Maximum likelihood, tropical models and matroids

UKBB, Basel

Raul Penaguiao

April 16th, 2024

Slides can be found at `raulpenaguiao.github.io/`
Joint work with Federico Ardila and Christopher Eur

Vita

- Bachelor in Lisbon University.
- Masters in ETH.
- PhD in UZH.
- SNF PostDoc grant.

ETH ZÜRICH

MASTERS THESIS

The Chromatic Symmetric Function on Random Graphs

Author: Raúl Penaguião

Supervisors: Valentin Féray, Angelika Steger

PATTERN HOPF ALGEBRAS

RAUL PENAGUIAO

University of Zurich, Institute of Mathematics

ABSTRACT. In this paper, we expand on the notion of *combinatorial presheaf*, first introduced explicitly by Aguiar and Mahajan in 2010 but already present in the literature in some other points of view. We do this by adapting the algebraic framework of species to the study of substructures in combinatorics. Afterwards, we consider functions that count the number of patterns of objects and endow the linear span of these functions with a product and a coproduct. In this way, any well behaved family of combinatorial objects that admits a notion of substructure generates a Hopf algebra, and this association is functorial. For example, the Hopf algebra on permutations studied by Vargas in 2014 and the Hopf algebra on symmetric functions are particular cases of this construction.

A specific family of pattern Hopf algebras of interest are the ones arising from *commutative combinatorial presheaves*. This includes the presheaves on graphs, posets and generalized permutahedra. Here, we show that all the pattern Hopf algebras corresponding to commutative presheaves are free.

We also study a remarkable non-commutative presheaf structure on marked permutations, *i.e.* permutations with a marked element. These objects have a natural product called inflation, which is an operation motivated by factorization theorems for permutations. In this paper we find new factorization theorems for marked permutations. We use these theorems to show that the pattern Hopf algebra for marked permutations is also free, using Lyndon words techniques.

THE FEASIBLE REGIONS FOR CONSECUTIVE PATTERNS OF PATTERN-AVOIDING PERMUTATIONS

JACOPO BORGA AND RAUL PENAGUIAO

ABSTRACT. We study the feasible region for consecutive patterns of pattern-avoiding permutations. More precisely, given a family $\mathcal{C}$ of permutations avoiding a fixed set of patterns, we consider the limit of proportions of consecutive patterns on large permutations of $\mathcal{C}$. These limits form a region, which we call the *consecutive patterns feasible region for* $\mathcal{C}$.

We determine the dimension of the consecutive patterns feasible region for all families $\mathcal{C}$ closed either for the direct sum or the skew sum. These families include for instance the ones avoiding a single pattern and all substitution-closed classes. We further show that these regions are always convex and we conjecture that they are always polytopes. We prove this conjecture when $\mathcal{C}$ is the family of τ-avoiding permutations, with either τ of size three or τ a monotone pattern. Furthermore, in these cases we give a full description of the vertices of these polytopes via cycle polytopes.

Along the way, we discuss connections of this work with the problem of packing patterns in pattern-avoiding permutations and to the study of local limits for pattern-avoiding permutations.

UNIVERSITY OF ZÜRICH

Algebraic and geometric studies of combinatorial substructures and chromatic invariants

by

Raul Penaguiao

Vita

- PostDoc at the Freie Universität Berlin.
- PostDoc at San Francisco state University.
- PostDoc at the Max Planck institute.
- Software engineer at Haeusler.

Freie Universität Berlin

San Francisco State University

MAX-PLANCK-GESELLSCHAFT

HAEUSLER
the forming factory

The tropical critical points of an affine matroid

Federico Ardila-Mantilla* Christopher Eur† Raul Penaguiao‡

Abstract

We prove that the number of tropical critical points of an affine matroid (M, e) is equal to the beta invariant of M. Motivated by the computation of maximum likelihood degrees, this number is defined to be the degree of the intersection of the Bergman fan of (M, e) and the inverted Bergman fan of $N = (M/e)^{\perp}$, where e is an element of M that is neither a loop nor a coloop. Equivalently, for a generic weight vector w on $E - e$, this is the number of ways to find weights $(0, x)$ on M and y on N with $x + y = w$ such that on each circuit of M (resp. N), the minimum x-weight (resp. y-weight) occurs at least twice. This answers a question of Sturmfels.

ANTIPODE FORMULAS FOR PATTERN HOPF ALGEBRAS

RAUL PENAGUIAO, YANNIC VARGAS

San Francisco State University

Technische Universität Graz

Abstract. The permutation pattern Hopf algebra is a commutative filtered and connected Hopf algebra. Its product structure stems from counting patterns of a permutation, interpreting the coefficients as permutation quasi-shuffles. The Hopf algebra was shown to be a free commutative algebra and to fit into a general framework of pattern Hopf algebras, via species with restrictions.

In this paper we introduce the cancellation-free and grouping-free formula for the antipode of the permutation pattern Hopf algebra. To obtain this formula, we use the popular sign-reversing involution method, by Benedetti and Sagan. This formula has applications on polynomial invariants on permutations, in particular for obtaining reciprocity theorems. On our way, we also introduce the packed word patterns Hopf algebra and present a formula for its antipode.

Other pattern algebras are discussed here, notably on parking functions, which recovers notions recently studied by Adeniran and Pudwell, and by Qiu and Remmel.

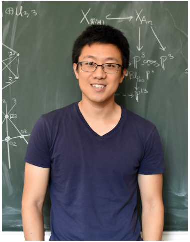
The dimension of the feasible region of pattern densities*

Frederik Garbe† Daniel Kráľ‡ Alexandru Malekshahian§
Raul Penaguiao¶

Abstract

A classical result of Erdős, Lovász and Spencer from the late 1970s asserts that the dimension of the feasible region of densities of graphs with at most k vertices in large graphs is equal to the number of non-trivial connected graphs with at most k vertices. Indecomposable permutations play the role of connected graphs in the realm of permutations, and Glebov et al. showed that pattern densities of indecomposable permutations are independent, i.e., the dimension of the feasible region of densities of permutation patterns of size at most k is at least the number of non-trivial indecomposable permutations of size at most k. However, this lower bound is not tight already for $k = 3$. We prove that the dimension of the feasible region of densities of permutation patterns of size at most k is equal to the number of non-trivial Lyndon permutations of size at most k. The proof exploits an interplay between algebra and combinatorics inherent to the study of Lyndon words.

Matroids and likelihood estimation



Left: Christopher Eur assistant professor at Carnegie Mellon University, **Right:** Federico Ardila, author of the principles of diversity and was given multiple awards.

A multinomial model

Is the dice fair?

Let $X \sim \mathrm{Mul}(4000, p, q, r, s)$ such that $p + q + r + s = 1$.

We see $X_1 = 980$ $X_2 = 991$, $X_3 = 1033$, and $X_4 = 996$.

$$\mathbb{P}[p, q, r, s] \propto p^{980} q^{991} r^{1033} s^{996}$$

Extremal points satisfy

$$\frac{\partial \mathbb{P}[p, q, r, s]}{\partial p} = \frac{\partial \mathbb{P}[p, q, r, s]}{\partial q} = \cdots$$

Optimal solution:

$$(p, q, r, s) = \left(\frac{980}{4000}, \frac{991}{4000}, \frac{1033}{4000}, \frac{996}{4000} \right).$$

Varieties as models - Equations

What is a variety?

Variety $\Leftrightarrow$ set of polynomial equations.

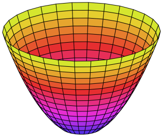


Figure: A variety given by the equation $x^2 + y^2 = z$.

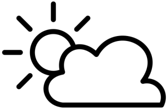
Varieties as a model - A toy model

A variety as a model

Modelling the weather using six variables: temperature (T), atmospheric pressure (A), wind (w), humidity (h), precipitation (P), and cloudiness (C). But they satisfy certain equations for instance

$$(P+A)^2 = C^2(T+A^2h)$$

$$(w+T)^3 = C + P^2A + P^3T$$



Each point in the variety is understood to be a **possible value** to occur.

Varieties as models - The strengths of algebra

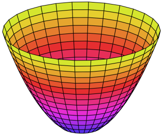


Figure: A variety given by the equation $x^2 + y^2 = z$.

Any line cuts through the parabola in two points.

Theorem (Bezout's theorem)

Intersecting a line through an equation of degree n gives us n different solutions, in general.

Varieties as models - Likelihood

Models

Variety as a model $\Leftrightarrow$ Equations + likelihood function

A variety as a model

Modelling the weather using six variables

$$(P+A)^2 = C^2(T+A^2h)$$

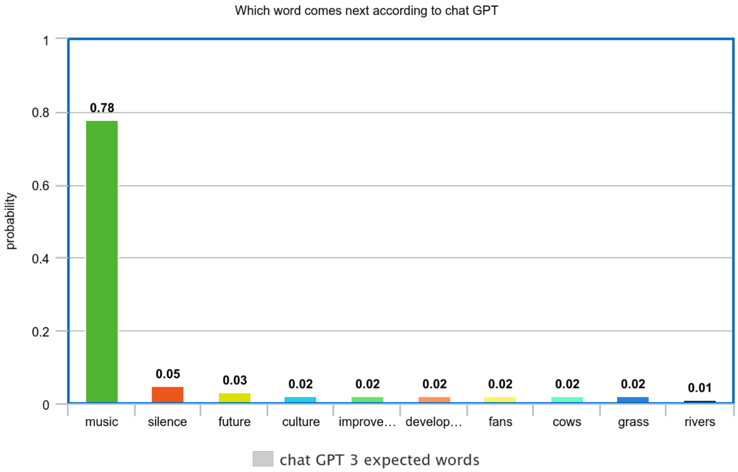
$$(w+T)^3 = C + P^2A + P^3T$$

$$\mathbb{P}[T, A, w, h, P, C] \propto PT^2CwA^3$$

Which word comes next? **The hills are alive with the sound of...**

Varieties as models - Likelihood

Which word comes next? **The hills are alive with the sound of...**



Maximum likelihood estimators

A variety as a model

Modelling the weather using six variables

$$(P+A)^2 = C^2(T+A^2h)$$

$$(w+T)^3 = C + P^2A + P^3T$$

$$\mathbb{P}[T,A,w,h,P,C] \propto PT^2CwA^3$$

How to compute optimal values? Find the points where $\dfrac{\partial}{\partial x_i}\mathbb{P}(x_i) = 0$.

These are called **critical points.**
What is the number of critical points of f? Depends on the choice of the likelihood function? No!
This number is called the **maximum likelihood degree** of a variety.

Tropical varieties

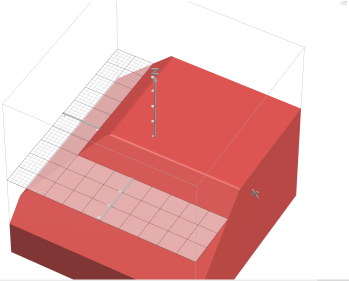
Tropicalization of polynomials

$$+ \Rightarrow min \text{ e.g. } 1 + x + y \Rightarrow min(1, x, y)\,.$$

$$\times \Rightarrow + \text{ e.g. } 3x^2y \Rightarrow 3 + 2x + y\,.$$

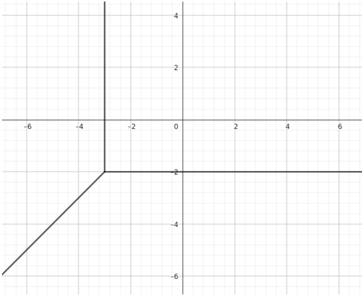
$$x^2 + 2x \Rightarrow min(2x, 2 + x)$$

$$3x + 2y + 1 \Rightarrow min(3 + x, y + 2, 1)$$

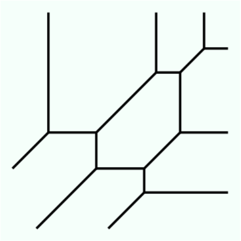


Tropical varieties

Tropical variety associated to $3x + 2y + 1$.

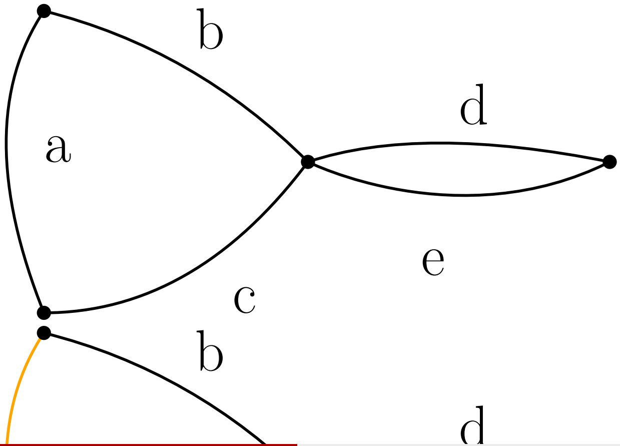


Tropical variety associated to $3x^3 - 2x^2y + xy^2 - 3y^3 - 2x^2 + xy - 3y^2 - 2x + y + 1$.

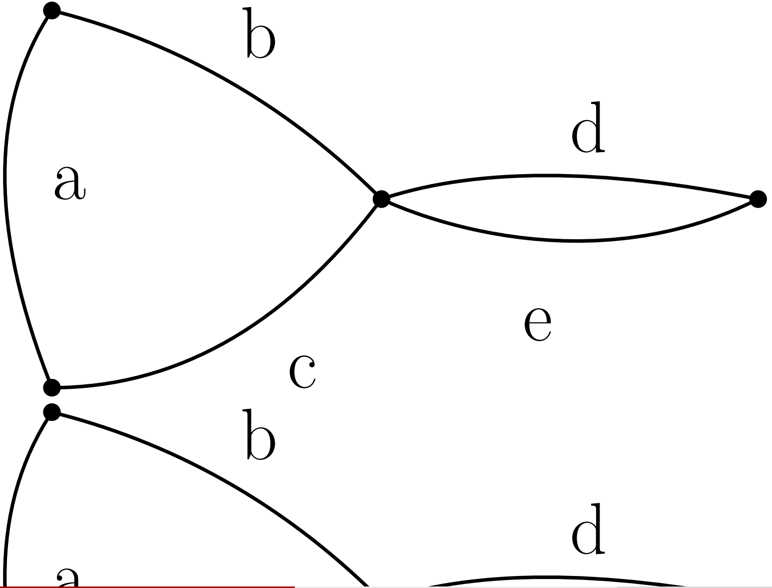


Matroids with therapies

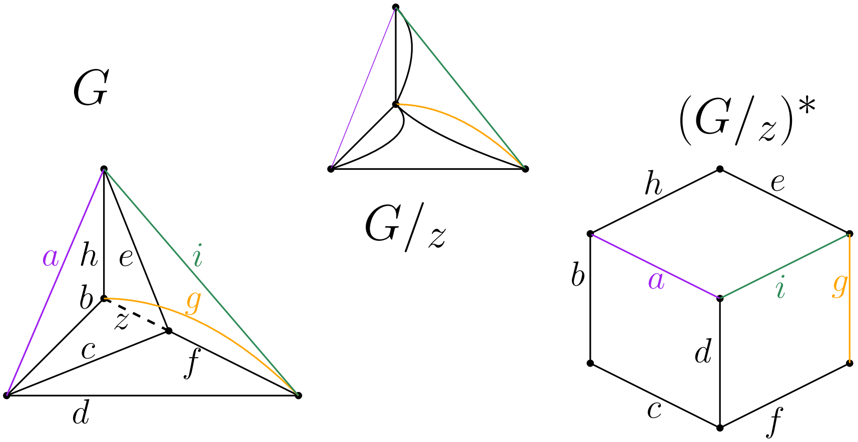
For a therapy of a pacient we are considering five different drugs.
$a, b, c \Rightarrow$ the pacient needs to take two of the three drugs.
$d, e \Rightarrow$ the pacient needs to take one of the two drugs.



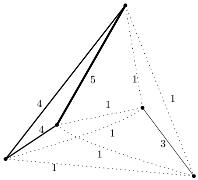
Dualizing matroids



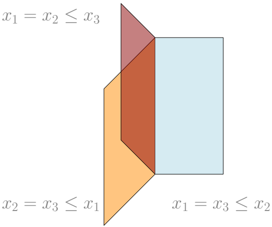
Dualizing matroids



The Bergman Fan



$$\Sigma_M \coloneqq \{\vec{x} \in \mathbb{R}^n | \forall \text{ circuits } C \text{ we have } \min_{c \in C} x_c \text{ is attained twice } \}.$$



Edge weight problem

Fix some vector $\mathbf{w} = (w_1, \ldots, w_n) \in \mathbb{Z}^n_{\geq 0}$.

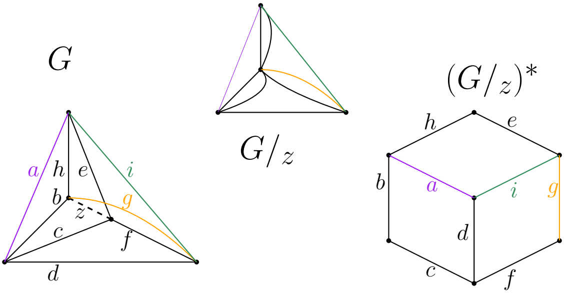
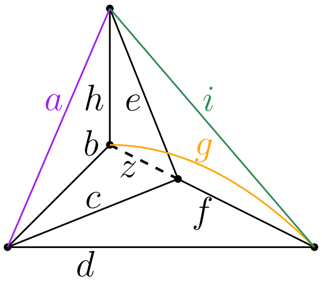


Figure: Find $\mathbf{x}$ and $\mathbf{y}$ edge weights that are *compatible* with G and $(G \setminus z)^*$.

- The sum of the weights is $\mathbf{w}$.
- **(Compatible)** Every cycle has at least two minimal edges.

Fix $\mathbf{w} = (0, 1, 1, 2, 2, 5, 3, 4, 7)$.



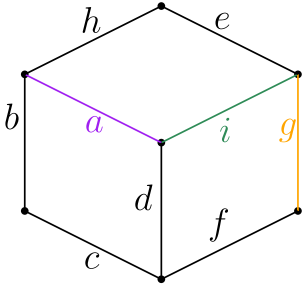


Figure: Find $\mathbf{x}$ and $\mathbf{y}$ edge weights that are *compatible* with G and $(G \setminus z)^*$.

	a	b	c	d	e	f	g	h	i
$\mathbf{x}$	00	01	00	00	02	03	00	00	02
$\mathbf{y}$	000	110	111	222	20	52	33	44	75
$\mathbf{w}$	0	1	1	2	2	5	3	4	7

Biblio

- Agostini D., Brysiewicz T., Fevola C., Kühne L., Sturmfels B., Telen S. (2021). *Likelihood Degenerations*. Motivation behind ML degree computations
- Adiprasito K., Huh J., Katz E. (2018). *Hodge Theory for combinatorial geometries* Degree computations in matroids
- Ardila F., Eur C., RP (2022) *The maximum likelihood of a matroid*.
- Baldwin, Elizabeth, and Paul Klemperer (2013) *Tropical geometry to analyse demand*.