{140} \frac{60}{54} = \frac{(2)^2(5)}{(2)^2(5)(7)} \frac{(2)^2(3)(5)}{(2)(3)^3} \longleftrightarrow \\ &\longleftrightarrow \{[2, 2, -2], [5, 1, -1], [7, 0, -1]\} + \\ &+ \{[2, 2, -1], [3, 1, -3], [5, 1, 0]\} = \\ &= \{[2, 4, -3], [3, 1, -3], [5, 2, -1], [7, 0, -1]\} = \\ &= \{[2, 1, 0], [3, 0, -2], [5, 1, 0], [7, 0, -1]\} \longleftrightarrow \\ &\longleftrightarrow \frac{a}{b} = \frac{(2)(5)}{(3)^2(7)} \end{aligned}$$

where the np-set representation in the penultimate line above was placed in its standard form, so that the fraction is irreducible.

Therefore, an interesting and useful association is established between the canonical representation of two positive rational numbers n and m and their np-set representation, which is summarized as follows:

The *product* (or *division*) of two positive rational numbers can be represented and performed in terms of the *sum* (or *difference*) of the np-sets associated to their canonical representations.

6 Least Common Denominator and Greatest Common Divider

Given two positive numbers n and m represented in terms of the respective np-set prime factor decomposition, their *greatest common divider* can be readily obtained as the *intersection* between those two np-set representations. As an example, consider $n = 360$ and $m = 420$. Then, we have:

$$\begin{aligned} \vec{c}(n) &= \{[2, 3, 0], [3, 2, 0], [5, 1, 0]\} \\ \vec{c}(m) &= \{[2, 2, 0], [3, 1, 0], [5, 1, 0], [7, 1, 0]\} \\ \vec{c}(n) \cap \vec{c}(m) &= \{[2, 2, 0], [3, 1, 0], [5, 1, 0]\} \implies \\ &\implies \gcd(n, m) = (2)^2 (3) (5) = 60 \end{aligned}$$

The least common denominator among two positive numbers n and m can be obtained from the *union* between their respective np-set prime factor representations. As an example respectively to the above n and m :

$$\begin{aligned} \vec{c}(n) \cup \vec{c}(m) &= \{[2, 3, 0], [3, 2, 0], [5, 1, 0], [7, 1, 0]\} \implies \\ &\implies \text{lcd}(n, m) = (2)^3 (3)^2 (5) (7) = 2520 \end{aligned}$$

In the case of more than two numbers, the two above methods can be used by employing the associative property of the np-set union and intersections.

The *least common denominator* (or *or greatest common divider*) of two positive rational numbers can be represented and performed in terms of the *union* (or *intersection*) of the np-sets associated to their canonical representations.

7 NP-Sets and Coprimes

Two positive integer numbers n and m are said to be mutually *coprimes* whenever they have only the number 1 as their greatest common divider. We readily have from the above developments that two numbers are coprime provided the intersection of their np-set representation contains only the element $[1, 1, 0]$. Prime numbers can then be understood as positive integers that are coprime with every other possible positive integer smaller than itself. Another interesting property is that a rational number having coprime numerator and denominator is necessarily irreducible.

Two coprime numbers n and m are associated to an interesting geometrical property, namely that when a discrete rectangle of height $n + 1$ and width $m + 1$ (or vice-versa) is constructed, the straight line connecting any of the two opposite vertices do not intersect any of the discrete points. This is depicted in Figure 3.

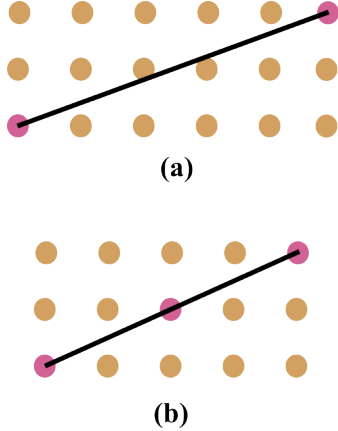


Figure 3: The grid (a) defined by the pair of coprimes $n = 2$ and $m = 5$. The shown diagonal does not intersect any of the grid points. As illustrated in (b), this is not the case for the pair of numbers $n = 2$ and $m = 4$, which are not coprimes.

Figure 4 illustrates the fact that the concept of prime number can be reformulated in terms of coprimality.

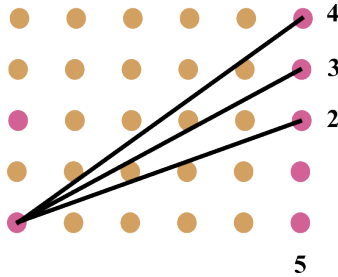


Figure 4: The number 5 is prime because it is coprime with each of integer numbers from 2 to 4.

The constructions in Figure 3 and 4 also hint at the fact that prime number structure is a particular aspect in *digital geometry* (e.g. [56, 57]), playing a central role in the concept of *digital lines* (e.g. [58, 59, 60]) mainly through the concept of *Farey sequences* (e.g. [61, 62], see also Section 14).

The divisibility between positive integers and rational numbers constitutes an interesting topic (e.g. [63]). Given that the above described approach also allows rational numbers to be expressed in terms of respective np-sets, it becomes possible to consider pairs of such numbers a and b so that the intersection of the n-set of their canonical representation only contains null element $[e, 0, 0]$. These pairs of rational numbers are henceforth said to *rational*

coprimes, therefore corresponding to a generalization of the concept of coprimality to the set of rational numbers.

As an example, the rational numbers $a = \frac{(2)^3(3)}{(5)^4}$ and $b = \frac{(7)^2}{(3)(11)}$ are coprimes because:

$$c_a \cap c_b = \{[2, 0, 0]; [3, 0, 0]; [5, 0, 0]; [7, 0, 0]; [11, 0, 0]\}$$

which only contains null np-set scalars.

However, the pair of rational numbers $a = \frac{(2)}{(3)^2}$ and $b = \frac{(2)}{(3)}$ are not coprime rationals because:

$$c_a \cap c_b = \{[2, 1, 0]; [3, 0, -1]\}$$

meaning that both 2 and 3 are divisors of a and b in the sense that their respective division or product result in rational numbers having the same prime factors as either a or b .

The extension of coprimality to rational numbers paves the way to possible respective generalizations of several related concepts and results in number theory, including *totients* and *Euler's theorem*.

Actually, the np-set approach to canonical representations of rational numbers allows the very concept of prime number to be extended to rational numbers as those numbers whose np-sets associated to their canonical representations have no intersection with np-sets of other canonical representations. However, there are two facts that need to be taken into account. First, we have that there is an infinite number of rational in any real interval, so that verifying primality in such a given interval would involve taking into account an infinite number of cases. Then, we have that the division between two rational numbers a and b , as understood in terms of intersections of np-set representations, actually corresponds to the product of a by the reciprocal of b , which is also a rational number. Because the product between two rational numbers actually corresponds to adding the np-sets associated to their canonical representation, a rational number will always be divisible by another rational. For instance, take $a = 2/3$ and $b = 1/7$. We then have that the division a/b is equal to $2/(37)$, which is a valid rational.

For the above reasons, a possible approach to prime rationals consists of considering the intersection between np-set canonical representations within a pre-specified discrete set Ω . Under these circumstances, given a finite set of rational number Ω , a *prime rational number* could be understood as follows:

Given a discrete set Ω of positive rational numbers, a positive rational number a is a *rational prime* in that set iff the intersection between the np-sets associated to the canonical representations of a and of every other number in Ω is the null np-set.

As an example, consider $\Omega = \{1/3, 2/3, 4/5, 5/7\}$. The rational numbers $11/17$ and $10/11$ are both rational primes in this set.

8 Prime Numbers and Power Law

Given a sequence S of values $\{s_0, s_1, s_2, \dots, s_i, \dots, s_{N-1}\}$, it is often interesting to try to identify a basic relationship underlying those values. For instance, is it possible to express, even if approximately, the values in S as a mapping from the set of natural (or positive integer) numbers? A first indication about this possible type of relationship can be readily obtained by visualizing the scatterplot obtained by the tuples $[i, s_i]$.

Figure 5 illustrates the scatterplot obtained as described above considering the prime numbers from $i = 100$ to 1000.

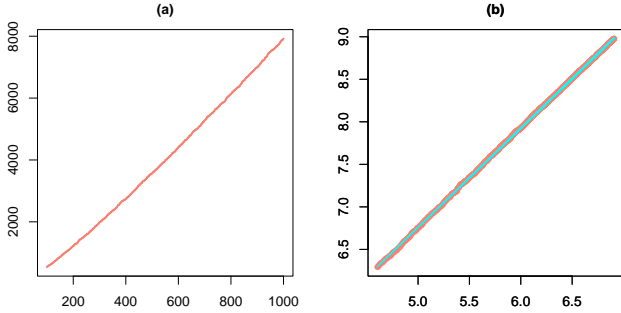


Figure 5: The mapping from the positive integer values $i = 0, 1, 2, \dots, N - 1$ to respective prime numbers for $i = 100$ to 1000, shown in linear (a) and logarithmic (b) scales. The relationship observed in the latter case approaches steadily a straight line, indicating an approximate power law relationship between i and the prime numbers. Also shown (in cyan) is the respective least mean square approximation of the observed relationship.

Interestingly, the considered sequence of prime numbers fits markedly well with a power mapping of the type:

$$\log(s_i) = b \log(i) + c \quad (16)$$

which can be rewritten as follows:

$$\log(s_i) = \log(i^b) + c \implies s_i = a i^b \quad (17)$$

with $a = e^c$ and where $\log()$ stands for the natural logarithm.

The exponent b can be estimated by least mean square regression on the logarithmic representation, yielding $c = 0.9328\dots$, $a = 2.541616\dots$, and $b = 1.1656\dots$ for the above considered case ($i = 100$ to 1000). The obtained fitting is expected to provide approximations not too far from the considered range of i . The average and standard deviation of the respectively obtained residuals were $-1.4074\dots e-18$

and $0.006416\dots$. The average of the absolute values of the residuals was $0.004834\dots$. This indicates that the residuals are evenly distributed above and below the respective average.

Figure 6 depicts the relationship between the approximated and original prime numbers in this case.

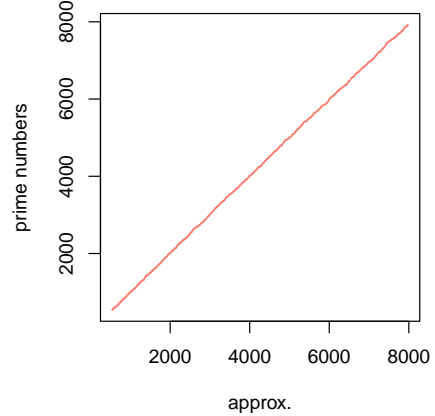


Figure 6: The relationship between the approximated and original prime numbers for $i = 100$ to 1000, obtained for the least squares method ($a = 2.541616\dots$, and $b = 1.1656\dots$).

Interestingly, the value of b being close to unit indicates a relatively uniform distribution of prime numbers within the considered range $i = 100$ to 1000. However, at the same time the value of b being larger than 1 implied a reduction of the density of obtained as i increased.

Though the above fitting can give an idea of the magnitude of the i -th prime number and suggests an approximated power relationship, it does not provide a particularly effective resource for predicting primes because the obtained error, in the average, tends to be larger than the respective average *prime gap* (i.e. the difference between a successive and predecessor prime number).

An interesting result providing a minimum bound to the i -th prime number has been developed [64] as:

$$s_i > B_i = i (\log(i) + \log(\log(i)) - 1). \quad (18)$$

For instance, for $i = 421$, we have from Equation 17 that $s_{421} \approx 2931$ and from Equation 18 that $s_{421} > B_i = 2880.26\dots$, while the 421-th prime is equal to 2909.

9 Prime Numbers Regularity and Complexity

Prime numbers are particularly important in number theory not only because they can be understood as corresponding to the basic ‘ingredients’ of other numbers (to

which they can be added through products, or removed from by divisions), but also because of their marked *irregularity* and *unpredictability*. These two properties are directly related to the potential *complexity* of prime numbers, in the sense that regular and/or predictable structures tend to be relatively simple, and not complex.

Despite continuing interest that prime numbers have received along centuries, relatively little has been discovered about eventual *regularity* or *patterns* underlying prime numbers. Actually, the identification of any related regularity or pattern would contribute to the prediction and identification of prime numbers, while also decreasing their complexity and unpredictability.

A simple approach to finding some underlying structure of prime numbers has been described in Section 8. Several other developments have been related in the literature, and two approaches can be highlighted here: (i) the identification of the number and density of prime numbers; and (ii) the identification of possible patterns underlying them.

Regarding the total number of prime numbers, as briefly discussed in Section 3, it is believed that the density of prime numbers tend to decrease with their size, so that there are fewer larger prime numbers than smaller ones. This subject constitutes one of the major research lines in number theory (e.g. [3, 4, 5, 6, 7, 8]), and therefore cannot be comprehensively revised or addressed here.

Ulam spiral [34, 35] consists of one of the few indications about pattern regularity of prime numbers. More specifically, when positive integers are arranged along a squared spiral, certain horizontal, vertical or diagonal lines with unit absolute slope tend to be denser of prime numbers, as illustrated in Figure 7.

Interestingly, the numbers belonging to the diagonals containing more primes have the general algebraic form:

$$p(n) = 4n^2 + bn + c \quad (19)$$

with $n = 0, 1, 2, \dots$, and b and c being integer constant values, b being *even*.

The respective density of prime numbers associated to these lines vary markedly with the parameters b and c , which have an effect on the respective types of solutions and respectively implied prime numbers. It remains a conjecture that some of these lines can generate infinitely many primes, or to have these numbers in a particularly high density. In particular, the quadratic polynomials expressed as follows:

$$p_{41}(n) = 4n^2 - 2n + 41 \quad (20)$$

$$p_{41}(n) = 4n^2 + 2n + 41 \quad (21)$$

seems to be particularly rich in positive integer roots that correspond to prime numbers. This line has been emphasized in cyan in the Ulam's spiral depicted in Figure 7.

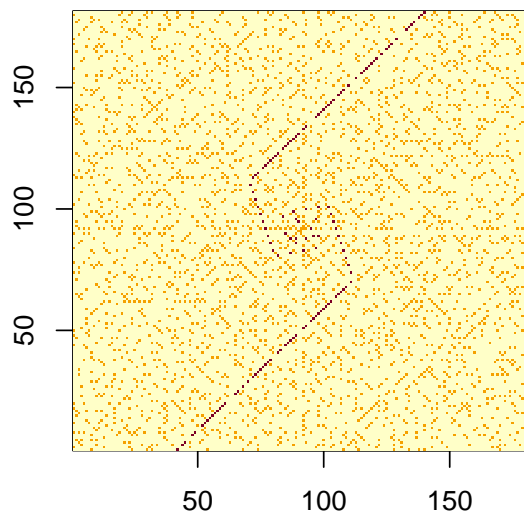


Figure 7: Ulam spiral for positive integers smaller than 181×181 . The composite numbers are shown in light yellow, and the prime numbers in orange. The prime numbers corresponding to the polynomials $p_{41}(n) = n^2 - 2n + 41$ and $p_{41}(n) = n^2 + 2n - 41$ are shown in black.

While prime numbers can be understood as presenting particular complexity as discussed above, the complexity of other positive integers can be approached in terms of the total number of their prime factors as well as the respective multiplicities. This possibility provides the main motivation for the developments described in the present work related to the possible quantification of the complexity of integer, rational and real numbers in terms of their prime factor structure revealed by their respective canonical representations.

10 Comparing Canonical Representations

When dealing with rational numbers, it would be interesting to compare their prime factor structure. Though the absolute value of their difference could be considered, the np-set generalization of the *Jaccard similarity index* (e.g. [65, 66, 67]) provides a promising approach to compare canonical representations of rational numbers since it considers both the prime factors and respective multiplicities.

The np-set generalization of the Jaccard similarity index (e.g. [67, 22]) between the canonical representations

\vec{a} and \vec{b} of two positive rational numbers a and b can be calculated as:

$$s(\vec{a}, \vec{b}) = s(\vec{b}, \vec{a}) = \frac{|\vec{a} \cap \vec{b}|}{|\vec{a} \cup \vec{b}|} = \frac{\sum_{i=1}^P (\min(m_{a_i}^p, m_{b_i}^p) - \max(m_{a_i}^p, m_{b_i}^p))}{\sum_{i=1}^P (\max(m_{a_i}^p, m_{b_i}^p) - \min(m_{a_i}^p, m_{b_i}^p))} \quad (22)$$

with $0 \leq s(\vec{a}, \vec{b}) \leq 1$. Observe that the np-set representations need to be in the standard form, which corresponds to the irreducible fractional representation.

The above developments lead to another interesting relationship between canonical representations and number theory, namely:

The *np-set Jaccard similarity index* between two positive rational numbers n and m corresponds to the quotient between the cardinalities of the np-set representations of their *greatest common divider* and *least common denominator*.

As an illustration, consider the case $a = 140/561$ and $b = 280/2805$. It follows that:

$$\begin{aligned} \vec{c}(a) &= \{[2, 2, 0], [3, 0, -1], [5, 1, 0], [7, 1, 0], [11, 0, -1]\} \\ \vec{c}(b) &= \{[2, 3, 0], [3, 0, -1], [5, 1, 0], [7, 1, 0], [11, 0, -2]\} \\ s(\vec{a}, \vec{b}) &= \frac{2 + 1 + 1 + 1 + 1}{3 + 1 + 1 + 1 + 2} = \frac{6}{8} = \frac{3}{4} \end{aligned}$$

The relatively high similarity value which has been obtained indicates that the two canonical representations are substantially similar, which is indeed the case.

It should be observed that a marked similarity between the canonical representation of two rational numbers does not necessarily imply that those numbers have similar magnitudes, and vice-versa. For instance, in the above example we have that $a \approx 0.24955$ and $b \approx 0.099822$, so that $a \gg b$ while $s(\vec{c}(a), \vec{c}(b))$ is relatively high.

From the perspective of the *canonical representation*, a rational number n having more prime factors (higher $P(n)$) and relatively small total multiplicity ($M(n)$) can be said to be comparatively more complex (from the perspective of minimal prime factor representation) than another rational number m with smaller $P(m)$ and larger $M(m)$. This interpretation is specific to the canonical representation of the compared rational numbers.

11 Prime and Multiplicity Signatures of Real Numbers

As discussed in the previous section, the complexity of rational numbers can be related to their prime factor struc-

ture, including the associated values of $P(n)$ and $M(n)$. That is possible because rational numbers have a finite rational representation, which is not the case with irrational numbers. In this section, an approach is developed for characterizing the progressive complexity of irrational numbers.

Given an irrational number a , we shall consider its successive decimal expansions a_i by $i = 1, 2, \dots, N$ digits. Since the described approach is generic to any real number, we henceforth refer to that more general situation. For each of the so obtained positive integer value, the respective canonical representation is calculated so that the associated $P(a_i)$ and $M(a_i)$ can be determined. These two sequences of values along i are henceforth denominated the *prime factor and multiplicity signatures* of the real value a . A real value a with identical signatures $P(a_i)$ and $M(a_i)$ can be understood to be relatively complex, because this implies that each prime factor is taken with multiplicity one, while the presence of relatively high multiplicities tends to make the number a comparatively simpler from the perspective of the canonical representations of its decimal expansions. Thus, henceforth we consider the np-set Jaccard similarity between the two signatures of a real value a , namely:

$$J(P(a_i), M(a_i)) = \left[\frac{\sum_{i=1}^N (\min(P(a_i), M(a_i)))}{\sum_{i=1}^N (\max(P(a_i), M(a_i)))} \right]^D \quad (23)$$

where D is a parameter controlling how strict the performed comparison is: the larger the value of D , the more strict the comparison.

The above Jaccard similarity index is henceforth called the *prime factor complexity* $C(a, N)$ of the real value a considering decimal expansions up to N digits. It is interesting to observe that the proportional type of comparison [68] implemented by the np-se Jaccard similarity will imply similar resolution concerning the comparison of two large or two small numbers. For instance, the comparison between 1 and 5 will yield the same Jaccard similarity as the comparison between 10 and 50.

A situation of particular interest, especially to be taken as a comparative reference, concerns the characterization of the Jaccard similarity between the prime factor and multiplicity signatures of random irrational numbers, namely numbers whose digits are drawn uniformly from the basic digits $0, 1, \dots, 9$. Figure 8 presents the histogram of the Jaccard similarity between their signatures.

This histogram is characterized by near bilateral symmetry, but also presenting some skewness. A relatively high average of 0.8053 has been obtained associated to a relatively small standard deviation of 0.06924. In principle, numbers with prime factor complexity larger than the obtained average tend to present a relatively more complex canonical structure.

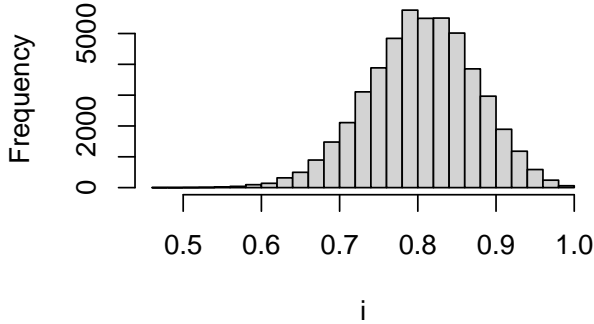


Figure 8: The histogram of the prime factor complexity (np-set Jaccard similarity index for $D = 1$) obtained by comparing the signatures $P(a_i)$ and $M(a_i)$ of 1000 random irrational numbers, with $i = 1, 2, \dots, 16$. The obtained histogram has a slightly skewed shape. The average and standard deviation similarities were 0.8053 ± 0.06924 .

Figure 9 illustrates several examples of the signatures $P(a_i)$ and $M(a_i)$ respectively to several fundamental irrational constants. Also shown are the prime factor complexities, corresponding to the respective np-set Jaccard similarity J between those signatures calculated by using the equation above.

12 Prime Complexity Networks

The similarity relationships among the prime factor complexity of a set of real values can be visualized in terms of respective *similarity networks* (e.g. [69]), which are graphs having each number represented as a node, while the similarity values are associated to respective pairwise links. This type of networks is henceforth denominated *prime complexity networks*.

Relationships between the prime structure of positive integer number in terms of graphs and complex networks have been approached before (e.g. [70, 71, 72, 73, 74, 75, 76, 77, 78, 63]). Here, we apply the concepts and methods described in the previous sections to obtain a graph interrelating the complexity of the prime structure of real some numbers, especially irrational values. All graphs in the present work have been visualized by using the Fruchterman-Reingold methodology [79].

Figure 10 illustrates the np-set Jaccard similarity network obtained while comparing the prime factor complexity of the eight primary irrational constants in Figure 9, while also incorporating a node corresponding to the case of random rational numbers composed by digits uniformly

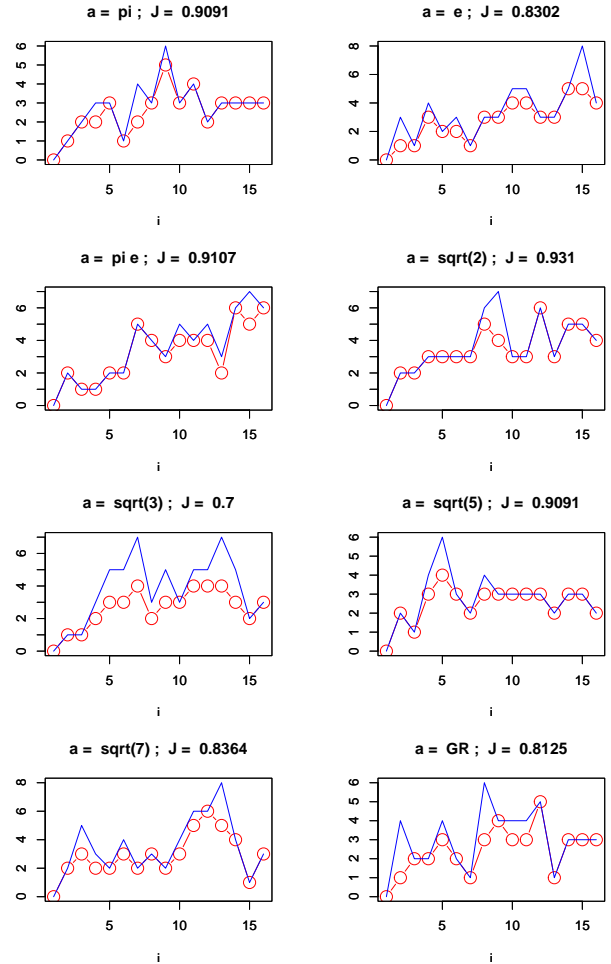


Figure 9: Illustration of prime factor and multiplicity signatures $P(a_i)$ and $M(a_i)$, in terms of $i = 1, 2, \dots, 16$, obtained for several irrational numbers, also including the respective np-set Jaccard similarity J . GR stands for the golden ratio. The prime factor signatures are shown in red, while the multiplicities signature are presented in blue. Interestingly, the fundamental constants π , e and $\sqrt{2}$ resulted with relatively larger values of signatures similarity, reflecting their more elaborated prime factor structure. The largest similarity was obtained for the constant $\sqrt{2}$.

chosen among the set $\{0, 1, \dots, 9\}$. The Jaccard similarity is calculated between the pairwise values of the np-set Jaccard similarities between the signatures.

Interestingly, two modules can be identified, corresponding to nodes more interconnected among themselves, as well as a relatively isolated node. The first module incorporates the constants π , $\sqrt{2}$, πe as well as $\sqrt{5}$. This group is characterized by the largest obtained prime factor complexities. The second module includes nodes e , $\sqrt{7}$, GR (golden ratio), and rnd (average of random fractionals). The isolated node corresponds to $\sqrt{3}$, which presented the smallest prime factor complexity.

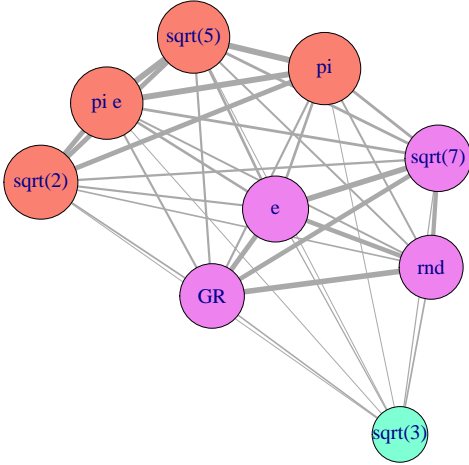


Figure 10: The prime complexity network (using Jaccard similarity for $D = 9$) obtained while considering the Jaccard similarity between the prime factor complexities of the irrational numbers whose signatures are illustrated in Fig. 9, as well as for the average complexity similarity obtained for random irrational numbers (rnd). The width of the links are related to the Jaccard similarity between the prime factor complexities $C(a, N)$ of the nine considered cases (considering $N = 16$), while the size of the nodes is proportional to the similarities between the respective signatures. Observe that the number πe is not known for certain to be irrational.

13 Factors-Multiplicity Similarity Networks

Though the prime complexity networks presented in the previous section can provide some insight about the complexity of the prime structure of decimal expansions of real numbers, the similarity between the number of primes and total multiplicities can only reflect the overall relationship between the respective prime structure complexities.

In this section, we consider similarity networks obtained not from the overall similarity between the factors and multiplicities signatures, but between tuples of numbers of prime factors and total multiplicity, therefore providing more information about the respective prime factor structure of the analyzed decimal expansion of real numbers. However, these networks refer to individual real numbers, and not to a set of real numbers as in the case of the prime complexity networks. More specifically, each node corresponds to one of the positive integer values obtained from the decimal expansion a_n of a same real number a , while the weights of the interconnections correspond to the pairwise (between two real numbers a and b) Jaccard

similarities between the respective tuples $(P(a_n), M(a_n))$ and $(P(b_n), M(b_n))$ where a_n is the decimal expansion of the real value a , $N P(a_n)$ is its number of prime factors and $M(a_n)$ corresponds to the total multiplicity. These networks are henceforth called *factors-multiplicity similarity networks*.

Before proceeding to the factors-multiplicity similarity of individual real numbers, we first illustrate this type of networks respectively to a set of integer values. In order to do so, the above described method is modified so that each node corresponds to a positive integer number, and the weights are defined as the respective pairwise Jaccard similarity between pairs of tuples $(P(u), M(u))$ and $(P(v), M(v))$ and where u is a positive integer, $P(u)$ is its number of prime factors, and $M(u)$ is its total multiplicity.

Figure 11 presents the factors-multiplicity similarity network obtained considering all integer numbers between 2 and 50. The widths of the interconnections are proportional to the similarity between pairs of tuples $(P(u), M(u))$, and the size of the nodes corresponds to the respective prime factor complexity weighted by the respective total number of prime factors.

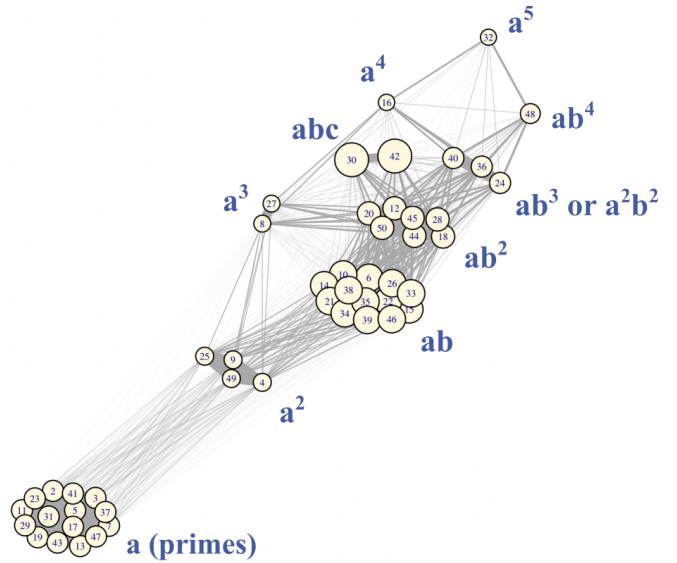


Figure 11: Factors-multiplicity network ($D = 7$) obtained for the positive integer values from 2 to 50. Each number is represented as a node, and the weight of the connections corresponds to the similarity between respective pairs of tuples $(P(u), M(u))$. The widths of the interconnections in the network are shown proportionally to the respective weights, while the size of the nodes are proportional to the weighted prime complexity of each number. Each of the obtained modules is associated to the respective type of canonical representation. For instance ab^2 indicates that the numbers in that group contains two prime factors (a and b), with total multiplicity 3.

The obtained network expresses to a good extent the overall prime factor structures of the considered numbers.

Well-defined modules have been obtained between numbers having the same $P(n)$ and $M(n)$. Observe that all interconnections within a same group, or between two same given groups, are identical. The width of the links reflect the jaccard similarity between the tuples $(P(u), M(u))$.

In addition to the well-defined modularity of the obtained network, it is also interesting to observe that its overall structure is elongated, defining an axis extending from the prime numbers at the lower-left group to the number 32 at the other extremity. The axis thus associated to the overall elongation of the network reflects the total multiplicity of the involved canonical representations. In addition, the numbers corresponding to powers of a unique prime factor resulted organizing along a chain along the upper portion of the obtained network. The two numbers having the largest prime factor complexity are 30 and 42, both of which type type abc , therefore involving three prime factors, all with unit multiplicity.

An alternative possibility to obtain similarity networks related to the prime factor structure of integer numbers consists of using the same method described above, but considering not the total multiplicity, but the respective sorted multiplicities. For instance, the number $u = 40$, with canonical representation $(2)^3(5)$, has total multiplicity 6, and sorted multiplicities 1, 5. The number 36 has canonical representation $(2)^3(3)$, total multiplicity 6 and sorted multiplicities 1, 5. Therefore, as these two numbers also have the same number of prime factors (equal to 2), they are associated to the same tuple $[2, 1, 5]$, so that the Jaccard similarity between these tuples is maximum and equal to 1.

Figure 12 depicts the same set of integer numbers as in Figure 11, but considering the sorted multiplicities instead of the total multiplicity of each considered integer number, as well as the number of prime factors.

All connections within the same module or between the same modules are identical. Compared to the network shown in Figure 11, more detail has been obtained, yielding more separated modules. This is a consequence of the fact that the adopted tuples provide additional information than only the number of prime factors and total multiplicity. In addition, the main axis associated to the overall elongation of the obtained network can be divided into two portions: one extending from the prime numbers to the number 32, and another extending from the prime numbers to the number 48. In both cases, the modules occur along increasing values of total multiplicities.

It is also interesting to observe that the similarity network shown in Figure 11 can be related to a folding of the network in Figure 12 around the prime numbers group. This is a consequence of the fact that the consideration of the number of prime factors and total multiplicity corresponds to a less detailed characterization of the respective

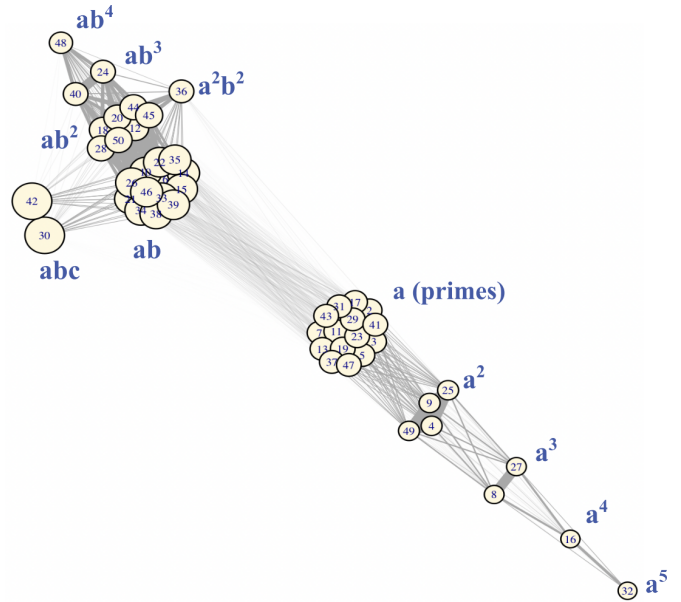


Figure 12: Similarity network obtained by considering the Jaccard similarity ($D = 7$) between tuples containing the total number of prime factors as well as the sorted multiplicities. A more detailed structure has been obtained than that shown in Fig. 11.

canonical representations, which tend to decrease the separation between the obtained modules and, more specifically, to the merging of modules having the same configurations of these two measurements.

The factors-multiplicity networks are now applied to the real values, namely those already considered in Figure 9. Markedly distinct structures have been obtained for each considered value. Several of the networks also presented considerable level of bilateral symmetry appearing along an elongated structure. This symmetry is mostly a consequence of the identical weights associated to links between nodes of same groups. The elongated axis, however, resulted associated to progressive variation of the prime factor complexity of each of the decimal expansions in each obtained network.

The networks obtained are characterized by marked similarity between the tuples number of prime factors and total multiplicity between most of its decimal expansions, which mostly involve prime factors with unit multiplicity. Though the network obtained for $\sqrt{3}$ is particularly compact, the prime factor complexity of its nodes (decimal expansions) are substantially smaller (observe that the sizes of nodes are not shown to the same scale across the networks).

14 Farey Sequences

Farey sequences (e.g. [15, 2]) are an interesting concept bringing together positive integer coprime numbers and

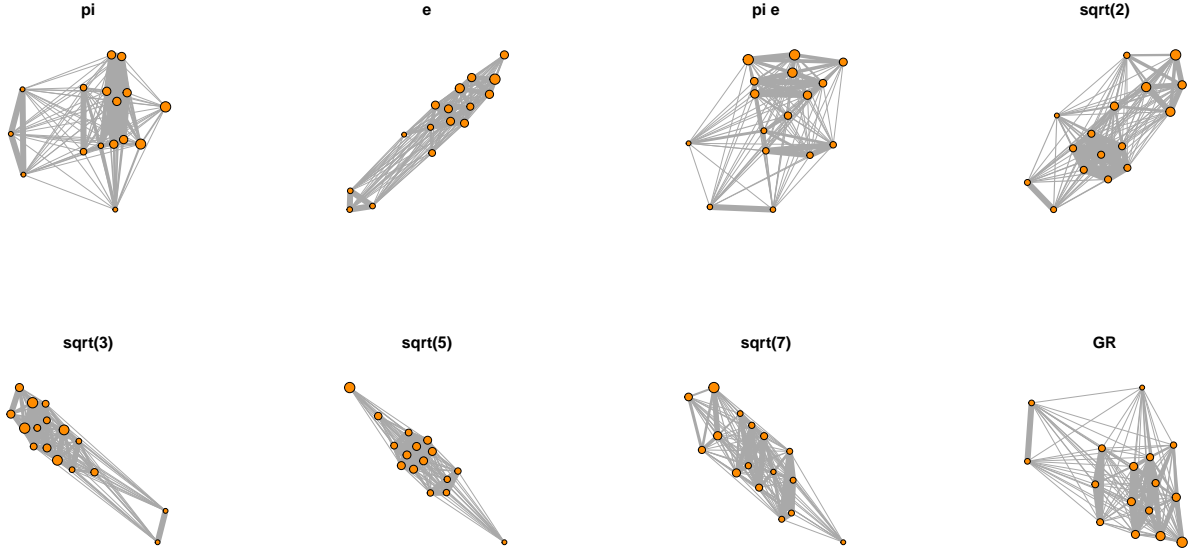


Figure 13: Factors-multiplicity networks obtained (considering similarity between tuples number of prime factors and total multiplicity, for $D = 5$) for the same real values as in Fig. 9. These networks provide reflect the relationships between the decimal expansions of each considered real value, not being directly related to their prime structure complexity of those real values. The size of the nodes in each network, however, correspond to the prime factor complexity of each respective decimal expansion.

rational numbers. Basically, a Farey sequence is a sequence of irreducible rational numbers, in increasing order, which can be either in the interval $[0, 1]$ or unconstrained. In addition, each element in this sequence has denominator smaller or equal to a pre-specified parameter n .

For instance, the Farey sequence, for $n = 5$ and restricted to the interval $[0, 1]$, is as follows:

$$F_5 = 0, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, 1$$

It follows from the above definition that each element in a Farey sequence has coprime numerator and denominator.

A Farey sequence F_n , where n is a positive integer, can also be understood as the set (no repetitions of elements), sorted in increasing order, of the elements of the respective uniformly spaced sequence:

$$U_n = 0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1 \quad (24)$$

placed in irreducible form.

The Farey sequence has been associated to several interesting geometric representations, including *sunburst* diagrams. These diagrams, illustrated in Figure 14 for $n = 5$, are obtained by plotting points at x- and y- coordinates corresponding to the numerator and denominator of each element in a Farey sequence, and then reflecting this con-

struction respectively to the identity function, and then on the other three quadrants of \mathbb{R}^2 .

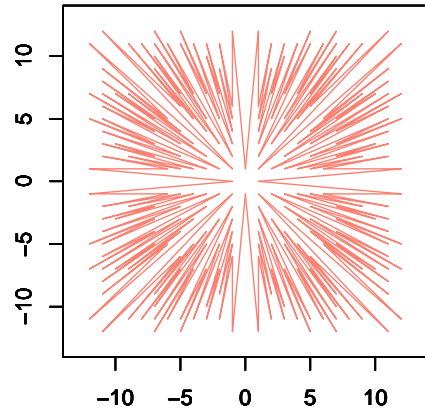


Figure 14: Farey sunburst diagram for $n = 12$.

It is also possible to obtain Jaccard similarity networks for the tuples [numerator, denominator] of a given Farey sequence, as illustrated in Figure 15 for $n = 12$.

The network depicted in Figure 15 reveals greater similarity (interconnectivity) among the most central ele-

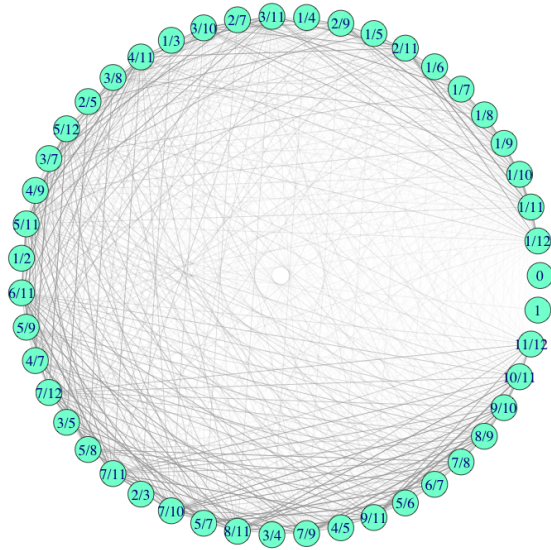


Figure 15: Farey similarity network for $n = 12$ represented in circular layout, with the number 0 represented as $0/1$. Observe that the elements of the Farey sequence have been shown sequentially in counter-clockwise sense.

ments of the Farey sequence (left-hand side of the obtained similarity network).

It is also possible to restrict a Farey sequence only to its elements having prime numbers as both numerator and denominator. This type of sequence is henceforth denominated *prime Farey sequence*. For instance, in the case of $n = 5$, we obtain:

$$F_5 = 0, \frac{1}{5}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, 1$$

Figure 16 illustrates this type of network obtained for $n = 20$, also incorporating the values 0 and 1.

15 ‘Prime’ Real Numbers

As discussed in Section 7, a possible approach to extending the concept of primes to rational numbers involves assuming a respective reference set Ω of rational numbers, so that a rational number a is said to be prime in Ω provided the intersection between the np-set associated to its canonical representation and the np-sets associated to the canonical representations of all numbers in Ω is the null np-set.

The Farey sequences discussed in Section 14 provides a natural choice for Ω , as it is composed of a finite set of rational numbers. For simplicity’s sake, the following

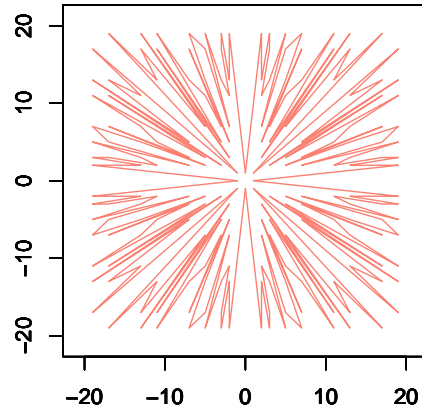


Figure 16: Prime Farey sunburst diagram for $n = 20$, also including the values 0 and 1. The obtained diagram resembles a snowflake.

developments will refer to the Farey sequence restricted to the interval $[0, 1]$.

First, observe that the Farey sequence for a given positive integer number n rational numbers whose numerators and denominators correspond to all positive integers from 0 to n . Therefore, a rational number $a = n/m$, where n and m are positive integers, can be understood as a *rational ‘prime’* in $\Omega = F_n$ if and only both n and m are coprimes with all the integer values in the set $\{2, 3, \dots, n\}$, which are represented as numerators or denominators of the respective Farey sequence. The use of quotes in ‘primes’ is adopted here in order to remind that, in traditional terms, prime numbers have been intrinsically related to natural numbers. The above reasoning leads to the possible definition of a *‘rational’ prime*:

A *‘prime’ rational number* respective to a restricted Farey sequence F_n is a rational number so that the intersection of the np-set associated with its canonical representation and the np-set associated to each of the elements of F_n is the null np-set.

It follows that only rational numbers of the form $1/p$, where p is a positive integer prime, can be rational ‘primes’, because forms like q/p , with $q \leq p$ as required of restricted Farey sequences, will have q appearing in the numerator and/or denominator of its elements, leading to the following:

The set of *‘prime’ rational numbers* are the numbers $1/p$, where p are the positive integer prime numbers.

The extension of the above concept to rational numbers larger than 1 is direct and involves normalizing (or considering only the fractional part) those numbers so as to fit the restricted Farey sequence.

Actually, the above approach to rational primes can also be readily extended to real numbers by understanding irrational numbers as real ‘prime’ numbers. Indeed, because an irrational number has infinite rational representation, the np-set associated to its canonical representation and those of the elements of any finite Farey sequence is the null np-set, from the perspective of the above assumptions and developments, leads to the following:

Every irrational number is a real ‘prime’ number. The set of *real ‘prime’ numbers* therefore consists of the irrationals and the rational primes with form $1/p$, where p is a positive integer prime.

The above developments are related to the concept of *irreducible numbers* in *integral domains* (e.g. [80]), the concept of factorization [81, 11], as well as the concept of divisibility in commutative rings (e.g. [82]). Informally speaking, an integral domain is a commutative ring (the product is commutative) so that the product of any pair of its elements is non-zero. Given that np-set intersections correspond to products of canonical representations of rational numbers, the above characterized real ‘primes’ can be understood to relate to irreducible numbers in the real integral domain.

16 Concluding Remarks

The present work has described an approach to study aspects related to the prime factor structure of real values. This has been done respectively to two main following ways.

First, we have shown how np-sets can be used to effectively express canonical representations of positive rational numbers of both terminating and repeating types. This approach allowed the identification of several interesting relationships between np-representations of canonical representations, including those between product and quotient of rational numbers and the sum and difference np-set operations of sum and difference, as well as between the concepts of least common denominator and greatest common divider and the np-set operations of union and intersection. A relationship has also been observed between the np-set Jaccard similarity index and the two latter concepts. The concept of coprime has also been extended to rational numbers by using the described np-set form of respective canonical representations.

The second approach described in the present work concerns the possibility of quantifying the prime factor structure of real numbers in terms of respective prime factor and multiplicity signatures obtained from the decimal expansions of the analyzed numbers. Several types of similarity networks have been described which can be used to reflect the prime factor structure of positive integer and real numbers. Preliminary examples have provided an indication of the potential of the described concepts and methods in number theoretical areas. These include the identification that the fundamental constants π and $\sqrt{2}$, which are both irrational values, have decimal expansions involving mostly prime factors with respective unit multiplicity. In addition, networks obtained by comparing the similarity between the number of primes and respective multiplicities allowed a set of numbers to be organized into a modular structure where each module corresponds to a possible type of canonical representation form.

Possible further studies include addressing other topics of number theory from the perspective of np-sets concepts and methods, as well as developing more comprehensive statistical characterization of the prime factor structure of rational and irrational real numbers.

Acknowledgments

Luciano da F. Costa thanks CNPq (grants no. 307085/2018-0 and 313505/2023-3) and FAPESP (grant 2022/15304-4)

Observations

As all other preprints by the author, the present work contains preliminary work subject to further revision and validation. Respective modification, commercial use, or distribution of any of its parts are not possible, as this work has author copyright. Many of the preprints by the author are also available in Hal and/or arXiv. This work can also be cited by using the DOI number or article identification link. It is also interesting to check for new versions of this work, as they may become available from time to time. Thanks for reading.

References

- [1] O. Ore. *Number theory and its history*. Courier Corporation, 2012.

- [2] G. H. Hardy and E. M. Wright. *An introduction to the theory of numbers*. Oxford Science Publication, 1979.
- [3] T. M. Apostol. *Introduction to Analytic Number Theory*. Springer, 1998.
- [4] I. M. Vidogradov. *Elements of Number Theory*. Dover, 2003.
- [5] Z. I. Borevich and I. R. Shafarevich. *Number theory*. Academic press, 1986.
- [6] G. Tenenbaum and M. M. France. *The prime numbers and their distribution*, volume 6. American Mathematical Soc., 2000.
- [7] G. J. O. Jameson. *The prime number theorem*. Number 53. Cambridge University Press, 2003.
- [8] L-K Hua. *Introduction to number theory*. Springer Science & Business Media, 2012.
- [9] R. Crandall and C. Pomerance. *Prime numbers*. Springer, 2001.
- [10] G. Marsaglia. The structure of linear congruential sequences. In *Applications of number theory to numerical analysis*, pages 249–285. Elsevier, 1972.
- [11] H. Riesel. *Prime numbers and computer methods for factorization*, volume 57. Springer Science & Business Media, 2013.
- [12] J. T. Machado and A. M. Lopes. Multidimensional scaling and visualization of patterns in prime numbers. *Communications in Nonlinear Science and Numerical Simulation*, 83:105128, 2020.
- [13] T. Nagell. *Introduction to number theory*, volume 163. American Mathematical Soc., 2021.
- [14] L. da F. Costa. A first glance at prime numbers. https://www.researchgate.net/publication/349466484_A_First_Glance_at_Prime_Numbers_CDT-55, 2021.
- [15] A. H. Beiler. *Recreations in the theory of numbers: The queen of mathematics entertains*. Courier Corporation, 1964.
- [16] M. Kraitchik. *Mathematical recreations*. Courier Corporation, 2006.
- [17] S. Singh. *Fermat’s enigma: The epic quest to solve the world’s greatest mathematical problem*. Anchor, 2017.
- [18] H. Darmon, F. Diamond, and R. Taylor. Fermat’s last theorem. *Current developments in mathematics*, 1995(1):1–154, 1995.
- [19] A. Wiles. Modular elliptic curves and fermat’s last theorem. *Annals of mathematics*, 141(3):443–551, 1995.
- [20] J. R. Myhill. A complete theory of natural, rational, and real numbers. *The Journal of Symbolic Logic*, 15(3):185–196, 1950.
- [21] A. Stakhov. The generalized golden proportions, a new theory of real numbers, and ternary mirror-symmetrical arithmetic. *Chaos, Solitons & Fractals*, 33(2):315–334, 2007.
- [22] L. da F. Costa. Integrating set theory and real vector spaces. https://www.researchgate.net/publication/383201213_Integrating_Set_Theory_and_Real_Vector_Spaces, 2024.
- [23] D. E. Knuth. *The Art of Computing*. Addison Wesley, 1998.
- [24] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36–66, 1989.
- [25] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.
- [26] K. Rana and et al. Mathematics of multisets. *International Journal for Research in Applied Sciences and Biotechnology*, 9(3):102–106, 2022.
- [27] I. Niven. *Irrational numbers*. Number 11. Cambridge University Press, 2005.
- [28] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE transactions on information theory*, 44(6):2743–2760, 1998.
- [29] P. D. Grünwald. *The minimum description length principle*. MIT press, 2007.
- [30] H. G. Mairson. Some new upper bounds on the generation of prime numbers. *Communications of the ACM*, 20(9):664–669, 1977.
- [31] L. M. Adleman. On distinguishing prime numbers from composite numbers. In *21st Annual Symposium on Foundations of Computer Science (sfcs 1980)*, pages 387–406. IEEE, 1980.
- [32] U. M. Maurer. Fast generation of prime numbers and secure public-key cryptographic parameters. *Journal of Cryptology*, 8:123–155, 1995.

- [33] M. Joye, P. Paillier, and S. Vaudenay. Efficient generation of prime numbers. In *Cryptographic Hardware and Embedded Systems—CHES 2000: Second International Workshop Worcester, MA, USA, August 17–18, 2000 Proceedings 2*, pages 340–354. Springer, 2000.
- [34] M. L. Stein, S. M. Ulam, and M. B. Wells. A visual display of some properties of the distribution of primes. *The American Mathematical Monthly*, 71(5):516–520, 1964.
- [35] M. Stein and S. M. Ulam. An observation on the distribution of primes. *The American Mathematical Monthly*, 74(1):43–44, 1967.
- [36] I. B. Kolossváry and I. T. Kolossváry. Distance between natural numbers based on their prime signature. *Journal of Number Theory*, 234:120–139, 2022.
- [37] R. Fitzpatrick. *Euclid’s elements of geometry*. 2008.
- [38] D. Kolba and T. Parks. A prime factor fft algorithm using high-speed convolution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(4):281–294, 1977.
- [39] P. Duhamel and M. Vetterli. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.
- [40] C. Burrus and P. Eschenbacher. An in-place, in-order prime factor fft algorithm. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(4):806–817, 1981.
- [41] J.-I. Guo, C.-M. Liu, C.-W. Jen, et al. A new array architecture for prime-length discrete cosine transform. *IEEE Transactions on Signal Processing*, 41(1):436–442, 1993.
- [42] E. Kranakis. *Primality and cryptography*. Springer-Verlag, 2013.
- [43] N. Koblitz. *A course in number theory and cryptography*, volume 114. Springer Science & Business Media, 1994.
- [44] J. Kraft and L. Washington. *An introduction to number theory with cryptography*. Chapman and Hall/CRC, 2018.
- [45] S. Y. Yan. *Computational number theory and modern cryptography*. John Wiley & Sons, 2013.
- [46] M. Kharbutli, Y. Solihin, and J. Lee. Eliminating conflict misses using prime number-based cache indexing. *IEEE Transactions on Computers*, 54(5):573–586, 2005.
- [47] S. Y. Yan. *Number theory for computing*. Springer Science & Business Media, 2013.
- [48] G. Wu, K. Zhang, C. Liu, and J. Li. Adapting prime number labeling scheme for directed acyclic graphs. In *Database Systems for Advanced Applications: 11th International Conference, DASFAA 2006, Singapore, April 12–15, 2006. Proceedings 11*, pages 787–796. Springer, 2006.
- [49] R. K. Bhullar, L. Pawar, V. Kumar, et al. A novel prime numbers based hashing technique for minimizing collisions. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pages 522–527. IEEE, 2016.
- [50] K. S. Williams and C. Simon. The ecology, behavior, and evolution of periodical cicadas. *Annual review of entomology*, 40(1):269–295, 1995.
- [51] J. Yoshimura. The evolutionary origins of periodical cicadas during ice ages. *The American Naturalist*, 149(1):112–124, 1997.
- [52] E. Goles, O. Schulz, and M. Markus. Prime number selection of cycles in a predator-prey model. *Complexity*, 6(4):33–38, 2001.
- [53] M. Loconsole and L. Regolin. Are prime numbers special? Insights from the life sciences. *Biology Direct*, 17(1):1–9, 2022.
- [54] P. Tarau. Towards a generic view of primality through multiset decompositions of natural numbers. *Theoretical Computer Science*, 537:105–124, 2014.
- [55] P. Tarau. Emulating primality with multiset representations of natural numbers. In *Theoretical Aspects of Computing—ICTAC 2011: 8th International Colloquium, Johannesburg, South Africa, August 31–September 2, 2011. Proceedings 8*, pages 218–238. Springer, 2011.
- [56] R. A. Melter and I. Tomescu. Metric bases in digital geometry. *Computer vision, graphics, and image Processing*, 25(1):113–121, 1984.
- [57] R. Klette and A. Rosenfeld. *Digital geometry: Geometric methods for digital picture analysis*. Elsevier, 2004.
- [58] C. E. Kim and A. Rosenfeld. Digital straight lines and convexity of digital regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):149–153, 1982.
- [59] M. D. McIlroy. A note on discrete representation of lines. *AT&T Technical Journal*, 64(2):481–490, 1985.

- [60] J. Koplowitz, M. Lindenbaum, and A. Bruckstein. The number of digital straight lines on an $n^* \times n$ grid. *IEEE Transactions on Information Theory*, 36(1):192–197, 1990.
- [61] M. Lindenbaum and J. Koplowitz. A new parameterization of digital straight lines. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(08):847–852, 1991.
- [62] D. Khoshnoudirad. Farey lines defining farey diagrams and application to some discrete structures. *Applicable Analysis and Discrete Mathematics*, pages 73–84, 2015.
- [63] P. A. Solares-Hernández, M. A. García-March, and J. A. Conejero. Divisibility networks of the rational numbers in the unit interval. *Symmetry*, 12(11):1879, 2020.
- [64] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois Journal of Mathematics*, 6(1):64–94, 1962.
- [65] P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37:547–549, 1901.
- [66] Wikipedia. Jaccard index. https://en.wikipedia.org/wiki/Jaccard_index.
- [67] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021.
- [68] A. Benatti and L. da F. Costa. Normalization in proportional feature spaces. https://www.researchgate.net/publication/383666067_Normalization_in_Proportional_Feature_Spaces, 2024.
- [69] L. da F. Costa. Coincidence complex networks. <https://iopscience.iop.org/article/10.1088/2632-072X/ac54c3>, 2022. *J. Phys.: Complexity*, (3): 015012.
- [70] C. Pomerance. The prime number graph. *Mathematics of Computation*, 33(145):399–408, 1979.
- [71] D. E. G. Malm. A graph of primes. *Mathematics Magazine*, 66(5):317–320, 1993.
- [72] G. Corso. Families and clustering in a natural numbers network. *Physical Review E*, 69(3):036106, 2004.
- [73] A. K. Chandra and S. Dasgupta. A small world network of prime numbers. *Physica A: Statistical Mechanics and its Applications*, 357(3-4):436–446, 2005.
- [74] D.-H. Shi and H.-J. Zhou. Natural number network and prime number theorem. *Complex Systems and Complexity Science*, 7(4):52–54, 2010.
- [75] G. Garcia-Perez, M. A. Serrano, and M. Boguná. Complex architecture of primes and natural numbers. *Physical Review E*, 90(2):022806, 2014.
- [76] S. M. Shekatkar, C. Bhagwat, and G. Ambika. Divisibility patterns of natural numbers on a complex network. *Scientific Reports*, 5(1):1–10, 2015.
- [77] X.-Y. Yan, W.-X. Wang, G.-R. Chen, and D.-H. Shi. Multiplex congruence network of natural numbers. *Scientific Reports*, 6(1):1–8, 2016.
- [78] L. da F. Costa. Prime factors networks. https://www.researchgate.net/publication/361687024_Prime_Factors_Networks, 2022.
- [79] T. M. Fruchterman, M. J. Thomas, and E. M. Reingold. Graph drawing by force-directed placement. *Software Practice and Experience*, 21:1129–1164, 1991.
- [80] D. W. Sharpe. *Rings and factorization*. CUP Archive, 1987.
- [81] D. D. Anderson, D. F. Anderson, and M. Zafrullah. Factorization in integral domains. *Journal of pure and applied algebra*, 69(1):1–19, 1990.
- [82] H. Matsumura. *Commutative ring theory*. Number 8. Cambridge university press, 1989.