Mémoire pour l'obtention de l'Habilitation à Diriger des Recherches

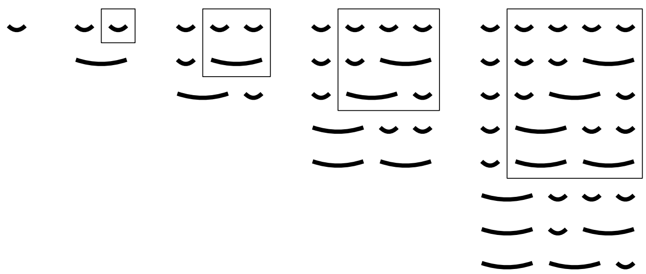
École doctorale n°37

Sciences Physiques pour l'Ingénieur et Microtechniques

par

Sergey Sergeevitch KIRGIZOV

Combinatorial Contemplations



Présenté et soutenu à Dijon, le 16 mars 2026

Composition du Jury :

Antoine Genitrini, Professeur à Sorbonne Université	Rapporteur
Renzo Pinzani, Professeur émérite à l'Université de Florence	Rapporteur
Eric Rivals, Directeur de Recherches CNRS, Université de Montpellier	Rapporteur
Olivier Bodini, Professeur à l'Université Sorbonne Paris Nord	Examinateur
Enrica Duchi, Professeure à l'Université Paris-Cité	Examinatrice
Alain Giorgetti, Professeur à l'Université Marie et Louis Pasteur	Examinateur
Elena Barcucci, Ancienne Professeure à l'Université de Florence	Invitée

Résumé

J'ai écrit ce texte pour mes pairs, les étudiants, les jeunes scientifiques et tous ceux qui s'intéressent à la combinatoire et aux mathématiques discrètes. Mon ouvrage est différent d'une monographie purement technique et contient plusieurs passages historiques, autobiographiques et philosophiques.

Le mémoire comprend trois chapitres. Le premier chapitre est une introduction, abordant mes réflexions philosophiques sur la nature du dénombrement et de l'énumération combinatoire, ainsi que sur les choses et leurs noms. Il décrit également la méthode classique de Goulden-Jackson, utilisée dans le deuxième chapitre. Ce deuxième chapitre, avec le troisième qui le suit, présentent mes contributions ainsi que certaines découvertes récentes de la littérature.

Plus précisément, le deuxième chapitre est consacré à la combinatoire de certains types de motifs dans la structure moléculaire des acides ribonucléiques (ARN, un des éléments les plus importants des organismes vivants). Il examine la distribution de ces motifs dans les structures d'ARN réelles et leurs modèles théoriques.

Le troisième chapitre aborde essentiellement deux sujets : une nouvelle restriction des chemins de Dyck, comptée par les nombres de Motzkin, et une nouvelle classe de mots, comptée selon Fibonacci. Outre des résultats purement scientifiques, ce chapitre fournit également un contexte autobiographique. Cela se conclut par la présentation des travaux connexes et des pistes possibles pour des recherches à venir, ainsi que plusieurs courts poèmes sur le processus fascinant de la traduction des pensées en langue de mots et de nombres.

Dans cet ouvrage, je me suis concentré principalement sur les aspects mathématiques des motifs apparaissant en combinatoire et en informatique. Mais n'oublions pas que c'est grâce à la langue que les connaissances mathématiques se transmettent de génération en génération, traversant l'univers.

Combinatorial contemplations

Узоров созерцание

Sergey Sergeevitch Kirgizov

February 21, 2026

for Yulia, Liza, Fedia, Sima, Vasia and Misha.

Ищущий да обрящет[1]

[1]Seek, and ye shall find — derived from Matthew 7:7–8.

Foretale

The word "pattern" is translated to russian word *узор*. It is pronounced like [uzor] and comes etymologycally from Old East Slavic *узоръ*. It is related to the verb *зреть* [zret'] (to see, to look at, to observe) prefixed by *у* [u] which is used to give a verb the perfective aspect. So, *узор* literally means "a thing that has been seen", but in contemporary language it is used to signify "a pattern, an ornament or a design".

I call the *Language* a tissue woven by history out of local languages and dialects. The Language contains a big part of our wisdom and the knowledge about the world. The reader's inquiring mind will undoubtedly find a lot of neat things and details in the Language, including many etymologycal connections between seemingly different words: matter and mothers; motifs and motivations; patterns, fathers and Jupiter... Eventually these beautiful details will create an overall vision of the whole canvas, a canvas full of patterns...

In this monograph, I will focus mainly on the mathematical aspects of patterns arising in combinatorics and computer science. But let's not forget that it is thanks to Language that mathematical knowledge is passed on from generation to generation, sailing across the universe.

I am writing this text for my peers, students, young scientists, and everyone who is interested in combinatorics and discrete mathematics. My treatise will be different from a purely technical monograph and will contain several historical, autobiographical and philosophical passages.

Contemplation is the key to comprehension.

Acknowledgments

I am endlessly grateful to Elena Barcucci, Olivier Bodini, Enrica Duchi, Antoine Genitrini, Alain Giorgetti, Renzo Pinzani and Eric Rivals — reviewers and jury members of my HDR thesis defense — for their participation, comments, advice, and helpful critique.

I can hardly imagine my scientific path without many conversations with my colleagues and teachers from Tselinograd, Akmola, Astana, Barnaul, Paris, Dijon and other cities. I greatly thank all of them, especially the members of the team founded by Jean-Marcel Pallo.

I warmly thank all my co-authors: Lobna Azaza, Jean-Luc Baril, Ian Basaille-Gahite, Alexis Baudin, Julien Bensmail, David Bevan, Célia Biane, Mireille Bousquet-Mélou, Alexander Burstein, Nadine Cullot, Maximilien Danisch, Romaric Duvignau, Sergey Dovgal, Rim Faiz, Alexander Frame, Abdelhamid Garah, Nicolas Gastineau, Elisabeth Gavignet, Richard Genestier, Marwan Ghanem, Benjamin Gras, Thierry Grison, Nathanaël Hassler, Greg Hampikian, Tatiana Kondrashova, Éric Leclercq, Clémence Magnien, Jacques Malenfant, Rémi Maréchal, Nader Mbarek, Amélie Medem, Olga Melekhova, Mehdi Naima, Khaydar Nurligareev, Mahmoud Omidvar, Armen Petrossian, Daniel Pinson, François Queyroi, José Luis Ramírez, Nicolas Rivierre, Marinette Savonnet, Fabien Tarissan, Isis Truck, Vincențiu Vajnovszki, Diego Villamizar, Michael Wallner.

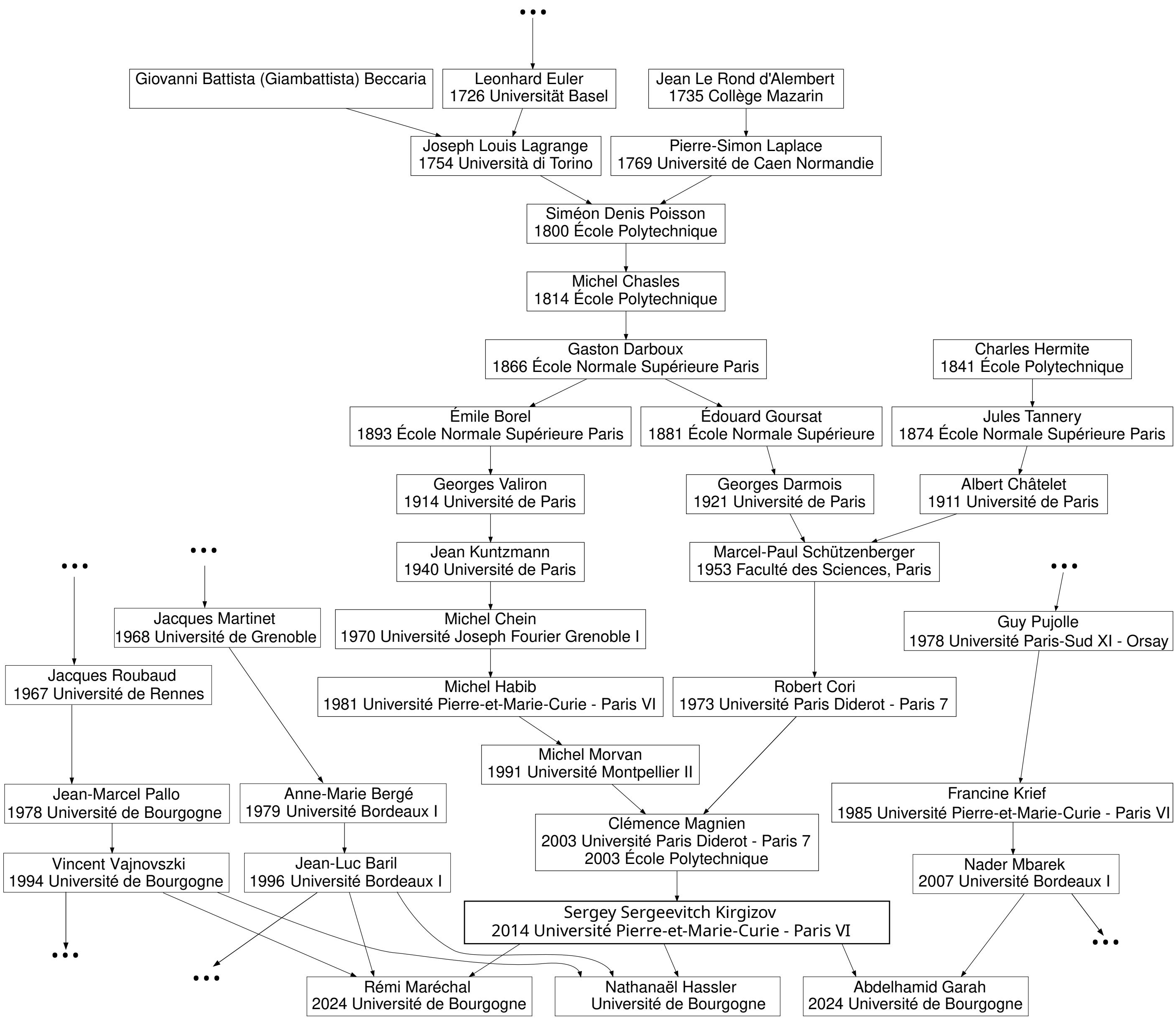
Special thanks to Dmitry "Sharkman" Akulinin for his constant support and philosophical conversations over a cup of tea, to Anaïs Vajnovszki for interesting chats on linguistics and languages.

I also thank all the poets, litterateurs and bards whose books and poems I read and whose songs I love.

I lovingly thank my family and friends.

A part of my academic genealogy

Drawn with the help of The Mathematics Genealogy Project
http://www.mathgenealogy.org
North Dakota State University, the American Mathematical Society;
and the libraries of Sorbonne University and University Burgundy Europe.



Contents

Commentary on the structure of this document

The monograph contains three Chapters. The first chapter is an introduction, outlining my philosophical views on the nature of counting, combinatorial enumeration, things and their names. It also contains a description of the classical Goulden-Jackson method which is used in the second Chapter. The second Chapter, together with the third, present my contributions as well as some recent findings of the literature.

More precisely, the second Chapter is focused on the combinatorics of certain types of patterns in the molecular structure of ribonucleic acids (RNAs, one of the most important elements of biological organisms). It examines the distribution of these patterns in the real-world RNA structures and their theoretical models.

The third Chapter essentially addresses two things: a new Motzkin-counted restriction of Dyck paths and a new class of Fibonacci-counted words. Not only does it provide purely scientific results, it also gives some autobiographical context. The third section of the Chapter 3 concludes the monograph by presenting a description of related works and possible directions for further research, as well as several short poems about the mesmerising process of translating thoughts into the language of words and numbers.

Funding sources A part of the research presented in this monograph was funded by the Agence Nationale de la Recherche (grants ANR JCJC Pics, ANR-22-CE48-0002, and ANR COREGRAPHIE, ANR-20-CE23-0002), and by Bourgogne-Franche-Comté region, project ANER ARTICO.

Chapter 1

Introduction and philosophical considerations

Until he had come up with a name, he was too pathetic to look at – a real idiot. But now that he had some label like graviconcentrate, he thought that he understood everything and life was a breeze.

— Arkady and Boris Strugatsky
Roadside Picnic
Translated by Antonina W. Bouis

Enumeration is perhaps one of the first mathematical activities that we begin to do at a fairly early age. We count: 1, 2, ... But how exactly we do that and how we compare sets of different objects having the same or different number of elements? This is a *hard question.* I hope that one day combinatorics as a part of mathematics and computer science will help to throw a light on this mental process. For now, discrete mathematical matters are quite far from hypothetical application to questions typically associated with biology, psychology and philosophy.

The word *enumeration* have several meanings. One of them is *listing*, when we mention, one by one, all objects in a set. The second meaning is *counting*. Meanwhile, the contemporary combinatorial enumeration is more about the study of structures and their formation under certain recursive laws than about sequences of numbers or listing objects *per se.* In combinatorics we also deal with different properties of integer sequences, with their asymptotic behavior, with algorithmic generation of elements of some sets, but typically it involves a detailed understanding of structures and deep comprehension of the recursive laws describing the birth of new objects of larger sizes from smaller, more primitive, elements.

It happens that means become more interesting than original goals.

1.1 Order and quantity

Which came first: the order or the quantity? The question is may seem strange at first glance, because we need a concept of order to be able to understand it. Without this concept of order how we can understand "first"? *Wovon man nicht sprechen kann, darüber muß man schweigen*[1] — once wrote Wittgenstein in his Tractatus [242]. *Of course* I won't follow Wittgenstein advice and *will* talk with you about this outlandish question. Remember that Wittgenstein, ten years after the publication of Tractatus, changed his mind about the logically dogmatic nature of philosophy and language. In his latter works he introduced the concept of *language games* and talked fondly about them [183]. We come back to *the question*: the quantity or the order, which is more important? This strange but surprisingly interesting question relates the very foundations of mathematics and computer science. Let me explain why I think so.

In 1921 Kuratowski [171] defined an ordered pair (a, b) as $\{\{a\}, \{a, b\}\}$ solving certain problems in a set theory of the early XX century. This definition is now well-known and accepted. It allows us to construct relations, functions and arrows using only the language of unordered sets. So, in many parts of mathematics order is no more considered as a primitive notion.

When we consider exactly how mathematics is put into practice, we see sequences of bits, we see programs built from those sequences, we see approximations of Turing machines with finite tapes, we see systems for automatically checking proofs and formalizing mathematics (Isabelle, HOL Light, Coq, Lean, Metamath...)—all are based on ordered sequences of bits. Unordered sets are realised and implemented as well-defined orders of bytes in data structures and associated procedures. Perhaps even the time itself is a non-symmetrical order of events.

So, simplifying slightly, I can say that the abstract mathematics and its physical sister—computer science—embrace the same thing from two opposite sides. Computer science construct unordered sets and other notions from ordered sequences, while the mathematics does the converse. Perhaps the significance of these opposite processes is that the essence is contained somewhere in the middle. Or maybe it is the very possibility to look at the world from different angles and from different sides that makes it so attractive to scientists. It could be also a manifestation of a bijective relationship between multiple interpretations of the same phenomenon

[1] Whereof one cannot speak, thereof one must be silent — Ludwig Wittgenstein [243] (translated by Frank P. Ramsey and Charles K. Ogden).

inaccessible to the direct examination.

1.2 Things and their names

Poincaré once wrote — *la Mathématique est l'art de donner le même nom à des choses différentes*[2]. I think that *entities, things and concepts* exists but the words are somehow "secondary", they form a kind of virtual descriptive reality. Eventually, words become entities and others words refer to them, braiding the filaments of cognition.

To see the essence of the world, it is necessary to consider the things and concepts from different angles, using different languages, and call them every so often by different names. This need seems to contradict Poincaré, but I think that it completes his ideas about the universal nature of the mathematical language and allows us, for instance, to see a link between enumeration and bijection; to suggest the existence of underlying *entity* that can be experienced by us as an enumeration or a bijection depending on the context. This *entity* may exist in the reality independently of thinking beings or appear as a form of physical or biological law guiding the thought processes in any brain engaged in mathematical reasoning, or even have divine or mystical nature.

Should we try to invent a new *universal* name to this *entity*, say *bijenumé*, or can we continue to consider it from different angles using different names? I'm more inclined towards the second option, because by inventing a new name and pretending to be universal, we merely add a new word to the set of already existing names of *the entity*. Every name have their own nuances of meaning, *bijenumé* will be associated with my current reasoning, which does not fully reflect the essence and future manifestations of *the entity*. Thus, I prefer to denounce to give the universal names. This, however, will not prevent me from talking about entities or concepts using already existing words or inventing new ones, but... without pretending to be universal.

1.3 Generating functions

Imagine you have to concisely package the information about number of certain things. If you have only five things, you can just write 5. But what happens if you have a finite set $\mathcal{A}$ consisting of things of different sizes:

[2] Mathematics is the art of giving the same name to different things [205] (translated by Francis Maitland).

1 thing of size 0, 2 things of size 1, and 3 things of size 2? In this case, we can use polynomials. We choose a formal variable, say x, whose degree will denote the size, and let monomial coefficients denote the number of things of a certain size:

$$A(x) = 1 + 2x + 3x^2.$$

We say that $A(x)$ is the *generating function* of the set $\mathcal{A}$. The union of two sets $\mathcal{A}$ and $\mathcal{B}$ having no common elements is called *(internal) disjoint union* and denoted by $\mathcal{A} \uplus \mathcal{B}$. The corresponding generating function is $A(x) + B(x)$, where $A(x)$ and $B(x)$ are generating functions for $\mathcal{A}$ and $\mathcal{B}$. Let $\mathcal{A} \times \mathcal{B}$ be the Cartesian product of two not necessarily disjoint sets $\mathcal{A}$ and $\mathcal{B}$, i.e. sets of all pairs (a, b) where $a \in A$ and $b \in \mathcal{B}$. It's quite easy to see that the generating function for the set $\mathcal{A} \times \mathcal{B}$ should be $A(x) \cdot B(x)$.

Leaving the discussion of the existence of potential or actual infinities for the future, we pretend that we have an infinite sequence of things, but every thing has a finite size. For example, let's look at binary words avoiding two adjacent ones. We have:

- 1 word of size 0, denoted by ε;
- 2 words of size 1: 0, 1;
- 3 words of size 2: 00, 01, 10;
- 5 words of size 3: 000, 001, 010, 100, 101;
- 8 words of size 4: 0000, 0001, 0010, 0100, 0101, 1000, 1001, 1010;
- etc.

Let's denote by $\mathcal{F}$ the set of such words, and try to write down the corresponding generating function:

$$F(x) = 1 + 2x + 3x^2 + 5x^3 + 8x^4 + \dots$$

Do we have enough paper to write all the terms of this "polynomial" where the number of terms is infinite? Of course, we can write something like

$$F(x) = \sum_{i=0}^{\infty} f_n x^n,$$

where f_n denotes the number of words of size n. But this does not give us any information about these numbers or a method of obtaining them. We need to understand the underlying structure, find rules to produce words and rules to compute f_n. In general, this is a tough problem (think about uncomputable sequences and numbers like Chaitin's constant [81]). It can, however, be solved in some special cases, sometimes by quite elegant and beautiful methods.

The set $\mathcal{F}$ of words avoiding 11 appears already in Knuth's book [165, p. 286] and admits a simple description. For two words a and b, let $a \cdot b$ denote their concatenation. For a word a and a set of words $\mathcal{B}$, we let $a \cdot \mathcal{B} = \{a \cdot b : b \in \mathcal{B}\}$. Remark that

$$\mathcal{F} = \{\varepsilon\} \uplus \{1\} \uplus 0 \cdot \mathcal{F} \uplus 10 \cdot \mathcal{F}, \tag{1.1}$$

i.e. any word avoiding two consecutive ones is either empty, equals to 1, starts with 0, or starts with 10, and if we remove starting 0 or 10 we obtain again a (smaller, shorter) word from the set $\mathcal{F}$. So, a generating function $F(x)$, corresponding to the set $\mathcal{F}$, should satisfy the functional equation

$$F(x) = 1 + x + x \cdot F(x) + x^2 \cdot F(x).$$

Solving it we obtain:

$$F(x) = \frac{1+x}{1-x-x^2}\,. \tag{1.2}$$

The Taylor expansion of this function around the point $x = 0$ gives the infinite power series $1 + 2x + 3x^2 + 5x^3 + 8x^4 + \ldots$, that corresponds to the Fibonacci sequence: $1, 2, 3, 5, 8, 13, \ldots$

Magic? At least something very similar.

Usually, when one works with power series, the questions of convergence arises. In our toy example, the generating function is analytic, and its radius of convergence is $\frac{\sqrt{5}-1}{2}$. Combinatorial sequences could be considered as "polynomials with infinite number of terms" and can be studied even if they are not convergent in the classical sense. In the general form, they are known as *formal power series* in the literature and provide a way to encode the sequences of numbers.

To properly define (or read/write) any sequence of numbers we need a notion of order (which number comes first, which number comes second, and so on). In the context of formal power series the order of the terms is no longer important, as it can be reconstructed from monomial powers. With the discussion in Subsection 1.1 in mind, I'm speculating that maybe the mathematical desire for unordered collection of elements was one of the unexposed motives behind the creation of the generating function theory.

Classical combinatorial books by Flajolet and Sedgewick [109], Wilf [241], and Stanley [225, 227] describe the ring of formal power series, possible operations with these series, and related questions in great detail. In this monograph, we use many elements from these books.

1.4 Patterns in combinatorics

What is a pattern? How to count their occurrences? Can they self-overlap? Which patterns are found frequently, and which are rare in large (random) structures? Is it important to avoid certain patterns and why? How can the combinatorial contemplation of patterns help us to characterize structures found in the borderlands and the frontier areas of mathematics and computer science?

We endeavor to answer these and other related questions. Of course, everything will depend on precise definitions and mathematical contexts. Many scientists are working in this field. Here is a photo (Figure 1.1) of some of them who came to Dijon for the 21st edition of the conference “Permutations patterns” organized by our team in 2023.

Figure 1.1: Permutation Patterns conference, Dijon, 2023.

Kitaev wrote a book about patterns in permutations and words [164]. Mansour [188] presented results on pattern avoidance and pattern enumeration in set partitions. A book [144] by Heubach and Mansour recounts methods and achievements in the field of patterns in compositions, words and related combinatorial structures. Asinowski [5] Bacher [16], Bean [48], Bernini [58], Bevan [61], Bilotta [64], Bouvel [71], Burstein [74], Cerbai [78], Cervetti [79], Duchi [100], Elizalde [104], Pierrot [203], Pudwell [207], Tenner [230], as well as many other passionate scientists, have done their dissertations on topics related to combinatorial patterns. There are also more general monographs on the subject. Flajolet and Sedgewick [109], Goulden and Jackson [130], Knuth [165, 166], Lothaire [178–180], Stanley [225–227], Bóna [75] and other authors tell us marvelous stories about patterns in a variety of settings on the pages of their now classic books.

Genitrini and Mailler [124] extended the Kozik's [169] pattern theory, which was originally introduced to study Boolean trees, to other tree-based models of satisfiability problems. Many other talented researchers study connections between patterns and logic, see for instance the work of Bendkowski, Bodini and Dovgal [50] on statistical properties of lambda terms; the treatise on dense digraphs (directed graphs) and 2-SAT formulae written by Dovgal and Nurligareev [98]; and the works of Zeilberger and Giorgetti [248], Bodini, Singh and Zeilberger [67, 68] which connect combinatorial maps and lambda terms. There are also numerous links with physics, see for example a recent work by Saniga, Holweck, Kelleher, Muller, Giorgetti and de Boutray [219] about finite geometric structures in the setting of quantum contextuality.

A myriad of scientists has been studying families of graphs that can be characterized by avoiding certain structures as (induced) subgraphs or minors. Perhaps the most famous result is due to Kuratowski [172], Frink and Smith [115], and Pontryagin. It states that a graph is planer if and only if it does not contain a subgraph that is a subdivision of $K_{3,3}$ or K_5 (see Figure 1.2). Wagner's [236] version of Kuratowski's theorem is formulated in terms of forbidden minors. Robertson–Seymour [215] theorem is a generalization of this result, actually they wrote a whole series of twenty papers, ≈500 pages.

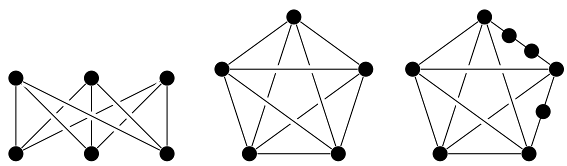


Figure 1.2: Graphs $K_{3,3}$, K_5 and a graph obtained by subdivision of certain edges of K_5.

Among the recent results on graphs, I would like to highlight the work of Ducoffe, Feuilloley, Habib and Pitois [101] that focuses on the following question: given an ordered graph of size n, how fast can we detect whether a fixed pattern of size k is present?

1.5 Classical Goulden-Jackson approach

First, let's consider binary words and consecutive patterns therein. Perhaps this is the simplest example we can show here, its essence has a pedagogical flavor. We want to show the reader (who may not be familiar with what follows) the basic ideas of the now classical Goulden-Jackson [129, 130] *cluster method*. A detailed description of this method for words over an arbitrary alphabet is given in several papers, see for instance works by Noonan and Zeilberger [198]; Bassino, Clément and Nicodème [45], together with Fayolle [44].

The Goulden-Jackson method is also related to Guibas-Odlyzko [136] work about words avoiding a given pattern and to Solov'ev's [224] paper about rare events. It is worth noting that the method, with the necessary modifications, is applicable not only to words and sequences of events, but also to other kinds of structures. Various examples can be found in a book by Flajolet and Sedgewick [109, Section III.7.4]; questions about consecutive pattern avoidance in lattice paths are treated in a series of articles by Asinowski, Bacher, Banderier, Gittenberger and Roitner [6–10]; Elizalde and Noy [105] use the method to enumerate consecutive patterns in permutations.

In Chapter 2 we will show how to adapt the classical Goulden-Jackson method in the context of endhered patterns in RNA secondary structures. Now let's go back to our simple example of binary words.

Denote by $\mathcal{B}_n$ the set of all binary words of size n. Let $\mathcal{B} = \bigcup_{n=0}^{\infty} \mathcal{B}_n$. In this context, we consider as patterns only consecutive non-empty subwords, also called *factors* in the literature. Given a pattern μ, we aim to provide a *bivariate generating function*

$$B_\mu(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} b_{n,k} x^n y^k, \tag{1.3}$$

where $b_{n,k}$ stands for the number of words of size n having exactly k occurrences of pattern μ.

How to get an expression for this function? An ad hoc method may work. In Section 1.3 we used one for the enumeration of the set $\mathcal{F}$ of binary words avoiding pattern $\mu = 11$, see Eq. (1.1). With a little tweaking, we can adapt it to obtain a full distribution of occurrences of $\mu = 11$. But we will go the other way.

We denote by $|\mu|$ the size of μ (number of letters inside μ, its length).

Definition 1. A pattern μ is *self-overlapping* if there is a word with size smaller than $2|\mu|$ that contains more than 1 occurrence of μ.

1.5.1 No self-overlap

At first, we only look at patterns that are non-self-overlapping. For instance, pattern 11 is self-overlapping, the word 111 contains 2 occurrences of 11. Patterns 010 and 0101 are also self-overlapping, but 01011 is not.

To describe binary words according to the number of occurrences of a given non-self-overlapping pattern μ, we start by showing the underlying sequential structure of any word. Denote by $\mathcal{S}$ the set of sequences, and by ❁ an atom, a general element in the sequence ❁ ❁ ❁ ... ❁. Any sequence is either empty or starts with an atom, which is followed by another sequence, so we have the following decomposition: $\mathcal{S} = \varepsilon \uplus$ ❁ $\cdot\, \mathcal{S}$. The corresponding generating function is, obviously, $S(x) = \frac{1}{1-x}$.

A binary word is a sequence wherein each atom is replaced by 0 or 1, its generating function is $S(2x)$. If we want to count them according to the size (marked by powers of x), the number of 1s (powers of y) and the number of 0s (powers of z), we compute the Taylor expansion of $S(xy + xz)$ near $x = 0$.

The variable y in the generating function $B_\mu(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} b_{n,k} x^n y^k$ marks an occurrence of pattern μ. A word m of length $|m|$ with k occurrences of μ corresponds to the term $x^{|m|} y^k$, which can be rewritten as

$$\underbrace{x \cdot x \cdot \ldots \cdot x}_{|m| \text{ times}} \cdot \underbrace{y \cdot y \cdot \cdots \cdot y}_{k \text{ times}}. \tag{1.4}$$

The coefficient $b_{|m|,k}$ gives us the number of such words.

Now let's label by the variable u some (maybe 0) occurrences of the pattern μ. We need to consider all possible labelling options. We can do this by replacing each y in (1.4) either by u (label) or by 1 (no label). Denote by $\mathcal{L}_\mu$ the set of such binary words with labels. Its generating function is

$$L_\mu(x,u) = B_\mu(x, 1+u). \tag{1.5}$$

For example, if $\mu = 01011$ the word 110101110101011001 of size 18 contributes as $x^{18}y^2$ to $B_\mu(x,y)$, but engenders four words in $\mathcal{L}_\mu$ and four terms in $L_\mu(x,u)$

$$\begin{aligned}
11\underline{01011}101\underline{01011}001 &\mapsto x^{18}u^2,\\
11\underline{01011}10101011001 &\mapsto x^{18}u,\\
110101110\underline{1}\underline{01011}001 &\mapsto x^{18}u,\\
110101110101011001 &\mapsto x^{18},
\end{aligned} \tag{1.6}$$

where the labelled occurrences of the pattern 01011 are underlined.

It seems we only make things harder, but, as is often the case, by enlarging the context, we reduce the complexity of the problem. It is possible to obtain the expression for $B_\mu(x, y)$ without cluster method. Usually, to construct binary words, we add a letter 0 or 1 to the beginning of a shorter word, and this can create a new occurrence of μ. If one wants to proceed in this direction, one has to consider words starting with certain sequences of letters. In perspective, this idea can be used to study consecutive patterns in arbitrary regular languages. But now let's get back to the spellbinding Goulden-Jackson method.

It is not hard to derive the generating function $L_\mu(x, u)$, as soon as we realise that a word in the set $\mathcal{L}_\mu$ is a sequence of letters from the alphabet $\{0, 1, \mu\}$ (do not forget that in our case the pattern μ is non-self-overlapping). We obtain

$$L_\mu(x, u) = S(x + x + ux^{|\mu|}) = \frac{1}{1 - 2x - ux^{|\mu|}}.$$

Yes! Letters can have several symbols inside! For example, "Dzs" is one letter of the Hungarian alphabet, Figure 1.3.

A	Á	B	C	Cs	D	Dz	Dzs	E	É	F
G	Gy	H	I	Í	J	K	L	Ly	M	N
Ny	O	Ó	Ö	Ő	P	Q	R	S	Sz	T
Ty	U	Ú	Ü	Ű	V	W	X	Y	Z	Zs

Figure 1.3: Hungarian alphabet.

From Eq. 1.5 we straightforwardly get

$$B_\mu(x, y) = L_\mu(x, y - 1) = \frac{1}{1 - 2x - (y - 1)x^{|\mu|}}. \tag{1.7}$$

This "y-1" trick is known as *symbolic inclusion-exclusion principle*. We keep only the words where all occurrences of μ are labelled by u. In particular, the example (1.6) is transformed as follows

$$\left.\begin{array}{r} 11\underline{01011}101\underline{01011}001 \mapsto x^{18}(y-1)^2 \\ + \\ 11\underline{01011}10101011001 \mapsto x^{18}(y-1) \\ + \\ 1101011101\underline{01011}001 \mapsto x^{18}(y-1) \\ + \\ 110101110101011001 \mapsto x^{18} \end{array}\right\} \Longrightarrow 11\underline{01011}101\underline{01011}001 \mapsto x^{18}y^2.$$

We conclude that the distribution of non-self-overlapping consecutive pattern in binary words depends only on the pattern size.

1.5.2 With self-overlap

A pattern may self-overlap in several possible ways, for instance, the pattern 0101010 has three possible self-overlappings, Figure 1.4

```
0101010        0101010        0101010
  0101010          0101010          0101010
```

Figure 1.4: Three possible self-overlappings of 0101010.

We need a way to encode the sizes of such overlappings.

Definition 2. An *autocorrelation polynomial* $A_\pi(x)$ for a pattern π of size n is defined as

$$A_\pi(x) = 1 + \sum_{k \in S} x^{n-k},$$

where S is the set of sizes of possible intersections of two different occurrences of the pattern π. In other words, x^{n-k} means that two occurrences of π have k letters in common.

Examples of autocorrelation polynomials:

$$\begin{aligned} A_{11}(x) &= 1 + x, \\ A_{111}(x) &= 1 + x + x^2, \\ A_{010}(x) &= 1 + x^2, \\ A_{0101}(x) &= 1 + x^2, \\ A_{01011}(x) &= 1, \\ A_{0101010}(x) &= 1 + x^2 + x^4 + x^6. \end{aligned}$$

For any non-self-overlapping pattern μ, we have $A_\mu(x) = 1$.

How many autocorrelation polynomials there are for binary words of size n?

Let's denote the answer by $\kappa(n)$. Sloane's On-Line Encyclopedia of Integer Sequences [200] contains the sequence $\{\kappa(n)\}_{n\in\mathbb{N}}$ under the reference A005434. It starts with 1, 2, 3, 4, 6, 8, 10, 13, 17, 21. Table 1.1 presents 6 classes of binary words of size 5 according to their autocorrelations.

Pattern	**Autocorrelation polynomial**
00001, 11110, 10000, 01111, 00011, 11100, 00111, 11000, 00101, 11010, 10100, 01011, 00110, 11001, 01100, 10011	1
01001, 10110, 10010, 01101	$1 + x^3$
00010, 11101, 01000, 10111, 01110, 10001	$1 + x^4$
01010, 10101	$1 + x^2 + x^4$
00100, 11011	$1 + x^3 + x^4$
00000, 11111	$1 + x + x^2 + x^3 + x^4$

Table 1.1: Autocorrelations of binary words of size 5.

In 1981 Guibas and Odlyzko [135] investigated this question in detail, gave upper and lower bounds of $\kappa(n)$, provide several equivalent characterizations of autocorrelations, studied enumeration of words with a given autocorrelation polynomial. They also proved a non-obvious statement:

Corollary 5.1 from [135]
The sequence A005434 is independent of the alphabet size: the number of possible autocorrelations is the same for binary, ternary and general k-ary words.

In 2001, Rivals and Rahmann [213] improved the lower bound on $\kappa(n)$ and, among other things, showed that corresponding autocorrelation polynomials form a *lattice*[3] under the term set inclusion: for two autocorrelation polynomials $A_\pi(x)$ and $A_\mu(x)$, we have $A_\pi(x) \subseteq A_\mu(x)$ if and only if all terms with coefficients 1 of $A_\pi(x)$ are also in $A_\mu(x)$ (see Figure 1.5).

In some cases, it is more convenient to encode autocorrelation polynomials as binary words, it is possible because such polynomials have 0,1 coefficients. In other cases, authors prefer to consider *period sets*, i.e. sets of powers of terms with coefficient 1. Using the notation from Definition 2 a period set is expressed as $\{n - k : n \in S\}$.

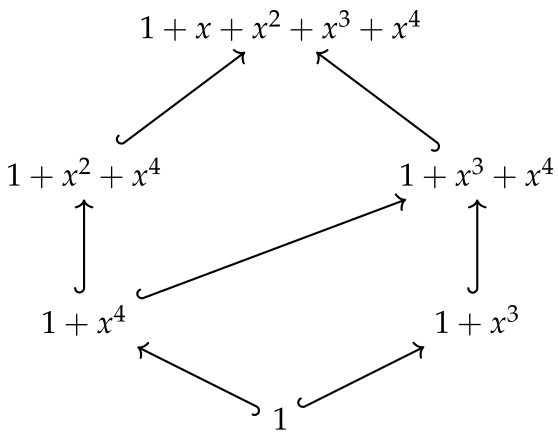


Figure 1.5: Lattice structure of 6 possible autocorrelations of binary words of size 5.

Guibas and Odlyzko conjectured that $\frac{\ln \kappa(n)}{\ln^2 n}$ converges as $n \to \infty$. In 2023, Eric Rivals, Michelle Sweering, and Pengfei Wang [214] proved the conjecture, 42 years after Guibas-Odlyzko work: $\lim_{n\to\infty} \frac{\ln \kappa(n)}{\ln^2 n} = \frac{1}{2\ln 2} =$ 0.72134752....

A non-trivial recursive predicate, proposed by Guibas and Odlyzko, can be used for exact enumeration and exhaustive generation of autocorrelation

[3] In order theory, a *lattice* is a partially ordered set (*a poset* for short) with an additional property: each pair of elements has a least upper bound and a greatest lower bound.

polynomials. An incremental algorithm has been provided by Rivals [212]. For more details in this area, we refer the reader to Chapter 8 (written by Mignosi and Restivo) of Lothaire's book [179], and to Wang's PhD thesis [238]. A C code, given by Sillke [222], also deserves the attention of connoisseurs.

Now we are going to explain how to use autocorrelation polynomials in order to obtain, in a relatively simple form, the bivariate generating function $B_\mu(x,y) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} b_{n,k}x^n y^k$ for the number of words of size n having exactly k occurrences of a self-overlapping pattern μ.

Consider for example the pattern $\mu = 01010$. In the word 010101010 there are 3 occurrences of μ, the corresponding term in $B_\mu(x,y)$ is x^9y^3. The word engenders 8 different words in the augmented set $\mathcal{L}_\mu$ of binary words, where some occurrences of the pattern are labelled. Again, all possible labelling options should be considered. These 8 words correspond to 8 terms in the bivariate generating function $L_\mu(x,u)$. We replace u by $y-1$ and, by symbolic inclusion-exclusion principle, obtain the only desired term x^9y^3:

$$\left.\begin{array}{r}
\underline{01010}1010 \mapsto x^9u \\
+ \\
01\underline{01010}10 \mapsto x^9u \\
+ \\
0101\underline{01010} \mapsto x^9u \\
+ \\
\underline{0101}\underline{\underline{0}}\underline{1010} \mapsto x^9u^2 \\
+ \\
\underline{01}\underline{\underline{010}}\underline{10}10 \mapsto x^9u^2 \\
+ \\
01\underline{01}\underline{\underline{010}}\underline{10} \mapsto x^9u^2 \\
+ \\
\underline{01}\underline{\underline{01}}\underline{\underline{\underline{0}}}\underline{\underline{10}}\underline{10} \mapsto x^9u^3 \\
+ \\
010101010 \mapsto x^9
\end{array}\right\} \Longrightarrow
\left.\begin{array}{r}
\underline{01010}1010 \mapsto x^9(y-1) \\
+ \\
01\underline{01010}10 \mapsto x^9(y-1) \\
+ \\
0101\underline{01010} \mapsto x^9(y-1) \\
+ \\
\underline{0101}\underline{\underline{0}}\underline{1010} \mapsto x^9(y-1)^2 \\
+ \\
\underline{01}\underline{\underline{010}}\underline{10}10 \mapsto x^9(y-1)^2 \\
+ \\
01\underline{01}\underline{\underline{010}}\underline{10} \mapsto x^9(y-1)^2 \\
+ \\
\underline{01}\underline{\underline{01}}\underline{\underline{\underline{0}}}\underline{\underline{10}}\underline{10} \mapsto x^9(y-1)^3 \\
+ \\
010101010 \mapsto x^9
\end{array}\right\} \Longrightarrow \underline{01}\underline{\underline{01}}\underline{\underline{\underline{0}}}\underline{\underline{10}}\underline{10} \mapsto x^9y^3.$$

As in the simple case of non-self-overlapping patterns, we have $B_\mu(x,y) = L_\mu(x, y-1)$. On further reflection, we realise that it's necessary to replace

$(y-1)x^{|\mu|}$ in Eq. (1.7) by something more general. The clusters of self-overlapping patterns can be of any size, they can be constructed as sequences of steps arising from the autocorrelation polynomial. So we have:

$$B_\mu(x,y) = L_\mu(x,y-1) = \frac{1}{1-2x-C_\mu(x,y-1)}, \tag{1.8}$$

where $C_\mu(x,u) = \frac{ux^{|\mu|}}{1-u(A_\mu(x)-1)}$ enumerates sequences of overlapping occurrences (clusters) of the pattern μ, with the specific occurrences labelled by u. This function is called *cluster generating function*. Comparing Eq. (1.7) and Eq. (1.8) we see that the case of non-self-overlapping patterns follows if $A_\mu(x) = 1$.

We conclude that the distribution of a self-overlapping consecutive pattern in binary words depends only on the associated autocorrelation polynomial and the size of the pattern in question.

1.5.3 Tracking several patterns at the same time

Is it possible to track several patterns at the same time? To provide a multivariate generating function

$$B_\Gamma(x,y_1,y_2,\ldots,y_{|\Gamma|}) = \sum_{x=0}^{\infty}\sum_{y_1=0}^{\infty}\cdots\sum_{y_{|\Gamma|}=0}^{\infty} b_{n,k_1,k_2,\ldots,k_{|\Gamma|}} x^n y_1^{k_1} y_2^{k_2}\cdots y_\Gamma^{k_{|\Gamma|}},$$

where $\Gamma = \{\mu_1,\mu_2,\ldots,\mu_{|\Gamma|}\}$ is a finite set of patterns of finite sizes?

This question is treated in detail by Bassino, Clément and Nicodème [45], together with Fayolle [44], see also a work of Noonan and Zeilberger [198]. In this introductory part, we only consider the case of binary words where no pattern in Γ is a proper factor of another pattern in Γ. For instance, 101 is a proper factor of 1101.

In Eq. 1.8, the cluster generating function is $C_\mu(x,u) = \frac{ux^{|\mu|}}{1-u(A_\mu(x)-1)}$. We need to replace it by something better suited to a set of patterns. Let's pretend for a moment, that we have such a generating function, and call it $C_\Gamma(x,u_1,u_2,\ldots)$. Since the overall structure is essentially the same—we still have the words that are sequences of letters 0, 1, or clusters of overlapping patterns—the following holds:

$$B_\Gamma(x,y_1,\ldots) = \frac{1}{1-2x-C_\Gamma(x,y_1-1,y_2-1\ldots)}. \tag{1.9}$$

Irish litterateur James Joyce once said, “For myself, I always write about Dublin, because if I can get to the heart of Dublin I can get to the heart of all the cities of the world. In the particular is contained the universal”. Joyce’s book “Finnegans Wake” helped Murray Gell-Mann to spell properly the word *quark*, which is now known to many, not just physicists and writers.

So, we will first look at two patterns 11 and 101, and then the general case will follow pretty naturally. Denote by $\mathcal{C}_{\{11,101\}}$ the set of clusters constructed by overlapping patterns 11 and 101, where the specific occurrences of 11 are labelled by u_1 and the specific occurrences of 101 are labelled by u_2. We decompose the set $\mathcal{C}_{\{11,101\}}$ into two disjoint subsets

$$\mathcal{C}_{\{11,101\}} = \mathcal{E}_{11} \uplus \mathcal{E}_{101},$$

where $\mathcal{E}_{11}$ contains only the clusters ending with 11, and $\mathcal{E}_{101}$ only the clusters ending with 101.

For a set of words $\mathcal{A}$ and a word b, we let $\mathcal{A} \cdot b = \{a \cdot b : a \in \mathcal{A}\}$. Applying a similar method that we used to derive Eq. 1.1 we obtain:

$$\begin{aligned} \mathcal{E}_{11} &= \{11\} \uplus \mathcal{E}_{11} \cdot 1 \uplus \mathcal{E}_{101} \cdot 1, \\ \mathcal{E}_{101} &= \{101\} \uplus \mathcal{E}_{11} \cdot 01 \uplus \mathcal{E}_{101} \cdot 01. \end{aligned}$$

It is translated into a system of functional equations as follows:

$$\begin{cases} E_{11}(x,u_1,u_2) = x^2u_1 + E_{11}(x,u_1,u_2) \cdot xu_1 + E_{101}(x,u_1,u_2) \cdot xu_1, \\ E_{101}(x,u_1,u_2) = x^3u_2 + E_{11}(x,u_1,u_2) \cdot x^2u_2 + E_{101}(x,u_1,u_2) \cdot x^2u_2. \end{cases}$$

Omitting some parentheses, we can write this system in matrix form:

$$\begin{pmatrix} E_{11} & E_{101} \end{pmatrix} = \begin{pmatrix} x^2u_1 & x^3u_2 \end{pmatrix} + \begin{pmatrix} E_{11} & E_{101} \end{pmatrix} \begin{pmatrix} xu_1 & x^2u_2 \\ xu_1 & x^2u_2 \end{pmatrix}$$

or

$$\begin{pmatrix} E_{11} & E_{101} \end{pmatrix} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} xu_1 & x^2u_2 \\ xu_1 & x^2u_2 \end{pmatrix} \right) = \begin{pmatrix} x^2u_1 & x^3u_2 \end{pmatrix}. \tag{1.10}$$

Solving this system we obtain:

$$E_{11} = \frac{u_1x^2}{1 - u_1x - u_2x^2}, \quad E_{101} = \frac{u_2x^3}{1 - u_1x - u_2x^2},$$

$$C_{\{11,101\}}(x,u_1,u_2) = E_{11} + E_{101} = \frac{u_1x^2 + u_2x^3}{1 - u_1x - u_2x^2},$$

$$B_{\{11,101\}}(x,y_1,y_2) = \frac{1}{1-2x-C_{\{11,101\}}(x,y_1-1,y_2-1)}.$$

We derived the generating function $B_{\{11,101\}}(x,y_1,y_2)$, where y_1 corresponds to 11, y_2 to 101. Now it is fairly easy to obtain the generating function $F(x,w) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty} f_{n,k}x^n w^k$ for n-size words avoiding pattern 11 and having k occurrences of 101:

$$F(x,w) = B_{\{11,101\}}(x,0,w).$$

Table 1.2 shows some values extracted from this function by Taylor expansion. As a matter of fact, we can also retrieve the generating function $F(x)$, see Eq. (1.2), as $F(x) = B_{\{11,101\}}(x,0,1)$. To enumerate words avoiding 11 and 101 (at the same time) we use $B_{\{11,101\}}(x,0,0)$.

k \ n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	2	3	4	6	9	13	19	28	41	60
1			1	2	3	6	11	18	30	50
2					1	2	3	7	14	24
3							1	2	3	8
4									1	2

Table 1.2: Number of binary words of size n avoiding 11 and having exactly k occurrences of 101.

The matrix $\begin{pmatrix} xu_1 & x^2u_2 \\ xu_1 & x^2u_2 \end{pmatrix}$ is called *correlation matrix* for patterns 11 and 101, it encodes all possible overlappings of these patterns. In general case we have several patterns: μ_1, μ_2, μ_3, ... The corresponding correlation matrix is

$$\begin{pmatrix} u_1\cdot R_{11} & u_2\cdot R_{12} & u_3\cdot R_{13} & \dots \\ u_1\cdot R_{21} & u_2\cdot R_{22} & u_3\cdot R_{23} & \dots \\ u_1\cdot R_{31} & u_2\cdot R_{32} & u_3\cdot R_{33} & \dots \\ \vdots & \vdots & \ddots & \dots \end{pmatrix},$$

where R_{ij} is a polynomial representation of possible overlappings of the pattern μ_i with the pattern μ_j (order is important). In particular, $R_{ii} = A_{\mu_i}(x) - 1$.

If we consider patterns 11, 100 and 001, the possible overlappings are

$$\begin{pmatrix} \begin{array}{l}\texttt{11}\\ \texttt{\ 11}\end{array} & \begin{array}{l}\texttt{11}\\ \texttt{\ 100}\end{array} & \emptyset \\ \\ \emptyset & \emptyset & \begin{array}{l}\texttt{100}\\ \texttt{\ 001}\end{array} \quad \begin{array}{l}\texttt{100}\\ \texttt{\ \ 001}\end{array} \\ \\ \begin{array}{l}\texttt{001}\\ \texttt{\ \ 11}\end{array} & \begin{array}{l}\texttt{001}\\ \texttt{\ \ 100}\end{array} & \emptyset \end{pmatrix},$$

and the correlation matrix is

$$\begin{pmatrix} u_1 x & u_2 x^2 & 0 \\ 0 & 0 & u_3(x + x^2) \\ u_1 x & u_2 x^2 & 0 \end{pmatrix},$$

Denote by I the identity matrix, by M the correlation matrix, by T the vector of starting patterns, and by E the vector of unknown generating functions. The general form of Eq. (1.10) is

$$E(I - M) = T, \tag{1.11}$$

and thus we have

$$E = T(I - M)^{-1}.$$

Notice that the matrix $(I - M)$ is always invertible. To compute the function C_Γ, the essential element of Eq. (1.9), we sum all entries of the vector E.

Chapter 2

Endhered patterns in RNA

To attain knowledge, add things every day.
To attain wisdom, remove things every day.
— Lǎo Zǐ, Dào Dé Jīng
Translated by Raymond B. Blakney

INSPIRED by Hampikian-Andersen work [138] on nullomers (absent short factors) in DNA sequences and an enthralling number of achievements related to patterns in the combinatorial literature, I want to present here some results on both theoretical and real-world data exploration aspects of *endhered (end-adhered) patterns* in RNA-related discrete structures.

An endhered pattern is a subset of arcs in a matching, such that the corresponding starting points are consecutive and the same holds for the ending points. Figure 2.1 shows a hopefully intuitive visual interpretation of a matching with occurrences of certain endhered patterns. The formal definitions will be presented to the reader a few pages later.

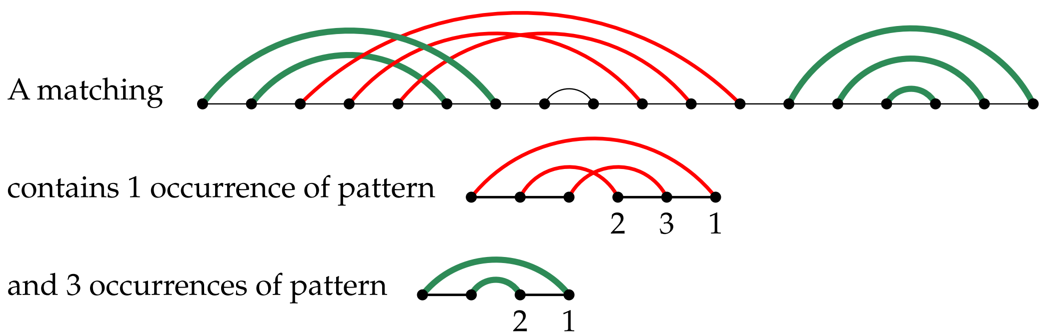


Figure 2.1: Endhered patterns in a matching.

This chapter is mostly based on the recently published paper co-authored with Célia Biane, Greg Hampikian and Khaydar Nurligareev [62]. I have added some figures and additional explanations. The chapter also includes several results from an upcoming paper with Khaydar Nurligareev. Another part of this work was presented in SeqBIM 2024, this part was developed together with Daniel Pinson [63], during his internship at our lab. We examine the distribution of endhered patterns in matchings and real-world RNA structures (with pseudoknots) derived from Protein Data Bank [55] using x3DNA-DSSR [181, 182][1] and FR3D-python [201, 220][2]. In the theoretical setting, we sometimes use ad hoc combinatorics, and the Goulden-Jackson cluster method allows us to solve the general case: derive the distribution of any given endhered pattern(s). I will omit some proofs and details, adding some thoughts about possible continuations of this work.

I would like also to express my gratitude to Matteo Cervetti, Yann Ponty and Eric Rivals for fruitful discussions and to Justin Masson for his help with Python code.

2.1 RNA secondary structures

Ribonucleic acids (RNAs) are macromolecules fulfilling many biological functions: they code for protein, are involved in the regulation of gene expression, can have catalytic activities and store the genetic information of certain viruses. The simplified structure of RNAs is defined at the primary level as sequences of four nucleotides: Adenine (A), Uracil (U), Guanine (G), and Cytosine (C). Note that if we want to have a more detailed description of the primary structure of RNA we need more letters, since several hundred nucleotide modifications have already been discovered in nature.

The secondary structure, studied in this chapter, abstracts from the nature of the nucleotide and considers only the bonds forming between nucleotides during the synthesis of RNAs and shaping how the molecules folds in space. Two types of bonds are formed during the RNA folding process: phospho-diester bonds (known as strong bonds) are formed between pairs of consecutive nucleotides in the sequence forming the RNA chain, and hydrogen bonds (also known as weak bonds) are formed between pairs of nucleotides distant in the sequence. The secondary structure represents an intermediate level between the primary sequence and the shape, and has the advantage of being both relevant from a biological perspective and tractable from a computational point of view.

[1] See https://x3dna.org/.

[2] Available at https://github.com/BGSU-RNA/fr3d-python.

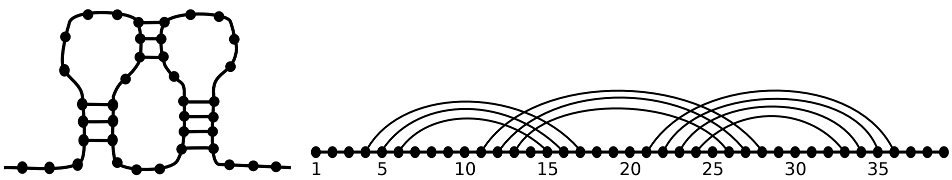
RNA secondary structures have been formalized as graphs primarily by Waterman [239]. Ponty [176] gives a variant of Waterman definition of RNA secondary structure without pseudoknots and having a minimal number θ of unpaired positions between pair positions. Formally, *Waterman-Ponty RNA secondary structure* S of size n is defined as a set of base-pairs $(i, j), 1 \leqslant i < j \leqslant n$, such that:

1. Each position is monogamous, $\forall (i, j) \neq (i', j') \in S : \{i, j\} \cap \{i', j'\} = \emptyset$.

2. Minimal distance θ between paired nucleotides, $\forall (i, j) \in S : j - i \geqslant \theta$.

3. No pseudoknot allowed, $\forall (i, j), (i', j') \in S, i < i' : (j' < j)$ or $(j < i')$.

These structures can be represented using the *dot-bracket notation*. A secondary structure of an n-nucleotide RNA is encoded as an n-length sequence of parentheses "(",")" and dots ".", where an open parenthesis corresponds to a nucleotide paired to another nucleotide represented by a closed parenthesis, and dots correspond to unpaired nucleotides. *The extended dot-bracket notation* includes also other types of parentheses: "[]", "{}", "<>", "aA", etc. The extended dot-bracket notation enables the representation of pairings in pseudoknots. Figure 2.2 shows 4 different representations of an example of RNA secondary structure.



(a) (b)

...(((....[[[.)))...((((.]]]....))))... (4,17),(5,16),(6,15),(11,28),(12,27)(13,26),(21,36),(22,35),(23,34),(24,33)

(c) (d)

Figure 2.2: A drawing (a), an arc diagram (b), an extended dot bracket notation (c), and a set of pairs (d) representing an example of an RNA secondary structure.

Other models of RNA secondary structures were studied from combinatorial point of view. Haslinger and Standler [140] examined enumerative and asymptotic properties of bi-secondary structures, i.e. arc diagrams with arcs both in upper and lower part of the plane but without arc intersections. The concept of RNA shape was introduced by Giegerich, Voss, and

Rehmsmeier [125] who developed an algorithm for the computation of minimum free energy RNA shapes. The number and asymptotics of RNA shapes was studied by Lorenz, Ponty, and Clote [177]. Reidys and Wang [210] considered a generalization of RNA shapes defined on k mutually-crossing arcs.

We look at RNA structures with no restrictions on the number of arc crossings.

2.2 Patterns in RNA secondary structures

At the biological level, common patterns have been observed in secondary structures: single strand regions, hairpin and internal loops, bulges and various computational tools exist for detecting these patterns from primary sequence, including the work of Macke *et al.* [185] More and more information is being gathered from three-dimensional reconstructions of RNA molecules paving the way to a better comprehension of the laws governing the RNA folding process and the formation of RNA patterns.

Rødland [218] proposed a classification of RNA secondary structures in four level of abstractions: nucleotide, ladder, stem and collapsed level, based on the considered internal patterns. In his work, the nucleotide level corresponded to structures with arc diagrams containing unpaired nucleotides, the ladder level corresponded to structures abstracted from unpaired nucleotides, the stem level was abstracted from bulges and internal loops, and the collapsed level was abstracted from nested loops. Rødland studied different kind of pseudoknot patterns of increasing complexity: H-pseudoknot, double hairpin pseudoknot and pseudotrefoil. He counted these patterns in RNA secondary structures of increasing complexity and studied their asymptotics. He also showed that the theoretical number of pseudoknots in secondary structures is higher than in real secondary structure of the Rfam [132, 133] and PseudoBase [235] databases. Note that Rødland's collapsed structures correspond to RNA shapes studied by Giegerich, Voss, and Rehmsmeier [125].

Quadrini [208] addressed the problem of searching a given structural pattern, defined as a sequence of crossing loops in a RNA secondary structure or shape and characterized by arbitrary number of pseudoknots. She proposed polynomial time algorithms for their identification. A paper by Gan, Pasquali, and Schlick [116] studied RNA structures and their patterns using graph-based representations. In the future, it would be interesting to compare the relationships between different kinds of patterns.

2.3 Endhered patterns in matchings

2.3.1 Basic definitions

By *matching* of size n, we mean the sequence of $2n$ points $(1, 2, 3, \ldots, 2n)$ endowed with a set of n arcs, such that every point is linked to one and only one another point. Figure 2.3a shows examples of matchings of small sizes.

A *permutation* of size n is a bijection from the set $\{1, 2, \ldots, n\}$ to itself. It can be represented as a sequence $\pi = \pi_1 \pi_2 \ldots \pi_n$, meaning that i is sended to π_i. The set of all size-n permutations equipped with their composition form a *symmetric group* S_n. A *transposition* of two elements, written as (ab), sends a to b and vice-versa. See books by Bóna [76], Kitaev [164] and Stanley [225, 227] for more details about the classical notations and results.

Matchings of size n can be considered as fixed-point-free involutions (self-inverses) in the symmetric group S_{2n}. Thus, the matching with 4 arcs at the bottom of Figure 2.3a can be represented by the permutation 3 6 1 5 4 2 8 7, which is the product of disjoint transpositions $(13)(26)(45)(78)$.

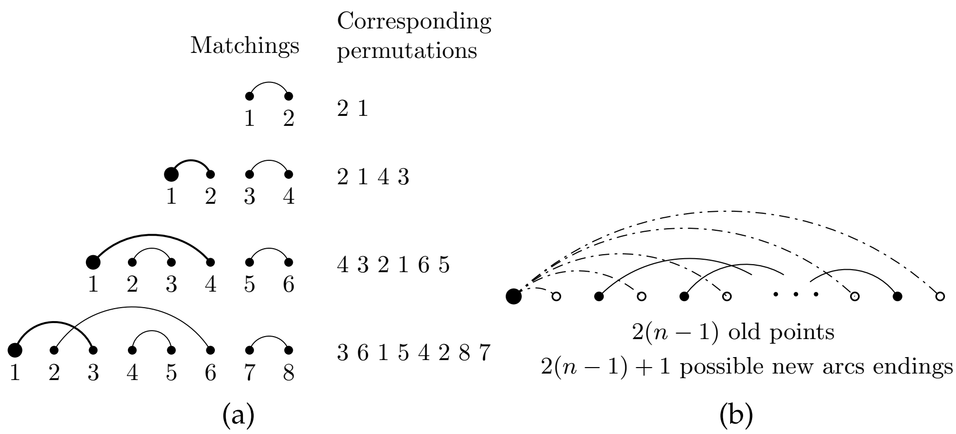


Figure 2.3: An example of a matching construction, corresponding permutations (a) and a schema of recursive construction of matchings (b).

For a positive integer n, any matching of size n can be uniquely constructed from some matching of size $n-1$ by the following procedure. We add a new arc starting at the left of the already existing $2(n-1)$ points and ending at some of $2(n-1)+1$ possible new positions, see Figures 2.3b and 2.4. This observation leads us to the following recurrence relation for the number of matchings of size n: $a_n = (2n-1)a_{n-1}$ with $a_0 = 1$. As a consequence, $a_n = (2n-1)!! = (2n-1) \cdot (2n-3) \cdot \ldots \cdot 3 \cdot 1$. The cor-

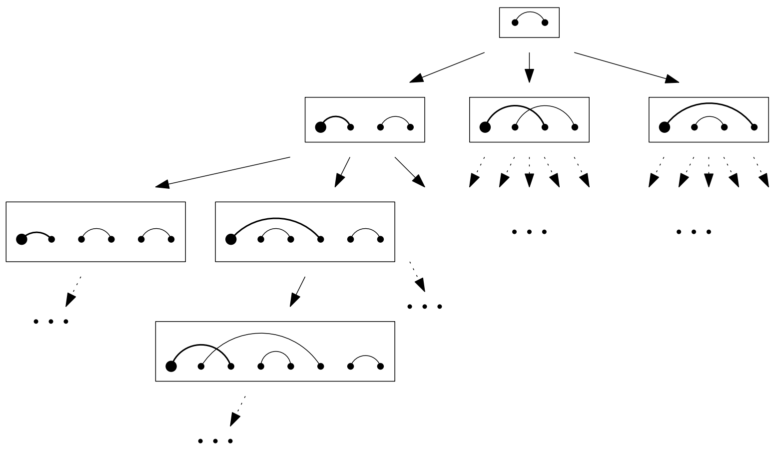


Figure 2.4: First levels of a recursive construction of matchings.

responding sequence starting with $1, 1, 3, 15, 105, 945, 10395$ is known as A1147 in Sloane's Encyclopedia [200].

In mathematical literature, matchings often appear in different contexts, from the representation theory of Lie algebras [77] to the geometry of moduli spaces of flat connections on surfaces [3]. Efficient generating algorithms for involutions with (without) fixed points were established by Vajnovszki [234]. Ardnt's book [4] also presents interesting generating algorithms for involutions and other combinatorial structures.

Different kinds of patterns in matchings are being actively studied in combinatorics for the past twenty years. Initially, the interest to this topic came from the rapidly developing study of permutation patterns, since a matching can be thought as a permutation of a specific form. From this perspective, we say that a matching σ is a *pattern* in a matching μ if σ can be obtained from μ by deleting some of its arcs (and consistently relabelling the remaining vertices). For instance, Chen, Deng, Du, Stanley, and Yan [82] studied distributions of crossings and nestings. Jelínek [151], as well as Bloom and Elizalde [66], considered pattern avoiding matchings in the case when σ is a permutational matching of size 3. An extension for more general patterns was elaborated by Cervetti and Ferrari [80], while other authors, such as Chen, Mansour and Yan [83], Jelínek and Mansour [152], considered partial patterns.

As we see, what matters in the above investigations is the relative positions of arcs that form a pattern. At the same time, the distances between starting and ending points of these arcs are not fixed. The main object of our study concerns specific restrictions imposed on the arcs. Namely, the starting points of a pattern, as well as its ending points, form an interval, while the distance between these two intervals may vary. We call such patterns *endhered* (end-adhered) to emphasize the nature of these restrictions.

Definition 3. An *endhered pattern* is a matching, such that the starting point of any of its arcs precedes the ending point of any other arc. In other words, a matching of size p written as a permutation $\sigma = \sigma_1 \dots \sigma_{2p}$ is an endhered pattern if $\pi = \sigma_{p+1} \dots \sigma_{2p}$ is a permutation of size p (such matchings are also called *permutational*). Figure 2.5a presents an example of an endhered pattern of size 3. In the following, we identify endhered patterns with the corresponding permutations.

We say that a matching $\mu = \mu_1 \dots \mu_{2n}$ *contains* an endhered pattern $\pi = \pi_1 \dots \pi_p$ at position $(i+1, j+1)$, where $i \geqslant 0$ and $i + p \leqslant j \leqslant n - p$, if

$$\mu_{s+i} = \pi_s^{-1} + j, \qquad s = 1, \dots, p$$

(here, $\pi^{-1} = \pi_1^{-1} \dots \pi_p^{-1}$ is the inverse to the permutation π). In other words, μ contains p arcs whose starting points are $i+1, \dots, i+p$, whose ending points are $j+1, \dots, j+p$, and that form the endhered pattern π. Figure 2.5b shows an example of a matching containing the pattern shown on Figure 2.5a.

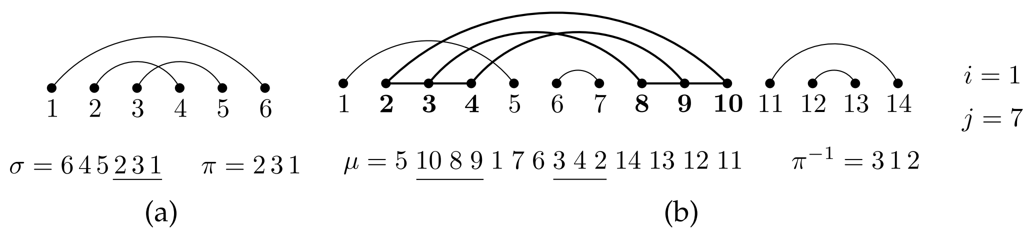


Figure 2.5: Endhered pattern 231 (a) and an example (b) of its occurrence.

The endhered patterns have rarely been studied in the literature, although they may shed light on the formation of collapsed RNA structures from Rødland's paper about pseudoknots [218]. The only work in this direction that we are aware of is the paper of Baril [25] who examined one of two endhered patterns of size 2 in his study on irreducible involutions and permutations.

2.3.2 Endhered twists

Given an endhered pattern π, let us denote by $a_{n,k}(\pi)$ the number of matchings of size n with k occurrences of π. If, additionally, τ is another endhered pattern, then we designate by $a_{n,k,m}(\pi, \tau)$ the number of matchings of size n with k and m occurrences of patterns π and τ, respectively. Certain patterns, for instance $\pi =$  and $\tau =$ , have the same distribution, meaning that $a_{n,k}(\pi) = a_{n,k}(\tau)$. Here we establish such equidistributed classes of endhered patterns with the help of bijections, i.e. without direct enumeration. To this end, we apply matching transformations that we call *endhered twists*.

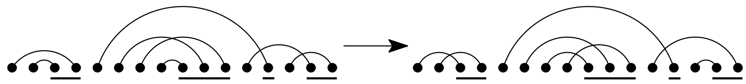


Figure 2.6: An example of the right endhered twist, runs of right points are underlined.

Definition 4. The *left endhered twist* (resp. *right endhered twist*) is a transformation that takes a matching μ and produces a matching $\texttt{letw}(\mu)$ (resp. $\texttt{retw}(\mu)$) such that all runs of consecutive starting (resp. ending) points are reversed. Figure 2.6 shows an example of the right twist.

Endhered twists of endhered patterns correspond to classical symmetries on permutations. Thus, the right endhered twist applied to an endhered pattern $\pi = \pi_1 \ldots \pi_p$ is its reverse: $\texttt{retw}(\pi) = \pi_p \ldots \pi_1$. At the same time, the left endhered twist is the complement: $\texttt{letw}(\pi) = (p + 1 - \pi_1) \ldots (p + 1 - \pi_p)$. For example,

$$\begin{aligned}
&\texttt{retw}(321) = 123, \qquad \texttt{retw}(231) = 132, \qquad \texttt{retw}(312) = 213, \\
&\texttt{letw}(321) = 123, \qquad \texttt{letw}(132) = 312, \qquad \texttt{letw}(213) = 231.
\end{aligned}$$

Geometrically, if we represent permutations as square tables, the endhered twists are axial symmetries (see Figure 2.7).

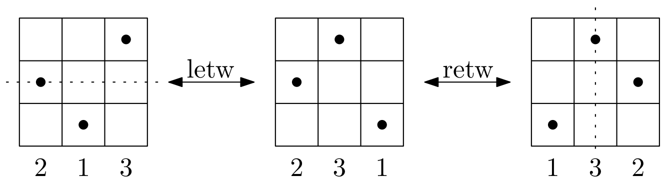


Figure 2.7: Geometrical meaning of endhered twists.

Lemma 5. *Two endhered patterns π and τ of the same size have the same joint distribution if they are identical under the left or right endhered twists. In other words, if $\pi = \texttt{letw}(\tau)$ or $\pi = \texttt{retw}(\tau)$, then*

$$a_{n,k,m}(\pi, \tau) = a_{n,k,m}(\tau, \pi) \tag{2.1}$$

for any integers n, k and m. In particular, $a_{n,k}(\pi) = a_{n,k}(\tau)$.

Proof. Suppose that π is obtained from τ by applying the left endhered twist. Let μ be a matching containing k occurrences of π and m occurrences of τ. Applying the left twist to the whole μ, we transform every occurrence of π to τ and vice versa. No other occurrence of π or τ are created. This implies relation (2.1). The cases of the right twist is obtained *mutatis mutandis*. □

Given a positive integer n and an endhered pattern π, the value $a_{n,0}(\pi)$ is the number of matchings of size n avoiding π. We say that two endhered patterns π and τ are *Wilf equivalent* if $a_{n,0}(\pi) = a_{n,0}(\tau)$ for every n. For example, Wilf equivalent patterns of size 2 and 3 are shown in Figure 2.8. Matchings avoiding the pattern  correspond to involutions with no fixed points and no successions considered by Baril [25].

For the endhered patterns of size 2 and 3, as we will see, their Wilf equivalence implies that the corresponding patterns are equidistributed. In other words, if $a_{n,0}(\pi) = a_{n,0}(\tau)$ for every n, then $a_{n,k}(\pi) = a_{n,k}(\tau)$ for all n and k. In the general case, this question is not trivial.

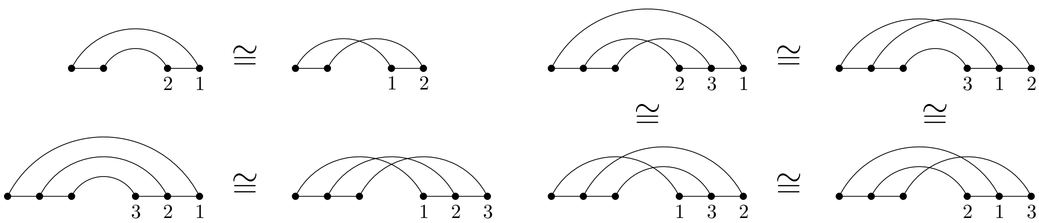


Figure 2.8: Wilf equivalence classes of endhered patterns of size 2 and 3.

Corollary 6. *The joint distribution of the endhered patterns* $\pi = 1 \dots p$ *and* $\tau = p \dots 1$ *is symmetric. In particular, the number of size-n matchings containing k*  *and m*  *equals the number of size-n matchings containing m*  *and k* .

Corollary 6 is consistent with the result of Chen, Deng, Du, Stanley, and Yan [82] who showed that, in the case of classical patterns, the joint distribution of crossings and nestings is symmetric.

2.3.3 Endhered patterns of size 2

It follows from Lemma 5 that patterns  and  have the same distribution, so we need only to enumerate one of them. Skipping the proofs, I cite here a few results from our paper [62]. For simplicity, let $a_{n,k}$ stand for $a_{n,k}(21)$.

Theorem 7. *For $n \geqslant 1$, the number of size-n matchings containing exactly k occurrences of pattern*  *(resp.* *) satisfies*

$$\begin{cases} a_{n+1,k} &= a_{n,k-1} + 2(n-k) \cdot a_{n,k} + 2(k+1) \cdot a_{n,k+1} \\ a_{1,0} &= 1 \\ a_{1,k} &= 0, \quad k \neq 0. \end{cases}$$

Formally, we allow negative values of k. It follows that $a_{n,k} = 0$ whenever $k < 0$.

Table 2.1 shows first values of $a_{n,k}$. With this table and those that follow in the text, we can see the big picture of the initial terms, intuitively understand how they grow, formulate hypotheses about the global (asymptotic) behavior of these numerical sequences and prove them [62].

k \ n	1	2	3	4	5	6	7	8	9	OEIS
0	1	2	10	68	604	6584	85048	1269680	21505552	A165968
1	0	1	4	30	272	3020	39504	595336	10157440	A179540
2	0	0	1	6	60	680	9060	138264	2381344	
3	0	0	0	1	8	100	1360	21140	368704	
4	0	0	0	0	1	10	150	2380	42280	
5	0	0	0	0	0	1	12	210	3808	
6	0	0	0	0	0	0	1	14	280	
7	0	0	0	0	0	0	0	1	16	
8	0	0	0	0	0	0	0	0	1	

Table 2.1: Distribution of pattern  (21).

For any $n > k \geqslant 0$, we have

$$a_{n,k} = \binom{n-1}{k} a_{n-k,0} \quad \text{and} \quad a_{n+1,0} = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} (2k+1)!!,$$

where $m!!$ denotes the *double factorial* of m,

$$m!! = \begin{cases} m(m-2)(m-4)\cdots 4 \cdot 2 & \text{if } m \text{ is even,} \\ m(m-2)(m-4)\cdots 3 \cdot 1 & \text{if } m \text{ is odd.} \end{cases}$$

The exponential generating function $B(z,u)$ defined as

$$B(z,u) := \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_{n+1,k} \frac{z^n}{n!} u^k,$$

satisfies

$$B(z,u) = \frac{e^{z(u-1)}}{\sqrt{(1-2z)^3}}.$$

The exponential generating function of the shifted k-th row of Table 2.1 is

$$[u^k]B(z,u) = \frac{z^k}{k!} \cdot \frac{e^{-z}}{\sqrt{(1-2z)^3}}.$$

Theorem 8. *The asymptotic behavior of the numbers of matchings with k occurrences of pattern*  *(resp.* *), as $n \to \infty$, is*

$$a_{n,k} \sim \frac{1}{2^k k!} \left(\frac{2}{e}\right)^{n+1/2} n^n.$$

Corollary 9. *The limit distribution of pattern*  *(resp.* *) in a uniform random matching of size n follows asymptotically a Poisson law with parameter 1/2.*

Corollary 10. *The ratio of numbers from the k-th row to the numbers of the $(k+1)$-th row of Table 2.1 tends to $2(k+1)$. In other words, for any $k \in \mathbb{N}$,*

$$\frac{a_{n,k}}{a_{n,k+1}} \sim 2(k+1).$$

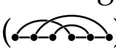
2.3.4 Patterns of size 3 and more

There are six endhered patterns of size 3. They are divided into two equivalence classes with respect to distributions of these patterns in matchings (see Fig. 2.8). We use the following notations:

$$c_{n,k} := a_{n,k}(321) = a_{n,k}(123)$$

to denote the number of matchings of size n containing k patterns 321 (), and

$$d_{n,k} := a_{n,k}(132) = a_{n,k}(213) = a_{n,k}(231) = a_{n,k}(312)$$

to designate the number of matchings of size n containing k patterns 132 (). Tables 2.2 and 2.3 show the first values.

In [62] we proved the following theorem.

k \ n	1	2	3	4	5	6	7	8	9
0	1	3	14	100	906	10022	130864	1969884	33583700
1	0	0	1	4	34	332	3866	52400	811248
2	0	0	0	1	4	36	362	4304	59256
3	0	0	0	0	1	4	38	392	4752
4	0	0	0	0	0	1	4	40	422
5	0	0	0	0	0	0	1	4	42
6	0	0	0	0	0	0	0	1	4
7	0	0	0	0	0	0	0	0	1

Table 2.2: Distribution of pattern  (321).

k \ n	1	2	3	4	5	6	7	8	9
0	1	3	14	99	900	9978	130455	1965285	33522915
1	0	0	1	6	45	414	4635	61110	927090
2	0	0	0	0	0	3	45	630	9405
3	0	0	0	0	0	0	0	0	15

Table 2.3: Distribution of pattern  (132).

Theorem 11. *For any $n > 0$ and $k \geqslant 0$,*

$$c_{n,0} = \sum_{s=0}^{\lfloor n/2 \rfloor} \binom{n-s}{s} a_{n-s,0},$$
$$c_{n,k} = \sum_{s=1}^{\lfloor (n-k)/2 \rfloor} \binom{k+s-1}{k} \binom{n-k-s}{s} a_{n-k-s,0} \quad \text{if } k > 0,$$

where $a_{n,0}$ is the number of matchings of size n avoiding pattern 21 ().

Theorem 11 is nice and even beautiful in a certain sense. But, ad hoc combinatorics gets tricky at this point. So, let me show by examples how to use the Goulden-Jackson method and the symbolic inclusion-exclusion principle in the case of endhered patterns.

We start by considering all possible matchings. The corresponding generating function is given by

$$F(z) = \sum_{n=0}^{\infty} (2n-1)!!\, z^n = 1 + z + 3z^2 + 15z^3 + 105z^4 + 945z^5 + \ldots$$

In every matching, we distinguish certain arcs, say, by coloring them violet. Algebraically, this is done by passing to the generating function

$$G(z,v) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} g_{n,k} z^n v^k = F(z+zv),$$

where $g_{n,k}$ is the number of matchings of size n with k violet arcs (marked by the variable v).

Take, for instance, a pattern . We replace violet arcs by "thick arcs" forming an occurrence of pattern  (see Figures 2.9 and 2.10). In general case, we should replace them by clusters consisting of intersected copies of a given pattern; pattern , however, admits no possible self-intersection. The corresponding generating function is

$$H(z,v) = G(z, z^2 v).$$

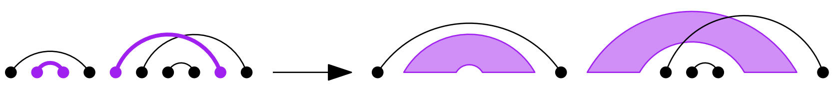


Figure 2.9: Schematic replacement of distinguished edges with patterns of interest.

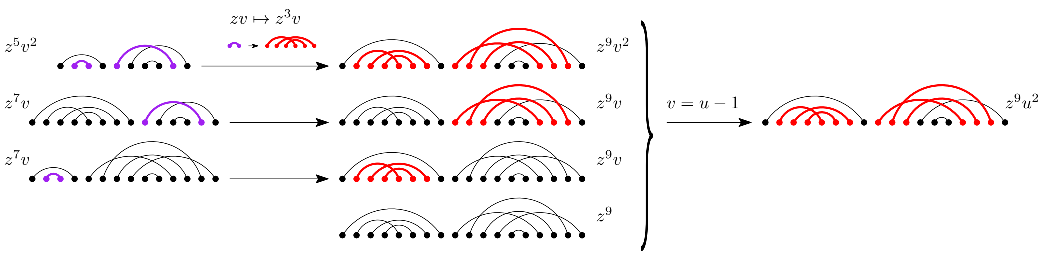


Figure 2.10: Cluster method and the inclusion-exclusion principle for pattern .

And finally, according to the inclusion-exclusion principle, we obtain the generating function

$$D(z,u) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} d_{n,k} z^n u^k = H(z, u-1)$$

whose coefficients $d_{n,k}$ enumerate the matchings of size n having k occurrences of the pattern . Figure 2.10 illustrates a part of this process. The following result takes place.

Theorem 12. *The bivariate generating function*

$$D(z,u) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} d_{n,k} z^n u^k,$$

where $d_{n,k}$ is the number of matchings of size n containing k patterns  *can be expressed as*

$$D(z,u) = F\big(z+(u-1)z^3\big) = \sum_{n=0}^{\infty}(2n-1)!!\,\big(z+(u-1)z^3\big)^n.$$

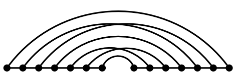
To obtain bivariate generating functions for patterns with possible self-intersections we need to generate clusters of them. And thus, we need to define autocorrelation polynomials in the context of endhered patterns.

Definition 13. An *autocorrelation polynomial* $A_\pi(z)$ for an endhered pattern π of size n is defined as

$$A_\pi(z) = 1 + \sum_{k\in S} z^{n-k},$$

where S is the set of sizes of possible intersections of two different occurrences of the pattern π in some matching. In other words, z^{n-k} means that two occurrences have k edges in common.

Some examples of autocorrelation polynomials:

Pattern	Autocorrelation	
	$A_{21}(z)$	$= 1 + z,$
	$A_{12}(z)$	$= 1 + z,$
	$A_{132}(z)$	$= 1,$
	$A_{321}(z)$	$= 1 + z + z^2,$
	$A_{3412}(z)$	$= 1 + z^2,$
	$A_{7564231}(z)$	$= 1 + z^3 + z^6.$

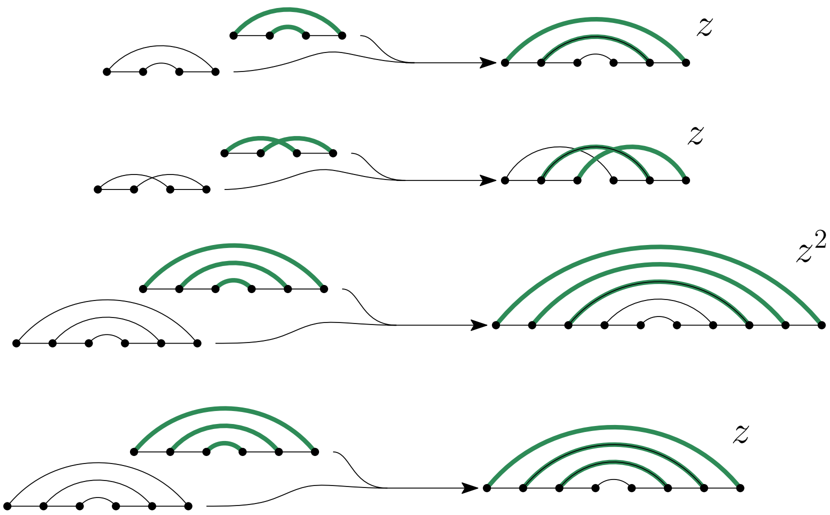


Figure 2.11: Examples of endhered pattern overlappings and corresponding terms of autocorrelation polynomials.

Certain overlappings are shown on Figure 2.11.

Clusters for the pattern  (21) looks like stacks of edges: we start by one occurrence of , then we add a sequence of m edges on top of it, creating m new occurrences of the pattern . The generating function for these clusters is $\frac{z^2}{1-vz}$. Replacing distinguished edges by a cluster, we calculate $G\left(z, \frac{vz}{1-vz}\right)$, and the inclusion-exclusion process goes like clockwork:

$$H_{21}(z, v) = G\left(z, \frac{vz}{1-vz}\right),$$
$$D_{21}(z, u) = H_{21}(z, u-1).$$

For  we have:

$$H_{321}(z, v) = G\left(z, \frac{vz^2}{1-v(z+z^2)}\right),$$
$$D_{321}(z, u) = H_{321}(z, u-1),$$

and Figure 2.12 illustrate the inclusion-exclusion principle in this case.

In general, for the pattern π of size ℓ, the generating function for clusters is $\frac{vz^\ell}{1-v(A_\pi(z)-1)}$, and we should replace any vz in $G(z, v)$ by $\frac{pz^\ell}{1-p(A_\pi(z)-1)}$, so we have the following result.

Theorem 14. *For a given endhered pattern π of size ℓ with autocorrelation polynomial $A_\pi(z)$, we have*

$$H_\pi(z,v) = G\left(z, \frac{vz^{\ell-1}}{1-v(A_\pi(z)-1)}\right),$$

$$D_\pi(z,u) = H_\pi(z,u-1) = F\left(z + \frac{(u-1)z^{\ell}}{1-(u-1)(A_\pi(z)-1)}\right) =$$

$$= \sum_{n=0}^{\infty}(2n-1)!!\left(z + \frac{(u-1)z^{\ell}}{1-(u-1)(A_\pi(z)-1)}\right)^n.$$

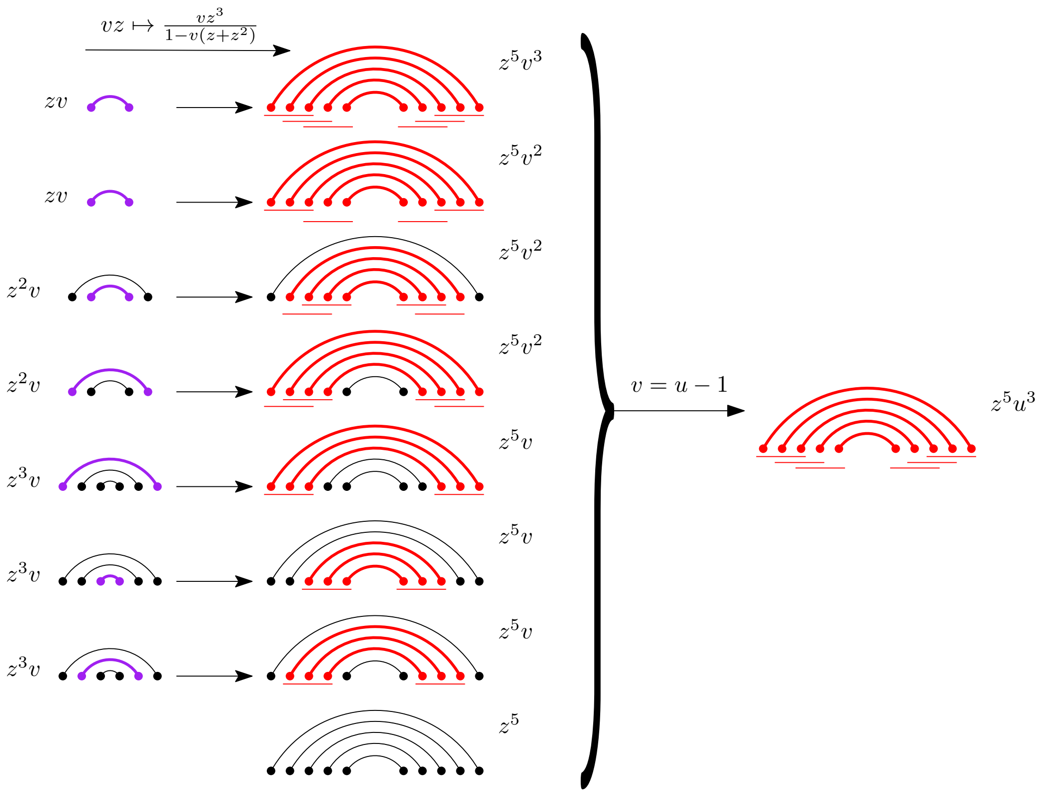


Figure 2.12: Clusters and the inclusion-exclusion principle for pattern . Replace every violet edge by a cluster of intersecting occurrences of the pattern.

For a pattern 7564231, depicted on Figure 2.13, the autocorrelation polynomial is $A(7564231) = 1 + z^3 + z^6$ and the first terms of its distribution are shown in Table 2.4. From Theorem 14 we obtain the following corollary.

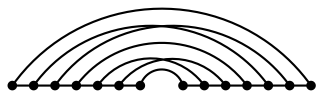


Figure 2.13: Endhered pattern 7564231.

k \ n	1	2	3	4	5	6	7	8	9	10
0	1	3	14	105	945	10395	135134	2027019	34459380	654728656
1	0	0	0	0	0	0	1	6	45	418
2	0	0	0	0	0	0	0	0	0	1

Table 2.4: Distribution of pattern 7564231.

Corollary 15. *The distribution of pattern the π depends only on its size and its autocorrelation polynomial. Two patterns, of the same size, having the same autocorrelation polynomial will have the same distribution.*

In order to generalize Lemma 5, we consider two patterns π and σ, such that neither of the two is contained in the other, and examine the joint distribution

$$D_{\pi,\sigma}(z,u,w) = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty} d_{n,k,m} z^n u^k w^m$$

whose coefficient $d_{n,k,m}$ is the number of matchings of size n containing k occurrences of π and m occurrences of σ.

Again, we replace violet edges in matchings by clusters of patterns and apply the inclusion-exclusion principle. The clusters in the two-pattern case look slightly different, now they are formed from π and σ, may contain both of them, and we need two variables p and s to distinguished π and σ. An occurrence of π can overlap with an occurrence of σ, so, in addition to autocorrelation polynomials $A_\pi(z)$ and $A_\sigma(z)$, we need some way to describe possible overlappings between these two patterns.

Definition 16. A *correlation polynomial* $A_{\pi,\sigma}(z)$ for two different endhered patterns π and σ, such that neither of the two is contained in the other, is defined as

$$A_{\pi,\sigma}(z) = \sum_{k\in S} z^{n-k},$$

where S is the set of sizes of possible intersections of π with σ. The order of patterns matters, we represent them as permutations to determine which patterns comes first in the cluster, see Figure 2.14 for visual explanations. In general, $A_{\pi,\sigma}(z) \neq A_{\sigma,\pi}(z)$.

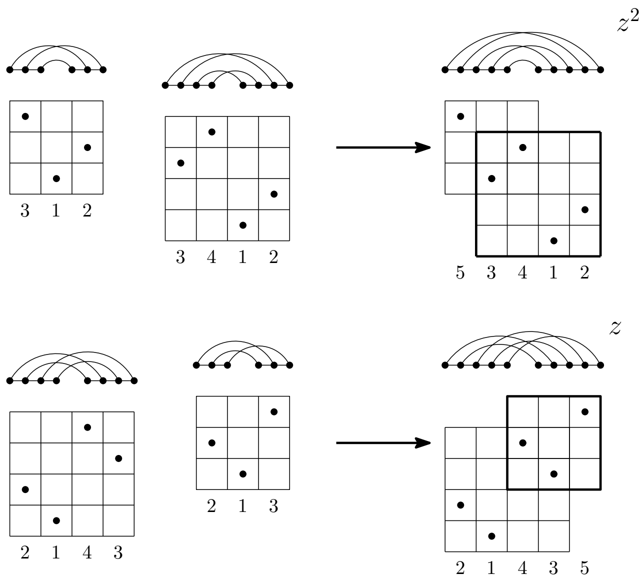
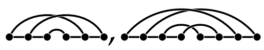
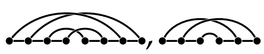


Figure 2.14: Order of overlapped endhered patterns in clusters and corresponding terms of autocorrelation polynomials.

Some examples of correlation polynomial:

Pattern	Correlation
	$A_{12,21}(z) \quad = 0,$
	$A_{21,12}(z) \quad = 0,$
	$A_{132,213}(z) \quad = z,$
	$A_{213,132}(z) \quad = z^2,$
	$A_{312,231}(z) \quad = z,$
	$A_{231,312}(z) \quad = z^2,$
	$A_{312,3412}(z) \quad = z^2,$
	$A_{3412,312}(z) \quad = 0.$

A cluster either ends with π or with σ, the corresponding generating functions are $C_\pi(z, p, s)$ and $C_\sigma(z, p, s)$, where z marks the size of cluster, p marks the labelled occurrences of π and s marks the labelled occurrences of σ. A cluster ending with π can be of three forms:

- constructed from a cluster ending with π by appending another occurrence of π according to the autocorrelation polynomial $A_\pi(z)$;
- constructed from a cluster ending with σ by appending a new occurrence of π according to the correlation polynomial $A_{\sigma,\pi}(z)$;

- contains only the occurrence of π, so we have the term $z^{|\pi|} \cdot p$, where $|\pi|$ denotes the size of π.

Clusters ending with σ are constructed in a similar manner. In order to find C_π and C_σ we need to solve the following system of functional equations (we don't write parentheses for functions on purpose to make it more compact):

$$\begin{cases} C_\pi = (A_\pi - 1) \cdot C_\pi \cdot p + A_{\sigma,\pi} \cdot C_\sigma \cdot p + z^{|\pi|} \cdot p, \\ C_\sigma = (A_\sigma - 1) \cdot C_\sigma \cdot s + A_{\pi,\sigma} \cdot C_\pi \cdot s + z^{|\sigma|} \cdot s. \end{cases}$$

Computing $C(z,p,s) = C_\pi(z,p,s) + C_\sigma(z,p,s)$ we obtain:

$$C(z,p,s) = \frac{ps \cdot \left(z^{|\pi|}(A_{\pi,\sigma} - A_\sigma + 1) + z^{|\sigma|}(A_{\sigma,\pi} - A_\pi + 1) \right) + pz^{|\pi|} + sz^{|\sigma|}}{ps \cdot \left(A_\pi A_\sigma - A_\pi - A_\sigma - A_{\pi,\sigma} A_{\sigma,\pi} + 1 \right) - A_\pi p - A_\sigma s + p + s + 1}.$$

And finally, we have:

$$H_{\pi,\sigma}(z,p,s) = G\left(z, \frac{C(z,p,s)}{z}\right),$$
$$D_{\pi,\sigma}(x,u,w) = H_{\pi,\sigma}(z, u-1, w-1).$$

Note that, if $|\pi| = |\sigma|$ and $A_\pi = A_\sigma$, the function $C(x,p,q)$ is symmetric under the change $p \leftrightarrow s$, so we obtain the following result, which can be regarded as a stronger version of Lemma 5.

Theorem 17. *Two patterns π and σ of the same size have the symmetric joint distribution if they have the same autocorrelation polynomial.*

A permutation that corresponds to endhered pattern whose occurrences cannot share any edges is called *strongly non-overlapping*. In a recent work with Khaydar Nurligareev [163] we show, among other things, that almost all permutations are strongly non-overlapping. Therefore, endhered patterns also almost always have no possible self-intersections, so the great majority of patterns of the same size have the same distribution, and almost always two patterns of the same size have the symmetric joint distribution.

With Khaydar Nurligareev we are currently working on a paper about exact and asymptotic distribution of endhered patterns in a general case. I'll show you some results here, without proofs.

For instance, the limit distributions of patterns  and  are different. In the case of , for a fixed $k \geqslant 0$, as $n \to \infty$, we have

$$c_{n,k} \sim \sqrt{2} \cdot C_k \cdot \frac{(2n)^{n-k}}{e^n},$$

where

$$C_0 = 1 \qquad \text{and} \qquad C_k = \sum_{s=1}^{k} \binom{k-1}{s-1} \frac{1}{2^s s!} \quad \text{if } k > 0.$$

On the other hand, for  we have

$$d_{n,k} \sim \frac{2^{n-2k+1/2}}{k! e^n} \cdot n^{n-k},$$

as $n \to \infty$. In other words, for large values of n, we have

$$\frac{c_{n,k}}{(2n-1)!!} \sim \frac{C_k}{2^k} \cdot \frac{1}{n^k} \qquad \text{and} \qquad \frac{d_{n,k}}{(2n-1)!!} \sim \frac{1}{2^{2k} k!} \cdot \frac{1}{n^k},$$

meaning that a large uniform random matching avoids both patterns with high probability.

2.4 Endhered patterns in real RNA

In paper [62] we also have conducted a study of native (real-world) RNA data and discuss how endhered patterns are distributed in various representations of the secondary RNA structures. These structures were obtained from Protein Data Bank [55][3] using our Python scripts[4]. The scripts depend on GEMMI [244][5] to parse mmCIF files from PDB and use two methods to detect base pairs in RNA molecules: a closed-source software x3DNA-DSSR [181, 182][6] and an open-source software FR3D-python [201, 220][7].

The software x3DNA-DSSR directly produces extended dot-bracket notations. To derive these notations x3DNA-DSSR uses only *canonical pairs*: A-U, C-G, wobble G-U, and A-T (in RNA-DNA hybrids) with cis. Watson-Crick/Watson-Crick interactions and without forming parallel mini-duplexes. FR3D-python gives a list of base pairs. We parse its results, filter canonical base pairs *à la* x3DNA-DSSR, and produce extended dot-bracket notations using a simple First-Come-First-Served method.

There are several hundred known, existing in nature, modifications of nucleotides. In the data they are denoted by special one-, two-, or three-letter codes, different from the classical 4 letters UACG. Modified nucleotides are mapped to short 1-letter nucleotide names. For example A23 is mapped A, EQ0 to G, and CCC to C. The Nucleic Acid KnowledgeBase [54, 173][8] contains details for these mappings. Following x3DNA-DSSR[9] we adapted one exception to these rules: pseudouridine (PSU) is mapped to P and not to U. This adaptation allows us to better compare the x3DNA-DSSR and FR3D-python results.

From the RNA secondary structures obtained with x3NA-DSSR and FR3D-python, we collapsed unpaired nucleotides in order to keep only paired ones. When keeping only RNA structures composed of one chain, this leads to 1501 RNA structures, in 933 (resp. 929) of them x3DNA-DSSR (resp. FR3D-python) has found at least one canonical base pair. The data has been accessed on August 29, 2024.

The structures in the extended dot-bracket notation are converted to matchings using an algorithm based on stacks. The algorithm works as follows: the word composed of parentheses is read from left to right, when

[3] PDB, https://www.rcsb.org
[4] Our scripts are available at https://gitlab.com/celiabiane/endhered_pattern
[5] https://github.com/project-gemmi/gemmi
[6] https://x3dna.org/
[7] https://github.com/BGSU-RNA/fr3d-python
[8] https://www.nakb.org/modifiednt.html
[9] See http://forum.x3dna.org/rna-structures/modified-nucleotides-incorrect

an opening character is met its index is stacked in a stack corresponding to the nature of the character. When a closing character is met, a pair corresponding to the last index in the corresponding stack and the current index is added to the matching, and the index of the opening parenthesis is unstacked. Figure 2.15 shows an example of this conversion.

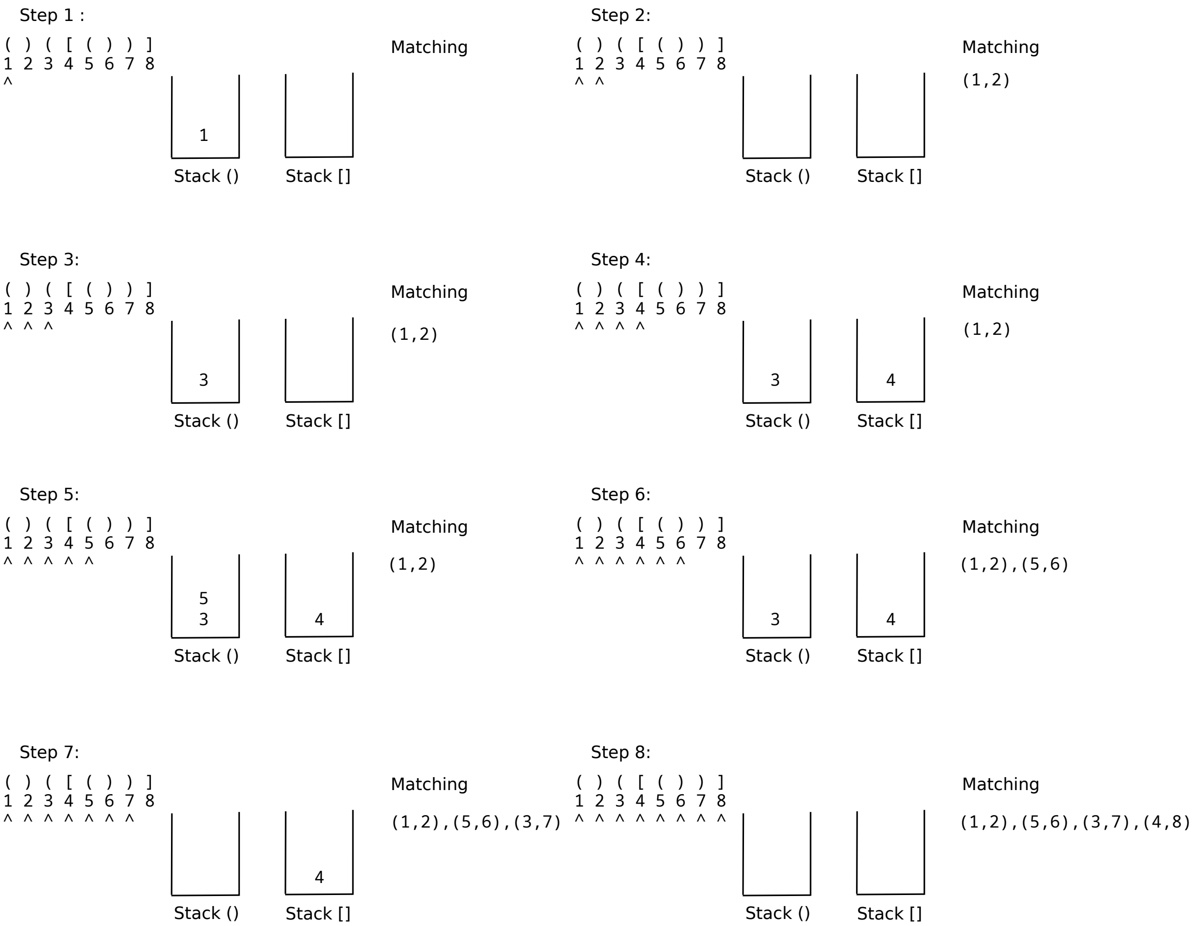


Figure 2.15: Example of conversion from dot-bracket notation to a matching.

The distribution of patterns  and  in RNA secondary structure is shown in Figure 2.16. We can observe that the majority of RNA secondary structures are of small size (between 0 and 100 nucleotides), and there seems to be a linear relation between the size of the matching and the number of occurrence of the pattern. This can be explained by the large number of hairpin patterns in RNA secondary structures.

Both x3DNA-DSSR and FR3D-python detect 2 RNA structures containing the  pattern: 4M4O, 5U3G. One of them, 4M4O, corresponds to

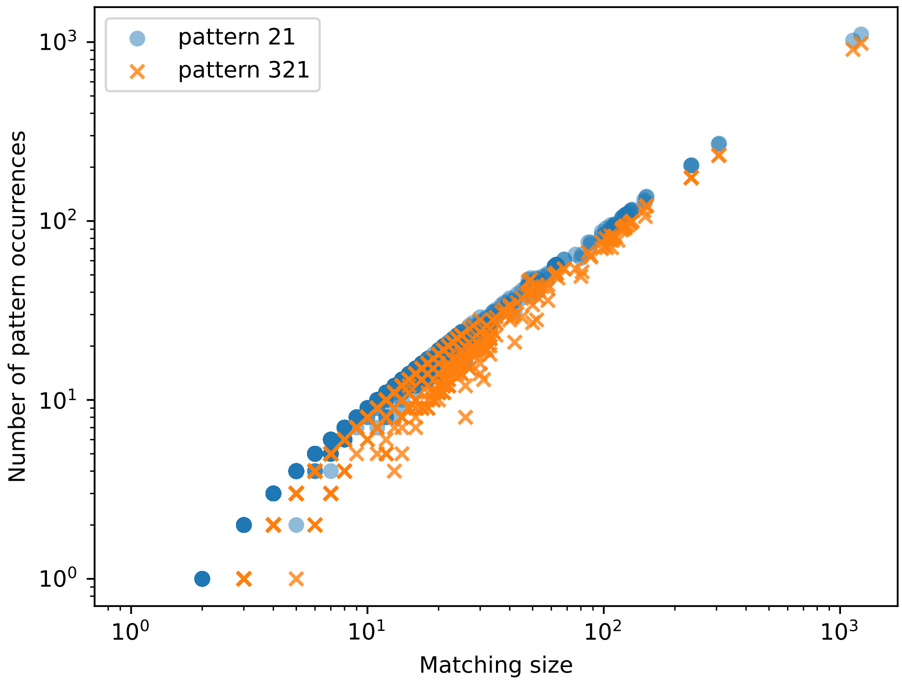


Figure 2.16: Scatter plot of the number of occurrences of patterns 21 and 321 as a function of the size of the matching. Axes are in logarithmic scale. RNA with no occurrences of patterns 21 and 321 are not displayed. Matchings are obtained using FR3D-python, the results for x3DNA-DSSR are similar and available on git repository `https://gitlab.com/celiabiane/endhered_pattern`.

the minE aptamer involved in a complex with a lysozyme. Both methods, based on x3DNA-DSSR and FR3D-python, give the following secondary structure:

((((((.(((((.((.......((((.(.(.[..)]..).)))))).))))).))))))

Another one, 5U3G, is the Dickeya dadantii ykkC riboswitch, which has the following secondary structure:

x3DNA-DSSR: ((((((..((...(((((....))))).))....))))))(((((((....))))))).(.[...(((.......))).)..]..
FR3D-python: .(((((..((...(((((....))))).))....))))).(((((((....))))))).(.[...(((.......))).)..]..

Diving into the data, we see that FR3D-python does not detect GTP-C pair on position (1,40) in this case.

The results of two methods differ in the case of molecule 7K16 (Tamana Bat Virus xrRNA1):

x3DNA-DSSR: {{...((((((((....))))))((((...[[[...)))))}}.))...]]]
FR3D-python: .(...[[((((((....))))))((((...{{{...)))))..]]...}}}

In this case x3DNA-DSSR detects, in addition to FR3D-python, a 5GP-C base pair on position (1,42). Removing this pair, we create a  pattern. Table 2.5 presents RNA secondary structures and RNA shapes (explained below) that contain at least one occurrence of a given pattern.

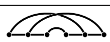
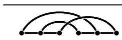
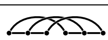
Pattern	RNA secondary structure		Extended RNA shape	
	x3DNA-DSSR	FR3D-python	x3DNA-DSSR	FR3D-python
	926	921	0	0
	2 (4M4O, 5U3G)	3 (4M4O, 5U3G, 8SH5)	59 (1A60, 1E95, 1KAJ, 1KPD, 1KPY, 1KPZ, 1L2X, 1L3D, 1RNK, 1YG3, 1YG4, 1YMO, 2A43, 2AP0, 2AP5, 2G1W, 2K95, 2K96, 2LC8, 2M58, 2M8K, 2RP0, 2RP1, 2TPK, 2XDB, 3IYQ, 3IYR, 3IZ4, 3U4M, 3U56, 3UMY, 437D, 4M4O, 4PQV, 4QG3, 4QVI, **4R8I,** 5KMZ, 5NPM, 5TPY, 5U3G, 6AGB, 6AHR, 6DLQ, 6DLR, 6DLS, 6DLT, 6DNR, 6E1T, 6E1V, 6VUH, 7K16, 7U4A, 7UO5, 8HIO, 8T29, 8T2A, 8T2B, 8T2O)	60 (1A60, 1E95, 1KAJ, 1KPD, 1KPY, 1KPZ, 1L2X, 1L3D, 1RNK, 1YG3, 1YG4, 1YMO, 2A43, 2AP0, 2AP5, 2G1W, 2K95, 2K96, 2LC8, 2M58, 2M8K, 2RP0, 2RP1, 2TPK, 2XDB, 3IYQ, 3IYR, 3IZ4, 3U4M, 3U56, 3UMY, 437D, 4M4O, 4PQV, 4QG3, 4QVI, 5KMZ, 5NPM, 5TPY, 5U3G, 6AGB, 6AHR, **6D3P,** 6DLQ, 6DLR, 6DLS, 6DLT, 6DNR, 6E1T, 6E1V, 6VUH, 7K16, 7U4A, 7UO5, 8HIO, **8SH5,** 8T29, 8T2A, 8T2B, 8T2O)
	1 (4M4O)	1 (4M4O)	8 (3U4M, 3U56, 3UMY, 4M4O, 4QG3, 4QVI, **4R8I**, 5NPM)	7 (3U4M, 3U56, 3UMY, 4M4O, 4QG3, 4QVI, 5NPM)
	1 (5U3G)	1 (5U3G)	13 (3U4M, 3U56, 3UMY, 4QG3, 4QVI, 5KMZ, 5NPM, 5U3G, 6DLQ, 6DLR, 6DLS, 6DLT, 6DNR)	13 (3U4M, 3U56, 3UMY, 4QG3, 4QVI, 5KMZ, 5NPM, 5U3G, 6DLQ, 6DLR, 6DLS, 6DLT, 6DNR)
	0	1 (7K16)	0	0
	918	908	0	0
 	0	0	0	0

Table 2.5: Biological RNA secondary structures (obtained from Protein Data Bank) and RNA shapes in which a given pattern occurs at least once. Differences are highlighted in bold.

It is surprising that pattern  appears so rarely in RNA structures while pseudoknots are thought to have important biological functions. We observe that occurrence of this pattern are "hidden" by the high frequency of nested bonds (see Figure 2.17). To neutralize this effect, we pass to RNA shapes.

Figure 2.17: Two examples of matchings with nesting bonds (occurrences of ) but without occurrences of .

The concept of RNA shapes have been introduced by Giegerich, Voss, and Rehmsmeier in 2004 [125]. In these shapes, no unpaired regions are included and nested bonds are combined. For instance, the secondary structure

..(((.((..(((....))).(((.....))))))))..

has the following RNA shape:

(()())

Originally, the RNA shapes have been defined in words with a single type of parenthesis. They are counted by Motzkin numbers [86, 93, 125] and correspond exactly to non-crossing matchings avoiding the endhered pattern .

We adapted the Giegerich-Voss-Rehmsmeier reduction to matchings with crossings represented by words with different types of parentheses. The result of this adaptation is what Rødland called *collapsed structures* [218]. This is done by keeping only (i, j) pairs in matching such that $(i+1, j-1)$ does not belong to the matching and then reindexing the pairs. The number of  patterns in resulting reduced matchings (RNA shapes) is obviously 0. Interestingly, the number of RNA shapes with at least one occurrence of pattern  increases up to 59 (60 with FR3D-python). Among those, 8 (7 with FR3D-python) RNA structures have pattern  and 13 have pattern . Other size 3 patterns (, , , ) are not detected in RNA shapes. It is expected for patterns , , and , as they contain the pattern  forbidden in RNA shapes.

We have considered only endhered patterns of size 2 and 3. But our approach can also be applied to longer patterns in native (real-world) data, as the function `count_pattern` in `count_visualisation.py` from our git repository can be applied to any endhered pattern.

Figure 2.18: https://rna.kirgizov.link - An interactive web application devoted to endhered patterns in real-world RNA structures (with pseudo-knots) derived from Protein Data Bank.

With Daniel Pinson [63] we have developed an interactive web application[10] dedicated to illustrate some features of the endhered pattern distributions in real-world RNA data that comes from Protein Data Bank. See Figure 2.18. This application allows us to explore distributions of longer endhered patterns, as well as other patterns corresponding to consecutive subwords in the extended dot-bracket notation.

2.5 Discussions and perspectives

Together with my colleagues [62], we have examined distributions of endhered patterns in matchings from theoretico-combinatorial and data-driven points of view. In matchings, patterns  and  have the same distribution. Six endhered patterns of size 3 are divided into 2 equidistributed groups: ,  and , , , . Moreover, the joint distribution of patterns of the same size is symmetric if the patterns are equivalent under the twist operation. We have also provided corresponding asymptotic behavior of these distributions for patterns of size 3 and 2. We are currently working on asymptotics in the general case,

[10]Available at https://rna.kirgizov.link.

adapting Borinsky's approach [69].

We have deliberately abstracted from the nucleic acid sequences, and modelled the secondary structures directly by matchings. Our results show that there is a big difference between observed and modelled pattern distributions. This means that non-restricted matchings are too permissive and new models should be developed to get closer to the observed pattern distributions. We wonder if it is possible to describe essential features of RNA secondary structures with pseudoknots, using pattern-based restrictions. The classical Waterman's definition of RNA structures and its generalizations (see Section 2.1) can be regarded as an example of pattern-based restriction used, among other things, to control the prediction algorithms for secondary structures from nucleic acid sequences. More insights into patterns frequencies in the native secondary RNA structures may guide us towards mixture of two definitions, complex enough to cover different pseudoknot-like structures presented in real data, but at the same time quite simple and neat to prevent the uncontrolled combinatorial explosion.

The non-existence of certain short sequences in genomic and protein sequences is a well-known fact [138, 143, 168, 204]. Applications include cancer research [1, 195] and forensic science [128]. Less is known about the forbidden secondary structures, although some interesting works about theoretical (im)possibility of inverse RNA folding have been published recently [247, 249]. One of the following research directions would be to determine what influence the distribution of endhered patterns in the native RNA. Some configurations are probably forbidden due to physicochemical constraints on the bending of RNA (something like Waterman-Ponty restrictions, but for the case that include pseudoknots), others may not be present because of biological reasons.

Are there some evolutionary mechanisms that divert the distribution of patterns from the theoretically observed in the equi-probabilistic model of matchings? How pattern distributions in secondary structures are related to RNA dynamics and function?

For any new combinatorial characterization of RNA structures, we need to develop a method for estimating their affinity with structures observed in native molecules. Perhaps one of such methods can be pattern-based: compare the distribution of patterns in native RNA with theoretically calculated distributions over matchings avoiding certain patterns. A possible way to extend Daniel Pinson's web application [63] might be the integration of such a method. Our exploratory study suggest, for instance, that the patterns  and  never appear in RNA.

Moreover, PDB references presented in Table 2.5 look impressive, especially the minE aptamer involved in a complex with a lysozyme, 4M4O; the Dickeya dadantii ykkC riboswitch, 5U3G; and Tamana Bat Virus xrRNA1, 7K16. It would be interesting to work together with biologists studying these structures.

Chapter 3

Patterns in other structures

霰まじる帷子雪は小紋かな
— 松尾 芭蕉

arare majiru
katabira-yuki wa
komon kana
— Matsuo Bashō

the hail on snow
and the patterned cloth
would you remind me...
to adore them both?
Free translation
by Kerzol and Sharkman

Structures that contain patterns of a given form are often beautiful. We can also observe amazing things if we forbid some patterns from arising while allowing all possible other patterns to appear. Completely unconstrained structures are fascinating in their own way, they allow us to understand what a random object looks like, what a white noise might look like.

One would say that by forbidding certain patterns or allowing only patterns from a specific set we create new structures, and that absolute randomness does not exist because it always depends on the space of possibilities, which in turn is determined by the description of structures. This line of thought raises questions about what kinds of structures can be characterized by patterns and what kinds cannot. For instance, words that do not contain factors from a given set are enumerated by rational generating functions and can be described by a regular grammar. However, there are regular grammars that are not described by a set of forbidden factors. Think for example of a grammar for binary words that contain only an even number of 1s. To describe such words, we should adapt another, more global, notion of patterns.

Transcending the science boundaries, we enter the world of art. Literature, and art in general, are tenderly attached to patterns. Let's look at some examples of constrained writing, a literary technique very common in poetry but also existing in prose:

- **Haiku**, a traditional Japanese poetry consisting of three phrases with prescribed numbers of morae[1] 5/7/5, and respecting other, more semantic, rules. Matsuo Bashō (XVII century) is known as the greatest haiku master. Haiku in those days were called hokku. Nowadays, such short poems exist in many languages, very different from Japanese. Poets no longer adhere strictly to original moraic patterns. Anyway, in non-Japanese languages it is not possible to follow 5/7/5-rule completely due to different rhythmic structures. However, the very essence and spirit of haiku are preserved even now.

- **Acrostic**, a poem (or other text composition) in which the first letters of each line (or each word) form a meaningful word or phrase. It is believed that acrostic was invented by Epicharmus Comicus of Syracuse (V century BC). Encrypting his name with the first letters, he signed the poems. Acrostics are common in Middle Ages and Renaissance, they can be found in modern times. The reader has probably seen such texts himself.

- **Lipogram**, a text that completely avoids some letters. The 300-page novel "La Disparition" by Georges Perec (XX century) contains no letter *e*, the most frequent letter in French. Perec also wrote "Les Revenentes" where every vowel is *e*. Ancient Greek poet Lasus of Hermione (VI century BC) did not like σ, so he wrote some poems avoiding it. Pindar and others followed his steps. The name of this poetic phenomenon is *asigmatism* in contrast to *sigmatism*, the sometimes excessive use of the letter σ in ancient Greek poetry, the letter which is already one of the most popular consonants in Greek. See works of Clayman [90], Scott [221] and Porter [206] for more details about the (a)sigmatism.

[1] *Mora* (plural *morae*) is an elementary word particle equal to a syllable or part of it in languages with long vowel and consonant sounds.

Occasionally writers and artists, who use certain pattern-based techniques, get together in groups. For example, in 1960 Raymond Queneau and François Le Lionnais founded Oulipo, a now famous group of writers and mathematicians. Oulipo members use constrained writing techniques to explore the potential of language. This movement became quite popular and other groups of artists chose similar names: Oupeinpo (visual arts), Oubapo (comics), Ouphopo (photography) and Outrapo (theatre).

In this Chapter, I would like to tell you a couple of stories about structures discovered while playing a forbid & allow game with *patterns*, understanding this word in a quite broad sense.

3.1 Constrained lattice paths

The first story of this Chapter will be about an unexpected yet fruitful discovery of a quaint restriction of classic *Dyck paths*, i.e. lattice paths that never go below the x-axis, start at $(0,0)$, end at $(2n,0)$, and consist of up steps $U=(1,1)$ and down steps $D=(1,-1)$.

One day, I was invited to present our joint work, with Vincent Vajnovszki and Jean-Luc Baril, about patterns in treeshelves [38] in Séminaire Algorithmique of GREYC team in Caen. On the evening of February 28, 2017, after the seminar, on the train back from Caen to Dijon via Paris, I was thinking about how to encode graphs (modulo isomorphism) of certain structure by Dyck paths. Halfway to that goal, and halfway to Dijon, I found a very interesting subclass of Dyck paths counted by Motzkin's numbers!

Definition 18. Denote by $\mathcal{D}^{h,\geqslant}$ the *subclass of Dyck paths constrained by height*. Any Dyck path from this subclass is either empty or can be represented as $U\alpha D\beta$ where α and β are two dyck paths from $\mathcal{D}^{h,\geqslant}$, while respecting $h(U\alpha D) \geqslant h(\beta)$, where h is a height of the path, i.e. the maximal ordinate reached by the path.

Figure 3.1 illustrates this height-based restriction. Note that it can be generalized by allowing a different operation instead of $\geqslant$ and a different kind of statistic instead of height.

This idea and subsequent discussions with Jean-Luc Baril and Armen Petrossian gave rise to our article entitled "Dyck paths with a first return decomposition constrained by height" that was published in 2018 in the journal *Discrete Mathematics* [35]. In 2019, we have extended our results to Motzkin paths [36]. In 2020, together with Richard Genestier, we provided [28] generating functions for the popularity and the distribution of

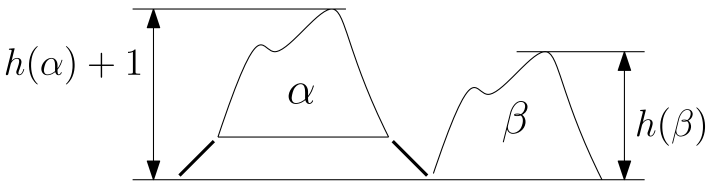


Figure 3.1: A Dyck path with a first return decomposition constrained by height $h(\alpha) + 1 \geqslant h(\beta)$

patterns of length at most three the set of such Dyck paths. We considered other variants of this restriction, in [31] the height is redefined with respect to a given pattern, and a general result is proved with applications to Dyck, Motzkin, skew Dyck, and skew Motzkin paths. Rigoberto Flórez, Leandro Junes and José L. Ramírez conducted a study of non-decreasing version of our paths in [111]. The original discovery was one of the main inspiration for my ANR JCJC project "Patterns in Combinatorics – PiCs"[2].

A little later, in 2023, together with Jean-Luc Baril, Vincent Vajnovszki and our PhD student Rémi Maréchal, we explored grand Dyck paths with air pockets [41] from various angles, including the above-mentioned height-dependent constraint. Rémi brilliantly defended his thesis in October 2024 [190].

In the following pages of this Section, I will recall the original problem related to graph community structures, restate one of the main bijective results, and talk about further possible directions in this area.

3.1.1 Mist of lattice paths and graph communities

Maximilien Danisch, a close friend of mine, who has sadly left this world several years ago, was working on his PhD thesis at the same time as me in Complex Networks[3] team of LiP6[4]. His thesis and the subsequent work were about community structures in graphs. Jean-Loup Guillaume and Bénédicte Le Grand were his PhD supervisors.

Community detection consists in finding groups of vertices more connected to each other than to vertices of other groups. There is a tremendous number of community detection methods: very popular Louvain [65]; clas-

[2]See https://anr.fr/Projet-ANR-22-CE48-0002

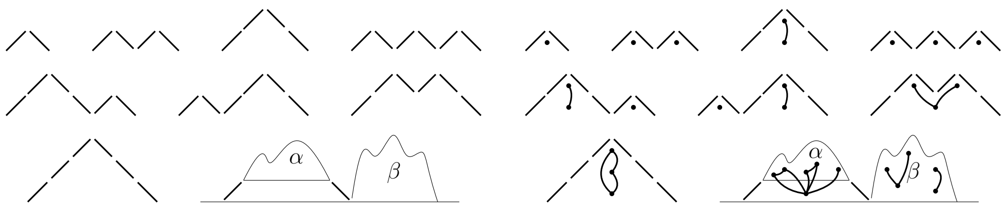
[3]See https://www.complexnetworks.fr/

[4]At the time of my thesis LiP6 was the laboratory of CNRS and Pierre and Marie Curie University (Paris VI). A bit later, Paris-Sorbonne University (Paris IV) and Paris VI merged to form Sorbonne University.

sical Girvan–Newman [127]; as well as a relatively new approach of Baudin, Danisch, Kirgizov, Magnien and Ghanem [46] involving a clique percolation method based on an efficient algorithm of k-clique listing (see a work [91] by Danisch, Balalau and Sozio). Fortunato [112], together with Hric [113], provided excellent surveys in this broad area.

There exist really a lot of different definitions and algorithms of community detection. One day I decided to add a coin to that pile. The idea was to somehow link Dyck paths and community structures. We know how Dyck paths are constructed. Any Dyck path is either empty, denoted by ε, or constructed from two Dyck paths α and β as $U\alpha D\beta$ (see Figure 3.2a).

We consider simple graphs, there are no multiple edges and no self-loops. A *clique* in a graph is a subset of vertices where all possible edges are present. Locally, cliques are the most connected subgraphs. The least connected parts of a graph have zero edges between them.



(a) (b)

Figure 3.2: (a) Dyck paths of small size and their general construction.
(b) Same, with associated graphs.

We construct Dyck paths by two operations: the *elevation* $U\alpha D$ and the *concatenation* $\alpha\beta$ where α and β are two smaller Dyck paths. Imagine, we have two graphs A and B, corresponding to paths α and β. It would be quite natural to think that the elevation of α corresponds to adding a vertex connected to all vertices of the graph A and the concatenation of α and β represents the union of two graphs with no new edges added. In this way, a path U^nD^n (power means repetition) corresponds to a complete graph K_n with n vertices and as many edges as possible, $n(n-1)/2$. The path $(UD)^n$ is just a set of n vertices without any edges. So it is possible to construct very connected graphs ("best communities"), their opposites and also some intermediate structures. Figure 3.2b shows some examples.

French TGV high-speed train attain 320 km/h. My thoughts aren't that quick, but even so, I realised that this construction has a few downsides. Not all graphs can be generated in this way, the graph  cannot. The savvy reader may notice that admissible graphs look a little woodsy, they can be bijectively mapped to trees. Furthermore, certain graphs are generated several times in different ways, see Figure 3.3.

Having accepted that not all graphs can be constructed, I tried to solve the non-uniqueness problem by allowing only Dyck paths from the set $\mathcal{D}^{h,\geqslant}$ (see Definition 18 and Figure 3.1). Any path of this set is either empty or have the first return decomposition $U\alpha D\beta$ that respects $h(U\alpha D) \geqslant h(\beta)$ where $\alpha, \beta \in \mathcal{D}^{h,\geqslant}$. With this restriction we keep the path 3.3a, while its reflection 3.3b is not allowed. But the paths 3.3c and 3.3d are still allowed and correspond to the same graph.

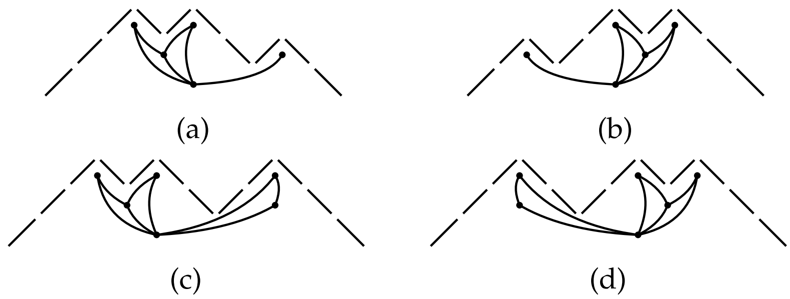


Figure 3.3: Generative non-uniqueness of some graphs.

3.1.2 Serpentine road between Dyck and Motzkin

The journey has not brought me any good definition of community structures in graphs. But when I tried to count what paths I got... a sequence starting just like the Motzkin[5] numbers: 1, 1, 2, 4, 9, 21. These numbers enumerate *Motzkin paths*, i.e. lattice paths from $(0,0)$ to $(n,0)$ that never go below the x-axis and consist of up steps $U = (1,1)$, flat steps $F = (1,0)$, and down steps $D = (1,-1)$.

Did I just found a new, previously undiscovered link between two classical lattice paths? I didn't even believe it right away. Next morning I rechecked the data using a computer and small values of n and showed the findings to my colleagues. Together we realised how to prove, present and extend these findings. In papers [31, 35, 36, 41, 111] the reader can find more results about related combinatorial objects, generating functions, recursively defined polynomials, continued fraction, transport of statistics, etc.

[5]Named after Theodore Samuel Motzkin. Many facts and conjectures about Motzkin numbers are gathered on entry number A1006 in Sloane's Encyclopedia.

Here I will only show the beautiful bijection φ, found with my co-authors Jean-Luc Baril and Armen Petrossian. For $P \in \mathcal{D}^{h,\geqslant}$, we set

$$\varphi(P) = \begin{cases} \varepsilon & \text{if } P = \varepsilon \quad \text{(empty)}, \\ \varphi(\alpha)F & \text{if } P = \alpha UD, \\ \varphi(\alpha)\varphi(\gamma)U\varphi(\beta)D & \text{if } P = \alpha UU\beta D\gamma D. \end{cases} \tag{3.1}$$

Denote by $\mathcal{D}_n^{h,\geqslant}$ the set of Dyck paths of length $2n$ from the set $\mathcal{D}^{h,\geqslant}$, and let $\mathcal{M}_n$ stand for the set of n-length Motzkin paths. Recall that $h(\alpha)$ is the maximal ordinate reached by the path α.

Theorem 19 (see Theorem 5 in [35]). *The map φ defined above is a bijection between $\mathcal{D}_n^{h,\geqslant}$ and $\mathcal{M}_n$. Moreover, for any $P \in \mathcal{D}_n^{h,\geqslant}$ we have $h(\varphi(P)) = \left\lfloor \frac{h(P)}{2} \right\rfloor$.*

Figures 3.4 and 3.5 illustrate how φ works.

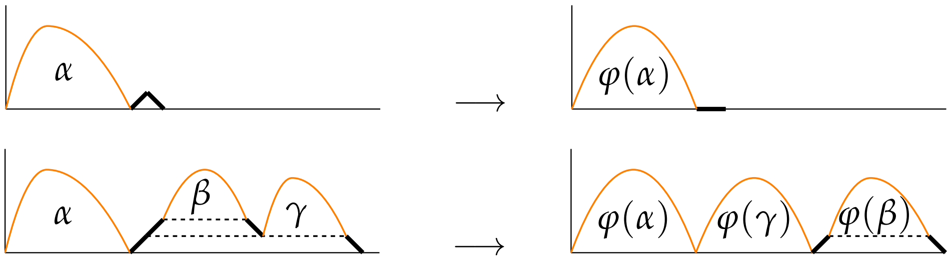


Figure 3.4: General form of the bijection between $\mathcal{D}_n^{h,\geqslant}$ and $\mathcal{M}_n$.

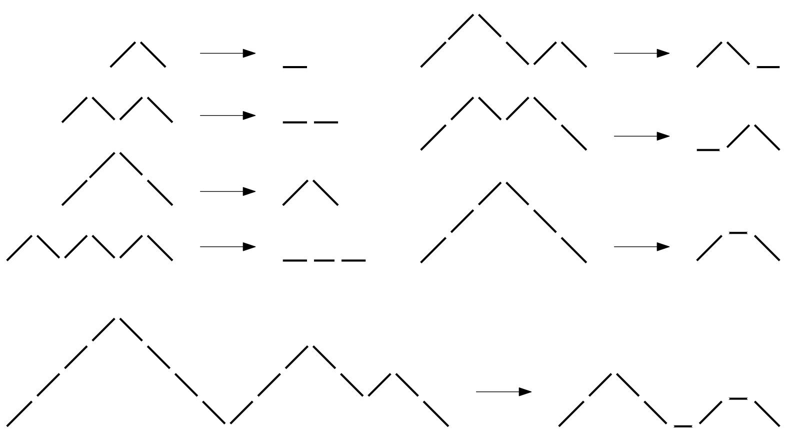


Figure 3.5: Examples of the bijection $\varphi : \mathcal{D}_n^{h,\geqslant} \to \mathcal{M}_n$.

3.1.3 Wonderland bijections and further directions

A Dyck path with air pockets, introduced in [32], is a nonempty lattice path in the first quadrant of $\mathbb{Z}^2$ starting at the origin, ending on the x-axis, and consisting of up-steps $U = (1, 1)$ and down-steps $D_k = (1, -k)$, $k \geqslant 1$, where two down steps cannot be consecutive. In [41, Section 8] we proved that the set of Dyck path with air pockets recursively restricted as in Definition 18 is equienumerated with the set of Motzkin paths avoiding consecutive patterns UF, FU and FF. Can we find a constructive bijection in this case?

Wonderland bijections are one-to-one correspondences that transform mirror-closed sets to sets lacking this mirror symmetry. In next Section, among other things, I will present another example of wonderland bijection for certain sets of pattern-constrained binary words. This bijection breaks the mirror symmetry and transforms a certain regular language into a non-regular one. Quite interesting, isn't it? It would be nice to find more examples of such correspondences.

3.2 Following the steps of Leonardo Pisano

1 2 3 5 8 13 21 ...

The Fibonacci sequence origins have been traced back to the works of ancient Indian mathematician Ācārya Piṅgala dealing with rhythmic structure patterns in Sanskrit poetry [166, p. 487], [223]. In Sanskrit and several other languages, we distinguish short and long syllables. A short syllable ⏑ is counted as one mora, and a long syllable ⏒ is counted as two morae. How many different poetic metres with n morae do we have? Fibonacci sequence gives the answer, see Figure 3.6. Appending a short or a long syllable, we feel the formula $a_n = a_{n-1} + a_{n-2}$.

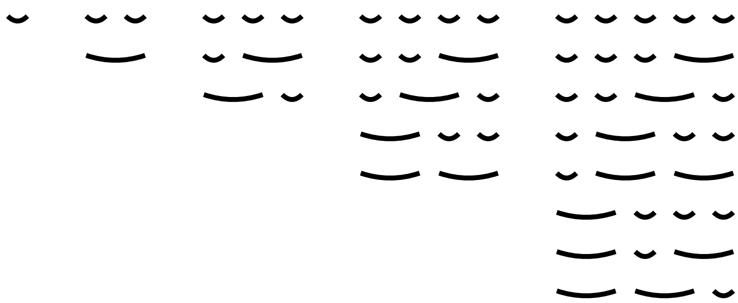


Figure 3.6: Sanskrit poetic metres with 1, 2, 3, 4 and 5 morae.

Tribonacci:	1	2	4	7	13	24	...
Tetranacci:	1	2	4	8	15	29	...
Pentanacci:	1	2	4	8	16	31	...

Multi-step generalization of Fibonacci numbers can be traced back to the works of Miles [193] and 14-year-old Feinberg [108]. A lot of different studies about these numbers appear after, including the works of Flores [110], Miller [194], Dubeau [99] and Wolfram [245]. A bunch of combinatorial objects are enumerated by these numbers. For instance, the Knuth's exercise [165, p. 286] shows that the set of length n binary words avoiding k consecutive 1s is enumerated by k-bonacci numbers respecting $a_n = a_{n-1} + a_{n-2} + \cdots + a_{n-k}$, with initial conditions $a_0 = 1, a_{-1} = 1$, and $a_j = 0$ for any $j < -1$.

In previous Section we constructed the set $\mathcal{D}^{h,\geqslant}$ of Dyck paths constrained by height. Every path of this set is either empty or have the first return decomposition $U\alpha D\beta$ respecting $h(U\alpha D) \geqslant h(\beta)$ with $\alpha, \beta \in \mathcal{D}^{h,\geqslant}$. Once upon a time, I wanted to prolong this definition or a similar restriction to other objects. I have tried various. The essential part of the construction is to compare values of some statistic (the height in case of Dyck paths) computed on two parts of the combinatorial object, and do the same for all substantial pairs of parts that constitute (recursively in case of Dyck paths) the object. Let's see how we can adapt such statistic comparison to words.

Any binary word can be viewed as a sequence of length-maximal occurrences of factors of the form $0^a1^b = \underbrace{000\cdots\cdots00}_{a}\underbrace{111\cdots11}_{b}$. The length-maximality of such occurrences should be taken in the local sense: they are not preceded by a 0 or followed by a 1. As it turns out the following definition yields amazing results.

Definition 20. A binary word is called *q-decreasing* if each its length-maximal occurrences of a factor of the form 0^a1^b, $a > 0$, satisfies $q \cdot a > b$.

Let $\mathcal{W}_{q,n}$ be the set of such words of length n and $\mathcal{W}_q = \bigcup_{n \in \mathbb{N}} \mathcal{W}_{q,n}$.

Examples

11100101011 is not 2-decreasing ($2 \cdot 2 > 1,\ 2 \cdot 1 > 1,\ 2 \cdot 1 \not> 2$)

01111 is not π-decreasing ($\pi \cdot 1 \not> 4$)

001111 is π-decreasing ($\pi \cdot 2 > 4$)

1101001 is φ-decreasing ($\varphi \cdot 1 > 1, \varphi \cdot 2 > 1$, where $\varphi = (1+\sqrt{5})/2$)

1100011000111111000 is not φ-decreasing ($\varphi \cdot 3 > 2, \varphi \cdot 3 \not> 5, \varphi \cdot 3 > 0$)

3.2.1 Gray code and Eğecioğlu-Iršič conjecture

In 2020, together with Jean-Luc Baril and Vincent Vajnovszki, we have been working on $\mathcal{W}_{q,n}$ where $q \in \mathbb{N}^+$. We showed constructively [40] that these words are in bijection with binary words having no occurrences of 1^{q+1}, and thus they are enumerated by the $(q+1)$-generalized Fibonacci numbers. We gave some enumerative results and reveal similarities between q-decreasing words and binary words having no occurrences of 1^{q+1} in terms of the frequency of 1-bits (see also [39]).

The *Hamming distance* between two binary words of the same length equals the number of positions at which they differ. A *k-Gray code* is an ordered list of words, such that the Hamming distance between any two consecutive words is at most k. George Robert Stibitz's and Frank Gray's patents [131, 229] discuss examples and applications of such a code for the set of n-length binary words. Three quarters of a century before Gray and Stibitz, Agathon-Louis Gros [134] considered the same code for Chinese rings puzzle solution. I recommend reading Jean-Paul Delahaye who writes beautifully about this subject [92]. Also, Émile Baudot [47] used a related code for letters in telegraphy.

In the same paper [40] we showed that q-decreasing words can be efficiently generated in lexicographic order and explain how the obtained generating algorithm can be turned into a 3-Gray code generating algorithm. Then, we gave a more intricate construction of a 1-Gray code for the particular case $q = 1$ (see Theorem 4 from [40]).

The day we finished this article and were ready to put it on the arXiv, I saw the Eğecioğlu-Iršič [106] fascinating work[6]. They introduced *run-constrained binary strings*. These are binary words, in which every run of 1s is immediately followed by a strictly longer run of 0s. The run-constrained binary strings are precisely the reverse of 1-decreasing words beginning with 0. Using these strings of length $n + 2$ as vertices, and connecting two vertices if they differ at only one position, the Eğecioğlu and Iršič form the *Fibonacci-run graph* $\mathcal{R}_n$ as the induced subgraph of the hypercube. As every non-empty run-constrained string must end with 00, authors of [106] actually drop these 2 zeros, but in $\mathcal{W}_{1,n}$ we do not. Figure 3.7 gives small examples.

In this light, our 1-Gray code $\mathbf{Z}_n$ from [40, Theorem 3] gives a Hamiltonian path in the Fibonacci-run graph. It settles Eğecioğlu-Iršič conjecture.

[6]It is interesting to note that first version of Eğecioğlu-Iršič paper [106] appeared on arXiv only a week before our work [40].

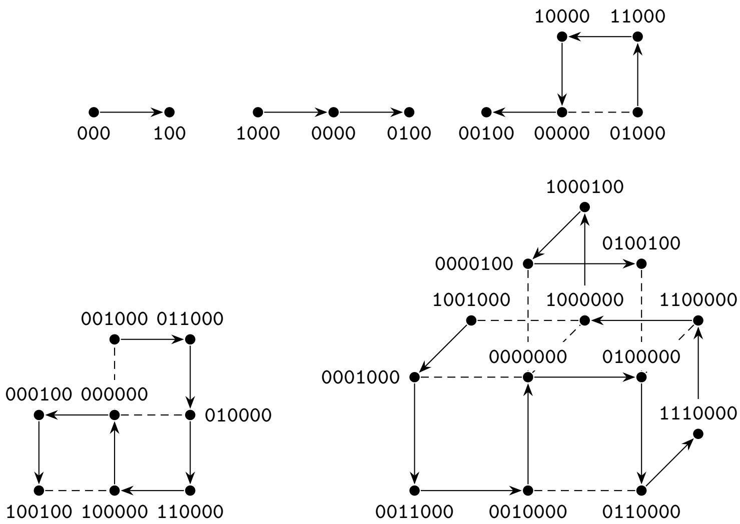


Figure 3.7: Fibonacci-run graphs for small values of n. Vertices correspond to the reverse of words from $\mathcal{W}_n^1$ beginning with 0. The Hamiltonian path is provided by Corollary 7 from [40].

Conjecture 21 (Asked by Eğecioğlu and Iršič, solved)**.** *For any $n \geqslant 1$, the Fibonacci-run graph $\mathcal{R}_n$ has a Hamiltonian path.*

Lemma 9.1 from [106] says that if $n \neq 1 \pmod 3$, then $\mathcal{R}_n$ does not contain a Hamiltonian cycle. Our method gives a Hamiltonian path, which is not a cycle. The question of whether there is a Hamiltonian cycle for the case $n = 1 \pmod 3$ remains open.

The validity of the parity condition (see Corollary 1 from [40]) and experimental investigations for small values, $0 \leqslant n \leqslant 5$ and $2 \leqslant q \leqslant 5$, suggest the following.

Conjecture 22. *For any $n, q \in \mathbb{N}^+$, there is a 1-Gray code for $\mathcal{W}_{q,n}$.*

The 3-Gray code exposed in [40] satisfies van Baronaigien and Ruskey's *constant amortized time* (*CAT*) *principle* [216]. Recently, Wong, Liu, Lam and Im [246] found a 2-Gray code (where consecutive words differs in at most 2 positions) for $\mathcal{W}_{q,n}$ for any real positive q. An extension of their approach yields cyclic 2-Gray codes for Lucas and Fibonacci words [175], the proposed algorithm works in constant amortized time per string and use $O(n^3)$ space. Note that for some values of n and $q \notin \mathbb{N}^+$ no 1-Gray code can exist. For example when $q = 2/3$ we have 12 words, 7 with odd number of 1s: $00001, 00100, 00010, 10000, 11001, 11100$, 11111; and 5 with even number of 1s $00000, 10010, 10001, 11000, 11110$. It is easy to check that there is no 1-Gray in this case (see also [154]).

3.2.2 Bijection breaking the language regularity

Considering the case of natural q, let's have a closer look at the bijection between the set of binary words avoiding $q+1$ consecutive 1s and the set of q-decreasing words. Both sets are enumerated by the $(q+1)$-generalized Fibonacci numbers. We will see that this bijection transforms a mirror-closed set into a set lacking this symmetry and converts a regular language into a non-regular! We provide the definitions below.

Let $\mathcal{B}_n$ denote the set of all binary words of length n, and denote by $\mathcal{B}_n(1^{q+1})$ those that avoid $q+1$ consecutive ones, i.e. have no 1^{q+1} factors.

Definition 23. For any $q \geqslant 1$, we define the map ψ^q

$$\psi^q(w) = \begin{cases} v001^q1^k & \text{if } w = v01^k, v \in \mathcal{B}, k \geqslant 0, \\ 1^{n+q+1} & \text{otherwise.} \end{cases}$$

Less formally, ψ^q inserts a factor 01^q immediately after the last occurrence of 0, and it adds the suffix 1^{q+1} to the word containing no 0. For example $\psi^1(0) = 0\underline{01}$, $\psi^1(00011) = 000\underline{01}11$, $\psi^2(0011101) = 001110\underline{011}1$ and $\psi^5(1) = 1\underline{111111}$. The value of q will be clear from the context, so by a slight abuse of notation ψ^q will be denoted ψ and the same convention is applied for the map φ defined below.

Definition 24. We define the map φ as

$$\varphi(w) = \begin{cases} 1^k & \text{if } w = 1^k \text{ and } k \in [0, q], \\ \psi(\varphi(v)) & \text{if } w = 1^q0v, \\ \varphi(v)01^k & \text{if } w = 1^k0v \text{ and } k \in [0, q-1]. \end{cases}$$

Table 3.1 gives some examples in case of $q = 1$. The following result is proved in [40, see Theorem 1].

Theorem 25. *For $n \geqslant 0, q \geqslant 1$, φ maps bijectively $\mathcal{B}_n(1^{q+1})$ into $\mathcal{W}_n^q$.*

Can we provide a nice graphical description of φ?

The bijection φ does not preserve *Graycodeness*, i.e. a Gray code for $\mathcal{B}_n(1^{q+1})$ is not necessarily mapped by φ to a Gray code for q-decreasing words, see Table 3.2.

Words avoiding 11, $\mathcal{B}_n(11)$				**1-decreasing words, $\mathcal{W}_{1,n}$**			
0	00	000	0000	0	00	000	0000
1	01	100	1000	1	10	001	0001
	10	001	0100		11	100	0010
		010	0001			110	1000
		101	1001			111	1001
			0010				1100
			0101				1110
			1010				1111

Table 3.1: Examples of φ for short words. Order is preserved, i.e. $\varphi(10) = 11$.

$u \in \mathcal{B}_4(111)$	$\varphi(u) \in \mathcal{W}_{2,4}$
1100	0011
1101	1111
1001	1001
1000	0001
1010	0101
1011	1101
0011	1100
0010	0100
0000	0000
0001	1000
0101	1010
0100	0010
0110	1110

Table 3.2: The images of words in $\mathcal{B}_4\,(111)$ under the bijection φ. Words in $\mathcal{B}_4\,(111)$ are listed in a BRGC-like order, called local reflected order in [233].

A *formal grammar* is a system to describe a set of words respecting a certain syntax. To specify the grammar, we need the following ingredients:

- an alphabet, a finite set Σ of terminal symbols;
- a finite set N of nonterminal symbols;
- a finite set P of production rules, i.e. rewriting rules of the form $x \to y$, where x and y are finite sequences of terminal or nonterminal symbols, and x is not empty, $x \neq \varepsilon$;
- a start symbol S, it is a special nonterminal symbol.

In *right-regular grammars* each production rule $x \to y$ have the following restrictions:

- the left-hand side, x, is a nonterminal symbol;
- there is at most one nonterminal symbol on the right-hand side. If there is one, it always appears at the end of y.

For instance, the following grammar describes the set of binary words avoiding 111.

$$\begin{array}{lll} S \to \varepsilon & B \to \varepsilon & C \to \varepsilon \\ S \to 0S & B \to 0S & C \to 0S \\ S \to 1B & B \to 1C & \end{array}$$

A language described by a right-regular grammar is called *regular*. Finite automatons can also be used to describe regular languages. Figure 3.8 shows an automaton that accepts only the words avoiding 111. For more information about regular languages, expressions and finite automatons see classic works by Chomsky and Schützenberger [87]; Allouche and Shallit [2]; Hopcroft, Motwani and Ullman [146].

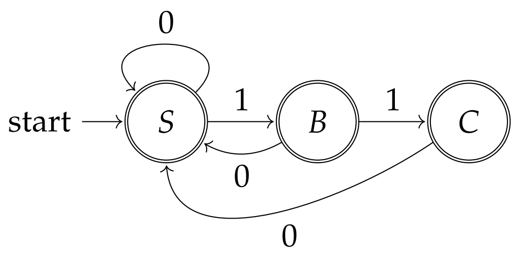


Figure 3.8: A finite automaton accepting binary words that avoid 111.

It is known that words avoiding a finite set of finite consecutive patterns (factors) always can be described by a regular language. We also know that words of length n from any regular language are enumerated by a rational generating function. The beautiful relationship between rational generating functions and regular languages is not, however, trivial. There are a lot of interesting questions and passages here. For instance, Barcucci, Del Lungo, Frosini, and Rinaldi [24] show how to construct a regular language for a given rational generating function.

One day, I sent an article about q-decreasing words (see Definition 20 on page 55) to the journal *Comptes Rendus Mathématique*. It was not accepted. But I'm grateful to the anonymous reviewer, he (or she) made a very important observation.

Theorem 26. *The language of q-decreasing words is not regular.*

Proof by an anonymous reviewer from Comptes Rendus Mathématique.

> Suppose the opposite, let our language be recognized by a finite automaton of n states. Consider a q-decreasing word $u = 0^a 1^b$, where $b > n$ (and $qa > b$ since the word is q-decreasing). Among the $n + 1$ last symbols of u (all of them are 1s), at least two correspond to the same state A of the automaton. If the difference between their positions is k, it means that we get to A after every k digits 1 starting from the first of them. So, for every i, the word $u1^{ki} = 0^a 1^{b+ki}$ is also accepted by the automaton, even when $b + ki > qa$. A contradiction.

□

The set of binary words avoiding $q + 1$ consecutive 1s form a regular language. But the equienumerated set of q-decreasing words is not! We obtain the following result:

Theorem 27. *Bijection φ transforms a regular language into a non-regular one.*

3.2.3 Q-bonacci

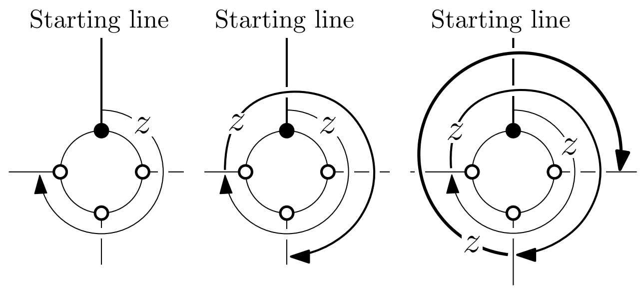
It took me almost one year to understand how we can deal with q-decreasing words when $q \in \mathbb{Q}^+$. In 2022, I presented this work on the Twentieth International Conference on Fibonacci Numbers and Their Applications [155], the accompanying paper is published in the Fibonacci Quarterly journal [154]. This construction uses *model polynomials* associated with positive rational numbers.

Definition 28. A polynomial

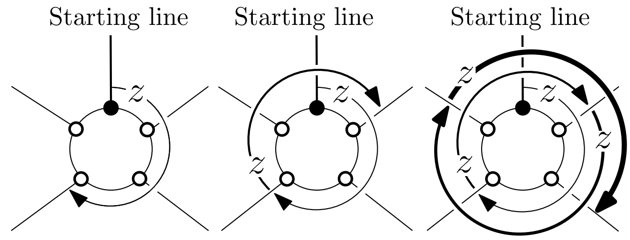
$$P_{q=\frac{c}{d}}(y,z) = \sum_{i=0}^{c-1} z^{1+\left\lfloor \frac{i}{q} \right\rfloor} y^i$$

is called *a model polynomial* of a positive rational number q represented by the irreducible fraction $q = \frac{c}{d}$.

For instance, $P_{\frac{2}{3}}(y,z) = z + z^2y$, $P_{\frac{3}{2}}(y,z) = z + zy + z^2y^2$, and $P_{1/k}(x) = z$ for any $k \in \mathbb{N}^+$. Figure 3.9 presents a graphical interpretation of model polynomials.



$$P_{q=\frac{3}{4}}(y,z) = \quad z \quad + \quad z^2y \quad + \quad z^3y^2$$



$$P_{q=\frac{3}{5}}(y,z) = \quad z \quad + \quad z^2y \quad + \quad z^4y^2$$

Figure 3.9: A graphical representation of model polynomial $P_{q=\frac{3}{4}} = z + z^2y + z^3y^2$. For $j > 0$, a term z^iy^j in a model polynomial means that one must make i arc-steps of the angle $2q\pi$ in order to cross the starting line j times.

In this paper I proved two following theorems.

Theorem 29. *Let $q \in \mathbb{Q}^+$ be represented by the irreducible fraction $q = \frac{c}{d}$. The generating function $W_q(y,z) = \sum_{r=0}^{\infty}\sum_{i=0}^{\infty} w_{r,i}z^ry^i$, where $w_{r,i}$ is the number of words from $\mathcal{W}_q$ of length $r+i$ containing exactly r zeros and i ones, is*

$$W_q(y,z) = \frac{1 - z^dy^c}{(1-y)\big(1 - z^dy^c - P_q(y,z)\big)}.$$

Theorem 30. *Let a positive rational number q be represented by the irreducible fraction $\frac{c}{d}$. The number of n-length binary words from $\mathcal{W}_{q,n}$, denoted by w_n, can be expressed as*

$$w_n = \sum_{j \in J} w_{n-j} + w_{n-(c+d)}, \tag{3.2}$$

where J is the set of powers from the model polynomial $P_{q=\frac{c}{d}}(x, x)$. For example, when $q = \frac{3}{2}$, we have $P_{\frac{3}{2}}(x, x) = x + x^2 + x^4$, and $J = \{1, 2, 4\}$.

Initial conditions $w_0, w_1, \cdots, w_{c+d-1}$ are obtained by setting $w_n = 0$ for $n < 0$, unrolling Equation (3.2) from left to right, while adding an extra 1 for every w_i for $0 \leqslant i < c + d$.

The total number of 0s (in other words, the *popularity* of 0s) in all words from $\mathcal{W}_{q=1,n}$ is enumerated by a shift of the sequence A6478 in Sloane's On-line Encyclopedia of Integer Sequences [200]. The corresponding generating function is obtained by evaluating $\frac{\partial W_1(x,xz)}{\partial z}|_{z=1}$. It is quite unexpected, but the sequence A6478 enumerates also the edges in the *Fibonacci hypercube* considered by Rispoli and Cosares [211]. A Fibonacci hypercube is a polytope determined by the convex hull of the *Fibonacci cube* which in turn is defined by Hsu in [148] as the graph whose vertices correspond to binary words of size n avoiding two consecutive 1s and where two vertices are connected if and only if the corresponding words differ at only one position.

Is it possible to give some kind of nice bijective construction between the edges of Fibonacci Hypercube and the 0s in words from $\mathcal{W}_{q=1,n}$? What can be said about the popularity of 0s in $\mathcal{W}_{q,n}$ with $q > 1$?

The generalized golden ratio is defined as $\varphi_k = \lim_{n \to \infty} a_{n+1}/a_n$, where a_{n+1} and a_n are two adjacent k-bonacci numbers. The golden ratio is $\varphi_2 = (1 + \sqrt{5})/2$, and $\varphi_3 = (1 + \sqrt[3]{19 + 3\sqrt{33}} + \sqrt[3]{19 - 3\sqrt{33}})/3$ is known as the Tribonacci constant. The Tetranacci constant φ_4 have quite a large expression in radicals. In general, φ_k is expressed as the largest root of the polynomial $x^k - x^{k-1} - \cdots - x - 1$. See Wolfram's paper [245] for full details. In the same paper, Wolfram conjectured that there is no expression in radicals for $k \geqslant 5$. By computing the Galois group, with the help of the computer algebra system Magma [70], he confirmed the conjecture for $5 \leqslant k \leqslant 11$. Martin [191] proved the case of even or prime k. Furthermore, Cipu and Luca [89] demonstrated the impossibility of the construction of φ_k by ruler and compass for $k \geqslant 3$. As far as I can tell, the question whether there is an expression in radicals remains open for odd non-prime $k > 11$. Dubeau [99] proved that φ_k approaches 2 when $k \to \infty$.

For natural q we have $\varphi_{q+1} = \lim_{n\to\infty} |\mathcal{W}_{q,n+1}|/|\mathcal{W}_{q,n}|$. Non-natural q, in some way, allows us to see what happens with the generalized golden ratio, when its parameter becomes non-natural. As the generating functions are rational in case $q \in \mathbb{Q}^+$, classical analytic combinatorics method can be used to find the limit. It equals to $1/\beta$, where β the smallest by modulus root of the denominator of the corresponding generating function $W_{q=\frac{c}{d}}(x) = \frac{1-x^{c+d}}{(1-x)\left(1-x^{c+d}-P_q(x,x)\right)}$ (see Theorem 29). Figure 3.10 presents some numerical estimations for the function $q \mapsto \lim_{n\to\infty} |\mathcal{W}_{q,n+1}|/|\mathcal{W}_{q,n}|$, where q takes rational values from $[0, 2.02]$ with step $1/50$.

I would like to ask a question, related to Wolfram conjecture:

For which rational values of q there is an expression in radicals for $\varphi_{q+1} = \lim_{n\to\infty} |\mathcal{W}_{q,n+1}|/|\mathcal{W}_{q,n}|$?

Remark, that the set $\mathcal{W}_{q,n}$ is well-defined even if we extend the domain of the parameter q to all positive real numbers. In the article [154], I proposed two conjectures in this realm. Both were solved in our work with Sergey Dovgal [97], and presented at the Permutation Pattern conference [96] in Dijon.

Conjecture 31 (solved in [97]). *For any given $r \in \mathbb{R}^+$, $\lim_{n\to\infty} |\mathcal{W}_{r,n+1}|/|\mathcal{W}_{r,n}|$ exists.*

Conjecture 32 (solved in [97]). *The function $r \mapsto \lim_{n\to\infty} |\mathcal{W}_{r,n+1}|/|\mathcal{W}_{r,n}|$ is strictly increasing over the interval $[0, +\infty)$ and discontinuous at every positive rational r.*

The next Subsection is devoted, in part, to this peculiar function.

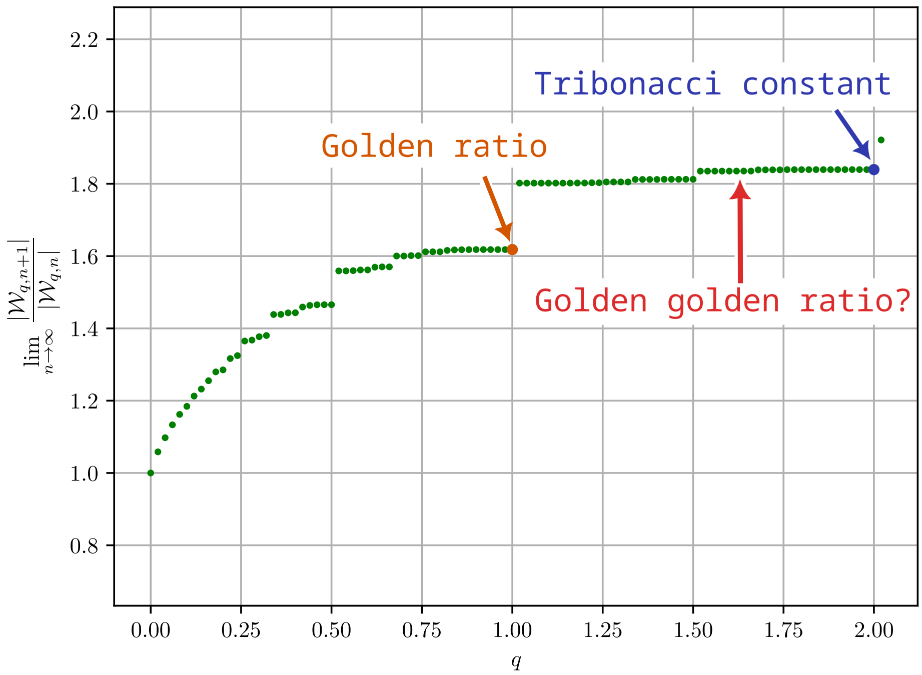


Figure 3.10: Numerical estimation of $\lim_{n\to\infty} |\mathcal{W}_{q,n+1}|/|\mathcal{W}_{q,n}|$ for several values of $q \in [0, 2.02]$, using a step 0.02.

3.2.4 Sturm, Minkowski and the $\mathbb{R}$-bonacci fractal

This Subsection is based on a recent paper [97], co-authored with Sergey Dovgal, where we extend the family of q-decreasing words, to cover all positive real numbers as possible values of the parameter q. Here, I will omit proofs, focusing on constructions and statements of results. Our results were presented at the Permutation Patterns 2023 [96] conference.

Ray cutting and Sturmian words

For any real $q > 0$, the *ray cutting word* $s(q)$ is defined as an intersection sequence of a straight half-line $y = qx$ for $x \in (0, \infty)$ with the lines of a square grid ($y = i$ or $x = i$ for $i \in \mathbb{N}^+$). Going along the half-line, starting from $(0, 0)$, we write `1` if the line intersects a horizontal edge and `0` in case of a vertical edge (see Figure 3.11), we write `01` (in this order) when crossing an intersection point of grid lines.

For any irrational slope q, the word $s(q)$ is aperiodic and Sturmian. In the general setting, *Sturmian words* are defined as cutting sequences of the line $y = ax + b$ for $x \in (0, \infty)$, irrational $a > 0$ and real $b \in [0, 1)$ or equivalently as binary words having exactly $n + 1$ factors (contiguous subwords) of length n. Sturmian words shine in several areas of mathematics: combinatorics, number theory, tilings, discrete dynamical systems. The structures similar to Sturmian words were already studied by Johann III Bernoulli [59] in 1771. Expositions of Sturmian words and related results can be found in Chapter 2 (written by Berstel and Séébold) of Lothaire's œuvre [179] and in the book of Allouche and Shallit [2]. For a rational slope q, the word $s(q)$ is periodic, its shortest factor f such that $s(q) = f \cdot f \cdot f \ldots$, where $\cdot$ means concatenation, corresponds to the Christoffel word of slope q [60, 88].

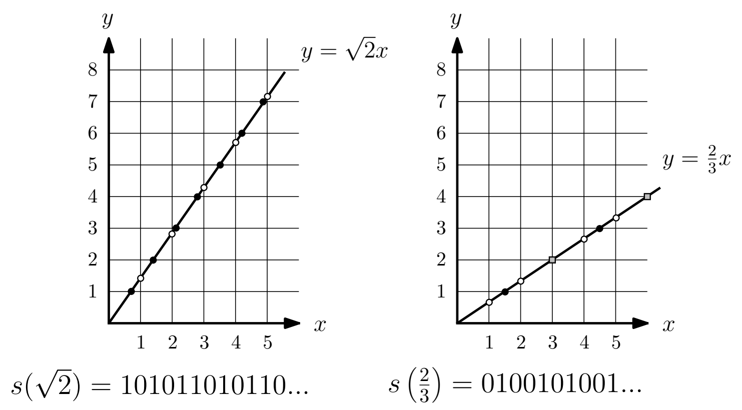


Figure 3.11: Cutting sequences with slopes $\sqrt{2}$ and $\frac{2}{3}$.

Two kinds of Fibonacci words

One paradigmatic example of Sturmian words is the *Fibonacci word* 010010100100101... which is characterized by a cutting sequence of the line with a slope $1/\varphi$, where $\varphi = (1+\sqrt{5})/\sqrt{2}$. It can also be obtained either by a recursive simultaneous application of substitution rules $\{0 \mapsto 01, 1 \mapsto 0\}$ to an initial string 0, or as a limit of recursive concatenations of strings $S_n = S_{n-1}S_{n-2}$, where $S_0 = 0$ and $S_1 = 01$.

Now consider another Fibonacci object, (or, more generally, k-bonacci), which is an ensemble of binary words of length n avoiding k consecutive 1s. The set of such words is in bijection with tilings of stripes of length $(n+1) \times 1$ with tiles of size 1×1 (monomers), 2×1 (dimers), $\ldots$, $k \times 1$ (k-mers), so it is convenient to call them *k-bonacci tilings*. The cardinality of the set of such words of length n is equal to nth k-bonacci number. See the beginning of the Section 3.2) for historical and prosodic notes.

These two Fibonacci objects belong to two seemingly different worlds. In [97] we proposed a link between these worlds. I recall here how certain subsets of prefixes of ray cutting words can be used as building blocks to construct generalized Fibonacci tilings.

Construction from ray cutting prefixes

We express q-decreasing words as sequences of ray cutting prefixes ending on 1. It is handy to use the Kleene star operator (it corresponds to SEQ operator in the Flajolet–Sedgewick book [109]), which constructs a disjoint union of finite concatenations from strings of a given family. For instance, $(\{0, 10\})^*$ provides all binary strings which are empty or end on 0 and do not contain two consecutive 1s. We also use the "$\cdot$" symbol to denote all possible pairwise concatenations between the elements of two families.

Theorem 33. *For $q \in \mathbb{R}^+$, the set $\mathcal{W}_q$ of q-decreasing words can be represented as*

$$\mathcal{W}_q = (\{1\})^* \cdot (\mathcal{S}_q)^*, \text{ where } \mathcal{S}_q = \cup_{i=0}^{\infty}\{0^{1+\lfloor i/q \rfloor}1^i\}.$$

For a binary word α containing n 0s and m 1s, we define a transformation $\kappa(\alpha) = 0^{n+1}1^m$, so that the empty word ε is mapped to the word 0. We provide a decomposition of q-decreasing words into partitions of certain ray cutting prefixes by using the above transformation.

Theorem 34. *For $q \in \mathbb{R}^+$, the transformation κ bijectively maps the set of prefixes ending with 1 of the ray cutting word $s(q)$ to the set $\mathcal{S}_q$, which is used in the construction of q-decreasing words.*

Tables 3.3 and 3.4 provide examples.

q	Ray cutting word	Factors from $\mathcal{S}_q$	Some q-decreasing words
$\sqrt{2}$	101011010110...	$\kappa(\varepsilon) = 0$, $\kappa(1) = 01$, $\kappa(101) = 0011$, $\kappa(10101) = 000111$, ...	111100011000111, 000111101000001, 100000101000001
$\frac{2}{3}$	010010100101...	$\kappa(\varepsilon) = 0$, $\kappa(01) = 001$, $\kappa(01001) = 000011$, $\kappa(0100101) = 00000111$, ...	111100001100001, 000011001000001, 001000000001111

Table 3.3: Illustration of the transformation κ. Prefixes ending with 1 of the ray cutting word $s(q)$ correspond to factors from the set $\mathcal{S}_q$.

q	Ray cutting word	Counting Sequence $\lvert\mathcal{W}_{q,n}\rvert$	OEIS
$\frac{1}{2}$	0010010010010010...	1, 2, 3, 4, 6, 9, 13, 19, 28, 41, 60, 88, 129, 189, ...	Narayana's cows, A930
$1/\varphi$	0100101001001010... Fibonacci word	1, 2, 3, 5, 8, 12, 19, 30, 47, 74, 116, 182, 286, 448, ...	NEW
$\frac{2}{3}$	0100101001010010...	1, 2, 3, 5, 8, 12, 19, 30, 47, 74, 116, 182, 286, 449, ...	Comp. into 1s, 3s and 5s, A60961
1	0101010101010101...	1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, ...	Fibonacci, A45
2	1011011011011011...	1, 2, 4, 7, 13, 24, 44, 81, 149, 274, 504, 927, 1705, 3136, ...	Tribonacci, A73
$\frac{3}{2}$	1010110101101011...	1, 2, 4, 7, 13, 23, 42, 76, 138, 250, 453, 821, 1488, 2697, ...	NEW
$\sqrt{2}$	1010110101101010...	1, 2, 4, 7, 13, 23, 42, 76, 138, 250, 453, 821, 1488, 2697, ...	NEW
φ	1011010110110101...	1, 2, 4, 7, 13, 24, 44, 81, 148, 272, 499, 916, 1681, 3085, ...	NEW
e	1101110111011011...	1, 2, 4, 8, 15, 29, 56, 108, 208, 401, 773, 1490, 2872, 5536, ...	NEW
π	1110111011101110...	1, 2, 4, 8, 16, 31, 61, 120, 236, 463, 910, 1788, 3513, 6901, ...	NEW

Table 3.4: Examples of ray cutting words and corresponding counting sequences for the cardinalities of q-decreasing words.

Rational discontinuity

Here we study the function $\Phi(q) = \lim_{n\to\infty} \frac{|\mathcal{W}_{q,n+1}|}{|\mathcal{W}_{q,n}|}$, whose graph is shown on Figure 3.10. Using the previously mentioned `SEQ` operator, Theorem 33 yields the generating function $W_q(x) = \sum_{n=0}^{\infty} |\mathcal{W}_{q,n}|x^n$ of the family $\mathcal{S}_q$ for any $q \in \mathbb{R}^+$:

$$W_q(x) = \frac{1}{(1-x)\left(1 - \sum_{i=0}^{\infty} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}\right)}. \tag{3.3}$$

The case where q is a positive *rational number* represented by an irreducible fraction $\frac{c}{d}$ is treated in [154] where the generating function $W_q(x)$ is expressed as

$$W_{q=\frac{c}{d}}(x) = \frac{1 - x^{c+d}}{(1-x)\left(1 - x^{c+d} - \sum_{i=0}^{c-1} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}\right)}. \tag{3.4}$$

In other words, Eq. (3.3) holds for any positive real q, and is more general, although simpler, form of Eq. (3.4) which is only valid for $q \in \mathbb{Q}^+$ (see also Theorem 29).

We have the following result about the asymptotic behavior of the coefficient $[x^n]W_q(x)$.

Theorem 35. *For any real $q > 0$, the number of q-decreasing words of length n grows as $C_q \cdot \Phi(q)^n$, where $1/\Phi(q)$ is the unique smallest in modulus root of $1 - \sum_{i=0}^{\infty} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}$, and*

$$C_q = -\frac{\Phi(q)}{\left((1-x)\left(1 - \sum_{i=0}^{\infty} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}\right)\right)'(1/\Phi(q))}.$$

The proof, presented in details in [97], uses a classical asymptotic analysis of meromorphic functions and a variant of Daffodil Lemma [109].

The equation $1 = \sum_{i=0}^{\infty} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}$ shares the smallest in modulus root with

$$A_q := 1 - \sum_{i=1}^{\infty} \sum_{j=0}^{\infty} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor + j},$$

the fact that has a nice geometrical interpretation. To see it, we have to decompose the set $\mathcal{W}_q$ in another way, different from what was given in Theorem 33. Here we use a set $\mathcal{F}_q$ of factors $0^a 1^b$ such that $qa > b$ and $b \geqslant 1$. With this, we decompose any word $w \in \mathcal{W}_q$ as a sequence of 1s, followed by a sequence of factors from $\mathcal{F}_q$, followed by a sequence of 0s. Any of these sequences can be empty. We have

$$w = \overbrace{1...1}^{\text{some ones}} \overbrace{f_1 f_2 .... f_k}^{f_\ell \in \mathcal{F}_q} \overbrace{0...0}^{\text{some zeros}}, \quad \text{where } \mathcal{F}_q = \bigcup_{i=1}^{\infty} \bigcup_{j=0}^{\infty} \{ \overbrace{0\ldots00}^{1+\left\lfloor \frac{i}{q} \right\rfloor + j \text{ zeros}} \underbrace{1\ldots11}_{i \text{ ones}} \}.$$

Now, we write the g.f. $W_q(x)$ as

$$W_q(x) = \frac{1}{1-x} \cdot \frac{1}{A_q} \cdot \frac{1}{1-x}.$$

Consider the grid $\mathbb{Z}^+ \times \mathbb{Z}^+$, and make every point (a,b) correspond to a factor $0^a 1^b$. The power series A_q sums over all points with positive integer coordinates found under the line $b = qa$. Figure 3.12 gives some examples.

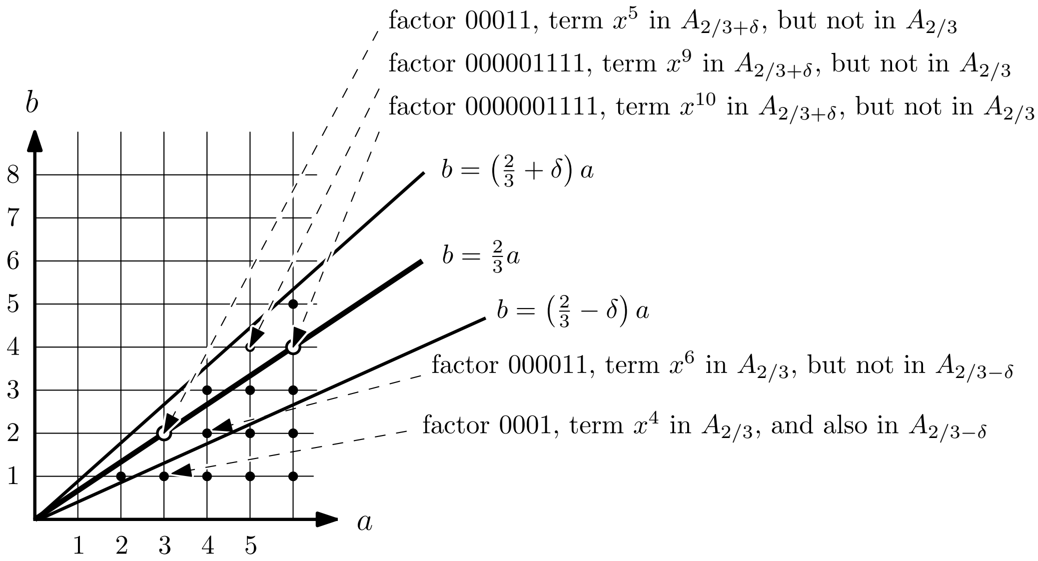


Figure 3.12: The line $b = \frac{2}{3}a$ and the geometrical interpretation of factors $0^a 1^b$ where $\frac{2}{3}a > b$.

From Theorem 35 we see that the function $\Phi(q)$ the function $\Phi(q)$ is well-defined as $\lim_{n\to\infty} \frac{|\mathcal{W}_{q,n+1}|}{|\mathcal{W}_{q,n}|}$. It behaves as follows.

Theorem 36. *The function* $\Phi(q) = \lim_{n\to\infty} \frac{|\mathcal{W}_{q,n+1}|}{|\mathcal{W}_{q,n}|}$ *is*

a) *strictly increasing over* $q \in [0, \infty)$*;*
b) *bounded,* $1 \leqslant \Phi(q) < 2$*, with* $\Phi(0) = 1$ *and* $\lim_{q\to\infty} \Phi(q) = 2$*;*
c) *left-continuous (and right-discontinuous) at every positive rational point;*
d) *continuous at every positive irrational point.*

The proof (see [97]) of Theorem 36 follows from the geometrical representation of the polynomial A_q. We can also calculate the jump sizes.

Theorem 37. *For any irreducible* $q = \frac{c}{d} \in \mathbb{Q}^+$, $\lim_{\delta \to 0^+} \Phi(q+\delta)$ *equals the reciprocal of the smallest in modulus root, denoted by* ρ_q^+, *of the polynomial*

$$\Pi_q^+ := 1 - (2-x)x^{c+d} - \sum_{i=0}^{c-1} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}.$$

We have

$$\lim_{\delta \to 0^+} \Phi(q+\delta) - \Phi(q) = \frac{1}{\rho_q^+} - \frac{1}{\rho_q},$$

where ρ_q *is the smallest in modulus root of* $\Pi_q := 1 - x^{c+d} - \sum_{i=0}^{c-1} x^{1+i+\left\lfloor \frac{i}{q} \right\rfloor}$.

Stern–Brocot tree, Minkowski's $?(x)$ and $\mathbb{R}$-bonacci fractal

As we can see on Figure 3.10 and 3.14a the graph of the function $\Phi(q)$ shows a certain amount of self-similarity. We first note that the fractal structure of $\Phi(q)$ appears more regular in 3.14b as we apply Minkowski's question-mark function over the x-axis. Then, we harvest the Stern–Brocot tree. It will give us fruits in the form of sequences of nested intervals to zoom into, to get a better sense of self-similarity. See Figures 3.15, 3.16 and 3.17 for an illustration of this *récolte*. Below we provide the basic definitions and constructions needed to state the results, detailed proofs can be found in our paper [97].

The Minkowski's question mark function $?(x)$ is a somewhat unusual trick, even its notation already hints at this. We must first discuss *mediants*, sometimes called the *freshman sum*, and the construction of the Stern–Brocot tree [72,73,228]. As Stern noted, it was actually his friend Eisenstein [103] who discovered this construction. He did it, as it later turned out, independently of Brocot. Mansuy wrote a beautiful article [189] about Brocot, who was a French clockmaker and amateur mathematician.

For two irreducible fractions a/b and c/d their *mediant* is defined as $(a + c)/(b + d)$. The root of the Stern–Brocot tree is $1/1$, which is the mediant of two conventionally irreducible fractions $1/0$ and $0/1$. To determine the left (resp. right) child of a node x/y of the level i we need to find the greatest (resp. smallest) fraction $x'/y' < x/y$ (resp. $x'/y' > x/y$) that appears in set of values of first i levels together with $1/0$ and $0/1$, and compute the mediant $(x + x')/(y + y')$. For instance, the left child of $2/3$ is $3/5$, it is calculated as the mediant of $1/2$ and $2/3$. Figure 3.13 illustrates this process. Stern–Brocot tree contains all rationals once.

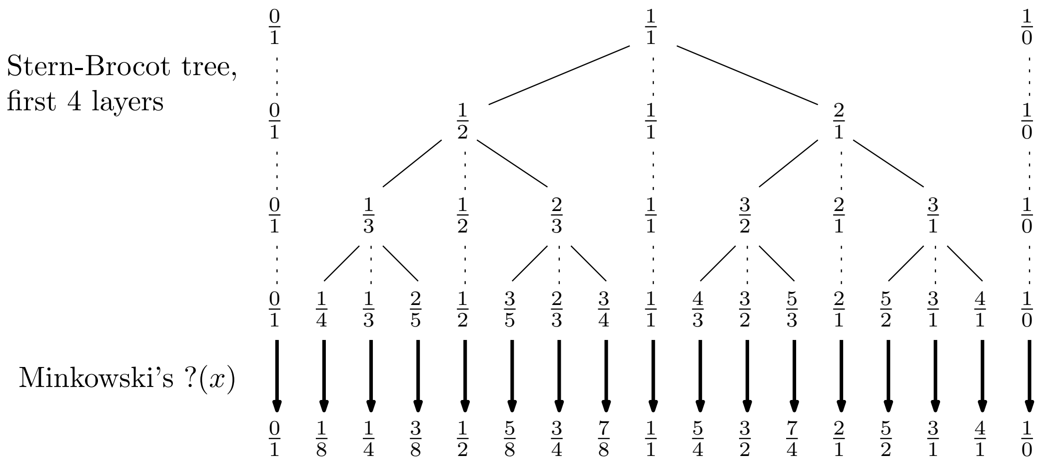


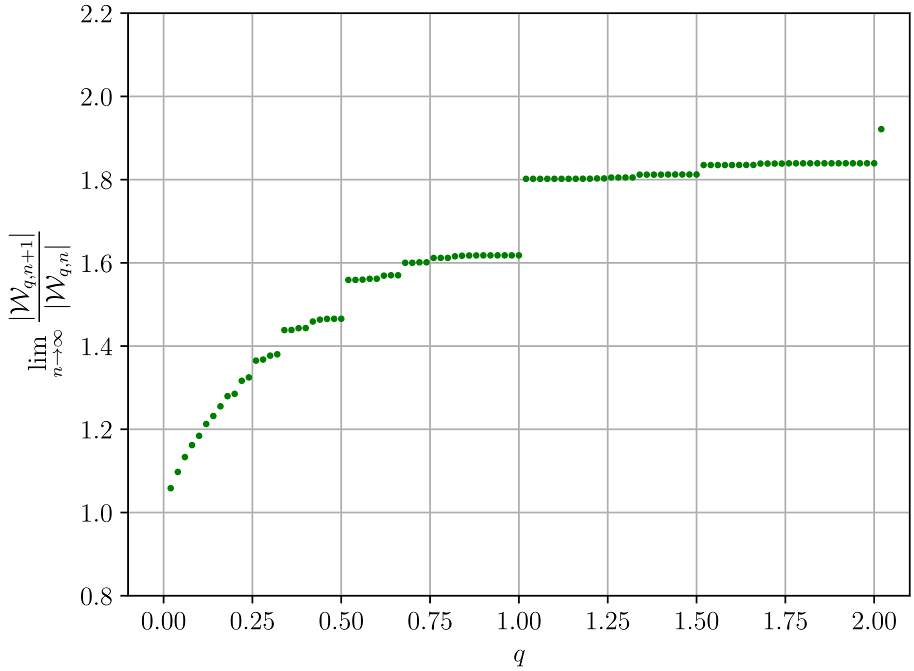
Figure 3.13: The Stern–Brocot tree and Minkowski's $?(x)$.

Minkowski's question-mark function, denoted by $?(x)$, maps a positive rational value x to a positive dyadic rational $a/2^k$ with $a, k \in \mathbb{N}$. By definition, $?(0) = 0$ and $?(1) = 1$. Whenever $x \in (0, 1)$ is a rational number represented by an irreducible fraction a/b, such that in the Stern–Brocot tree it is constructed via taking a mediant of two fractions p/r and p'/r', its image under Minkowski's function is defined as

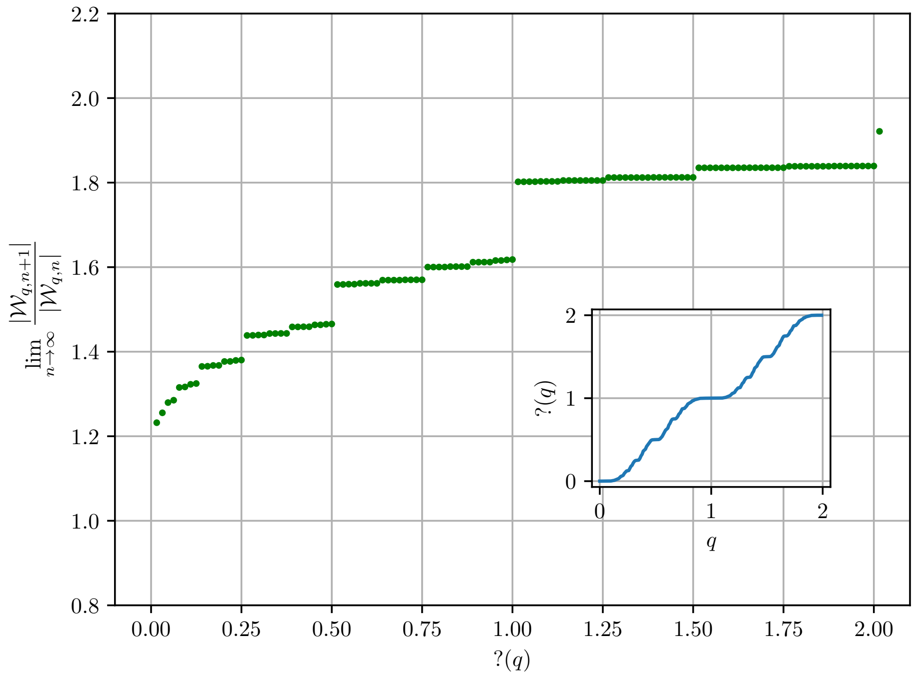
$$? (x) = ? \left(\frac{p + p'}{r + r'} \right) := \frac{1}{2} \left(? \left(\frac{p}{r} \right) + ? \left(\frac{p'}{r'} \right) \right).$$

In other words, we descend the Stern–Brocot tree in search of the a/b, and "in parallel" construct a resulting value by applying the mean instead of the mediant. For $x > 1$, Minkowski's function is defined as $?(x + 1) =?(x) + 1$. In general, $?(x)$ is monotonically increasing, and can be defined on all $\mathbb{R}^+$ [94].

Figure 3.16c is obtained by applying the simple rescaling on the vertical axis and Minkowski's question-mark function followed by the simple rescaling on the horizontal axis for intervals $(k/(k + 1), 1]$ and their images. Figure 3.17 illustrates this for intervals $(\frac{1+k}{3+2k}, \frac{1}{2}]$.



(a)



(b)

Figure 3.14: The function $\Phi(q) = \lim_{n\to\infty} \frac{|\mathcal{W}_{q,n+1}|}{|\mathcal{W}_{q,n}|}$ before Minkowski's rescaling (a) and after (b).

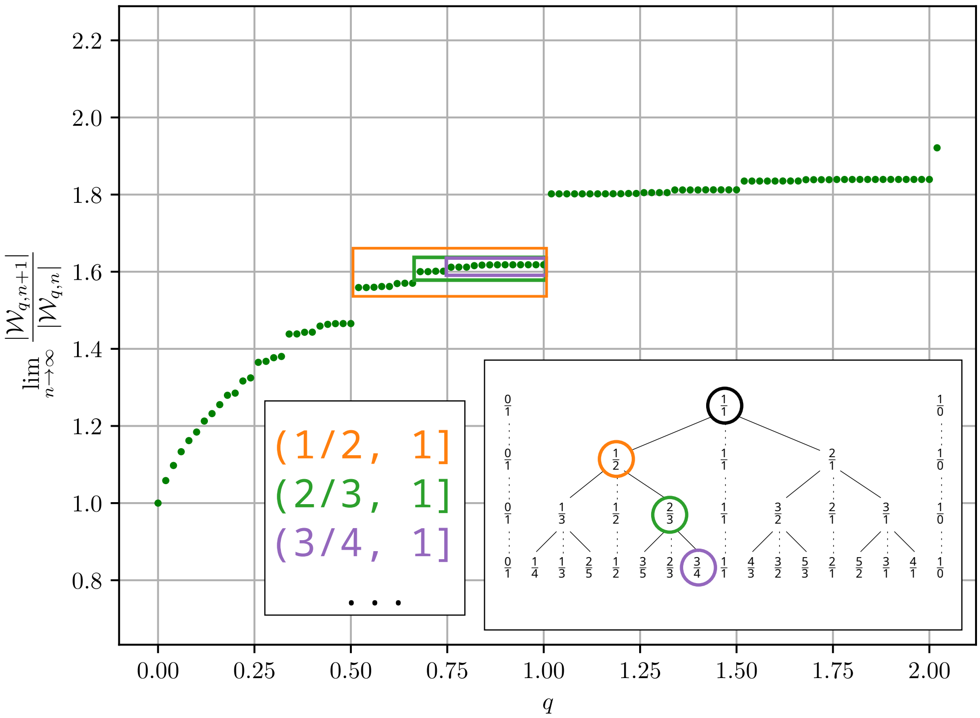
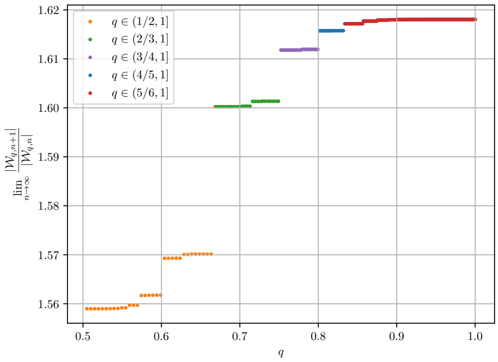
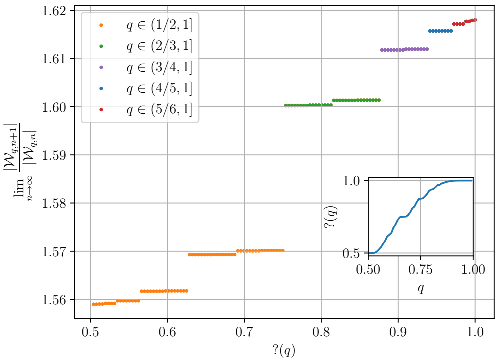


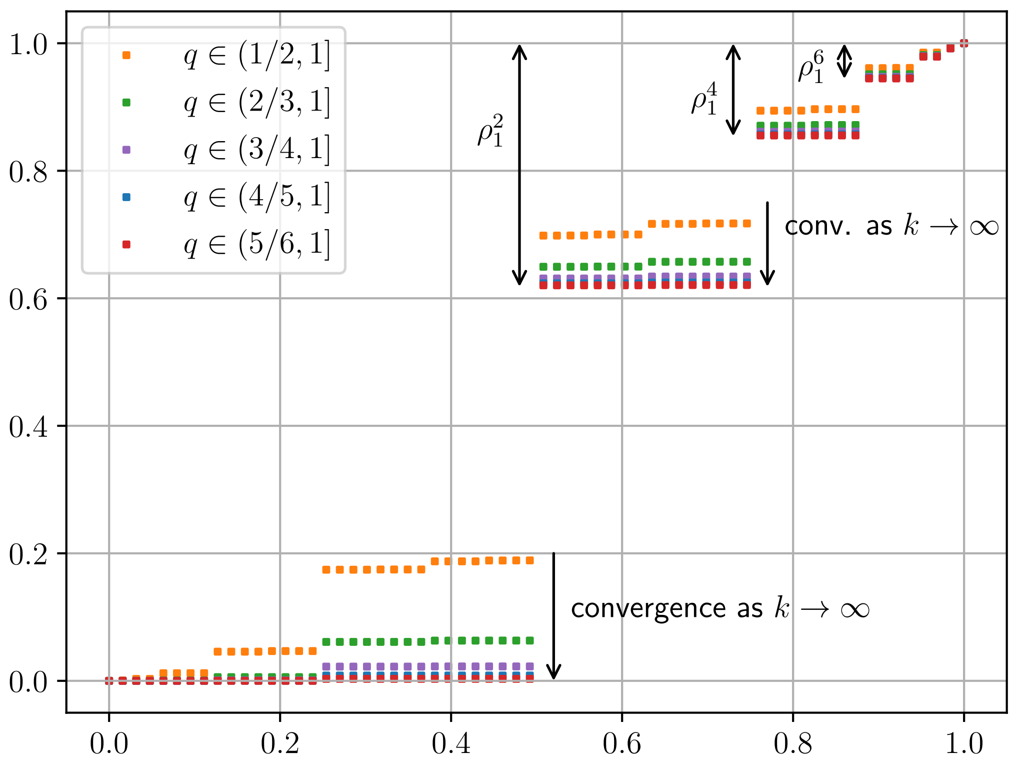
Figure 3.15: Zoom regions for intervals $(\frac{k-1}{k}, 1]$.



(a)

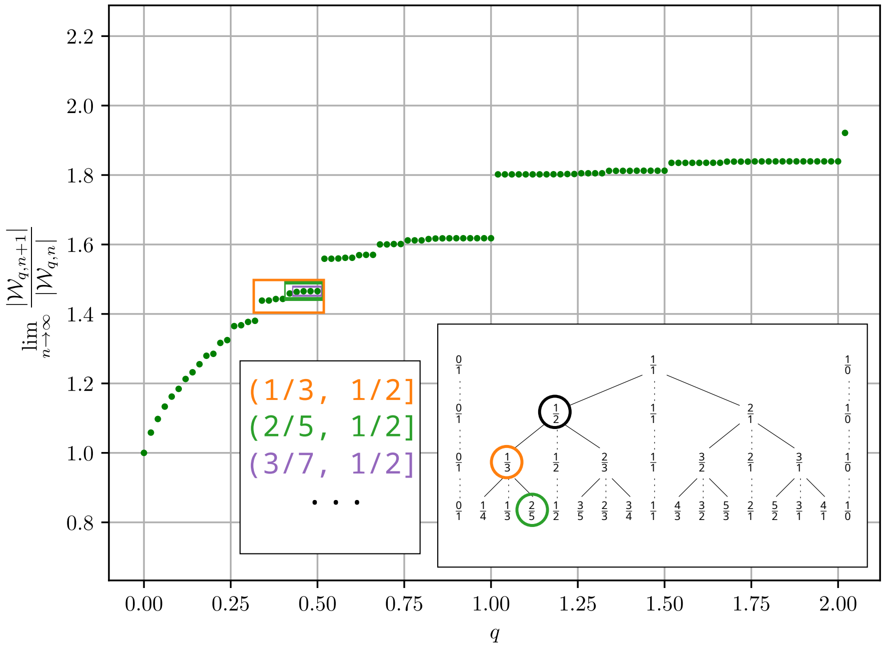


(b)

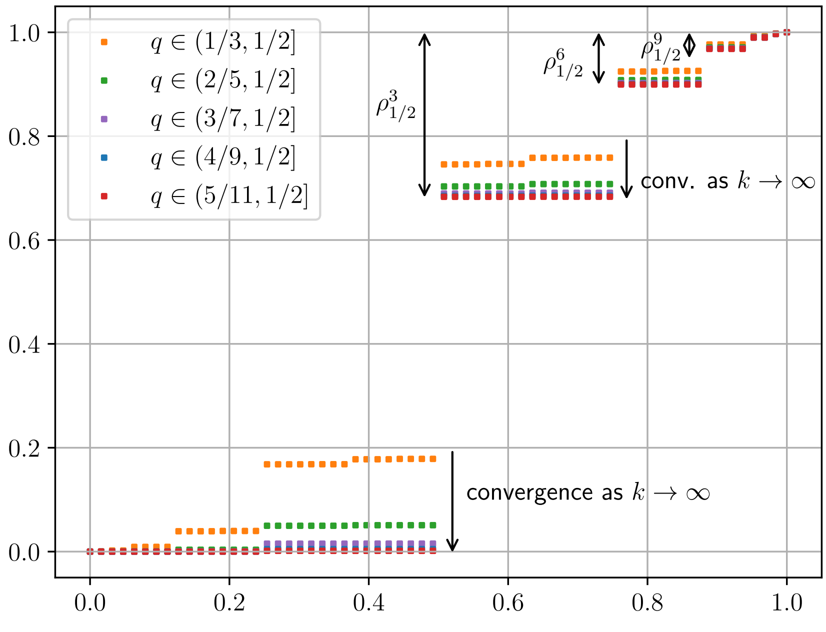


(c)

Figure 3.16: Regions for intervals $(\frac{k-1}{k}, 1]$: before Minkowski's rescaling (a), after the rescaling (b), rescaled and superimposed (c).



(a)



(b)

Figure 3.17: Regions for intervals $(\frac{1+k}{3+2k}, \frac{1}{2}]$: before Minkowski's rescaling (a), rescaled and superimposed (c).

Let q denote $\frac{c}{d}$ and let q_k^* denote $\frac{p+ck}{r+dk}$, i.e. the kth left approximation of q in the Stern–Brocot tree. In [97] we proved the following.

Theorem 38. *The graph of the function $\Phi(q)$ rescaled on the intervals $(q_k^*, q]$ tends to a constant function on the semi-open intervals $(q_{k-1}^*, q_k^*]$:*

$$\lim_{k\to\infty} \frac{\Phi(q) - \lim_{\delta\to 0^+} \Phi(q_{k-1}^* + \delta)}{\Phi(q) - \Phi(q_k^*)} = 1.$$

The ratio between consecutive constants on the rescaled graph tends to ρ_q^{c+d}:

$$\lim_{k\to\infty} \frac{\Phi(q) - \Phi(q_k)}{\Phi(q) - \Phi(q_{k-1})} = \rho_q^{c+d}.$$

The proof is based on Pick's Theorem [202], a geometrical representation of polynomials Π_q and some polynomial gymnastics. Using our technique, it is possible to derive a similar result for the right approximation of any rational point $q = c/d$.

Where do the jumps go higher?

The function $\Phi(q)$ has the highest jump at point $q = 1$, the second-highest jump is at $q = 1/2$, the third-highest jump appears when $q = 2$, see Figure 3.14a. The sequence of positive rational numbers ordered by corresponding jumps of the function $\Phi(q) = \lim_{n\to\infty} |\mathcal{W}_{q,n+1}|/|\mathcal{W}_{q,n}|$ starts with

$$1, \frac{1}{2}, 2, \frac{1}{3}, \frac{1}{4}, 3, \frac{2}{3}, \frac{1}{5}, \frac{1}{6}, \frac{3}{2}, \frac{1}{7}, 4, \frac{2}{5}, \frac{1}{8}, \frac{1}{9}, \frac{1}{10}, \frac{3}{4}, \frac{1}{11}, \frac{2}{7}, \frac{1}{12}, 5, \frac{3}{5}, \frac{1}{13}, \frac{4}{3}, \frac{1}{14}, \frac{2}{9}, \frac{1}{15}, \frac{1}{16}, \frac{5}{2}, \frac{1}{17}, \ldots$$

> Is it possible to explain this sequence without polynomial root calculations?

3.3 Related works and perspectives

Here I list some open research pathways, in addition to what has already been said in this Section, in Section 2.5 about RNA structures and in Subsection 3.1.3 about the wonderland bijection quest. I will also briefly mention my other works from master, PhD and postdoc periods.

3.3.1 Gray codes for q-decreasing structures

The recent Wong-Liu-Lam-Im progress [246] in the resolution of the extended version of Eğecioğlu-Iršič conjecture about the existence of 1-Gray code for q-decreasing words with natural q indicates that this problem may be solved in the future, perhaps by a method much more elegant than our solution for the case $q = 1$.

Another thought-intriguing direction is to explore the power of machine-verifiable proofs using, for instance, a Why3-and-Coq-based approach described by Giorgetti, Dubois, and Lazarini [126].

The discovery of q-decreasing words and related Gray codes motivated further study of similar objects. Hassler, Vajnovszki and Wong [142] provided, among other things, captivating Gray codes for the set of length n binary words of weight k (i.e. with exactly k 1s) with the property that any prefix contains at least p times as many 0s as 1s, for any $p \in \mathbb{R}^+$.

3.3.2 From Fibonacci to Catalan and back again

In 2022, Barcucci, Bernini and Pinzani invited us to *un viaggio combinatorio di esploratori* connecting Fibonacci, Catalan and various related combinatorial objects [22, 23]. Intriguing bijective and enumerative links between certain subsets of Dyck paths, integer compositions and q-decreasing words have been discovered by them along with Bilotta [19, 21]. They found restricted Dyck path interpretations of the q-decreasing words. Barcucci, Bernini, Bilotta and Pinzani in another paper [20] give a bijection between q-decreasing words and a new special class of pattern avoiding binary words. It would be interesting to study Grey codes for these new Fibonacci objects and find new interesting bijections.

3.3.3 Pattern transferring

In 2020, with Richard Genestier and Jean-Luc Baril we investigated [28] pattern transfer between constrained Dyck paths, discussed in Section 3.1, and Motzkin paths. Because of the intricate structure of the constrained

Dyck paths, the direct study of their patterns seems complicated. Thanks to our bijection which "converts" patterns in constrained Dyck paths to patterns in usual Motzkin paths, we managed to enumerate patterns in constrained Dyck paths by enumerating patterns in their bijective Motzkin images. Let me explain a little more about this method and notations used.

A *statistic* on a set S of combinatorial objects is an association of an integer in $\mathbb{Z}$ to each object in S. For instance the map that returns the number of valleys DU in a Dyck paths is a statistic. We denote by $\hat{\mathbf{1}}$ (resp. $\hat{\mathbf{2}}, \hat{\mathbf{n}}$) the constant statistic that sends any object to 1 (resp. $\hat{\mathbf{2}}, \hat{\mathbf{n}}$). Let $\mathcal{S}$ be the set of all statistics on a set S. For $\mathbf{X}, \mathbf{Y} \in \mathcal{S}$, we define the statistic $\mathbf{X} + \mathbf{Y}$ so that $(\mathbf{X} + \mathbf{Y})(P) = \mathbf{X}(P) + \mathbf{Y}(P)$ for any $P \in S$, which endows $\mathcal{S}$ with a $\mathbb{Z}$-module structure. Let S and T be two sets of combinatorial objects, and let $\mathcal{S}$ and $\mathcal{T}$ be the associated statistic sets. Two statistics $\mathbf{X} \in \mathcal{S}$ and $\mathbf{Y} \in \mathcal{T}$ have the *same distribution* if and only if there exists a bijection f from S to T such that for any $P \in S$ we have $\mathbf{X}(P) = \mathbf{Y}(f(P))$. In this case, we say that f transports the statistic $\mathbf{X}$ into $\mathbf{Y}$, which can be shortly written with the statistic equation $f(\mathbf{X}) = \mathbf{Y}$ (or $\mathbf{X} = \mathbf{Y}$ whenever f is the identity). To a given pattern P, we associate the pattern statistic $\mathbf{P} : \mathcal{A} \to \mathbb{N}$ such that $\mathbf{P}(a)$ is the number of occurrences of the pattern P in the object $a \in \mathcal{A}$. For instance, for Dyck paths of semilength n we have $\mathbf{U} + \mathbf{D} = 2 \cdot \hat{\mathbf{n}}$.

I present below the first Theorem from [28]. Other results are more involved, in some cases the systems of such pattern equations appear, and they are proved to be useful (see Section 3 of [28]).

Theorem 39. *For $n \geqslant 0$, the bijection φ (see Eq.(3.1) on page 53, Subsection 3.1.2) from $\mathcal{D}_n^{h,\geqslant}$ to $\mathcal{M}_n$ transports statistics associated to patterns of length two as follows:*

$$\varphi(\mathbf{UD}) = \mathbf{F} + \mathbf{UD}, \tag{3.5}$$

$$\varphi(\mathbf{UU}) = \varphi(\mathbf{DD}) = \mathbf{U} + \mathbf{UU} + \mathbf{UF}, \tag{3.6}$$

$$\varphi(\mathbf{DU}) = \mathbf{FF} + \mathbf{FU} + \mathbf{DF} + \mathbf{DU}. \tag{3.7}$$

The line (3.5) means that every peak UD in Dyck path $a \in \mathcal{D}_n^{h,\geqslant}$ corresponds to a flat step F or to a peak UD in the Motzkin image $\varphi(a)$ and vice versa. See Figure 3.5 on page 53 for some examples.

Although the scientists use different notations to convey information about how one pattern metamorphoses into another, the essence of this method imbues the combinatorial literature. It is related to the great study of equidistributed statistics, started by a seminal work of MacMahon [186]. I would even say that a pattern-transferring bijection is in some sense similar to the notion of homomorphism. Knowledge or hypotheses about pattern and statistic transfers, in some cases, help us find the desired constructive bijections! Consider the FindStat[7], an on-line database of combinatorial statistics initiated by Chris Berg and Christian Stump; maintained by Martin Rubey, Tilman Möller and others [53]. As of April 17, 2025, it contains 329 maps and 1964 statistics, which can be searched by keywords or by examples. See also the Combinatorial Object Server[8], first launched by Frank Ruskey in 2000, and relaunched in 2018 by Torsten Mütze, Joe Sawada and Aaron Williams [217]. It allows us to generate combinatorial objects, explore combinatorial algorithms and download their free source code.

During the lockdown, on June 2020, I presented our work entitled "Pattern distribution in faro words and permutations" [29] at the on-line edition of the conference "Permutation Patterns". In this work, in collaboration with Jean-Luc Baril, we have identified a new type of permutations and words, and studied bijective correspondences to other objects. We conjectured that a certain subset of our words of length n is in bijection with Dumont permutations of length $2n$ avoiding pattern 2143. After my presentation, during a session of questions, Alexander Burstein became interested in our conjecture. Later we published together an article resolving the conjecture [27]. The same paper also presents examples of pattern correspondences. Let me say a little more about this.

A k-ary word $w = w_1 w_2 \dots w_n$ avoids a classical pattern (resp. consecutive pattern) $p = p_1\text{-}p_2\text{-}\cdots\text{-}p_k$ (resp. $p = p_1 p_2 \dots p_k$) if there does not exist a strictly increasing sequence of indices $i_1 i_2 \dots i_k$ (resp. with $i_{j+1} = i_j + 1$ for $1 \leqslant j \leqslant k-1$) such that $w_{i_1} w_{i_2} \dots w_{i_k}$ is order-isomorphic to p (see Kitaev's book [164] for more information).

Faro permutations are those permutations (words of length n consisting of all distinct letters from $[1, n]$) that avoid three consecutive patterns 231, 321 and 312, which is equivalent to the fact that there is no i such that $\pi_i > \pi_{i+2}$. A dispersed Dyck path [137], closely related to famous Bertrand's ballot problem, is a lattice path never going below the x-axis, starting at $(0, 0)$, ending at $(n, 0)$, consisting of up steps $U = (1, 1)$, down steps $D = (1, -1)$ and flat steps $F = (1, 0)$ lying on the x-axis. In [27] we constructed a bijection between faro permutations and dispersed Dyck paths. Let's have a look at this bijection for small faro permutations, where U, F, D steps are depicted respectively by /, _, and \:

```
                                          /\                             /\
_____   ___/\   __/\_   _/\__   _/\/\   _/  \   /\___   /\_/\   /\/\_   /  \_
12345   12354   12435   13245   13254   14253   21345   21354   21435   31425
```

[7] https://www.findstat.org/
[8] http://combos.org/

The bijection g from [27, Theorem 3.4] transfers consecutive permutation patterns to consecutive patterns in paths. For instance, any occurrence of the consecutive pattern 21 (a descent) in a faro permutation corresponds to the pattern U in the path, obtained from the permutation applying the bijection; every occurrence of a consecutive pattern 132 corresponds to an occurrence of FU, UU or DU, so we have the following equation:

$$g(\mathbf{132}) = \mathbf{DU} + \mathbf{UU} + \mathbf{FU}.$$

Occurrences of path patterns are literally equal, while occurrences of permutation patterns are equal only modulo order isomorphism. A path $FUDUD$ contains 1 occurrence of DU and 1 occurrence of FU, and the corresponding permutation 13254 contains 2 consecutive occurrences of 132, videlicet 132 and 254.

Similar analysis can be done for other patterns. For example, we have

$$g(\mathbf{123}) = \hat{\mathbf{n}} - \hat{\mathbf{2}} - \mathbf{FU} - \mathbf{UU} - 2 \cdot \mathbf{DU} - \mathbf{DF} - \mathbf{DD},$$

where n is the length of the corresponding lattice path. In fact, the existing notations for pattern equations in the literature, including our own works [27,28], are usually more sophisticated and use additional symbols. It increases the technical accuracy but obscures the conceptual. Inspired by Occam's Razor, I seek for simplest possible yet powerful notation, reflecting only the important information.

A result from [27] implies that the expected number of the occurrences of the consecutive pattern 21 (respectively 12, 132 and 213) in a randomly selected faro permutation of length n is asymptotically equivalent to $n/2$. Surprisingly, the expectation of 123 is asymptotically equivalent to $\sqrt{\pi n/2}$. And we can conclude the probability of an occurrence of 123 at a random position in a random faro permutation approaches 0 as n grows. Inspired by this result, with Nathanaël, we are currently working on a new article about emerging consecutive pattern avoidance in permutations, the results of which were presented at the conference Permutation Patterns, held in St Andrews, Scotland, in July 2025 [141].

Other examples of pattern transfers can be found in our paper about Dyck paths with air pockets [32], which exposes also a twirly link with Fibonacci meanders, a kind of closed smooth self-overlapping curves. See Luschny's [184] and Wienand's [240] posts in OEIS Wiki about such meanders. You may also consider our work about a transformation *à la Foata* for special kinds of descents and excedances in permutations [30] and related paper by Han, Mao and Zeng [139] who prove, among other things, Vajnovszki's conjecture from [30]. An engaging trinomial relation [37] concerning the number of cells lying at different levels of Motzkin polyominoes of size n and first n terms of the expanded $(1 + x + x^2)^n$ can also be seen as an example of pattern transferring, or statistic transferring method.

It is worth continuing to look both for new results on pattern transfers and for convenient ways of presenting them.

3.3.4 Pattern-aware structural decompositions

Many works in the young field of pattern enumeration use *ad-hoc* (in the good sense of the word) methods. This is due to the fact that structures and their properties are highly variable, and a small change in the initial conditions, in the shape of the *basic bricks*, in the rules of combinatorial construction, often implies huge differences in the large structures. So we need a very high level of precision and impeccable rigor to find the right results.

There are, however, certain similarities in the different questions and the *ad-hoc* methods developed to solve them, we are starting to see a bigger and more coherent picture. Here are some examples of works in this direction:

- Guibas and Odlyzko [136] proposed a generic method for enumerating words avoiding a given pattern.

- Banderier and Flajolet [18] developed a unified theory of lattice paths in half- and quarter-planes, focusing on enumerative and asymptotic aspects. Extending the ideas of this article, Asinowski, Bacher, Banderier, Gittenberger and Roitner in a series of recent papers [6–10] presented a substantial study about enumeration of consecutive patterns in lattice paths.

- Using a recursive approach, Bean, Bernini, Cervetti and Ferrari [49] provided an algorithm for finding a combinatorial specification and a generating function for Motzkin paths avoiding a given set of not necessarily consecutive patterns. In the case of a single pattern, they also showed that obtained generating function is rational over x and $\frac{1-\sqrt{1-4x^2}}{2x^2}$. Their work is a natural continuation of the work by Bacher, Bernini, Ferrari, Gunby, Pinzani and West [17] about patterns in Dyck paths.

- Classical Goulden-Jackson cluster method [129, 130] is used to study the distribution of patterns in objects represented by sequences of other objects. The works of Wang [237] and Zhuang [250] are examples of applications of this method in the context of lattice paths. See also works by Noonan and Zeilberger [198]; Bassino, Clément and Nicodème [45], together with Fayolle [44]; Elizalde and Noy [105].

I would like to try to contribute to this process of pattern understanding, for example by combining the theory of formal languages with the combinatorial study of patterns. Formal grammars can be used to generate combinatorial objects.

If we can generate, we should be able to understand,

understand not only the objects themselves
but also how the patterns are formed therewithin.

It would be nice to find a method to automatically enumerate patterns. We could extend the existing studies by considering, at first, consecutive patterns in even binary words (with an even number of 1s, such language is an example of regular language that cannot be characterized by a pattern avoidance, thus, we should do more than the classical Goulden-Jackson method allows). Next, extend the results to any regular language, and thereafter look at context-free and more complex grammars, for consecutive and not necessarily consecutive patterns, with or without order isomorphism.

3.3.5 Structure and enumeration of lattice intervals

The word *lattice*, aptly, have several meanings. When we talk about lattice paths, we mean by lattice a repeating arrangement of points. In case of Dyck and Motzkin paths it is $\mathbb{N}^2$. In order theory, by lattice we understand a partially ordered set (*a poset* for short) with an additional property: each pair of elements has a least upper bound and a greatest lower bound.

If the elements of such partially ordered sets are lattice paths... we should speak about *lattice path lattices*, which reminds me Wittgenstein's language games [183] and his understanding of *grammar*.

Different order structures can be defined over Dyck paths. Perhaps the most known of them is the Tamari lattice [114, 150]. Suppose α is a Dyck path, possibly empty. The Tamari cover relation transforms a Dyck path $\beta DU\alpha D\gamma$ into $\beta U\alpha DD\gamma$. For illustrations, consider Figures 3.18 and 3.19. The lattice order relation is the transitive closure of this cover relation. The Tamari lattice can be described in many equivalent ways, using other objects instead of Dyck paths, see a beautiful book edited by Müller, Pallo and Stasheff [196].

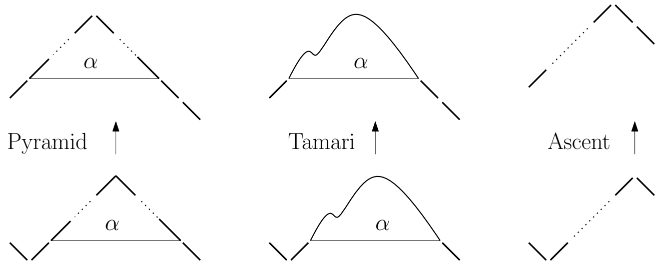


Figure 3.18: Pyramid, Tamari and ascent cover relations.

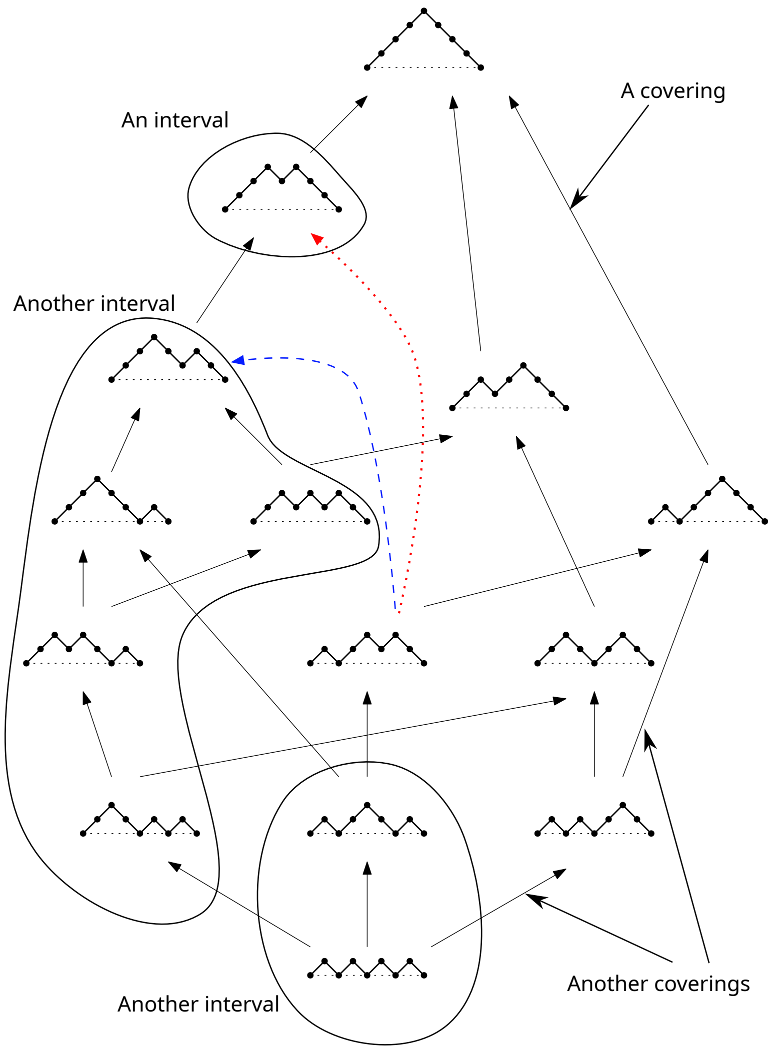


Figure 3.19: Hasse diagram for Pyramid, Tamari (with the dotted red edge) and ascent (with dashed blue edge) lattices of Dyck paths of semilength 4.

In a poset $\mathcal{A}$, for any two comparable elements a and b, the set $\{c \in \mathcal{A}, a \leqslant c \leqslant b\}$ is called an *interval*. The *join* of two elements from $\mathcal{A}$ is their least upper bound. Similarly, the *meet* of two elements from $\mathcal{A}$ is their greatest lower bound. A *cover relation (covering)* between a and b takes place if $a \leqslant b$ and $a \neq b$ and there are no other elements between them.

Together with Jean-Luc Baril and Mehdi Naima we introduced [33, 34] another lattice structure on Dyck paths. In this case, a cover relation transports pyramids: $DU^kD^k \mapsto U^kD^kD, k \geqslant 0$. It is a sublattice of Tamari lattice, i.e. any comparable elements in the pyramid lattice are also comparable in Tamari way, but the converse is not true (see Figures 3.18 and 3.19). We explained this fact and presented other results: about the number of intervals; coverings; meet and join irreducible elements; outgoing and incoming edges in the Hasse diagram (see Fig. 3.19 for examples). We also gave an involution on Dyck paths that transports the bistatistic of the numbers of outgoing and incoming edges to its reverse. Of the several questions posed in the article, I would like to point out one that we are currently working on with Enrica Duchi, Jean-Luc Baril, Mehdi Naima and Khaydar Nurligareev:

> Algebraically, we know that the number of intervals in the pyramid lattice of Dyck paths of semilength n is equal to the number of Dyck paths of semilength n where each up-step not at ground level comes in two colors, OEIS sequence A64062:
>
> $$1, 3, 13, 67, 381, \ldots$$
>
> Is it possible to find a combinatorial bijection between these two sets?

Yet another cover relation, $DU^kD \mapsto U^kDD, k \geqslant 0$, engenders *the ascent lattice*, which was considered by Clément Chenevière under the name “Greedy Dyck lattice”, GDyck_n, in his PhD thesis [85, Section 7.2], following a suggestion of Philippe Nadeau. In our recent work [26] with Mireille Bousquet-Mélou, Jean-Luc Baril and Mehdi Naima we look at the ascent lattice and its generalizations.

The generating function for the number of intervals in the ascent lattice is found to be algebraic of degree 3. The proof is based on a recursive decomposition of intervals involving two catalytic parameters. The solution of the corresponding functional equation is inspired by recent work on the enumeration of walks confined to a quadrant related to invariant approach based on ideas developed by Tutte in his enumeration of coloured maps [231, 232]. See also a captivating treatise by Bernardi, Bousquet-Mélou and Raschel [57] devoted to the invariant approach.

In the same paper, we consider order structures over m-Dyck paths of semilength mn in which all ascent lengths are multiples of m, and the similar structures on mirrored m-Dyck paths. We show that the first poset $\mathbb{D}_{m,n}$ is a lattice for any m, while the second poset $\mathbb{D}'_{m,n}$ is only a join semilattice when $m > 1$. In both cases, the enumeration of intervals is still described by an equation in two catalytic variables.

Consider a word of comparable elements and read it from right to left in order to construct a binary search tree. The rightmost element is the root. Next, recursively, larger elements go to the right subtree and smaller to the left. What to do with the elements that we have already placed in the tree? In case of a right strict binary search tree, considered by Hivert, Novelli and Thibon [145], the elements already seen in sequence go to the left. We can obtain the same trees, see Figure 3.20 for instance.

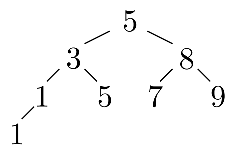


Figure 3.20: A right strict binary search tree constructed from words 15139785 and 97851135.

Words that give rise to the same tree are said to lie in the same *sylvester class*[9]. The *sylvester congruence*, on words is generated by the commutation relations

$$ac \cdots b \equiv ca \cdots b, \quad a \leqslant b < c.$$

It is known that the words $w = w_1 \cdots w_n$ avoiding the not necessarily consecutive patterns 121 and 132, form a set of representatives of sylvester classes [199, Sec. 2.7]: every class contains such a word, and two distinct words of this type are never congruent. Here the pattern avoidance is considered modulo order isomorphism, i.e. we cannot find $1 \leqslant i < j < k \leqslant n$ such that either $w_i = w_k < w_j$ or $w_i < w_k < w_j$.

In paper [26] we present a bijection between intervals in $\mathbb{D}'_{m,n}$ and sylvester classes of m-parking functions of lengths n (certain kinds of words), considered in [145, 199]. We also explain connections with Nadeau-Tewari [197] poset NT. Certain parts of NT are isomorphic to $\mathbb{D}_{m,n}$ and $\mathbb{D}'_{m,n}$, some intervals in NT correspond to sylvester classes of certain words.

Furthermore, we provide bijections between intervals in $\mathbb{D}_{m,n}$ and $\mathbb{D}'_{m,n}$ and certain quadrant walks. Combining these bijections with probabilistic

[9]Adjective *sylvestre* comes from Latin *silva* ("woods, forest"), an indirect tribute to the great algebraic combinatorist James Joseph Sylvester.

results on such walks we give asymptotic estimates of the number of intervals in $\mathbb{D}_{m,n}$ and $\mathbb{D}'_{m,n}$. Their form implies that the generating functions of intervals are no longer algebraic, nor even D-finite (i.e. not a solution of a system of linear differential equations with polynomial coefficients), when $m > 1$.

There are many other poset structures on Dyck paths, for example: Stanley lattice (or Dyck lattice [85]), the Kreweras lattice [56,170], the greedy version of Tamari lattice due to Dermenjian [95], the alt-Tamari lattices of Chenevière [84]. Figure 3.21 presents the subposet-inclusion structure of these posets.

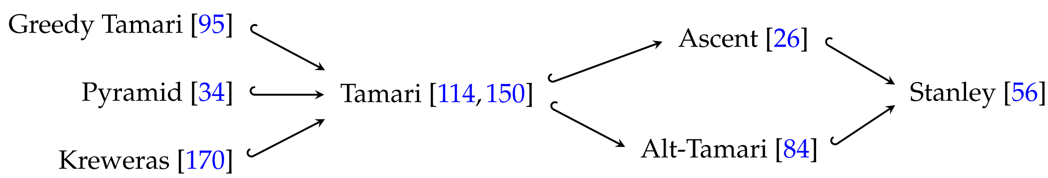


Figure 3.21: Subposet-inclusion structure of some orders on Dyck paths.

In an ongong work, together with Nathanaël Hassler and Jean-Luc Baril we are exploring the intervals in a new lattice induced by a bitwise comparison on q-decreasing words from Section 3.2.

It is a good idea to think about whether we can better understand the structure and count the intervals in the *autocorrelation lattice* from Section 1.5 (see Figure 1.5 on page 13). It is also interesting to consider the poset of autocorrelation polynomials for endhered patterns from Section 2.3.

3.3.6 On graphs and networks

There are several connections between the domain of telecommunications networks and combinatorics, see the seminal papers by Hsu [147,148] about Fibonacci cube, his work together with Page and Liu [149] and a book by Eğecioğlu, Klavžar and Mollard [107].

Together with Nader Mbarek and our PhD student Abdelhamid Garah, during his thesis, we published a paper [120] in the *Internet of Things* journal and made several conference presentations (SITIS 2022 [117], LCN 2023 [118] and DASC 2023 [119]) about the autonomous management of security services in the context of Internet of Things. Two other communications were recently presented at NTMS 2025 and IWCMC 2025 [121, 122]. It was an interesting collaboration, which helped me to learn more about the IoT domain. Many combinatorial research questions have yet to be asked at the borderland between practice and theory.

During my master's internships, I have been working on reinforcement learning for autonomic resource allocation in clouds, together with Dutreilh, Melekhova, Malenfant, Rivierre and Truck [102].

At the time of my PhD, I've been studying the Internet topology dynamics from network theoretical and statistical points of view [153]. See our works with Magnien, Medem and Tarissan [160–162, 187, 192].

I've had the opportunity to work Queyroi [209] on hierarchical graph clustering, with Bensmail and Duvignau on the complexity of deciding whether a graph admits an orientation with fixed weak diameter [51, 52]; with Gastineau, Gras and Omidvar [123] on packing coloring and subsets preserving path distance.

At one time I was also involved in the study of social networks and the discourses within them, see my works with Kondrashova and Frame [167]; with Azaza, Savonnet, Leclercq and Faiz [12, 13, 15]; with Azaza, Savonnet, Leclercq, Gastineau and Faiz [14]; with Azaza, Leclercq, Savonnet and Frame [11]; with Leclercq, Savonnet, Frame and Basaille [159]; with Leclercq, Savonnet, Grison and Basaille [174]; with Leclercq, Savonnet, Cullot, Grison and Gavignet [43]; with Danisch and Leclercq [158]; with Basaille, Leclercq, Savonnet and Cullot [42]. See also our works [156, 157] about temporal density of complex networks and ego-community dynamics.

I keep communicating with my colleagues from different branches of science, recently we published a paper about a clique percolation method for community detection, co-authored with Baudin, Danisch, Magnien and Ghanem [46]. My website contains papers, slides of talks and posters: https://kirgizov.link.

3.3.7 Verses about the language and combinatorics

Суть

Все проявленья человека,
Включая слова и числа,
Являют собой преображенья,
Карикатуры смысла.

Essence

Toutes nos expressions,
Les mots et les quantités,
Ne sont que des réflexions,
Des caricatures de nos idées.

Crux

All human expressions,
Including numbers and words,
Are like bizarre reflections,
Caricatures of thoughts.

Bibliography

[1] N. Ali, C. Wolf, S. Kanchan, S. R. Veerabhadraiah, L. Bond, M. W. Turner, C. L. Jorcyk, and G. Hampikian. 9S1R nullomer peptide induces mitochondrial pathology, metabolic suppression, and enhanced immune cell infiltration, in triple-negative breast cancer mouse model. *Biomedicine & Pharmacotherapy*, 170:115997, 2024.

[2] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[3] J. E. Andersen, J. Mattes, and N. Reshetikhin. The poisson structure on the moduli space of flat connections and chord diagrams. *Topology*, 35(4):1069–1083, 1996.

[4] J. Arndt. *Matters Computational: Ideas, Algorithms, Source Code*. Springer, Berlin, Heidelberg, 2011.

[5] A. Asinowski. *Geometric Permutations in the Plane and in Euclidean Spaces of Higher Dimension*. PhD thesis, Technion-Israel Institute of Technology, 2005.

[6] A. Asinowski, A. Bacher, C. Banderier, and B. Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns: Asymptotic aspects and Borges's theorem. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018)*, volume 110 of *LIPIcs*, pages 10:1–10:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[7] A. Asinowski, A. Bacher, C. Banderier, and B. Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns: enumerative aspects. In *International Conference on Language and Automata Theory and Applications*, volume 10792 of *Lecture Notes in Computer Science*, pages 195–206. Springer, 2018.

[8] A. Asinowski, A. Bacher, C. Banderier, and B. Gittenberger. Analytic combinatorics of lattice paths with forbidden patterns, the vectorial kernel method, and generating functions for pushdown automata. *Algorithmica*, 82(3):386–428, 2020.

[9] A. Asinowski and C. Banderier. On lattice paths with marked patterns: Generating functions and multivariate gaussian distribution. In M. Drmota and C. Heuberger, editors, *31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2020)*, volume 159 of *LIPIcs*, pages 1:1–1:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[10] A. Asinowski, C. Banderier, and V. Roitner. Generating functions for lattice paths with several forbidden patterns. In *Proceedings of Formal Power Series and Algebraic Combibatorics (FPSAC'20)*, volume 84B of *Séminaire Lotharingien de Combinatoire*, 2020.

[11] L. Azaza, S. Kirgizov, E. Leclercq, M. Savonnet, and A. Frame. Evaluation de l'influence sur twitter: Application au projet "twitter aux elections européennes 2014". In *Etudier le Web politique: regards croisés*, 2015.

[12] L. Azaza, S. Kirgizov, M. Savonnet, E. Leclercq, and R. Faiz. Influence assessment in twitter multi-relational network. In *Signal Image Technologies and Internet-Based System (SITIS), Eleventh International IEEE Conference on*. IEEE, 2015.

[13] L. Azaza, S. Kirgizov, M. Savonnet, E. Leclercq, and R. Faiz. Évaluation de l'influence dans un réseau multi-relationnel : le cas de twitter. In *INFORSID'2016, Le congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision)*, 06 2016.

[14] L. Azaza, S. Kirgizov, M. Savonnet, É. Leclercq, N. Gastineau, and R. Faiz. Information fusion-based approach for studying influence on twitter using belief theory. *Computational Social Networks*, 3(1):5, 2016.

[15] L. Azaza, M. Savonent, E. Leclercq, S. Kirgizov, and R. Faiz. Évaluation de l'influence polarisée dans un réseau multirelationnel. Application à twitter. *Document Numérique*, 20(1):67–100, 2017.

[16] A. Bacher. *Chemins et animaux : applications de la théorie des empilements de pièces*. PhD thesis, Université Bordeaux 1, 2011.

[17] A. Bacher, A. Bernini, L. Ferrari, B. Gunby, R. Pinzani, and J. West. The Dyck pattern poset. *Discrete Mathematics*, 321:12–23, 2014.

[18] C. Banderier and P. Flajolet. Basic analytic combinatorics of directed lattice paths. *Theoretical Computer Science*, 281(1-2):37–80, 2002.

[19] E. Barcucci, A. Bernini, S. Bilotta, and R. Pinzani. Dyck paths enumerated by the $\mathbb{Q}$-bonacci numbers. In S. Brlek and L. Ferrari, editors, Proceedings of the 13th edition of the conference on *Random Generation of Combinatorial Structures. Polyominoes and Tilings*, Bordeaux, France, 24-28th June 2024, volume 403 of *Electronic Proceedings in Theoretical Computer Science*, pages 49–53. Open Publishing Association, 2024.

[20] E. Barcucci, A. Bernini, S. Bilotta, and R. Pinzani. Pattern avoiding and q-decreasing binary words. *RAIRO - Theoretical Informatics and Applications*, 59:13, 2025.

[21] E. Barcucci, A. Bernini, S. Bilotta, and R. Pinzani. Rational Dyck paths. *Journal of Integer Sequences*, 28:Article 25.3.2, 2025.

[22] E. Barcucci, A. Bernini, and R. Pinzani. From the Fibonacci to Pell numbers and beyond via Dyck paths. *Pure Mathematics and Applications (PU.M.A.)*, 30:17–22, 2022.

[23] E. Barcucci, A. Bernini, and R. Pinzani. Sequences from Fibonacci to Catalan: A combinatorial interpretation *via* Dyck paths. *RAIRO - Theoretical Informatics and Applications*, 58:8, 2024.

[24] E. Barcucci, A. Del Lungo, A. Frosini, and S. Rinaldi. A technology for reverse-engineering a combinatorial problem from a rational generating function. *Advances in Applied Mathematics*, 26(2):129–153, 2001.

[25] J.-L. Baril. Avoiding patterns in irreducible permutations. *Discrete Mathematics & Theoretical Computer Science*, 17(3):2158, 2016.

[26] J.-L. Baril, M. Bousquet-Mélou, S. Kirgizov, and M. Naima. The ascent lattice on Dyck paths. *The Electronic Journal of Combinatorics*, 32(2):P2.36, 2025.

[27] J.-L. Baril, A. Burstein, and S. Kirgizov. Pattern statistics in faro words and permutations. *Discrete Math.*, 344(8):Paper No. 112464, 17, 2021.

[28] J.-L. Baril, R. Genestier, and S. Kirgizov. Pattern distributions in Dyck paths with a first return decomposition constrained by height. *Discrete Math.*, 343(9):111995, 9, 2020.

[29] J.-L. Baril and S. Kirgizov. Pattern distribution in faro words and permutations. Permutation Patterns conference. Valparaiso University, 2020.

[30] J.-L. Baril and S. Kirgizov. Transformation à la Foata for special kinds of descents and excedances. *Enumerative Combinatorics and Applications*, 1(3):Article #S2R19, Mar 2021.

[31] J.-L. Baril and S. Kirgizov. Lattice paths with a first return decomposition constrained by the maximal height of a pattern. *Discrete Mathematics Letters*, 8, 2022.

[32] J.-L. Baril, S. Kirgizov, R. Maréchal, and V. Vajnovszki. Enumeration of Dyck paths with air pockets. *Journal of Integer Sequences*, 26, 2023.

[33] J.-L. Baril, S. Kirgizov, and M. Naima. A lattice on Dyck paths close to the Tamari lattice. In *Computational Logic and Applications*, Kraków, Poland, Dec 2023. Jagiellonian University.

[34] J.-L. Baril, S. Kirgizov, and M. Naima. A lattice on Dyck paths close to the Tamari lattice. *Séminaire Lotharingien de Combinatoire*, 89, 2025.

[35] J.-L. Baril, S. Kirgizov, and A. Petrossian. Dyck paths with a first return decomposition constrained by height. *Discrete Mathematics*, 341(6):1620–1628, 2018.

[36] J.-L. Baril, S. Kirgizov, and A. Petrossian. Motzkin paths with a restricted first return decomposition. *Integers*, 19:Paper No. A46, 19, 2019.

[37] J.-L. Baril, S. Kirgizov, J. L. Ramírez, and D. Villamizar. The combinatorics of motzkin polyominoes. *Discrete Applied Mathematics*, 364:1–15, 2025.

[38] J.-L. Baril, S. Kirgizov, and V. Vajnovszki. Patterns in treeshelves. *Discrete Mathematics*, 340(12):2946–2954, 2017.

[39] J.-L. Baril, S. Kirgizov, and V. Vajnovszki. Asymptotic bit frequency in Fibonacci words. *Pure Mathematics and Applications (PU.M.A.)*, 30(1):23–30, 2022.

[40] J.-L. Baril, S. Kirgizov, and V. Vajnovszki. Gray codes for Fibonacci q-decreasing words. *Theoretical Computer Science*, 927:120–132, 2022.

[41] J.-L. Baril, R. Maréchal, S. Kirgizov, and V. Vajnovszki. Grand Dyck paths with air pockets. *The Art of Discrete and Applied Mathematics*, 7(1):#P1.07, Sep 2023.

[42] I. Basaille, S. Kirgizov, E. Leclercq, M. Savonnet, and N. Cullot. Towards a Twitter Observatory : A multi-paradigm framework for collecting, storing and analysing tweets. In *RCIS 2016, IEEE Tenth International Conference on Research Challenges in Information Science*, pages 1–10, 2016.

[43] I. Basaille, E. Leclercq, M. Savonnet, N. Cullot, S. Kirgizov, T. Grison, and E. Gavignet. Un observatoire pour la modélisation et l'analyse des réseaux multi-relationnels. une application à l'étude du discours politique sur twitter. *Document Numérique*, 20(1):101–136, 2017.

[44] F. Bassino, J. Clément, J. Fayolle, and P. Nicodème. Counting occurrences for a finite set of words: an inclusion-exclusion approach. In P. Jacquet, editor, *2007 Conference on Analysis of Algorithms, AofA 07*, volume AH of *DMTCS Proceedings*, pages 31–46, Juan les Pins, France, 2007. Discrete Mathematics and Theoretical Computer Science.

[45] F. Bassino, J. Clément, and P. Nicodème. Counting occurrences for a finite set of words: combinatorial methods. *ACM Transactions on Algorithms*, 8:31:1–31:28, Jul 2012.

[46] A. Baudin, M. Danisch, S. Kirgizov, C. Magnien, and M. Ghanem. Clique percolation method: Memory efficient almost exact communities. In B. Li, L. Yue, J. Jiang, W. Chen, X. Li, G. Long, F. Fang, and H. Yu, editors, *Advanced Data Mining and Applications*, pages 113–127, Cham, 2022. Springer International Publishing.

[47] J.-M.-E. Baudot. Système de télégraphie rapide, June 1874. Patent Brevet 103,898.

[48] C. Bean. *Finding structure in permutation sets*. PhD thesis, Háskólinn í Reykjavík, Reykjavik University, 2018.

[49] C. Bean, A. Bernini, M. Cervetti, and L. Ferrari. On the generating functions of pattern-avoiding Motzkin paths. *Journal of Symbolic Computation*, 113:126–138, 2022.

[50] M. Bendkowski, O. Bodini, and S. Dovgal. Matchings and partial patterns. *The Electronic Journal of Combinatorics*, 26(4):P4.1, 2019.

[51] J. Bensmail, R. Duvignau, and S. Kirgizov. On the complexity of turning a graph into the analogue of a clique. 9th International colloquium on graph theory and combinatorics, 2014.

[52] J. Bensmail, R. Duvignau, and S. Kirgizov. The complexity of deciding whether a graph admits an orientation with fixed weak diameter. *Discrete Mathematics and Theoretical Computer Science*, 17(3):31, 2016.

[53] C. Berg, M. Rubey, T. Stump, Christian Möller, et al. FindStat - The combinatorial statistics database, 2011. http://www.FindStat.org. Accessed: 17 April 2025.

[54] H. M. Berman, C. L. Lawson, and B. Schneider. Developing community resources for nucleic acid structures. *Life*, 12(4):540, 2022.

[55] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[56] O. Bernardi and N. Bonichon. Intervals in Catalan lattices and realizers of triangulations. *Journal of Combinatorial Theory, Series A*, 116(1):55–75, 2009.

[57] O. Bernardi, M. Bousquet-Mélou, and K. Raschel. Counting quadrant walks via Tutte's invariant method. *Combinatorial Theory*, 1, 2021.

[58] A. Bernini. *Some properties of pattern avoiding permutations*. PhD thesis, Università degli Studi di Firenze, 2006.

[59] J. I. Bernoulli. Sur une nouvelle espece de calcul. *Recueil pour les Astronomes*, 1:255–284, 1771.

[60] J. Berstel, A. Lauve, C. Reutenauer, and F. V. Saliola. *Combinatorics on words: Christoffel words and repetitions in words*, volume 27 of *CRM Monograph Series*. American Mathematical Society, 2008.

[61] D. I. Bevan. *On the growth of permutation classes*. PhD thesis, The Open University, 2015.

[62] C. Biane, G. Hampikian, S. Kirgizov, and K. Nurligareev. Endhered patterns in matchings and RNA. *Journal of Computational Biology*, 32(1):28–46, 2025.

[63] C. Biane, G. Hampikian, S. Kirgizov, K. Nurligareev, and D. Pinson. Distribution of endhered patterns in RNA-related secondary structures. In *SeqBIM*. GDR BIMMM, GDR IFM, Rennes, 2025.

[64] S. Bilotta. *Pattern avoiding words and cross-bifix-free sets: a general approach*. PhD thesis, Università degli Studi di Firenze, 2013.

[65] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.

[66] J. Bloom and S. Elizalde. Pattern avoidance in matchings and partitions. *The Electronic Journal of Combinatorics*, 20(2), 2013.

[67] O. Bodini, A. Singh, and N. Zeilberger. Sampling β-normal linear λ-terms. *Pure Mathematics and Applications (PU.M.A), Proceedings of GASCom 2022: Generation Aleatoire de Structures Combinatoires*, 30(1):45–55, 2022.

[68] O. Bodini, A. Singh, and N. Zeilberger. Asymptotic distribution of parameters in trivalent maps and linear lambda terms. *Combinatorial Theory*, 5(2):#13, 2025.

[69] M. Borinsky. Generating asymptotics for factorially divergent sequences. *Electronic Journal of Combinatorics*, 25(4):p4.1, 32, 2018.

[70] W. Bosma, J. Cannon, and C. Playoust. The magma algebra system i: The user language. *Journal of Symbolic Computation*, 24(3):235–265, 1997.

[71] M. Bouvel. *Quelques problèmes combinatoires et algorithmiques sur les classes de permutations*. PhD thesis, Paris 7 — Université Paris Diderot, Università degli Studi di Firenze, 2009.

[72] A. Brocot. Calcul des rouages par approximation, nouvelle méthode. *Revue Chronométrique*, 3, 1861.

[73] A. Brocot. *Calcul des rouages par approximation, nouvelle méthode*. 1862.

[74] A. Burstein. *Enumeration of words with forbidden patterns*. PhD thesis, University of Pennsylvania, 1998.

[75] M. Bóna. *A Walk Through Combinatorics: An Introduction to Enumeration and Graph Theory*. World Scientific, 2002.

[76] M. Bóna. *Combinatorics of Permutations*. Chapman and Hall/CRC, 3 edition, 2022.

[77] R. Campoamor-Stursberg and V. O. Manturov. Invariant tensor formulas via chord diagrams. *Journal of Mathematical Sciences*, 128(4):3018–3029, 2005.

[78] G. Cerbai. *Sorting permutations with pattern-avoiding machines*. PhD thesis, Università degli Studi di Firenze, 2021.

[79] M. Cervetti. *Pattern posets: enumerative, algebraic and algorithmic issues*. PhD thesis, University of Trento, 2021.

[80] M. Cervetti and L. Ferrari. Enumeration of some classes of pattern avoiding matchings, with a glimpse into the matching pattern poset. *Annals of Combinatorics*, 26(4):971–995, 2022.

[81] G. J. Chaitin. A theory of program size formally identical to information theory. *J. ACM*, 22(3):329–340, 1975.

[82] W. Y. Chen, E. Y. Deng, R. R. Du, R. P. Stanley, and C. H. Yan. Crossings and nestings of matchings and partitions. *Transactions of the American Mathematical Society*, 359(4):1555–1575, 2007.

[83] W. Y. Chen, T. Mansour, and S. H. Yan. Matchings avoiding partial patterns. *The Electronic Journal of Combinatorics*, 13(1):R112, 2006.

[84] C. Chenevière. Linear intervals in the Tamari and the Dyck lattices and in the alt-Tamari posets, 2022. https://arxiv.org/abs/2209.00418.

[85] C. Chenevière. *Enumerative study of intervals in lattices of Tamari type*. PhD thesis, IRMA, Université de Strasbourg, 2023. https://publication-theses.unistra.fr/public/theses_doctorat/2023/CHENEVIERE_Clement_2023_ED269.pdf.

[86] S. K. Choi, C. Rim, and H. Um. Narayana Number, Chebyshev Polynomial and Motzkin Path on RNA Abstract Shapes. In J. de Gier, C. E. Praeger, and T. Tao, editors, *2017 MATRIX Annals*, MATRIX Book Series, pages 153–166. Springer International Publishing, Cham, 2019.

[87] N. Chomsky and M.-P. Schützenberger. The algebraic theory of context-free languages. In *Computer programming and formal systems*, pages 118–161. 1963.

[88] E. B. Christoffel. Observatio arithmetica. *Annali di Matematica Pura ed Applicata*, 6:148–152, 1873.

[89] M. Cipu and F. Luca. On the Galois group of the generalized Fibonacci polynomial. *Analele ştiinţifice ale Universităţii "Ovidius" Constanţa. Seria Matematică*, 9(1):27–38, 2001.

[90] D. L. Clayman. Sigmatism in greek poetry. *Transactions of the American Philological Association (1974-)*, 117:69–84, 1987.

[91] M. Danisch, O. Balalau, and M. Sozio. Listing k-cliques in sparse real-world graphs. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 589–598. International World Wide Web Conferences Steering Committee, 2018.

[92] J.-P. Delahaye. Des suites exotiques pour écrire les nombres. *Pour La Science*, 568:75–76, 2025.

[93] M.-P. Delest and G. Viennot. Algebraic languages and polyominoes enumeration. *Theoretical Computer Science*, 34(1):169–206, 1984.

[94] A. Denjoy. Sur une fonction réelle de Minkowski. *Journal de Mathématiques Pures et Appliquées. Neuvième Série*, 17(1-4):105–151, 1938.

[95] A. Dermenjian. Maximal degree subposets of ν-Tamari lattices. *The Electronic Journal of Combinatorics*, 30(2):P2.43, 2023.

[96] S. Dovgal and S. Kirgizov. Sturm meets Fibonacci in Minkowski's fractal bar. Permutation Patterns Conference. University of Burgundy, Dijon, France, 2023.

[97] S. Dovgal and S. Kirgizov. Structure and growth of $\mathbb{R}$-bonacci words. *The Electronic Journal of Combinatorics*, 32(3):P3.32, 2025.

[98] S. Dovgal and K. Nurligareev. Asymptotics for graphically divergent series: dense digraphs and 2-SAT formulae, 2023. http://arxiv.org/abs/2310.01213.

[99] F. Dubeau. On r-generalized Fibonacci numbers. *Fibonacci Quarterly*, 27(3):221–229, 1989.

[100] E. Duchi. *Polyominoes, permutominoes and permutations*. HDR thesis, Paris 7 — Université Paris Diderot, 2018.

[101] G. Ducoffe, L. Feuilloley, M. Habib, and F. Pitois. Pattern detection in ordered graphs, 2023. https://arxiv.org/abs/2302.11619.

[102] X. Dutreilh, S. Kirgizov, O. Melekhova, J. Malenfant, N. Rivierre, and I. Truck. Using reinforcement learning for autonomic resource allocation in clouds: Towards a fully automated workflow. In *ICAS 2011, The Seventh International Conference on Autonomic and Autonomous Systems*, pages 67–74, 2011.

[103] G. Eisenstein. Eine neue Gattung zahlentheoretischer Funktionen, welche von zwei Elementen abhängen und durch gewisse lineare Funktional-Gleichungen definiert werden. *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königl. Preuss. Akademie der Wissenschaften zu Berlin*, pages 36–42, 1850.

[104] S. Elizalde. *Statistics on pattern-avoiding permutations*. PhD thesis, Massachusetts Institute of Technology, 2004.

[105] S. Elizalde and M. Noy. Clusters, generating functions and asymptotics for consecutive patterns in permutations. *Advances in Applied Mathematics*, 49(3):351–374, 2012.

[106] O. Eğecioğlu and V. Iršič. Fibonacci-run graphs I: Basic properties. *Discrete Applied Mathematics*, 295:70–84, 2021.

[107] O. Eğecioğlu, S. Klavžar, and M. Mollard. *Fibonacci Cubes with Applications and Variations*. World Scientific, 2023.

[108] M. Feinberg. Fibonacci-tribonacci. *Fibonacci Quarterly*, 1(3):71–74, 1963.

[109] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[110] I. Flores. Direct calculation of k-generalized Fibonacci numbers. *Fibonacci Quarterly*, 5(3):259–266, 1967.

[111] R. Flórez, L. Junes, and J. L. Ramírez. Enumerating several aspects of non-decreasing Dyck paths. *Discrete Mathematics*, 342(11):3079–3097, 2019.

[112] S. Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[113] S. Fortunato and D. Hric. Community detection in networks: A user guide. *Physics reports*, 659:1–44, 2016.

[114] H. Friedman and D. Tamari. Problèmes d'associativité: Une structure de treillis finis induite par une loi demi-associative. *Journal of Combinatorial Theory*, 2(3):215–242, 1967.

[115] O. Frink and P. A. Smith. Irreducible non-planar graphs, 1930. Announcement N. 179.

[116] H. H. Gan, S. Pasquali, and T. Schlick. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Research*, 31(11):2926–2943, 2003.

[117] A. Garah, N. Mbarek, and S. Kirgizov. An architecture for confidentiality self-management in the internet of things. In K. Yétongnon, A. Dipanda, and L. Gallo, editors, *16th International Conference on Signal-Image Technology & Internet-Based Systems, SITIS*, pages 472–478, Dijon, France, 2022. IEEE.

[118] A. Garah, N. Mbarek, and S. Kirgizov. Decision tree-based confidentiality self-management in the Internet of things. In S. Oteafy, E. Bulut, F. Tschorsch, and K. Thilakarathna, editors, *IEEE 48th Conference on Local Computer Networks (LCN)*, pages 1–4, Daytona Beach, FL, USA, 2023.

[119] A. Garah, N. Mbarek, and S. Kirgizov. IoT data confidentiality self-management. In *IEEE Intl. Conf. on Dependable, Autonomic and Secure Computing, Intl. Conf. on Pervasive Intelligence and Computing, Intl. Conf. on Cloud and Big Data Computing, Intl. Conf. on Cyber Science and Technology Congress, DASC/PiCom/CBDCom/CyberSciTech*, pages 395–400, Abu Dhabi, United Arab Emirates, 2023. IEEE.

[120] A. Garah, N. Mbarek, and S. Kirgizov. Enhancing iot data confidentiality and energy efficiency through decision tree-based self-management. *Internet of Things*, 26:101219, 2024.

[121] A. Garah, N. Mbarek, and S. Kirgizov. DBSCAN-based IoT object integrity self-management. In *The 21st International Wireless Communications & Mobile Computing Conference (IWCMC*, pages 395–400, Abu Dhabi, United Arab Emirates, 2025. IEEE.

[122] A. Garah, N. Mbarek, and S. Kirgizov. Fuzzy logic-based IoT object integrity self-management. In *12th IFIP International Conference on*

New Technologies, Mobility and Security (NTMS), Paris, France, 2025. IEEE.

[123] N. Gastineau, B. Gras, S. Kirgizov, and M. Omidvar. Packing coloring and subsets preserving path distance. Les 18es Journées Graphes et Algorithmes (JGA), 11 2016.

[124] A. Genitrini and C. Mailler. Generalised and quotient models for random and/or trees and application to satisfiability. *Algorithmica*, 1:1–33, 2016.

[125] R. Giegerich, B. Voss, and M. Rehmsmeier. Abstract shapes of RNA. *Nucleic Acids Research*, 32(16):4843–4851, 2004.

[126] A. Giorgetti, C. Dubois, and R. Lazarini. Combinatoire formelle avec why3 et coq. In *JFLA 2019 : Journées Francophones des Langages Applicatifs*, Les Rousses, France, 2019.

[127] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[128] J. Goswami, M. C. Davis, T. Andersen, A. Alileche, and G. Hampikian. Safeguarding forensic DNA reference samples with nullomer barcodes. *Journal of Forensic and Legal Medicine*, 20(5):513–519, 2013.

[129] I. P. Goulden and D. M. Jackson. An inversion theorem for cluster decompositions of sequences with distinguished subsequences. *Journal of the London Mathematical Society*, s2-20(3):567–576, 1979.

[130] I. P. Goulden and D. M. Jackson. *Combinatorial Enumeration*. Wiley, 1983.

[131] F. Gray. Pulse code communication, Mar 17 1953. US Patent 2,632,058.

[132] S. Griffiths-Jones, A. Bateman, M. Marshall, A. Khanna, and S. R. Eddy. Rfam: an RNA family database. *Nucleic Acids Research*, 31(1):439–441, 2003.

[133] S. Griffiths-Jones, S. Moxon, M. Marshall, A. Khanna, S. R. Eddy, and A. Bateman. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research*, 33(suppl_1):D121–D124, 2004.

[134] L.-A.-L. Gros. *Théorie du baguenodier par un clerc de notaire lyonnais*. Aimé Vingtrinier, Lyon, 1872.

[135] L. J. Guibas and A. M. Odlyzko. Periods in strings. *Journal of Combinatorial Theory, Series A*, 30(1):19–42, 1981.

[136] L. J. Guibas and A. M. Odlyzko. String overlaps, pattern matching, and nontransitive games. *Journal of Combinatorial Theory, Series A*, 30(2):183–208, 1981.

[137] B. Hackl, C. Heuberger, and H. Prodinger. Counting Ascents in Generalized Dyck Paths. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018)*, volume 110, pages 26:1–26:15, Dagstuhl, Germany, 2018.

[138] G. Hampikian and T. Andersen. Absent sequences: nullomers and primes. In *Biocomputing*. World Scientific, 2007.

[139] B. Han, J. Mao, and J. Zeng. Equidistributions around special kinds of descents and excedances. *SIAM Journal on Discrete Mathematics*, 35(4):2858–2879, 2021.

[140] C. Haslinger and P. F. Stadler. RNA structures with pseudo-knots: graph-theoretical, combinatorial, and statistical properties. *Bulletin of Mathematical Biology*, 61:437–467, 1999.

[141] N. Hassler and S. Kirgizov. Emerging consecutive pattern avoidance. Permutation Patterns Conference. University of St Andrews, Scotland, 2025.

[142] N. Hassler, V. Vajnovszki, and D. Wong. Greedy gray codes for some restricted classes of binary words. In S. Brlek and L. Ferrari, editors, Proc. of the 13th edition of the conference on *Random Generation of Combinatorial Structures. Polyominoes and Tilings (GASCom 2024)*, volume 403 of *Electronic Proceedings in Theoretical Computer Science*, pages 108–112. Open Publishing Association, 2024.

[143] J. Herold, S. Kurtz, and R. Giegerich. Efficient computation of absent words in genomic sequences. *BMC Bioinformatics*, 9(1):167, 2008.

[144] S. Heubach and T. Mansour. *Combinatorics of Compositions and Words*. Chapman and Hall/CRC, 2009.

[145] F. Hivert, J.-C. Novelli, and J.-Y. Thibon. The algebra of binary search trees. *Theoretical Computer Science*, 339(1):129–165, 2005.

[146] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation (3rd Edition)*. Addison-Wesley Longman Publishing Co., Inc., 2006.

[147] W.-J. Hsu. Fibonacci cubes: Properties of an asymmetric interconnection topology. In *ICPP (1)*, pages 722–723, Austin, Texas, USA, 1991.

[148] W.-J. Hsu. Fibonacci cubes-a new interconnection topology. *IEEE Transactions on Parallel and Distributed Systems*, 4(1):3–12, Jan 1993.

[149] W.-J. Hsu, C. Page, and J.-S. Liu. Fibonacci cubes-a class of self-similar graphs. *Fibonacci Quarterly*, 31(1):65–71, 1993.

[150] S. Huang and D. Tamari. Problems of associativity: A simple proof for the lattice property of systems ordered by a semi-associative law. *Journal of Combinatorial Theory, Series A*, 13(1):7–13, 1972.

[151] V. Jelínek. Dyck paths and pattern-avoiding matchings. *European Journal of Combinatorics*, 28(1):202–213, 2007.

[152] V. Jelínek and T. Mansour. Matchings and partial patterns. *The Electronic Journal of Combinatorics*, 17:R158, 2010.

[153] S. Kirgizov. *Analyse empirique et modélisation de la dynamique de la topologie de l'internet*. PhD thesis, Paris 6, 2014.

[154] S. Kirgizov. $\mathbb{Q}$-bonacci words and numbers. *Fibonacci Quarterly*, 60(5), 2022.

[155] S. Kirgizov. $\mathbb{Q}$-bonacci words and numbers. 20th Fibonacci conference, Sarajevo, Bosnia and Herzegovina, 2022.

[156] S. Kirgizov and E. Leclercq. Temporal density of complex networks and ego-community dynamics. Abstract presented at Conference on Complex Systems (ECCS or CCS), 2016.

[157] S. Kirgizov and E. Leclercq. Temporal density of community structure. MARAMI, 2019.

[158] S. Kirgizov, E. Leclercq, and M. Danisch. A la recherche des mini-publics : un problème de communautés, de singularités et de sémantique. Journée Données Participatives et Sociales. Conférence Extraction et Gestion des Connaissances, EGC), 2016.

[159] S. Kirgizov, E. Leclercq, M. Savonnet, A. Frame, and I. Basaille-Gahitte. A web application for event detection and exploratory data analysis for twitter data. Abstract presented at "Twitter at the European Elections 2014 : International Perspectives on a Political, Communication Tool" conference, 2015.

[160] S. Kirgizov and C. Magnien. Internet topology dynamics: stochastic process estimation from partial observations. Journée jointe des GDR ISIS et Phénix "Analyse et inférence pour les réseaux", 2013.

[161] S. Kirgizov, C. Magnien, F. Tarissan, and A. Liu. Studying the impact of measurement frequency on the ip-level routing topology dynamics. In *24-ème édition du colloque Gretsi*, September 2013.

[162] S. Kirgizov, A. Medem, C. Magnien, and F. Tarissan. Vers une modélisation réaliste de la dynamique de la topologie de routage au niveau ip. Journées Automnales, ResCom, 2012.

[163] S. Kirgizov and K. Nurligareev. Asymptotics of self-overlapping permutations. *Discrete Mathematics*, 348(5):114400, 2025.

[164] S. Kitaev. *Patterns in permutations and words*. Springer Science & Business Media, 2011.

[165] D. E. Knuth. *The art of computer programming. Volume 3: Sorting and searching*. Addison-Wesley, 2 edition, 1998.

[166] D. E. Knuth. *The art of computer portland. Volume 4A: Combinatorial Algorithms, Part 1*. Addison-Wesley, 2011.

[167] T. Kondrashova, A. Frame, and S. Kirgizov. (re)constuire la temporalité d'un événement médiatique sur twitter : une étude contrastive. (re)constructing the temporality of media events on twitter: a contrastive study. XXe Congrès de la SFSIC: Temps, temporalités et information-communication, 06 2016.

[168] G. Koulouras and M. C. Frith. Significant non-existence of sequences in genomes and proteomes. *Nucleic Acids Research*, 49(6):3139–3155, 2021.

[169] J. Kozik. Subcritical pattern languages for and/or trees. *Discrete Mathematics & Theoretical Computer Science*, DMTCS Proceedings vol. AI, Fifth Colloquium on Mathematics and Computer Science, 2008.

[170] G. Kreweras. Sur les partitions non croisées d'un cycle. *Discrete Mathematics*, 1(4):333–350, 1972.

[171] C. Kuratowski. Sur la notion de l'ordre dans la théorie des ensembles. *Fundamenta Mathematicae*, 2(1):161–171, 1921.

[172] C. Kuratowski. Sur le problème des courbes gauches en topologie. *Fundamenta Mathematicae*, 15:271–283, 1930.

[173] C. L. Lawson, H. M. Berman, L. Chen, B. Vallat, and C. L. Zirbel. The nucleic acid knowledgebase: a new portal for 3d structural information about nucleic acids. *Nucleic Acids Research*, 52(D1):D245–D254, 2023.

[174] E. Leclercq, M. Savonnet, T. Grison, S. Kirgizov, and I. Basaille-Gahitte. SNFreezer: a platform for harvesting and storing tweets in a big data context. In A. Frame, A. Mercier, G. Brachotte, and C. Thimm, editors, *Twitter and the European Parliamentary Elections: researching political uses of microblogging*, pages 19–33. Peter Lang, 2016.

[175] B. Liu, D. Wong, C.-T. Lam, and S.-K. Im. Generating a cyclic 2-gray code for lucas words in constant amortized time. In P. Bonizzoni and V. Mäkinen, editors, *36th Annual Symposium on Combinatorial Pattern Matching (CPM 2025)*, volume 331 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:18, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[176] W. A. Lorenz, Y. Ponty, and P. Clote. Asymptotics of RNA shapes. *Journal of Computational Biology*, 15(1):31–63, 2008.

[177] W. A. Lorenz, Y. Ponty, and P. Clote. Asymptotics of RNA shapes. *Journal of Computational Biology*, 15(1):31–63, 2008.

[178] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, 1997.

[179] M. Lothaire. *Algebraic Combinatorics on Words*. Cambridge University Press, 2002.

[180] M. Lothaire. *Applied Combinatorics on Words*. Cambridge University Press, 2005.

[181] X.-J. Lu, H. J. Bussemaker, and W. K. Olson. DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Research*, 43(21):e142, 2015.

[182] X.-J. Lu and W. K. Olson. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research*, 31(17):5108–5121, 2003.

[183] W. Ludwig. *Blue and Brown Books*. Basil Blackwell, 1935, 1958.

[184] P. Luschny. Fibonacci meanders, 2011. http://oeis.org/wiki/User:Peter_Luschny/FibonacciMeanders. Accessed: 18 April 2025.

[185] T. J. Macke, D. J. Ecker, R. R. Gutell, D. Gautheret, D. A. Case, and R. Sampath. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research*, 29(22):4724–4735, 2001.

[186] P. A. MacMahon. Two applications of general theorems in combinatory analysis: (1) to the theory of inversions of permutations; (2) to the ascertainment of the numbers of terms in the development of a determinant which has amongst its elements an arbitrary number of zeros. *Proceedings of the London Mathematical Society*, s2-15(1):314–321, 01 1917.

[187] C. Magnien, A. Medem, S. Kirgizov, and F. Tarissan. Towards realistic modeling of ip-level routing topology dynamics. *Networking science*, 3(1):24–33, 2013.

[188] T. Mansour. *Combinatorics of set partitions*. Chapman and Hall/CRC, 2013.

[189] R. Mansuy. Achille Brocot, mathématicien à ses heures. *Culture Math*, 2008.

[190] R. Maréchal. *Combinatoire énumérative et bijective de différentes familles de chemins de Dyck avec trous d'air*. Theses, Université de Bourgogne, Oct 2024.

[191] P. A. Martin. The galois group of $x^n - x^{n-1} - \ldots - x - 1$. *Journal of Pure and Applied Algebra*, 190(1):213–223, 2004.

[192] A. Medem, C. Magnien, and F. Tarissan. Impact of power-law topology on IP-level routing dynamics: simulation results. In *Proceedings of the Fourth International Workshop on Network Science for Communication Networks (NetSciCom)*, 2012.

[193] E. Miles Jr. Generalized Fibonacci numbers and associated matrices. *The American Mathematical Monthly*, 67(8):745–752, 1960.

[194] M. Miller. On generalized Fibonacci numbers. *The American Mathematical Monthly*, 78(10):1108–1109, 1971.

[195] A. Montgomery, G. C. Tsiatsianis, I. Mouratidis, C. S. Y. Chan, M. Athanasiou, A. D. Papanastasiou, V. Kantere, N. Syrigos, I. Vathiotis, K. Syrigos, N. S. Yee, and I. Georgakopoulos-Soares. Utilizing nullomers in cell-free RNA for early cancer detection. *Cancer Gene Therapy*, 2024.

[196] F. Müller-Hoissen, J. M. Pallo, and J. Stasheff, editors. *Associahedra, Tamari lattices and related structures: Tamari memorial Festschrift*, volume 299 of *Progress in Mathematics*. Birkhäuser Basel, 2012.

[197] P. Nadeau and V. Tewari. p-partitions with flags and back stable quasisymmetric functions. *Combinatorial Theory*, 4(2), 2024.

[198] J. Noonan and D. Zeilberger. The Goulden—Jackson cluster method: extensions, applications and implementations. *Journal of Difference Equations and Applications*, 5(4-5):355–377, 1999.

[199] J.-C. Novelli and J.-Y. Thibon. Hopf algebras of m-permutations, $(m+1)$-ary trees, and m-parking functions. *Advances in Applied Mathematics*, 117:102019, 2020.

[200] OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences, 2025. http://oeis.org.

[201] A. I. Petrov, C. L. Zirbel, and N. B. Leontis. WebFR3D–a server for finding, aligning and analyzing recurrent RNA 3D motifs. *Nucleic Acids Research*, 39(suppl_2):W50–W55, 2011.

[202] G. A. Pick. Geometrisches zur Zahlenlehre. *Sitzungsberichte des deutschen naturwissenschaftlich-medicinischen Vereines für Böhmen "Lotos" in Prag.*, pages 311–319, 1899.

[203] A. Pierrot. *Combinatoire et algorithmique dans les classes de permutations*. PhD thesis, Paris 7 — Université Paris Diderot, 2013.

[204] A. J. Pinho, P. J. Ferreira, S. P. Garcia, and J. M. Rodrigues. On finding minimal absent words. *BMC Bioinformatics*, 10(1), 2009.

[205] H. Poincare. *Science et Méthode*. 1920.

[206] J. I. Porter. Lasus of hermione, pindar and the riddle of s. *The Classical Quarterly*, 57(1):1–21, 2007.

[207] L. K. Pudwell. *Enumeration schemes for pattern-avoiding words and permutations*. PhD thesis, Rutgers University, 2008.

[208] M. Quadrini. Structural relation matching: an algorithm to identify structural patterns into RNAs and their interactions. *Journal of Integrative Bioinformatics*, 18(2):111–126, 2021.

[209] F. Queyroi and S. Kirgizov. Suppression distance computation for hierarchical clusterings. *Information Processing Letters*, 115(9):689–693, 2015.

[210] C. M. Reidys and R. R. Wang. Shapes of RNA pseudoknot structures. *Journal of Computational Biology*, 17(11):1575–1590, 2010.

[211] F. J. Rispoli and S. Cosares. The Fibonacci hypercube. *Australasian journal of combinatorics*, 40:187–196, 2008.

[212] E. Rivals. Incremental computation of the set of period sets. In R. Královič and V. Kůrková, editors, *SOFSEM 2025: Theory and Practice of Computer Science*, pages 254–268, Cham, 2025. Springer Nature Switzerland. https://arxiv.org/abs/2410.12077.

[213] E. Rivals and S. Rahmann. Combinatorics of periods in strings. *Journal of Combinatorial Theory, Series A*, 104(1):95–113, 2003.

[214] E. Rivals, M. Sweering, and P. Wang. Convergence of the Number of Period Sets in Strings. In K. Etessami, U. Feige, and G. Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 100:1–100:14, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[215] N. Robertson and P. D. Seymour. Graph minors. xx. wagner's conjecture. *Journal of Combinatorial Theory, Series B*, 92(2):325–357, 2004.

[216] D. Roelants van Baronaigien and F. Ruskey. Generating subsets with a given sum. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 14:87–96, 1993.

[217] F. Ruskey, T. Mütze, J. Sawada, A. Williams, et al. COS++. The Combinatorial Object Server, 2000, 2018. http://combos.org. Accessed: 17 April 2025.

[218] E. A. Rødland. Pseudoknots in RNA secondary structures: representation, enumeration, and prevalence. *Journal of Computational Biology*, 13(6):1197–1213, 2006.

[219] M. Saniga, F. Holweck, C. Kelleher, A. Muller, A. Giorgetti, and H. d. Boutray. Hexagons govern three-qubit contextuality. *Quantum*, 9:1601, 2025.

[220] M. Sarver, C. L. Zirbel, J. Stombaugh, A. Mokdad, and N. B. Leontis. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of Mathematical Biology*, 56(1–2):215–252, 2007.

[221] J. A. Scott. Sigmatism in greek dramatic poetry. *The American Journal of Philology*, 29(1):69–77, 1908.

[222] T. Sillke. Program: Autocorrelation-range, 2001. http://www.mathematik.uni-bielefeld.de/~sillke/SEQUENCES/autocorrelation-range.c.

[223] P. Singh. The so-called Fibonacci numbers in ancient and medieval India. *Historia Mathematica*, 12(3):229–244, 1985.

[224] A. D. Solov'ev. A combinatorial identity and its application to the problem concerning the first occurrence of a rare event. *Theory of Probability & Its Applications*, 11(2):276–282, 1966.

[225] R. P. Stanley. *Enumerative Combinatorics, volume 1*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2011.

[226] R. P. Stanley. *Catalan numbers*. Cambridge University Press, 2015.

[227] R. P. Stanley. *Enumerative Combinatorics, volume 2*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2023.

[228] M. A. Stern. Ueber eine zahlentheoretische funktion. *Journal für die reine und angewandte Mathematik*, 55:193–220, 1858.

[229] G. R. Stibitz. Binary counter, Jan 12 1943. US Patent 2,307,868.

[230] B. E. Tenner. *The combinatorics of reduced decompositions*. PhD thesis, Massachusetts Institute of Technology, 2006.

[231] W. T. Tutte. Dichromatic sums for rooted planar maps. In *Combinatorics (Proc. Sympos. Pure Math., Vol. XIX, Univ. California, Los Angeles, Calif., 1968)*, volume 19 of *Proceedings of Symposia in Pure Mathematics*, pages 235–245. American Mathematical Society, Providence, RI, 1971.

[232] W. T. Tutte. Chromatic sums for rooted planar triangulations: The cases $\lambda = 1$ and $\lambda = 2$. *Canadian Journal of Mathematics*, 25(2):426–447, 1973.

[233] V. Vajnovszki. A loopless generation of bitstrings without p consecutive ones. In *Combinatorics, computability and logic*, pages 227–240. Springer, 2001.

[234] V. Vajnovszki. Generating involutions, derangements, and relatives by ECO. *Discrete Mathematics & Theoretical Computer Science*, 12(1):479, 2010.

[235] F. H. D. van Batenburg, A. P. Gultyaev, C. W. A. Pleij, J. Ng, and J. Oliehoek. PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Research*, 28(1):201–204, 2000.

[236] K. Wagner. Über eine Eigenschaft der ebenen Komplexe. *Mathematische Annalen*, 114:570–590, 1937.

[237] C.-J. Wang. *Applications of the Goulden-Jackson cluster method to counting Dyck paths by occurrences of subwords*. PhD thesis, Brandeis University, 2011.

[238] P. Wang. *On some computational questions regarding string overlaps*. PhD thesis, LIRMM, Université Montpellier, CNRS, 2024.

[239] M. S. Waterman. Secondary structure of single-stranded nucleic acids. *Advances in Mathematics Supplementary Studies*, 1:167–212, 1978.

[240] S. Wienand. Personal page on the oeis wiki, 2011. http://oeis.org/wiki/User:Susanne_Wienand#Definition_of_meanders_concerned_by_these_sequences. Accessed: 18 April 2025.

[241] H. S. Wilf. *generatingfunctionology*. A. K. Peters, Ltd., 2006.

[242] L. Wittgenstein. *Logisch-Philosophische Abhandlung*. W. Ostwald's Annalen der Naturphilosophie, 1921.

[243] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Kegan Paul, 1922. Translated by C. K. Ogden and F. P. Ramsey.

[244] M. Wojdyr. GEMMI: a library for structural biology. *Journal of Open Source Software*, 7(73):4200, 2022.

[245] D. A. Wolfram. Solving generalized Fibonacci recurrences. *Fibonacci Quarterly*, 36(2):129–145, 1998.

[246] D. Wong, B. Liu, C.-T. Lam, and M. Im. Generating cyclic 2-Gray codes for Fibonacci q-decreasing words. In R. Uehara, K. Yamanaka, and H.-C. Yen, editors, *WALCOM: Algorithms and Computation*, pages 91–102, Singapore, 2024. Springer Nature Singapore.

[247] H.-T. Yao, C. Chauve, M. Regnier, and Y. Ponty. Exponentially few RNA structures are designable. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB'19, page 289–298. Association for Computing Machinery, 2019.

[248] N. Zeilberger and A. Giorgetti. A correspondence between rooted planar maps and normal planar lambda terms. *Logical Methods in Computer Science*, 11(3), 2015.

[249] T. Zhou, W. Y. Tang, D. H. Mathews, and L. Huang. Undesignable RNA structure identification via rival structure generation and structure decomposition. In J. Ma, editor, *Research in Computational Molecular Biology (RECOMB)*, pages 270–287, Springer, Cham, 2024.

[250] Y. Zhuang. A generalized Goulden–Jackson cluster method and lattice path enumeration. *Discrete Mathematics*, 341(2):358–379, 2018.

Curriculum vitæ
Sergey Kirgizov

Born 5th June, 1988
Married, with five children

Languages: Russian, French and English

Academic Positions

- Maître de conférences (Associate Professor), Université Bourgogne Europe, 2019–present
- Research and Teaching Fellow (Vacataire), Université de Bourgogne, 2018-2019
- Research and Teaching Fellow (ATER), Université de Bourgogne, 2015-2017
- Postdoc Research Fellow, Université de Bourgogne, 2014-2015

Industry Positions

- R&D engineer, Page Up, Dijon, Paris, France, 2017-2019
- Software engineer and architect, Scientific Saas, Barnaul, Russia, 2010-2013
- Software engineer and team leader, Altay State University, Barnaul, Russia 2007-2010
- Web-site administrator, Barnaul, Russia, 2009
- System and network administrator, Barnaul, Russia, 2006-2007

Education

- PhD in Computer Science, Université Pierre et Marie Curie, 2011-2014
- Master degree in Applied Mathematics and Informatics, Altay State University, 2009-2011
- Master Internship 2, Université Pierre et Marie Curie, 2011
- Master Internship 1, Université Pierre et Marie Curie, 2010
- B.S. in Mathematics, Altay State University, 2005-2009
- Lyceum of Information Technologies "Kvant", Barnaul, Russia, 2003-2005

Research projects

- Participant of the ANR "COMETA-GAE, Enumerative combinatorics and applications: random and exhaustive generation", 2025-2030
- Project leader of the ANR "PICS — Patterns in Combinatorics", 2022-2026
- Project leader of the "Artico — L'ART de la COmbinatoire et ses Interactions", Bourgogne-Franche-Comté region, 2021-2023
- Participant of the ANR "COREGRAPHIE", 2021-2025
- Principal participant of the PEPS MoMIS SEISME, 2015-2016

- Participant of the "TEE 2014" project
- Participant of the ANR "DynGraph", 2010-2013
- Participant of the FP7 FIRE project EULER, 2010-2014
- Participant of the project funded by EIT-ICT Labs action line FNS, 2014

Distinctions

- Best Start-Up award. Our start-up "Scientific Saas" won the competition "IT-Start", 2012
- Best Paper Award, The Seventh International Conference on Autonomic and Autonomous Systems, ICAS, 2011
- Prize "For the fastest solution of the most difficult task", All-Russia Oracle Database Olympiad, 2009
- Scholarship for Excellence in Studies, 2009-2011
- Gold Medal for Excellence in Studies, 2005
- Barnaul City Mayor's Special Scholarship for Excellence in Studies, 2004-2005

PhD Students

- Abdelhamid Garah, "Gestion autonome des services de sécurité dans l'Internet des objets", joint supervision with Nader Mbarek, Université de Bourgogne, 2021-2024

- Rémi Maréchal, "Combinatoire énumérative et bijective de différentes familles de chemins de Dyck avec trous d'air", joint supervision with Jean-Luc Baril and Vincent Vajnovszki, Université de Bourgogne, 2021-2024

- Nathanaël Hassler, "Motifs dans les Permutations et les Mots", joint supervision with Jean-Luc Baril and Vincent Vajnovszki, Université de Bourgogne, 2023-ongoing

Postdocs

- Célia Biane, ANR "PICS — Patterns in Combinatorics" project, 2023-2024
- Khaydar Nurligareev, ANR "PICS — Patterns in Combinatorics" project, 2023-2024

Organization of Conferences

- Les 26e Journées Graphes et Algorithmes (JGA), 2024
- 21th International Conference on Permutation Patterns, 2023
- 16th International Conference on Signal Image Technology & Internet based Systems, SITIS, 2022

- 7th Workshop on Complex Networks, CompleNet, 2016
- ALGOTEL conference, 2015
- MARAMI conference, 2014
- Journée thématique "Dynamique des graphes", 2012
- Journées Automnales 2012 ResCom à Paris 6

Administrative responsibilites

- Member of the Bureau de la Commission de Proposition (BCP) of section 27 of the University of Burgundy, since 2023
- Member of the LIB laboratory council, since 2023
- Co-organiser of LIB laboratory seminars, since 2022
- Coordinator of internship for 4th year internships at Polytech Dijon, Computer Science and Electronics Department, since 2021
- Member of the Polytech Dijon School Council, since 2021
- Member of the Polytech Dijon Orientation Council, since 2021
- Coordinator of student projects for the 4th year of Polytech Dijon, Computer Science and Electronics Department, Since 2020

Other activities

- Participation in art-science exhibitions:
  - *Les concepts mathématiques vus par Yulia Kirgizova*. "Le Cortex" library of the University of Burgundy, 2023
  - *Les Réseaunautes.* Travelling exhibition: Darcy garden, secondary schools in the BFC region, "Le Cortex" library, 2020-2022
  - *Résomes*. The Atheneum, Cultural Centre of the University of Burgundy, 2020
- Founder of Papers$^{\gamma}$ — discussing board for scientific papers.
- Reviewer for

  FPSAC; The Electronic Journal of Combinatorics; Discrete Mathematics; SIAM Journal on Discrete Mathematics; CPM 2025: 36th Annual Symposium on Combinatorial Pattern Matching; The Australasian Journal of Combinatorics; GASCom; ALGOTEL; IEEE Communications Letters; CompleNet; International Journal of Computer Mathematics; Journal of Algebraic Combinatorics; Journal of Automata, Languages and Combinatorics; Discrete Mathematics & Theoretical Computer Science; Integers; Annales Mathematicae Silesianae; Filomat; Computing in Geometry and Topology; Fibonacci Quarterly; Journal of Integer Sequences; Symposium on Computers and Communications; Conference on

Standards for Communications and Networking; Discrete Mathematics Letters; RAIRO Informatique théorique et applications, Theoretical Informatics an Applications

Contact Information
LIB UR 7534
Université Bourgogne Europe
9 av. Alain Savary
BP 47 870, 21078 Dijon Cedex, FRANCE
Email: sergey.kirgizov@u-bourgogne.fr
Site web: https://kirgizov.link

Publications

International Scientific Journals (32)

1. **Structure and growth of $\mathbb{R}$-bonacci words**
   Sergey Dovgal and Sergey Kirgizov
   The Electronic Journal of Combinatorics, 32(3), Article P3.32, 2025
2. **A lattice on Dyck paths close to the Tamari lattice**
   Jean-Luc Baril, Sergey Kirgizov and Mehdi Naima
   In memory of Jean Pallo
   Séminaire Lotharingiens de Combinatoire, 89, 2025
3. **The ascent lattice on Dyck paths**
   Jean-Luc Baril, Mireille Bousquet-Mélou, Sergey Kirgizov, and Mehdi Naima
   The Electronic Journal of Combinatorics, 32(2), P2.36, 2025
4. **Asymptotics of self-overlapping permutations**
   Sergey Kirgizov and Khaydar Nurligareev
   Discrete Mathematics, 348(5), 2025
5. **Endhered patterns in matchings and RNA**
   Célia Biane, Greg Hampikian, Sergey Kirgizov, and Khaydar Nurligareev
   Journal of Computational Biology, 32(1), 2025
6. **The combinatorics of Motzkin polyominoes**
   Jean-Luc Baril, Sergey Kirgizov, José Luis Ramírez, and Diego Villamizar
   Discrete Applied Mathematics, 364, 2025
7. **Grand zigzag knight's paths**
   Jean-Luc Baril, Nathanaël Hassler, Sergey Kirgizov and José Luis Ramírez
   Enumerative combinatorics and applications, 5(1), 2025

8. **Enhancing IoT data confidentiality and energy efficiency through decision tree-based self-management**
   Abdelhamid Garah, Nader Mbarek, and Sergey Kirgizov
   Internet of Things, 26, 2024

9. **Grand Dyck paths with air pockets**
   Jean-Luc Baril, Sergey Kirgizov, Rémi Maréchal, and Vincent Vajnovszki
   The Art of Discrete and Applied Mathematics, 7(1), 2023

10. **Enumeration of Dyck Paths with Air Pockets**
    Jean-Luc Baril, Sergey Kirgizov, Rémi Maréchal, and Vincent Vajnovszki
    Journal of Integer Sequences, 26(3), 2023

11. **$\mathbb{Q}$-bonacci words and numbers**
    Sergey Kirgizov
    The Fibonacci Quarterly, 60(5), 2022

12. **Polyominoes and graphs built from Fibonacci words**
    Sergey Kirgizov and José Luis Ramírez
    The Fibonacci Quarterly, 60(5), 2022

13. **Dyck paths with catastrophes modulo the positions of a given pattern**
    Jean-Luc Baril, Sergey Kirgizov, Armen Petrossian
    Australasian Journal of Combinatorics, 84(3), 2022

14. **Asymptotic bit frequency in Fibonacci words**
    Jean-Luc Baril, Sergey Kirgizov and Vincent Vajnovszki
    Pure Mathematics and Applications, 30(1), 2022

15. **Gray codes for Fibonacci q-decreasing words**
    Jean-Luc Baril, Sergey Kirgizov and Vincent Vajnovszki
    Theoretical Computer Science, 927, Pages 120-132, 2022

16. **Lattice paths with a first return decomposition constrained by the maximal height of a pattern**
    Jean-Luc Baril and Sergey Kirgizov
    Discrete Mathematics Letters, 8, 2022

17. **Pattern statistics in faro words and permutations**
    Jean-Luc Baril, Alexander Burstein and Sergey Kirgizov
    Discrete Mathematics, 344(8), 2021

18. **Bijections from Dyck and Motzkin meanders with catastrophes to pattern avoiding Dyck paths**
    Jean-Luc Baril and Sergey Kirgizov
    Discrete Mathematics Letters, 7, 2021

19. **Transformation à la Foata for special kinds of descents and excedances**
Jean-Luc Baril and Sergey Kirgizov
Enumerative Combinatorics and Applications, ECA 1:3, Article S2R19, 2021

20. **Pattern distributions in Dyck paths with a first return decomposition constrained by height**
Jean-luc Baril, Richard Genestier and Sergey Kirgizov
Discrete Mathematics, 343(9), 2020

21. **Bijections between directed animals, multisets and Grand-Dyck paths**
Jean-luc Baril, David Bevan and Sergey Kirgizov
The Electronic Journal of Combinatorics, 27(2), Article P2.10, 2020

22. **Motzkin paths with a restricted first return decomposition**
Jean-luc Baril, Sergey Kirgizov and Armen Petrossian
INTEGERS, 19, 2019

23. **Enumeration of Lukasiewicz paths modulo some patterns**
Jean-luc Baril, Sergey Kirgizov and Armen Petrossian
Discrete Mathematics, 342(4), 2019

24. **Descent distribution on Catalan words avoiding a pattern of length at most three**
Jean-luc Baril, Sergey Kirgizov and Vincent Vajnovszki
Discrete Mathematics, 341(9), Pages 2608-2615, 2018

25. **Dyck paths with a first return decomposition constrained by height**
June 2017, Jean-luc Baril, Sergey Kirgizov and Armen Petrossian
Discrete Mathematics, 341(6), Pages 1620-1628, 2018

26. **Forests and pattern avoiding permutations modulo pure descents**
Jean-luc Baril, Sergey Kirgizov and Armen Petrossian
Pure Mathematics and Applications (PUMA), 27(1), 2018

27. **Patterns in treeshelves**
Jean-luc Baril, Sergey Kirgizov and Vincent Vajnovszki
Discrete Mathematics, 340 (12), Pages 2946-2954, 2017

28. **The pure descent statistic on permutations**
Jean-luc Baril, Sergey Kirgizov *Discrete Mathematics*, 340 (10), Pages 2550-2558, 2017

29. **Information fusion-based approach for studying influence on twitter using belief theory**
Lobna Azaza, Sergey Kirgizov, Marinette Savonnet, Éric Leclercq, Nicolas Gastineau, and Rim Faiz
Computational Social Networks, 3, 2016, pp. 1–26, 2016

30. **The complexity of deciding whether a graph admits an orientation with fixed weak diameter**
Julien Bensmail, Romaric Duvignau, Sergey Kirgizov
Discrete Mathematics and Theoretical Computer Science (DMTCS), 17(3):31-42, 2016.
31. **Suppression Distance Computation for Set Covers and Hierarchies**
François Queyroi and Sergey Kirgizov
Information Processing Letters, 115.9 (2015): 689-693, 2015
32. **Towards realistic modeling of IP-level routing topology dynamics**
Clémence Magnien, Amélie Medem, Sergey Kirgizov and Fabien Tarissan
Networking Science, 3 (1-4), pp.24-33, 2013

French Scientific Journals (2)

1. **Évaluation de l'influence polarisée dans un réseau multi-relationnel : application à twitter**
Lobna Azaza, Marinette Savonnet, Eric Leclercq, Sergey Kirgizov, and Rim Faiz
Revue Document Numérique, 2017/1 (Vol. 20)
2. **Un observatoire pour la modélisation et l'analyse des réseaux multirelationnels : une application à l'étude du discours politique sur twitter**
Ian Basaille, Sergey Kirgizov, Éric Leclercq, Marinette Savonnet, Nadine Cullot, Thierry Grison, and Elisabeth Gavignet
Revue Document Numérique, 2017/1 (Vol. 20)

Book chapters (1)

1. **Tweets from the Campaign Trail: Researching Candidates' Use of Twitter during the European Parliamentary Elections**, co-auteur du chapitre *"SNFreezer: a Platform for Harvesting and Storing Tweets in a Big Data Context"*, éditeurs Alexander Frame, Arnaud Mercier, Gilles Brachotte, and Caja Thimm. Pages 19-33. Peter Lang. 2016

Peer-Reviewed International Conferences (28)

1. **Emerging consecutive pattern avoidance**
Nathanaël Hassler and Sergey Kirgizov
Presented by Nathanaël Hassler, *Permutation Patterns*, St Andrews, Scotland, 2025
2. **Fuzzy logic-based IoT object integrity self-management**
Abdelhamid Garah, Nader Mbarek and Sergey Kirgizov
Presented by Abdelhamid Garah, *The 12th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Paris, France, 2025.

3. **DBSCAN-based IoT object integrity self-management**
   Abdelhamid Garah, Nader Mbarek and Sergey Kirgizov
   Presented by Abdelhamid Garah, *The 21st International Wireless Communications & Mobile Computing Conference (IWCMC)*, Abu Dhabi, UAE, 2025

4. **Brick wall excursions**
   Sergey Kirgizov, Khaydar Nurligareev, and Michael Wallner
   Presented by Khaydar at *ALEA Days, Journées ALEA*, Marseille, 17-21 March, 2025

5. **Distribution of endhered patterns in RNA-related secondary structures**
   Célia Biane, Greg Hampikian, Sergey Kirgizov, Khaydar Nurligareev, and Daniel Pinson
   Presented by Sk at *SeqBIM 2024*, Rennes, 28-29 November, 2024

6. **The combinatorics of Motzkin polyominoes abstract**
   Jean-Luc Baril, Sergey Kirgizov, José Luis Ramírez, and Diego Villamizar
   Presented by Diego Villamizar from Universidad Sergio Arboleda (Colombia) *ICECA, International Conference Enumerative Combinatorics and Applications*, University of Haifa, Israel, 26-28 August, 2024

7. **Asymptotics of endhered patterns in perfect matchings**
   Célia Biane, Khaydar Nurligareev, and Sergey Kirgizov
   Presented by Khaydar at *ALEA Days, Journées ALEA*, Marseille, March 11-15, 2024

8. **Grand zigzag knight's paths**
   Jean-Luc Baril, Nathanaël Hassler, Sergey Kirgizov and José Luis Ramírez
   Presented by Nathanaël at *ALEA Days, Journées ALEA*, Marseille, March 11-15, 2024

9. **A lattice on Dyck paths close to the Tamari Lattice**
   Jean-Luc Baril, Sergey Kirgizov and Mehdi Naima
   Presented by Mehdi at *Computational Logic and Applications*, Jagiellonian University, Kraków, December 14-15, 2023

10. **IoT Data Confidentiality Self-Management**
    Abdelhamid Garah, Nader Mbarek and Sergey Kirgizov
    Presented by Abdelhamid at IEEE DASC, Abu Dhabi, UAE, November 14-17, 2023

11. **Decision Tree-Based Confidentiality Self-Management in the Internet of Things**
    Abdelhamid Garah, Nader Mbarek, and Sergey Kirgizov
    Presented by Nader at LCN 2023, Daytona Beach, Florida, USA, October 1-5, 2023

12. **Sturm meets Fibonacci in Minkowski's fractal bar**
    Sergey Dovgal and Sergey Kirgizov
    Permutation Patterns 2023, Dijon, France, July 3-7, 2023
13. **An introduction to Dyck paths with air pockets**
    Jean-Luc Baril, Sergey Kirgizov, Rémi Maréchal and Vincent Vajnovszki
    ALEA Days, Journées ALEA, March 13-17, 2023
14. **Fibonacci q-decreasing words: enumerative results and Gray codes**
    Sergey Kirgizov, Jean-Luc Baril and Vincent Vajnovszki
    AUA-UAEU Workshop on Graph Theory, Combinatorics and Applications (GTCA), Al Ain, UAE, November 13-15, 2022
15. **An Architecture for Confidentiality Self-management in the Internet of Things**
    Abdelhamid Garah, Nader Mbarek, and Sergey Kirgizov
    International Workshop on IoT Security and Quality of Service, SITIS 2022, 21 October, 2022
16. **Q-bonacci words and numbers**
    Sergey Kirgizov
    The Twentieth International Conference on Fibonacci Numbers and Their Applications, Sarajevo, Bosnia and Herzegovina, 25-29 July, 2022
17. **Asymptotic bit frequency in Fibonacci words**
    Sergey Kirgizov, Jean-Luc Baril and Vincent Vajnovszki
    GASCom 2022, Villa Toeplitz, Varese, Italy, June 13-15, 2022
18. **Clique percolation method: memory efficient almost exact communities**
    Alexis Baudin, Maximilien Danisch, Sergey Kirgizov, Clémence Magnien and Marwan Ghanem
    17th anniversary of the International Conference on Advanced Data Mining and Applications, (ADMA'21)
    Sydney, Australia, 2-4 February, 2022
19. **Qubonacci words**
    Jean-Luc Baril, Sergey Kirgizov and Vincent Vajnovszki
    Permutations patterns 2021, Virtual, University of Strathclyde, June 15-16
20. **Pattern distribution in faro words and permutations**
    Jean-luc Baril and Sergey Kirgizov
    Permutations patterns 2020, Virtual, Valparaiso University, June 30 and July 1
21. **Temporal density of community structure**
    Sergey Kirgizov and Éric Leclercq
    The 10th Conference on Network Modeling and Analysis : MARAMI, Dijon, France, 2019

22. **Pattern avoiding permutations modulo pure descents**
Jean-luc Baril, Sergey Kirgizov and Armen Petrossian
Permutation Patterns Conference, Reykjavik University, Iceland, 2017

23. **Temporal density of complex networks and ego-community dynamics**
Sergey Kirgizov, Eric Leclercq
Annual Conference on Complex Systems (ECCS or CCS), Amsterdam, 2016

24. **Towards a Twitter Observatory: A multi-paradigm framework for collecting, storing and analysing tweets**
Ian Basaille, Sergey Kirgizov, Éric Leclercq, Marinette Savonnet, et Nadine Cullot
RCIS 2016, IEEE Tenth International Conference on Research Challenges in Information Science, Grenoble, France, 1-3 June 2016

25. **A web application for event detection and exploratory data analysis for Twitter data** Sergey Kirgizov, Eric Leclercq, Marinette Savonnet, Alexander Frame, Ian Basaille-Gahite
Twitter at the European Elections 2014: International Perspectives on a Political Communication Tool, Dijon, 2015

26. **Influence Assessment in Twitter Multi-Relational Network** Lobna Azaza, Sergey Kirgizov, Marinette Savonnet, Éric Leclercq, Rim Faiz
Eleventh International IEEE Conference on Signal Image Technologies and Internet-Based System (SITIS), Bangkok, Thailand: 436-443, 2015

27. **On the complexity of turning a graph into the analogue of a clique**
Julien Bensmail, Romaric Duvignau, Sergey Kirgizov
9th International colloquium on graph theory and combinatorics (ICGT), 2014

28. **Using Reinforcement Learning for Autonomic Resource Allocation in Clouds: towards a fully automated workflow**
Xavier Dutreilh, Sergey Kirgizov, Olga Melekhova, Jacques Malenfant, Nicolas Rivierre, and Isis Truck
In The 7th International Conference on Autonomic and Autonomous Systems (ICAS'2011), best paper award, 2011

Peer-Reviewed French National Conferences (12)

1. **L'art et la combinatoire**
Sergey and Yulia Kirgizova
Journée Arts Design et Sciences, Dijon, 30 mai, 2024

2. **Enumeration of Dyck paths with air pockets**
Jean-Luc Baril, Sergey Kirgizov, Rémi Maréchal and Vincent Vajnovszki

Poster *Les journées JNIM (Journées Nationales de l'Informatique Mathématique)*, Villeneuve d'Ascq (France), 29 mars – 1 avril, 2022

3. **Packing coloring and subsets preserving path distance**
   Nicolas Gastineau, Benjamin Gras, Sergey Kirgizov, Mahmoud Omidvar
   Journées Graphes et Algorithmes (JGA), Paris, 2016
4. **(Re)constuire la temporalité d'un événement médiatique sur Twitter : une étude contrastive**
   Tatiana Kondrashova, Alexander Frame, and Sergey Kirgizov
   XXe congrès de la SFSIC : Temps, temporalités et information-communication, 8-10 juin 2016 Metz (France)
5. **Évaluation de l'influence dans un réseau multi-relationnel : le cas de Twitter**
   Lobna Azaza, Sergey Kirgizov, Marinette Savonnet, Éric Leclercq, Rim Faiz
   INFORSID'2016, Le congrès INFORSID (INFormatique des ORganisations et Systèmes d'Information et de Décision, du 31 mai au 3 juin, à Grenoble, France
6. **A la recherche des mini-publics : un problème de communautés, de singularités et de sémantique**
   Données participatives et sociales
   Eric Leclercq, Sergey Kirgizov, Maximilien Danisch
   16ème conférence francophone sur l'Extraction et la Gestion des Connaissances (EGC 2016): l'atelier Données participatives et sociales, Reims, Janvier 2016
7. **Evaluation de l'influence sur Twitter : Application au projet "Twitter aux Elections Européennes 2014"**
   Sergey Kirgizov, Lobna Azaza, Éric Leclercq, Marinette Savonnet et Alexander Frame
   Journées d'étude "Etudier le Web politique : Regards croisés", Lyon, 12-13 Mai 2015
8. **Papers$^\gamma$, Discussing board for scientific papers**
   Conference SO Data 3, Paris, 26 Mars 2015
9. **Internet Topology Dynamics: stochastic process estimation from partial observations**
   with Clémence Magnien
   Journée jointe des GDR ISIS et Phénix "Analyse et inférence pour les réseaux", Paris, 2013
10. **Distribution multimodale de la taille du sous-graphe des plus courts chemins dans un graphe aléatoire**
    Journées Graphes et Algorithmes (JGA), Orsay, 2013
11. **Dynamique de la topologie de l'internet : impact de la fréquence de mesure sur les observations**
    Sergey Kirgizov, Clémence Magnien, Fabien Tarissan and Azhu Liu
    24ème colloque GRETSI, Brest, France, 2013

12. **Vers une modélisation réaliste de la dynamique de la topologie de routage au niveau IP**
Sergey Kirgizov, Amélie Medem, Clémence Magnien, and Fabien Tarissan
Journées Automnales 2012 ResCom, Paris, 2012

Teaching

I give lectures, organise laboratory works and student projects. I teach mostly in French, but some of the courses are in English and Russian.

- Polytech Dijon, École polytechnique universitaire de Bourgogne
  Université Bourgogne Europe, 2019–current
  - Systèmes UNIX
  - Langue vivante 2, Русский язык
  - Big Data
  - Ingénierie des systèmes d'information
  - Systèmes d'information avancés
  - Hash & Crypto
  - Sécurité et intégrité cryptographique des données
  - Machine learning: reinforcement learning
- ESIREM, Université de Bourgogne, 2018–2019, Vacataire
  - Systèmes UNIX
  - Systèmes d'information avancés
- Université de Bourgogne, 2015-2017, ATER
  - Algorithmique avancée
  - Initiation à la programmation en Java
  - Systèmes d'exploitation
  - Langages C et C++
  - OpenMP
  - MPI
- Université de Bourgogne, from January to August 2015, Postdoc
  - Interfaces visuelles (Java)
- Université Pierre et Marie Curie, doctorant-vacataire, 2012-2014
  - Initiation à la compilation et aux machines virtuelles (Scheme et C)
  - Programmation récursive (Scheme)
  - Structures discrètes