Glimpses into Recent Developments in Subword Combinatorics and Binomial Coefficients

Markus A. Whiteland *

Monday 30 June, 10:10

Abstract

Subsequences of strings play a crucial role in combinatorics on words, string algorithmics, formal language theory, and various applied fields such as bioinformatics. They provide rich algebraic structures through concepts such as Simon's congruence, the binomial equivalence relation, and M-equivalence.

The main focus of this talk is on binomial coefficients of words: for two words u and v, the quantity $\binom{u}{v}$ denotes the number of times v appears as a scattered subword in u. These quantities appear, for example, in (generalised) Parikh-matrices. Recently, a q-analogue of binomial coefficients of words has been introduced. This analogue defines a polynomial in q such that setting $q = 1$ recovers the classical binomial coefficient of words. This polynomial formulation enables a finer analysis of subword counting and I will discuss their combinatorial interpretations, as well as their connection to Eğecioğlu's q-deformations of Parikh-matrices.

Additionally, binomial complexities of infinite words have been the subject of extensive recent research. In this talk, I will review recent results concerning k-binomial complexity functions of certain families of infinite words. If time permits, I will discuss some ongoing work regarding the automaticity of these functions in the context of fixed points of Parikh-collinear morphisms. These results shed new light on the combinatorial behavior of infinite words and their structural constraints.

Overall, this talk aims to provide a glimpse of recent developments of binomial coefficients of words and associated complexity functions. It is based on joint work with M. Golafshan, A. Renard, M. Rigo, M. Stipulanti, and N. Wingate.

*Department of Computer Science, Loughborough University, Loughborough, United Kingdom

Maximal 2-dimensional binary words of bounded degree

Alexandre Blondin Massé [*] Alain Goupil [†] Raphael L'Heureux [‡]
Louis Marin [§]

Monday 30 June, 11:40

Abstract

Let $d \in \{0, 1, 2, 3, 4\}$ and W be a 2-dimensional word of dimensions $h \times w$ on the binary alphabet $\{\square, \blacksquare\}$, where $h, w \in \mathbb{Z}_{>0}$. Assume that each occurrence of the letter $\blacksquare$ in W is adjacent to at most d letters $\blacksquare$ and let $|W|_{\blacksquare}$ be the number of letters $\blacksquare$ in W. We provide an exact formula for the maximum value of $|W|_{\blacksquare}$ for fixed (h, w). As a byproduct, we deduce an upper bound on the length of maximum snake polyominoes contained in a $h \times w$ rectangle.

[*]Université du Québec à Montréal, Montréal, Canada; and LACIM, Montréal, Canada
[†]Université du Québec à Trois-Rivières, Trois-Rivières, Canada; and LACIM, Montréal, Canada
[‡]Université du Québec à Trois-Rivières, Trois-Rivières, Canada
[§]Université Gustave Eiffel, LIGM, Champs-sur-Marne, France

Binomial Coefficients of Multidimensional Arrays

Mehdi Golafshan * Michel Rigo †

Monday 30 June, 12:20

Abstract

Motivated by Parikh matrices of picture arrays introduced in combinatorial image analysis, we propose a generalization of binomial coefficients of words to multidimensional arrays. These coefficients recursively count prescribed patterns occurring in an array. The base case is the one of binomial coefficients of words.

With our definition we extend Pascal's rule, the Chu–Vandermonde identity and therefore, the concept of Parikh matrices, in a natural way. We further present some more binomial-related identities and introduce (q,t)-deformations, i.e., multivariate polynomials whose evaluation at $(q,t) = (1,1)$ recovers the value of the classical coefficients. We explain the additional combinatorial information encoded in the coefficients of these (q,t)-polynomials compared to their integer-valued counterparts.

*Univ. of Liège, Dept. of Mathematics, Allée de la découverte 12 (B37), B-4000 Liège, Belgium

†Univ. of Liège, Dept. of Mathematics, Allée de la découverte 12 (B37), B-4000 Liège, Belgium

Supported by the FNRS Research grant T.196.23 (PDR)

Circularity and repetitiveness in non-injective DF0L systems

Herman Goulet-Ouellet [*] Karel Klouda [†] Štěpán Starosta [‡]

Monday 30 June, 14:30

Abstract

We study circularity in DF0L systems, a generalization of D0L systems. We focus on two different types of circularity, called weak and strong circularity. When the morphism is injective on the language of the system, the two notions are equivalent, but they may differ otherwise. Our main result shows that failure of weak circularity implies unbounded repetitiveness, and that unbounded repetitiveness implies failure of strong circularity. This extends previous work by the second and third authors for injective systems. To help motivate this work, we also give examples of non-injective but strongly circular systems.

[*]Czech Technical University in Prague, Faculty of Information Technology
This author was supported by the CTU Global Postdoc Fellowship program.
[†]Czech Technical University in Prague, Faculty of Information Technology
[‡]Czech Technical University in Prague, Faculty of Information Technology

And now there are four: Another brick in the wall of the optimal upper bound on the MP-ratio

Kristina Ago * Bojan Bašić †

Monday 30 June, 15:10

Abstract

The so-called *MP-ratio* is a kind of measure of how "packed with palindromes" a given word is. The lower bound on the MP-ratio for the set of all n-ary words is (trivially) 1, while the best possible upper bound is an open problem in the general case. It is solved for $n = 2$ (where the optimal upper bound is 4) and for $n = 3$ (where the optimal upper bound is 6). Also, it is known that in the n-ary case the optimal bound is between $2n$ and the order of the growth $n2^{\frac{n}{2}}$. In this article we solve this problem for quaternary words, for which we show that the best possible upper bound on the MP-ratio equals 8. We believe that this is the last case in which the result is $2n$, that is, we believe that for $n \geqslant 5$ there are words whose MP-ratio is strictly larger than $2n$.

*Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

†Department of Mathematics and Informatics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia

Digital Convexity and Combinatorics on Words

Alessandro De Luca * Gabriele Fici † Andrea Frosini ‡

Monday 30 June, 16:20

Abstract

An upward (resp. downward) digitally convex word is a binary word that best approximates from below (resp. from above) an upward (resp. downward) convex curve in the plane. We study these words from the combinatorial point of view, formalizing their geometric properties and highlighting connections with Christoffel words and finite Sturmian words. In particular, we study from the combinatorial perspective the operations of inflation and deflation on digitally convex words.

*DIETI, Università di Napoli Federico II, Italy
†Dipartimento di Matematica e Informatica, Università di Palermo, Italy
‡Dipartimento di Matematica e Informatica, Università di Firenze, Italy

S-adic expansions, invariant measures and substitutional subshifts

Martin Lustig *

Tuesday 1 July, 9:30

Abstract

In the first part of my talk I will present joint work with N. Bédaride and A. Hilion on (generalized) S-adic developments of a subshift X, and a related very convenient method ("vector towers") to describe invariant measures on X. In particular I will focus on our construction of d distinct ergodic probability measures for certain minimal subshifts X_d, for any given alphabet rank $r(X_d) = d \geq 1$, and the issuing construction of a fairly large family of minimal subshifts with topological entropy 0 and infinitely many distinct ergodic probability measures.

In a second part I will discuss the relation between substitutions on one hand and morphisms of free groups on the other, and describe a new invariant for primitive substitutional subshifts X which is expected to characterize X up to recognizable morphisms (work in progress).

*Institut de Mathématiques de Marseille, Aix-Marseille Université, Marseille, France

Factorizations and Monoids

Fabio Burderi *

Tuseday 1 July, 10:30

Abstract

In this paper we introduce the notion of generalized factorization for an arbitrary submonoid $M \subseteq A^*$, where A^* is the free monoid generated by an alphabet A, generalizing, in this way, the notion of factorization of A^*. Then we give a characterization of the free product of two submonoids of A^* in terms of unambiguous products of monoids. To do this we make use of the notion of coding partition of a set $X \subseteq A^+$, where A^+ is the free semigroup generated by an alphabet A. Moreover, given a coding partition of a set $X \subseteq A^+$, we will show how to construct a generalized factorization of X^*.

*Dipartimento d'Ingegneria, Università degli studi di Palermo, 90128 Palermo, Italy

Free Product of Formal Series

Fabio Burderi *

Tuseday 1 July, 11:40

Abstract

In this paper we introduce the notion of free product of formal series. Using this notion, we can characterize the free product of submonoids of A^*, where A^* is the free monoid generated by an alphabet A. Moreover given a set $X \subseteq A^+$, where A^+ is the free semigroup generated by an alphabet A, we can characterize a partition of X by the free product of the formal series associated to the classes of the partition.

*Dipartimento d'Ingegneria, Università degli studi di Palermo, 90128 Palermo, Italy

Note on dissecting power of regular languages

Josef Rukavicka *

Tuesday 1 July, 12:20

Abstract

Let $c > 1$ be a real constant. We say that a language L is *c-constantly growing* if for every word $u \in L$ there is a word $v \in L$ with $|u| < |v| \leq c + |u|$. We say that a language L is *c-geometrically growing* if for every word $u \in L$ there is a word $v \in L$ with $|u| < |v| \leq c|u|$. Given a language L, we say that L is REG-*dissectible* if there is a regular language R such that $|L \setminus R| = \infty$ and $|L \cap R| = \infty$. In 2013, it was shown that every c-constantly growing language L is REG-dissectible. In 2023, the following open question has been presented: "Is the family of geometrically growing languages REG-dissectible?"

For every $c > 1$, we construct a c-geometrically growing language L that is not REG-dissectible. Hence we answer negatively to the open question.

*Department of Mathematics, Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague

On the abelian complexity of infinite words

Idrissa Kaboré *

Wednesday 2 July, 9:30

Abstract

The abelian complexity is a combinatorial tool intervening in the study of infinite words. It was formally introduced in 2009 by Richomme et al. Since then it has been intensively studied. In this talk we give a survey on the developments on the subject with a focus on general results (joint work with Cassaigne in 2016) concerning the links between abelian complexity with other notions on infinite words. We conclude with a few questions.

*Université Polytechnique de Bobo-Dioulasso, Burkina Faso

Purely automatic sequences with the uniform distribution property

Shuo Li * Narad Rampersad †

Wednesday 2 July, 10:30

Abstract

An infinite sequence with finitely many distinct letters is said to have the *uniform distribution property* if all letters in the alphabet of the sequence have the same density in all arithmetical progressions. This property was first studied by Gelfond for *sum of digits* functions. In this note, we characterize a larger class of purely automatic sequences with the same property.

*Department of Mathematics, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9

†Department of Mathematics, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9

Shuffle squares and nest-free graphs

Jarosław Grytczuk [*†] Bartłomiej Pawlik [‡] Andrzej Ruciński [§¶]

Wednesday 2 July, 11:40

Abstract

A *shuffle square* is a word consisting of two shuffled copies of the same word. For instance, the French word tuteurer is a shuffle square, as it can be split into two copies of the word tuer. An *ordered graph* is a graph with a fixed linear order of vertices.

We propose a representation of shuffle squares in terms of special nest-free ordered graphs and demonstrate the usefulness of this approach by applying it to several problems. Among others, we prove that binary words of the type $(1001)^n$, n odd, are not shuffle squares and, moreover, they are the only such words among all binary words whose every 1-run has length one or two, while every 0-run has length two. We also provide a counterexample to a believable stipulation that binary words of the form $1^n0^{n-2}1^{n-4}\cdots$, n odd, are far from being shuffle squares (the distance measured by the minimum number of letters one has to delete in order to turn a word into a shuffle square).

[*]Supported by Narodowe Centrum Nauki, grant 2020/37/B/ST1/03298.

[†]Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-662 Warsaw, Poland

[‡]Institute of Mathematics, Silesian University of Technology, 44-100 Gliwice, Poland

[§]Supported by Narodowe Centrum Nauki, grant 2024/53/B/ST1/00164.

[¶]Faculty of Mathematics and Computer Science, Adam Mickiewicz University, 61-614 Poznań, Poland

Clustering of Return Words in Languages of Interval Exchanges

Francesco Dolce * Christian B. Hughes †

Wednesday 2 July, 12:20

Abstract

A word over an ordered alphabet is said to be *clustering* if identical letters appear adjacently in its Burrows-Wheeler transform. Such words are strictly related to (discrete) interval exchange transformations. We use an extended version of the well-known Rauzy induction to show that every return word in the language generated by a regular interval exchange transformation is clustering, partially answering a question of Lapointe (2021).

*Czech Technical University in Prague (Czech Republic)
†Czech Technical University in Prague (Czech Republic)

Combinatorial problems on closed and privileged words

Daniel Gabrić *

Thursday 3 July, 9:30

Abstract

A word w has a *border* u if u is a non-empty proper prefix and suffix of w. A word w is said to be *closed* if w is of length at most 1 or if w has a border that occurs exactly twice in w. A word w is said to be *privileged* if w is of length at most 1 or if w has a privileged border that occurs exactly twice in w. In this talk, we discuss recent combinatorial results on closed and privileged words, including their enumeration and generation. We also highlight key open problems and directions for future research.

*School of Computer Science, University of Guelph, Guelph, Ontario, Canada

The Heinis spectrum has non-empty interior

Harold Erazo * Carlos Gustavo Moreira †

Thursday 3 July, 10:30

Abstract

The Heinis spectrum Ω is the set of all pairs (α_u, β_u) such that $\alpha_u = \liminf_{n\to\infty} \frac{p_u(n)}{n}$ and $\beta_u = \limsup_{n\to\infty} \frac{p_u(n)}{n}$ for some infinite word u. In this paper, we demonstrate that there exists a closed connected set with non-empty interior contained in Ω. Furthermore, every point in this set can be represented as the pair (α_u, β_u) for some recurrent word u. The construction is explicit, algorithmic in nature and is based on constructing certain "Cantor sets of integers", whose "gaps" correspond to blocks of zeros.

*IMPA, Estrada Dona Castorina 110, 22460-320, Rio de Janeiro, Brazil
†SUSTech International Center for Mathematics, Shenzhen, Guangdong, People's Republic of China

On the closed-rich constant of infinite words

Anuran Maity * Svetlana Puzynina †

Thursday 3 July, 11:40

Abstract

A finite word w is called *closed* if it has length at most 1 or it contains a proper factor that occurs both as a prefix and as a suffix but does not have internal occurrences. An infinite word u is called *closed-rich* if the infimum of all possible ratios between the number of closed factors within any factor w of u and square of the length of w exists and is positive. We define this infimum as the closed-rich constant C_u of the infinite closed-rich word u. Puzynina and Parshina (2024) proved that infinite closed-rich words exist. In this paper, we estimate possible values of C_u for an infinite closed-rich word u, and apply these results to estimate the supremum C_{sup} of the closed-rich constants of infinite closed-rich words. We show that $0.0952 < C_{sup} \leq 0.165964$, where the lower bound comes from the Fibonacci word. [1]

*Indian Institute of Technology Guwahati, Department of Mathematics, Guwahati, India
Saint Petersburg State University, 7/9 Universitetskaya nab., 199034 St. Petersburg, Russia

†Saint Petersburg State University, 7/9 Universitetskaya nab., 199034 St. Petersburg, Russia

[1]The results from Section 3 were supported by the Russian Science Foundation, project 23-11-00133.

Symmetries of Rich Sequences with Minimum Critical Exponent

L'ubomíra Dvořáková [†] Edita Pelantová [‡]

Thursday 3 July, 12:20

Abstract

We state a conjecture on the repetition threshold of rich sequences over alphabet of any size. It is known to hold for binary and ternary alphabets. We provide two main contributions that may be helpful for the proof on larger alphabets. First we show that the ternary rich sequence with minimum critical exponent is a morphic image of a fixed point, i.e., an HD0L sequence. Second we draw attention to the fact that the rich sequences having the minimum critical exponent show a large degree of symmetry, i.e., they are G-rich with respect to a group G generated by more than one antimorphism. The notion of G-richness generalizes the notion of richness in palindromes which is based on one antimorphism, namely the reversal mapping.

[†]FNSPE Czech Technical University in Prague
[‡]FNSPE Czech Technical University in Prague

Words avoiding half-flips

James Currie * Narad Rampersad †

Thursday 3 July, 14:30

Abstract

We say that a word w contains a *half-flip* if it contains non-empty factors uv and vu where $|u| = |v|$. Fici reports a non-constructive proof of the existence of an infinite word over a finite alphabet avoiding half-flips and asks the size of the smallest alphabet over which half-flips may be avoided.

Half-flips are unavoidable over a 4-letter alphabet. Over an 8-letter alphabet we give a 3-uniform D0L avoiding half-flips; over a 5-letter alphabet we give a messier HD0L construction. We show how the method of templates, introduced by the authors, could potentially be used to analyse half-flip avoidance by this HD0L.

*The University of Winnipeg, Winnipeg MB R3B 2E9, Canada.
†The University of Winnipeg, Winnipeg MB R3B 2E9, Canada

Avoiding abelian and additive powers in rich words

Jonathan Andrade * Lucas Mol †

Thursday 3 July, 15:10

Abstract

This paper concerns the avoidability of abelian and additive powers in infinite rich words. In particular, we construct an infinite additive 5-power-free rich word over $\{0,1\}$ and an infinite additive 4-power-free rich word over $\{0,1,2\}$. The alphabet sizes are as small as possible in both cases, even for abelian powers.

*Department of Mathematics and Statistics, University of Victoria, Victoria, BC, Canada
†Department of Mathematics and Statistics, Thompson Rivers University, Kamloops, BC, Canada

A Characterization of Algebraic Multivariate Power Series with Sparse Support

Seda Albayrak *

Thursday 3 July, 16:20

Abstract

Christol's theorem characterizes algebraic formal power series over finite fields in terms of automatic sequences, establishing a fundamental link between algebraicity and computability. A refinement of this result by Albayrak and Bell provides an algebraic characterization of formal power series whose support is a sparse automatic set, showing that sparseness can be characterized in terms of certain key transformations such as the Frobenius map, multiplicative scaling, and power transformations. In this paper, we extend these results to the multivariate setting, building on Salon's generalization of Christol's theorem. By using a characterization of sparse regular languages, we look at algebraic multivariate formal power series whose supports are sparse subsets of $\mathbb{N}^d$ and characterize them in terms of above-mentioned transformations, providing a structural understanding of the interplay between algebraicity and sparseness in multiple variables.

*Simon Fraser University, Burnaby BC V5A 1S6, Canada

The link between return words and extensions of factors

France Gheeraert *

Friday 4 July, 9:30

Abstract

Extensions of factors, defined as the letters that can surround factors, were introduced and studied for their connection with factor complexity. On the other hand, return words, which are the words separating two occurrences of the same factor, originate in the symbolic dynamics approach of words and are mostly known for their link with S-adic representations. Throughout the years, many results showed that information on the number or structure of one of the two could be used to study the other. In this talk, I will present a survey of some of these results.

*Laboratoire Amiénois de Mathématique Fondamentale et Appliquée, Université Picardie Jules Verne, Amiens, France

A Succinct Study of Positionality for Dumont–Thomas Numeration Systems

Savinien Kreczman * Sébastien Labbé † Manon Stipulanti ‡

Friday 4 July, 10:30

Abstract

Numeration systems are maps between a set of numbers and a set of words that act as representations of these numbers. One desirable property is positionality: the ability to relate positions in the words to values of the numbers. In general, positionality is hard to decide. In this article, we obtain a criterion to decide the positionality of so-called Dumont–Thomas numeration systems, arising from substitutions. Then, we particularize this criterion to some well-behaved classes of substitutions, allowing us to link the related systems to existing literature.

*Department of Mathematics, ULiège, Liège, Belgium
†LaBRI, Université de Bordeaux, Talence, France
‡Department of Mathematics, ULiège, Liège, Belgium

Linear Recurrence Sequence Automata and the Addition of Abstract Numeration Systems

Olivier Carton * Jean-Michel Couvreur † Martin Delacourt ‡
Nicolas Ollinger §

Friday 4 July, 11:40

Abstract

Abstract numeration systems encode natural numbers using radix ordered words of an infinite regular language and linear recurrence sequences play a key role in their valuation. Sequence automata, which are deterministic finite automata with an additional linear recurrence sequence on each transition, are introduced to compute various $\mathbb{Z}$-rational non commutative formal series in abstract numeration systems. Under certain Pisot conditions on the recurrence sequences, the support of these series is regular. This property can be leveraged to derive various synchronized relations including a deterministic finite automaton that computes the addition relation of various Dumont-Thomas numeration systems and deterministic finite automata converting between various numeration systems. A practical implementation for Walnut is provided.

*Université Paris Cité, CNRS, IRIF, F-75013, Paris, France
†Université d'Orléans, INSA CVL, LIFO, UR 4022, Orléans, France
‡Université d'Orléans, INSA CVL, LIFO, UR 4022, Orléans, France
§Université d'Orléans, INSA CVL, LIFO, UR 4022, Orléans, France

About Δ-numeration

Bastien Laboureix * Eric Domenjoud †

Friday 4 July, 12:20

Abstract

In this article we study a numeration system previously used to prove combinatorial properties in discrete geometry: the Δ-numeration. Since this system, introduced via the fully subtractive algorithm, has been seen mainly as a tool, we propose here to study it from the point of view of numeration systems. In particular, we make the link with β-numeration and Cantor real bases. We reintroduce the rewriting system introduced to calculate in Δ-numeration. This systems is based on the properties of the fully subtractive algorithm and is normalising. Finally, we study the ultimately periodic case, a special case of alternate bases, and show that the ultimately periodic words represent exactly the elements of $\mathbb{Q}[\beta]$ where β is the inverse of a Pisot number.

*Centre Borelli, ENS Paris-Saclay, Université Paris-Saclay
†Université de Lorraine, CNRS, LORIA, 54000 Nancy, France