



**HAL**  
open science

# Big data approach to Kazhdan-Lusztig polynomials

Abel Lacabanne, Daniel Tubbenhauer, Pedro Vaz

► **To cite this version:**

Abel Lacabanne, Daniel Tubbenhauer, Pedro Vaz. Big data approach to Kazhdan-Lusztig polynomials. 2024. hal-04817065

**HAL Id: hal-04817065**

**<https://hal.science/hal-04817065v1>**

Preprint submitted on 3 Dec 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

# BIG DATA APPROACH TO KAZHDAN–LUSZTIG POLYNOMIALS

ABEL LACABANNE, DANIEL TUBBENHAUER AND PEDRO VAZ

ABSTRACT. We investigate the structure of Kazhdan–Lusztig polynomials of the symmetric group by leveraging computational approaches from big data, including exploratory and topological data analysis, applied to the polynomials for symmetric groups of up to 11 strands.

## CONTENTS

1. Introduction	1
2. Preliminaries in a nutshell and notation	3
3. Percentage of nonzero KL polynomials	5
4. Number of KL polynomials	8
5. Growth of KL polynomials	9
6. Unimodality of KL polynomials	12
7. Roots of KL polynomials: all roots	14
8. Roots of KL polynomials: Perron–Frobenius root	15
9. The KL ballmapper	17
References	20

## 1. INTRODUCTION

We investigate a data set of Kazhdan–Lusztig (KL) polynomials using techniques traditionally employed in data science.

**1A. Background and ideas.** The *Kazhdan–Lusztig (KL) polynomials* are fundamental yet enigmatic objects in combinatorial representation theory. First introduced in [KL79] for arbitrary Coxeter groups, this paper focuses on the type A Coxeter group, the symmetric group  $S_n$  on  $\{1, \dots, n\}$ . In this context, KL polynomials represent entries in a nonnegatively graded change-of-basis matrix between simple and Verma modules of  $\mathfrak{sl}_n$ . As a result, these polynomials are either zero or belong to  $1 + v\mathbb{Z}_{\geq 0}[v]$ , where  $v$  denotes the grading variable.

Although a main focus of research in combinatorics, geometry, and representation theory alike, not much is known about these polynomials. The starting point of this work is the observation that KL polynomials exhibit patterns akin to statistical distributions. By analyzing these distributions, we aim to uncover structures that remain elusive through traditional methods, such as combinatorial or geometric approaches. Our methodology involves systematic data processing and analysis, often referred to as “*big data*,” utilizing techniques such as data visualization, exploratory data analysis (EDA), and topological data analysis (TDA). Our approach is inspired by work in knot theory as, for example, [LHS22, DGS24] and differs from the deep learning approaches to KL theory as in e.g. [DVB<sup>+</sup>21, Wil24].

Using these methods, we will discuss several conjectures about KL polynomials and their distribution, and for some of them we give an indication of how to prove these conjectures.

*Remark 1A.1.* A critical aspect of this study is the scale of the data: KL polynomials are indexed by pairs of elements in  $S_n$ , resulting in  $(n!)^2$  entries. Although computable, their computation for  $S_{11}$ , with one permutation fixed as trivial, required approximately 60 days using the program in [War11] on the servers of Laboratoire de Mathématiques Blaise Pascal, Université Clermont Auvergne. For  $S_{12}$ , we anticipate significantly higher computational costs, at least 12 times longer, although this is a very conservative lower bound that completely excludes potential memory limitations.  $\diamond$

**Notation 1A.2.** In this paper, we have *conjectures* and *speculations*. Conjectures are presented in their standard sense, while speculations refer to preliminary hypotheses that lack full support from the data. We

2020 *Mathematics Subject Classification*. Primary: 05E10, 62R07, secondary: 20C08, 68P05.

*Key words and phrases*. Visualization, exploratory data analysis, topological data analysis, conjecturing, Kazhdan–Lusztig polynomials.

hope that both serve as an inspiration to prove or, **equally exciting**, disprove the corresponding statements. We also have **questions** meant in the standard way of the word.  $\diamond$

*Remark 1A.3.* All the data files and supplementary material (such as higher resolution pictures, code, a possible empty Erratum etc.) can be found online at [LTV24a].  $\diamond$

**1B. Future directions.** The immediate candidates that come to mind to apply big data methods are graph and knot polynomials, but this section focuses on potential representation-theoretical applications of big data methods. We outline several avenues for exploration:

- (I) *Generalizations to other KL polynomials.* An immediate extension is to study variations of KL polynomials, such as other types, antispherical or spherical KL polynomials, which have deep connections to representation theory, cf. [Soe97]. Data for such cases can potentially be generated using existing software, e.g. [dC06, Gib21]. Another intriguing direction is to investigate the ***p*-canonical (*p*KL) polynomials**, cf. [JW17], though their computational complexity presents a significant challenge. Special cases, such as those in [Bai24], may provide a tractable starting point for data collection and analysis.
- (II) *Canonical bases and change-of-basis polynomials.* Similarly and related, cf. [FKK98], **canonical (*a.k.a.* crystal) bases** of Lie algebras and their change-of-basis matrices give interesting polynomials. This should already be interesting for small ranks, see for example [KK99, MPT14], or related quantum base-changes such as [RT19, RT22]. Generating sufficient data for these cases appears feasible, though computational tools for these specific problems are not yet widely available as far as we are aware while writing this paper.
- (III) *Graded representation theory.* The simple representations of the symmetric group in finite characteristic have a nontrivial grading by [BK09]. By taking **graded dimensions**, one gets a polynomial invariant associated to (certain) partitions of  $n$ . Similar constructions arise in related algebras, such as KLR algebras, cf. [KL09, Rou08], graded Temperley–Lieb algebras, cf. [PRH14], and KLRW algebras, cf. [Web19, MT23]. The computational tools in [EM24] may provide a valuable starting point for data generation, with ongoing efforts (as communicated to us by the second author of [EM24]) to integrate these structures into SageMath.
- (IV) *Web spaces invariants.* Underlying the quantum knot polynomials are spaces of representation theoretical origin, often called **webs** (spiders [Kup96] or birdtracks [Cvi08]), provide a rich setting for studying polynomial invariants. Early examples include [RTW32, Yam89]. Evaluation of such webs yields polynomials akin to quantum knot polynomials. While the type A exterior case has a closed formula, cf. [LTV24b], suggesting limited complexity, extensions to other settings, such as  $SO(3)$  invariants and their connections to chromatic polynomials, cf. [Yam89, FK10], remain largely unexplored. The chromatic polynomial has recently been studied via TDA [SS24], and the quantum version of this category as, e.g., in [MPS17, Tub23] might give some quantization of [SS24]. Computational complexity here is expected to align with that of canonical bases as in (II).
- (V) *Tensor categories and knot/three-manifold invariants.* Building on [LHS22, DGS24], a promising direction is to analyze knot and three-manifold invariants arising from **tensor categories** (in the sense of [EGNO15]). Examples include invariants derived from Deligne categories, cf. [FL21], mixed quantum groups, cf. [AK92, STWZ23], subfactor theory, following the exposition in [Jon86], or exotic asymptotic algebras (see, for example, [MMM<sup>+</sup>23] for a summary). Currently, there are no well-developed computational approaches in this area as far as we know, making it an exciting frontier for future research.
- (VI) *Many other exciting direction.* For example, homology and Ext groups naturally yield polynomial invariants, opening another avenue for investigation. Exploring these connections may provide novel insights into representation theory and beyond.

As a side note, many of these and other representation theoretical problems could benefit from AI or deep learning approaches, but it is essential to distinguish these techniques from the big data methodologies emphasized in this work.

**Acknowledgments.** We are deeply grateful to Joe Baine, Andrew Mathas, Geordie Williamson, and Victor Zhang for their valuable insights and many stimulating discussions. Special thanks to the organizers of the exceptional conference “Diagrammatic Intuition and Deep Learning in Mathematics” (York, July 2024) and to Radmila Sazdanovic, whose inspiring talk at that conference started this project. DT would like to thank the knot PD[X[3, 1, 4, 32], X[1, 7, 2, 6], X[7, 3, 8, 2], X[11, 5, 12, 4], X[5, 11, 6, 10], X[19, 8, 20, 9], X[9, 18, 10, 19], X[25, 12, 26, 13], X[13, 24, 14, 25], X[31, 15, 32, 14], X[15, 21, 16, 20], X[27, 16, 28, 17], X[17, 26, 18, 27], X[21, 29, 22, 28], X[29, 23, 30, 22], X[23, 31, 24, 30]]—you are not just one of many.

AL is grateful for the support and hospitality of the Sydney Mathematical Research Institute (SMRI). AL and PV were supported by a PHC Tournesol Wallonie Bruxelles grant, and DT was sponsored by the ARC Future Fellowship FT230100489.

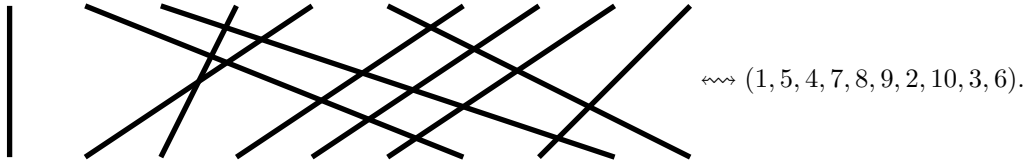
In this work, we utilized ChatGPT for coding, proofreading, generating conjectures, and assisting with proof development (and this sentence). Most of the computations were performed using the resources of the Laboratoire de Mathématiques Blaise Pascal.

## 2. PRELIMINARIES IN A NUTSHELL AND NOTATION

Fix the Coxeter presentation of the symmetric group  $S_n$  given by simple transpositions as Coxeter generators. That is, for  $m(i, i) = 1$ ,  $m(i, j) = 3$  if  $|i - j| = 1$  and  $m(i, j) = 2$  if  $|i - j| > 1$ :

$$S_n = \langle s_1, \dots, s_{n-1} \mid (s_i s_j)^{m(i,j)} = 1 \rangle, \quad s_i \rightsquigarrow (i, i+1).$$

The unit is denoted 1. We will illustrate elements of  $S_n$  either using *one-line notation* or *string diagrams* (read from bottom to top), with the prototypical example being:



The *length*  $\ell(w)$  of  $w \in S_n$  is then defined using the above Coxeter presentation, and the reduced expressions are also defined with respect to this presentation. Two elements  $v, w \in S_n$  are in *Bruhat order* with  $v$  being smaller and  $w$  being larger, denoted by  $v \leq_B w$ , if some substring of some reduced word for  $w$  is a reduced word for  $v$ . The Bruhat order is a partial order on  $S_n$  and we will use expressions like  $v >_B w$  with the evident meaning. Finally, *left and right descent sets* of  $w \in S_n$  are  $\text{lides}(w) = \{s_i \mid s_i w <_B w\}$  and  $\text{rdes}(w) = \{s_i \mid w s_i <_B w\}$ , which are used to define *extremal pairs* as  $\text{EP}(n) = \{v \leq_B w \in S_n \times S_n \mid \text{lides}(w) \subset \text{lides}(v), \text{rdes}(w) \subset \text{rdes}(v)\}$ . We also say  $v \leq_B w$  form an extremal pair if they are in  $\text{EP}(n)$ .

The Bruhat order can be refined into a total order, called *one-line order*, with  $v$  being smaller and  $w$  being bigger, denoted by  $v \leq_1 w$ , if  $v$  is smaller than  $w$  in one-line notation read lexicographically. We also use e.g.  $v >_1 w$  as in the Bruhat order.

Let  $\mathbf{v}$  denote some formal variable. [KL79] associates a polynomial in  $\mathbf{v}$ , the *KL polynomial*, with pairs of elements of  $S_n$ . That is, one can show that we have

$$v, w \in S_n \mapsto P_{v,w} = P_{v,w}(\mathbf{v}) \in (1 + \mathbf{v}\mathbb{Z}_{\geq 0}[\mathbf{v}]) \cup \{0\}.$$

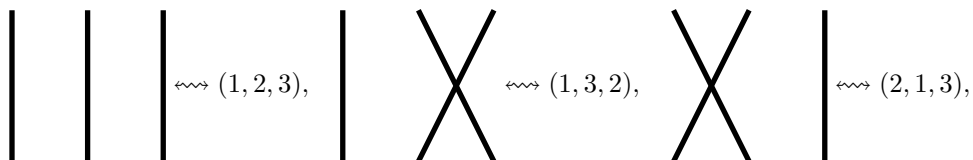
We will not repeat here the definition of the KL polynomials but rather list some properties:

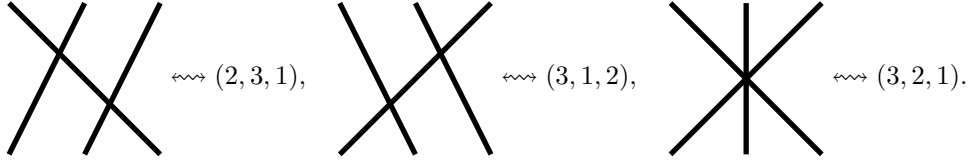
- (a)  $P_{w,w} = 1$ , and  $P_{v,w} \neq 0$  if and only if  $v \leq_B w$ .
- (b) For  $v <_B w$  we have  $\deg P_{v,w} \leq \frac{1}{2}(\ell(w) - \ell(v) - 1)$ , and the coefficients of  $\mathbf{v}^{1/2(\ell(w) - \ell(v) - 1)}$  is called the  $\mu$  *coefficient*  $\mu(v, w)$ . This coefficient is nonzero only if  $v \leq_B w$  form an extremal pair.
- (c) We have  $P_{u,w} - P_{v,w} \in \mathbb{Z}_{\geq 0}[\mathbf{v}]$  for all  $u, v, w \in S_n$  with  $u \leq_B v \leq_B w$ . In particular,  $P_{v,w}(1) \leq P_{1,w}(1)$  for all  $v, w \in S_n$ .
- (d) The maximum degree is  $\max\{\deg P_{v,w} \mid v, w \in S_n\} \leq \lfloor \frac{1}{2}\ell(w_0) \rfloor = \lfloor \frac{1}{4}(n-1)n \rfloor$ .
- (e) The computer program of our choice to compute these polynomials is the one from [War11]. This program is also available on [LTV24a]. Other good programs to compute KL polynomials (which are slower in our experience but can compute them for other Coxeter groups as well) are du Cloux's program [dC06] and Gibson's Magma programs [Gib21].

All of these facts can be found in the usual literature on KL polynomials, e.g. [KL79, BB05, Lus03].

*Remark 2.1.* Our data set seems too small to say anything beyond [War11] about the  $\mu$  coefficients, but we still mentioned them via (b), as they have attracted a lot of interest in the past.  $\diamond$

**Example 2.2.** The following example fixes our conventions. Let  $n = 3$ , and





The one-line ordering on these is  $(1, 2, 3) \leq_1 (1, 3, 2) \leq_1 (2, 1, 3) \leq_1 (2, 3, 1) \leq_1 (3, 1, 2) \leq_1 (3, 2, 1)$ . In this ordering we can put the KL polynomials  $P_{v,w}$  into a matrix:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here  $v$  indexes the rows and  $w$  the columns. The first row is the case of  $P_w = P_{1,w}$ , which we will focus on in some of the sections below.  $\diamond$

**Notation 2.3.** As another piece of notation, we collect the KL polynomials as *lists* (ordered, repetitions allowed, notation  $[-]$ ), *multisets* (not ordered, repetitions allowed, notation  $(-)$ ) or *sets* (not ordered, no repetitions, notation  $\{-\}$ ):

$$\text{list: } [P_{v,w}|v, w \in S_n], \quad \text{multiset: } (P_{v,w}|v, w \in S_n), \quad \text{set: } \{P_{v,w}|v, w \in S_n\}.$$

For a multiset we use superscripts to encode multiplicities. We then use the standard computer science notation to work with these, e.g.  $\sum$  for the sum over all entries,  $\max$  for the maximal entry etc.  $\diamond$

**Example 2.4.** In [Example 2.2](#) we have

$$(P_{v,w}|v, w \in S_3) = (0^{17}, 1^{19}), \quad \{P_{v,w}|v, w \in S_3\} = \{0, 1\},$$

and the list of KL polynomials is the given flattened matrix in that example. We have

$$\#(P_{v,w}|v, w \in S_3, P_{v,w} \neq 0) / \#(P_{v,w}|v, w \in S_3) = \frac{19}{36}, \quad \#\{P_{v,w}|v, w \in S_3, P_{v,w} \neq 0\} / \#\{P_{v,w}|v, w \in S_3\} = \frac{1}{2},$$

where, as throughout, we use  $\#$  for the number of elements, counted with multiplicities.  $\diamond$

We now briefly explain the setting we use throughout. If not stated otherwise, we will identify a KL polynomial with a vector as follows. We use some upper bound  $d$  for the degree of polynomials appearing, e.g.  $n^2$ . Then, for  $P_{v,w} \neq 0$ :

$$P_{v,w} = 1 + a_1 v^1 + \dots + a_d v^d \rightsquigarrow [1, a_1, \dots, a_d] \in \mathbb{Z}_{\geq 0}^{d+1} \subset \mathbb{R}^{d+1}.$$

For  $P_{v,w} = 0$ , we assign the zero vector. In this way, all KL polynomials are in the same vector space. In the language of computer science, we create the list of coefficients  $[1, a_1, \dots, a_d]$  padded with zeros until we get the desired length  $d$ . Moreover, *span* is  $\text{span } P_{v,w} = \deg P_{v,w} + 1$ .

**Example 2.5.** Note that our vectors are padded with zeros, so the span is not necessarily equal to  $d$ . Explicitly, for  $P_{v,w} = 1 + 3v^5 + 4v^9$  we have  $\text{span } P_{v,w} = 10$ , independent of  $d$ .  $\diamond$

**Example 2.6.** Using the vector notation, we have  $\sum P_{v,w} = P_{v,w}(1)$ .  $\diamond$

**Example 2.7.** For  $S_{10}$ , we can choose  $d = 22$ . There is a KL polynomial of the form  $P_{v,w} = 1 + 13v + 78v^2 + 282v^3 + 666v^4 + 1068v^5 + 1176v^6 + 879v^7 + 432v^8 + 132v^9 + 22v^{10} + v^{11}$ . We get

$$P_{v,w} = [1, 13, 78, 282, 666, 1068, 1176, 879, 432, 132, 22, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

as our coefficient list. We have  $\text{span } P_{v,w} = 12$ ,  $\sum P_{v,w} = 4750$ , and  $\max P_{v,w} = 1176$ .  $\diamond$

*Remark 2.8.* Recall that we focus on the symmetric group only; the construction of [\[KL79\]](#) in fact works for any Coxeter group. It would be interesting to redo the big data analysis for other Coxeter groups.  $\diamond$

Throughout, we are interested in large  $n$  behavior. We now fix some notation that we will use.

**Notation 2.9.** For functions  $f, g: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}_{>0}$  we use:

$$f \sim g \Leftrightarrow \forall \varepsilon > 0, \exists n_0 \text{ such that } \left| \frac{f(n)}{g(n)} - 1 \right| < \varepsilon, \forall n > n_0,$$

$$f \in \Omega(g) \Leftrightarrow \exists C > 0, \exists n_0 \text{ such that } |f(n)| \geq C \cdot g(n), \forall n > n_0,$$

$$f \in O(g) \Leftrightarrow \exists C > 0, \exists n_0 \text{ such that } |f(n)| \leq C \cdot g(n), \forall n > n_0.$$

We use these, in order, as *asymptotic behavior*, *lower and upper bounds*.  $\diamond$

**Notation 2.10.** To analyze functions that behave like  $n^a$  for some  $a \in \mathbb{R}$ , we use *effective exponent*. For a function  $f: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}_{>0}$ , this is the function  $n \mapsto \frac{\ln(f(n)) - \ln(f(n+1))}{\ln(n) - \ln(n+1)}$ . (Note that this is  $a$  for  $f(n) = n^a$ , hence the name.)

For functions that behave like  $a^n$  for some  $a \in \mathbb{R}_{>1}$  we use *successive quotients*. For a function  $f: \mathbb{Z}_{\geq 0} \rightarrow \mathbb{R}_{>0}$ , this is the function  $n \mapsto \frac{f(n+1)}{f(n)}$ . (Note that this is  $a$  for  $f(n) = a^n$ .)  $\diamond$

**Notation 2.11.** We sometimes use *log plots*, which means that we put a logarithmic scale on the y-axis.  $\diamond$

### 3. PERCENTAGE OF NONZERO KL POLYNOMIALS

Recall that  $P_{v,w} \neq 0$  if and only if  $v \leq_B w$ . So, two questions that one might ask are the following:

(3.i) What is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{den}_n = \#(P_{v,w}|v, w \in S_n, P_{v,w} \neq 0) / \#(P_{v,w}|v, w \in S_n)?$$

(This is the density.) That is, what is the asymptotic for the *percentage* of nonzero KL polynomials.

(3.ii) What is the asymptotic behavior  $n \rightarrow \infty$  of

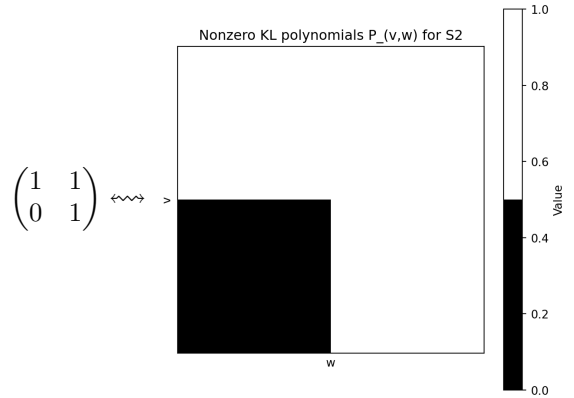
$$\text{den}_n^{av} = \sum(\sum P_{v,w}|v, w \in S_n) / \#(P_{v,w}|v, w \in S_n)?$$

That is a similar question as before, but now we ask about the *average* values of the KL polynomials at  $v = 1$ .

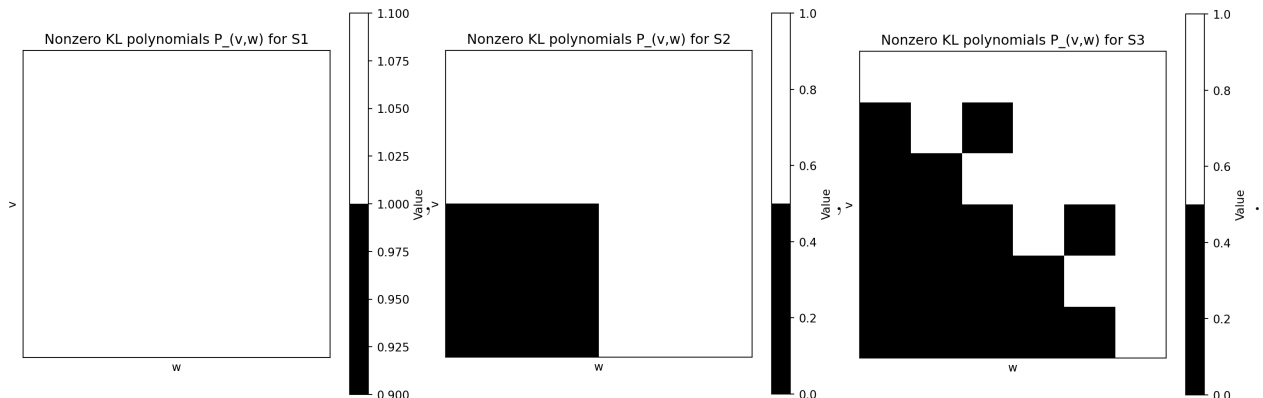
(3.iii) To calculate how ‘fiery’ the pictures is (this will make sense if you read ahead), we use the so-called *mean (pixel) intensity*  $\text{fire}_n$ : covert the picture to gray scale (by using  $n \mapsto \frac{n - \min}{\max - \min}$ ) and compute the average. What is

$$\lim_{n \rightarrow \infty} \text{fire}_n?$$

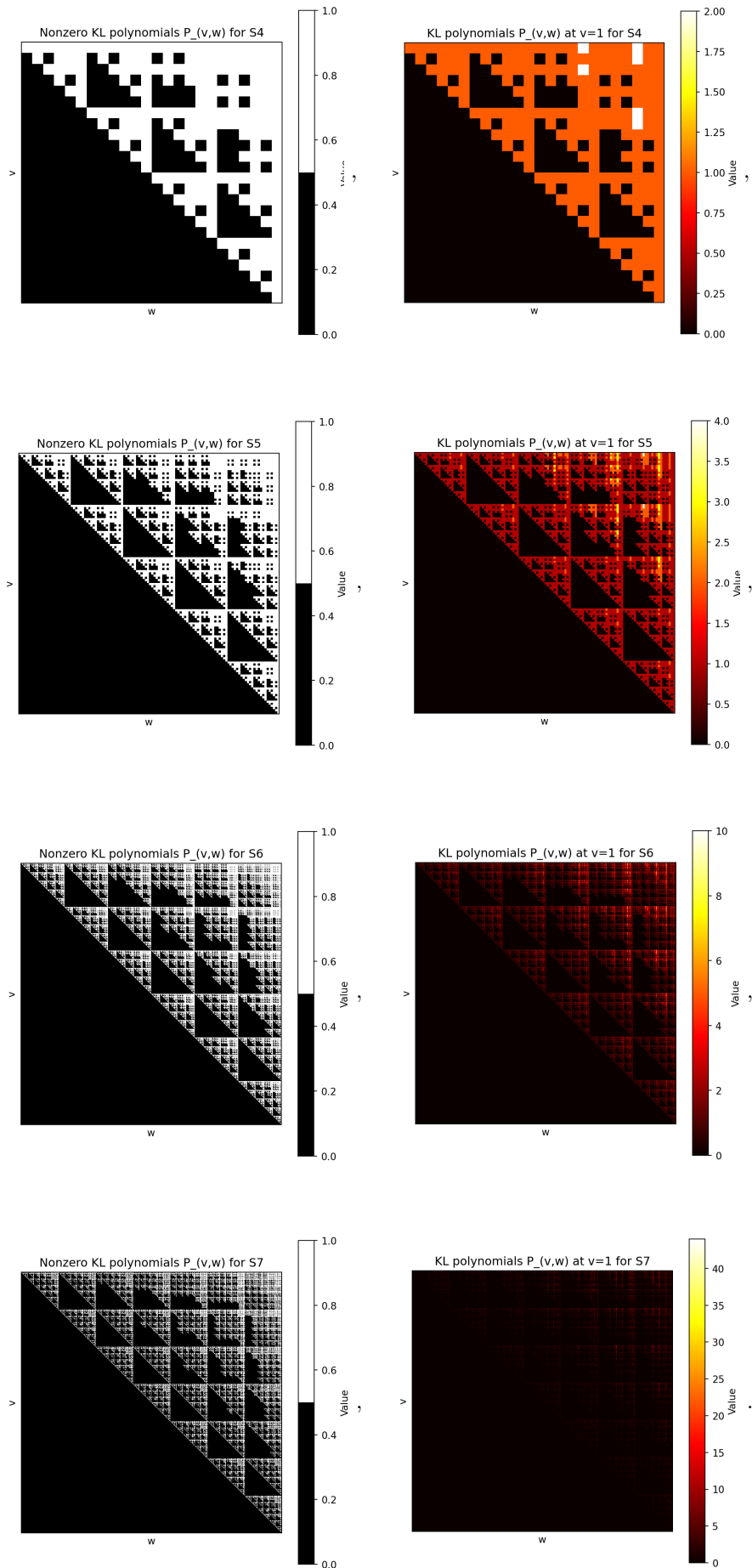
Before explaining the data, let us explain *grid plots of matrices*: One visualizes a matrix with  $\mathbb{R}_{\geq 0}$  entries as is with a heat map, where the temperature corresponds to the size of the entry. We use black as the coldest=smallest entry and white as the hottest=biggest entry. For example,



Using the lexicographical order in one-line notation, we can form a matrix  $(P_{v,w}(v = 1))_{v,w \in S_n} \in \mathbb{Z}_{\geq 0}^2$ , and for  $n = 2$  the above picture arises. Here are a few other ranks ( $n = 3$  should be compared with [Example 2.2](#)), with the convention that if all entries are zero or one, then there is only one plot:

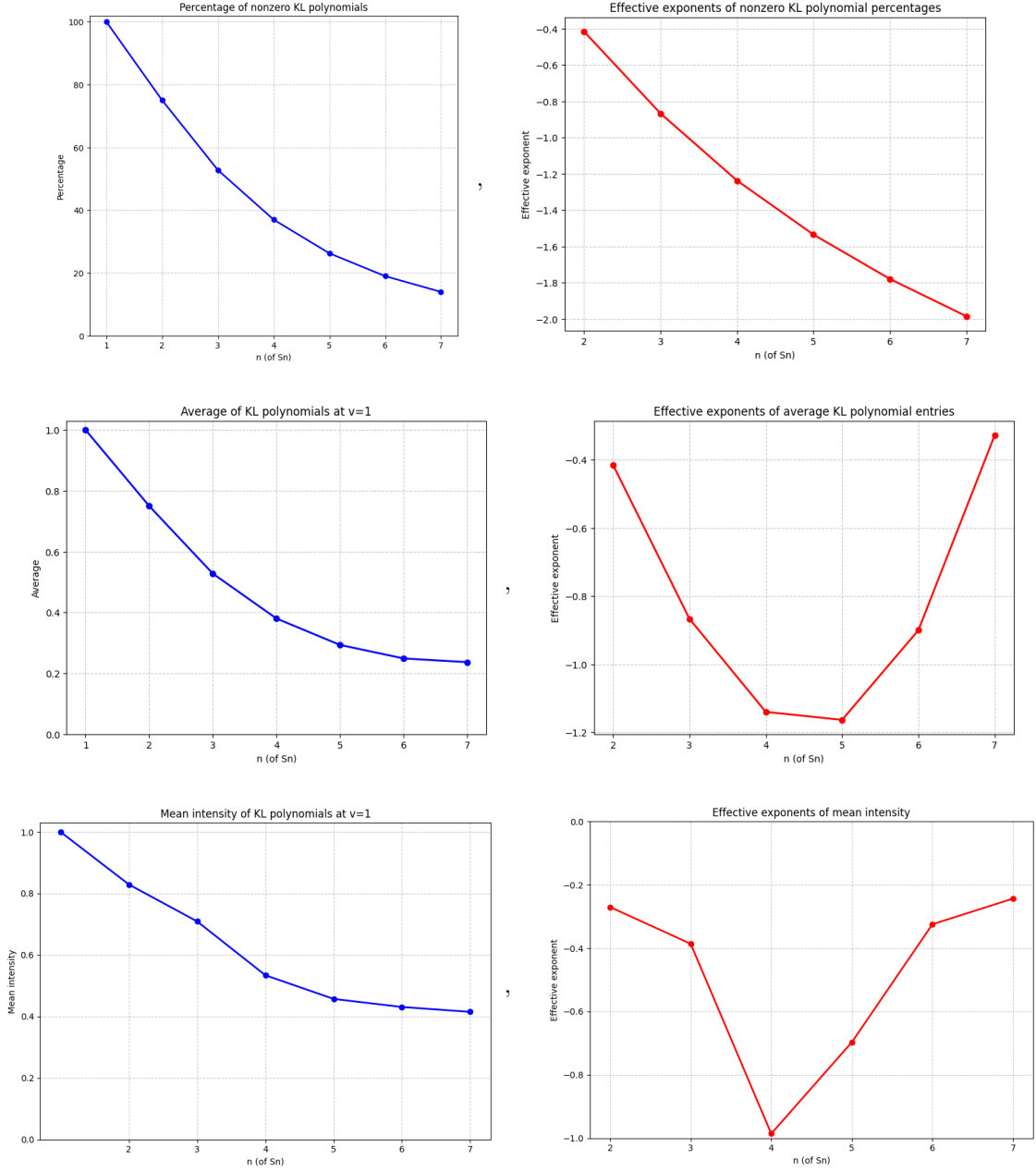


The following now have two plots:





The percentage of nonzero entries, the average entry, and the intensity are:



$n$	1	2	3	4	5	6	7
$\% \neq 0$	100	75	52.78	36.98	26.26	18.98	13.98
av	1	0.75	0.5278	0.3802	0.2933	0.2489	0.2366
in	1	0.8295	0.7095	0.5346	0.4573	0.4307	0.4140

Note that the average value and the percentage of nonzero entries are the same for  $n \in \{1, 2, 3\}$ .

The following conjecture comes from additional data that one gets using (a) in Section 2 as the Bruhat order is much easier to compute in large scale than KL polynomials. One can also see this in the effective exponent curve above, which starts to slow down in its decay. For a more refined version of Speculation 3.3 we would need more data as we cannot now avoid calculating KL polynomials.

**Conjecture 3.1.** *We have the asymptotic behavior*

$$\text{den}_n \sim n^{-(2+a)} \text{ for some } a \in [0, 1].$$

We can prove the following.

**Theorem 3.2.** *We have*

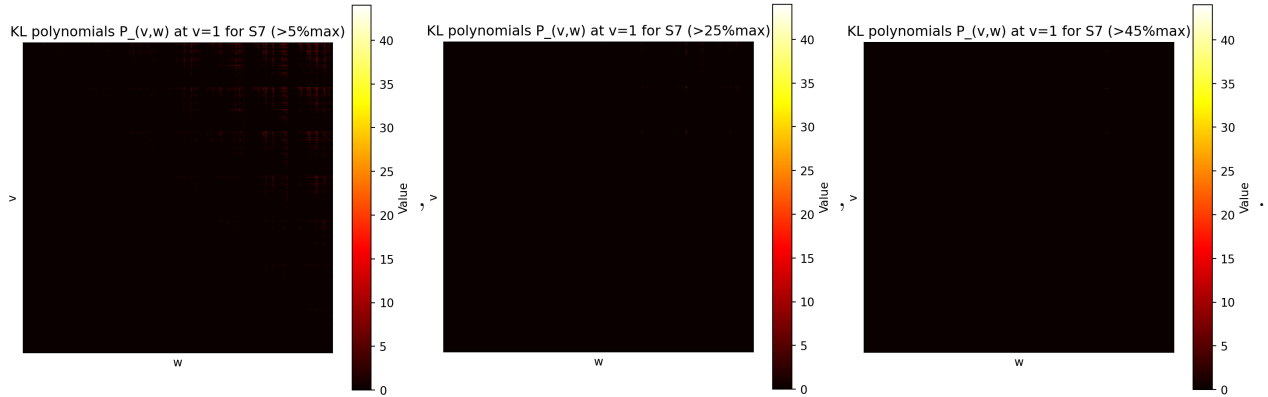
$$\Omega((0.708\dots - \epsilon)^n) \ni \text{den}_n \in O(n^{-2}) \text{ for all } \epsilon \in \mathbb{R}_{>0}.$$

Here  $0.708\dots = \sqrt[11]{25497938851324213335/22!}$ .



*Proof.* This follows from the properties of the KL polynomials that we have recalled in [Section 2](#) combined with [\[HP08, Theorem 1.1\]](#) (the precise value that the authors of [\[HP08, Theorem 1.1\]](#) recalled above can be found in [\[HP08, Section 3\]](#)).  $\square$

In summary, some KL polynomials are much larger than the rest. The following pictures, showing only the largest KL polynomials, illustrate this. Basically, there is nothing that one can see:



This motivates:

**Speculation 3.3.** *We have*

$$\text{den}_n^{av} \in \Omega(n^a) \text{ for some } a \in \mathbb{R}_{>0}, \quad \lim_{n \rightarrow \infty} \text{fire}_n = 0.$$

(The fire goes out, but some KL polynomials get very large.)

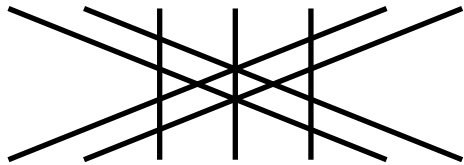
For a potential way to attack [Speculation 3.3](#), see the end of [Section 5](#).

*Remark 3.4.* There are efficient algorithms to check [Conjecture 3.1](#) because it is equivalent to the study of the proportion of elements  $u, w \in S_n$  that are in Bruhat order; see [\[HP08\]](#) for a discussion of the latter. The weighted version [Speculation 3.3](#) of [Conjecture 3.1](#) would need more information on how large KL polynomials are. There are a few references on this, for example, [\[BM01\]](#) (which proves [Section 2.\(c\)](#)) and [\[BB03\]](#), but not much seems to be known.  $\diamond$

#### 4. NUMBER OF KL POLYNOMIALS

From now on, we only look at  $P_w = P_{1,w} = [1, a_1, \dots, a_d] \in \mathbb{Z}_{\geq 0}^{d+1}$  for  $w \in S_n$ , which corresponds to the first row in the pictures of [Section 3](#).

*Remark 4.1.* Note that we lose almost all larger KL polynomials when focusing on the first row; see the final plot in [Section 3](#). However, the maximal value in this final plot is 44 and appears in the first row with the second entry being the permutation



which is index 4265. The same value appears in two other rows and the same column. This is true in general:  $P_{1,w}$  evaluated at  $v = 1$  is always larger than  $P_{v,w}$  evaluated at  $v = 1$ , for all  $v, w \in S_n$ , see (c) in [Section 2](#).  $\diamond$

We stress that there is a huge difference between the sets of KL polynomials and the multisets. So, the first questions we ask are:

(4.i) What is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{num}_n = \#\{P_w | w \in S_n\}?$$

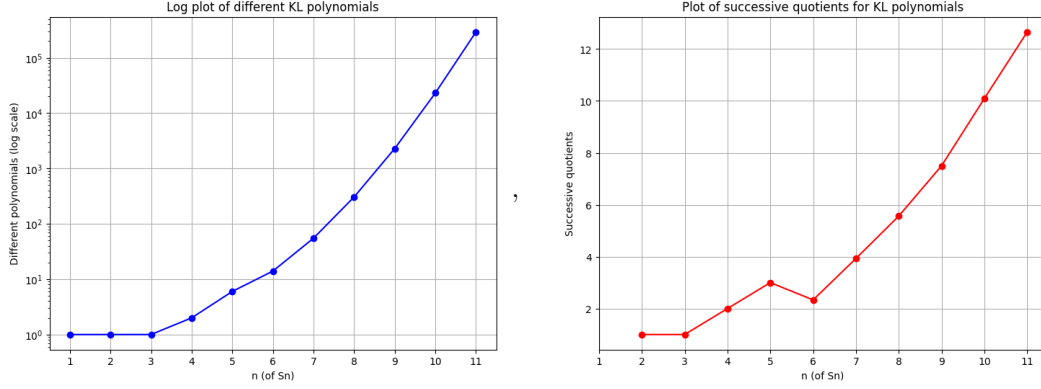
In other words, we ask about the asymptotic number of different KL polynomials.

(4.ii) What is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{num}_n^{\%} = \#\{P_w | w \in S_n\} / \#(P_w | w \in S_n)?$$

This asks for the percentage of different KL polynomials.

Here are the data:



$n$	1	2	3	4	5	6	7	8	9	10	11
#	1	1	1	2	6	14	55	306	2295	23163	293189

The graph on the left is a log plot.

**Conjecture 4.2.** We have **superexponential growth**, i.e.:

$$\text{num}_n \in \Omega(\gamma^n) \text{ for all } \gamma \in \mathbb{R}_{>1}.$$

Moreover, we have **exponential decay**, i.e.:

$$\text{num}_n \in O(\delta^n) \text{ for some } \delta \in (0, 1).$$

We do not know how to prove [Conjecture 4.2](#). In fact, even when we assume the combinatorial invariance conjecture (see e.g. [\[BBD<sup>+</sup>22\]](#) for an overview and some indication that it is true), we would still be left with understanding the Bruhat intervals of the symmetric group. Counting these is related to pattern avoidance, see, for example, [\[Ten07\]](#), and appears to be a very difficult problem.

### 5. GROWTH OF KL POLYNOMIALS

We are interested in questions of the form:

(5.i) What is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{ev}_n = \max\{\sum P_w | w \in S_n\}?$$

That is, we ask how fast the KL polynomials evaluated at  $v = 1$  grow in the rank  $n$ .

(5.ii) Similarly, what is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{coeff}_n = \max\{\max P_w | w \in S_n\}?$$

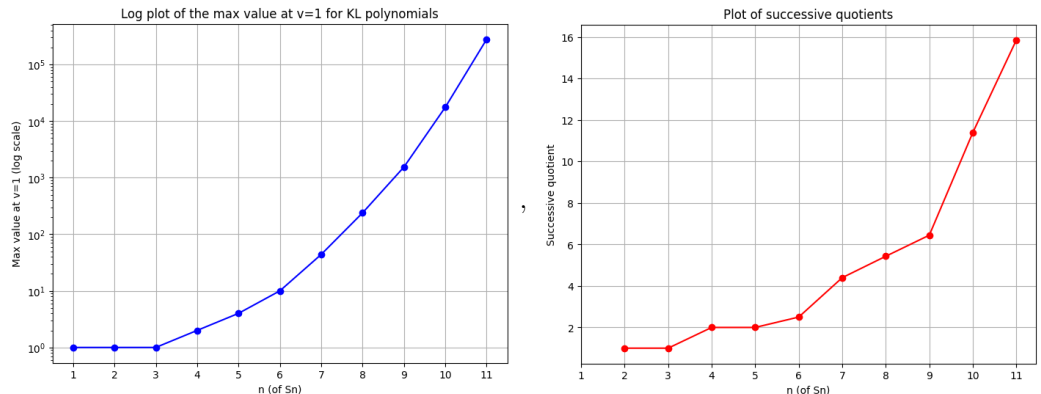
That is, we ask how fast the coefficients of KL polynomials grow in the rank  $n$ .

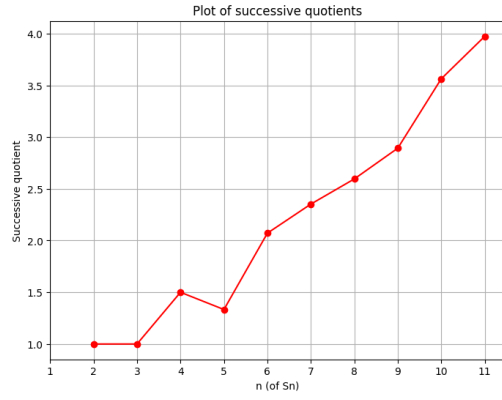
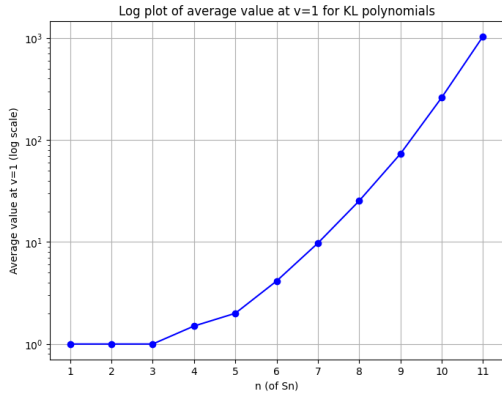
(5.iii) We can ask the same questions as in (5.i) and (5.ii) but for the average  $\text{ev}_n^{av}$  and  $\text{coeff}_n^{av}$  instead of the maximum. Here we mean **average** in sense of

$$\text{ev}_n^{av} = \sum\{\sum P_w | w \in S_n\} / \#\{P_w | w \in S_n\}, \quad \text{coeff}_n^{av} = \sum\{\sum P_w / \text{span } P_w | w \in S_n\} / \#\{P_w | w \in S_n\}.$$

Note that we do not count the padded zeros for the average, but only those zeros that are in the span of the KL polynomials.

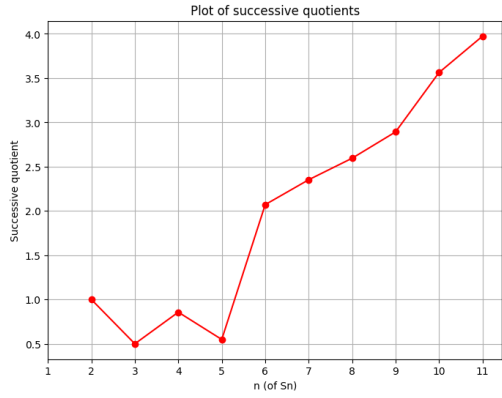
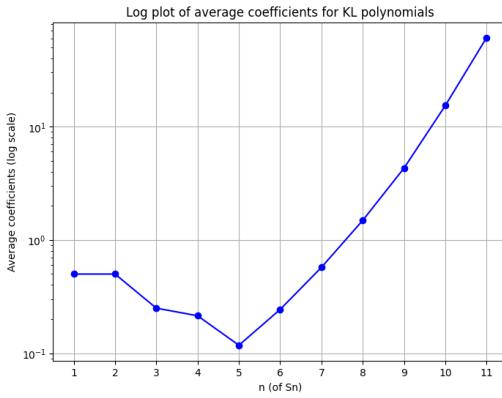
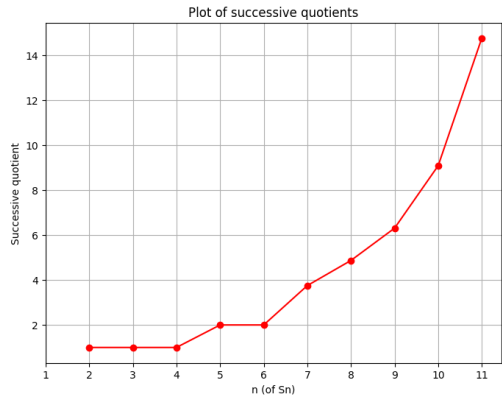
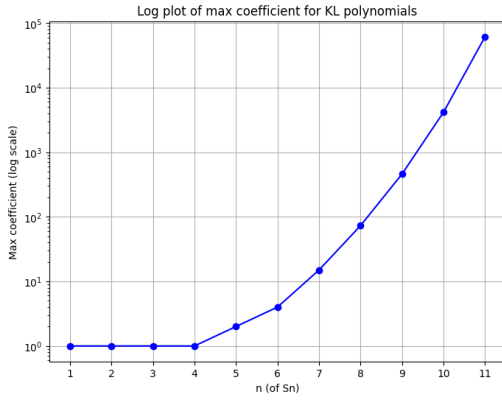
Here are the data that we collected. First, for the evaluation at  $v = 1$ :





$n$	1	2	3	4	5	6	7	8	9	10	11
max	1	1	1	2	4	10	44	239	1541	17566	278576
av	1	1	1	1.5	2.4	4.85	10	25.30	72.71	259.03	1035.78

Next, for the maximal coefficient:



$n$	1	2	3	4	5	6	7	8	9	10	11
max	1	1	1	1	2	4	15	73	460	4176	61582
av	1	0.5	0.25	0.21	0.12	0.24	0.57	1.49	4.30	15.33	39.84

In both cases, note that the graph on the left is a log plot.

**Conjecture 5.1.** *We have superexponential growth, i.e.:*

$$\text{coeff}_n \in \Omega(\gamma^n) \text{ for all } \gamma \in \mathbb{R}_{>1}.$$

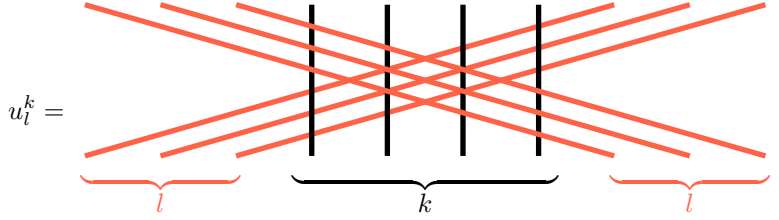
Since  $ev_n \geq \text{coeff}_n$ , the same holds for  $ev_n$ .

**Speculation 5.2.** *We have at least exponential growth, i.e.:*

$$\text{coeff}_n^{av} \in \Omega(\gamma^n) \text{ for some } \gamma \in \mathbb{R}_{>1}.$$

Since  $ev_n^{av} \geq \text{coeff}_n^{av}$ , the same holds for  $ev_n^{av}$ .

We now discuss a proof strategy for [Conjecture 5.1](#). For  $k + 2l = n$ , we consider

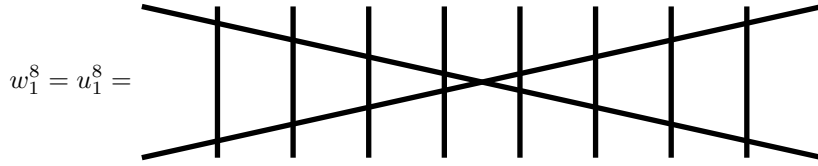


which is the element that pulls the first  $l$  strings to the last  $l$  positions, fixes  $k$  strings in the middle, and pulls the last  $l$  strings to the first  $l$  positions. In formulas  $u_l^k = (n - l + 1, \dots, n, l + 1, \dots, n - l, 1, \dots, l)$  (one-line notation). We use this element to define

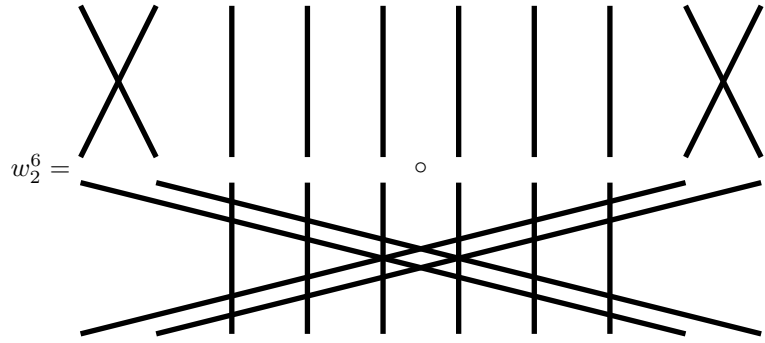
$$w_l^k = (w_0^l \otimes \text{id}_k \otimes w_0^l) \circ u_l^k,$$

which takes  $u_l^k$  and stacks the longest words  $w_0^l$  in  $S_l$  on the first and last  $l$  strings.

**Example 5.3.** For  $l = 1$  we have  $w_1^k = u_1^k$ , for example,



For  $l > 1$  we have  $w_l^k \neq u_l^k$ , for example,



where the bottom diagram is  $u_2^6$  and the top diagram is  $w_0^2 \otimes \text{id}_6 \otimes w_0^2$ . ◇

Here is a more specific conjecture:

**Conjecture 5.4.** We have that  $P_{w_l^k}$  is the Hilbert–Poincaré polynomial of  $(\mathbb{CP}^l)^{\times(k-1)}$ :

$$P_{w_l^k} = (1 + v + \dots + v^l)^{k-1}.$$

**Theorem 5.5.**

(a) [Conjecture 5.4](#) implies [Conjecture 5.1](#).

(b) [Conjecture 5.4](#) is true for  $l = 0$  and  $l = 1$ . This implies that  $\text{ev}_n$  and  $\text{coeff}_n$  grow exponentially.

*Proof.* (a). The middle a-nomial coefficients satisfy

$$\binom{ba}{a, \dots, a} \sim \sqrt{\frac{6}{(a^2-1)\pi}} \cdot b^{-1/2} \cdot a^b,$$

as follows from Stirling’s approximation. Since we can choose  $l$ , the conjecture follows.

(b). For  $l = 0$  this is clear, while the  $l = 1$  case follows from [\[SSV98, Theorem 2\]](#). In particular, as in (a), Stirling’s approximation gives  $\text{ev}_n \geq \text{coeff}_n \in \Omega(n^{-1/2} \cdot 2^n)$ . □

Let us recall the proof of [\[SSV98, Theorem 2\]](#), as one might hope that it generalizes to prove [Conjecture 5.4](#) in general. The authors of [\[SSV98\]](#) construct a geometric space, a flag variety, and show that this space admits a small resolution of singularities. This, in turn, shows that the computation of KL polynomials boils down to computing ordinary cohomology, which in the  $l = 1$  case is  $(1 + v)^b$ : the cohomology of a  $b$  times product of the sphere  $S^2$ , or equivalently, of  $\mathbb{CP}^1$ .

Replacing  $\mathbb{CP}^1$  by  $\mathbb{CP}^l$ , one could expect something similar to work, in general. However, in this case one does not have a small resolution of singularities, so some other argument is needed.

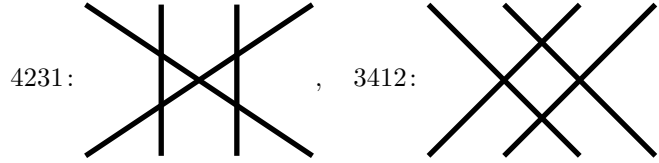
*Remark 5.6.* We first made the naive guess that  $u_l^k$  can be used to generalize [\[SSV98, Theorem 2\]](#). However, the KL polynomial of  $u_l^k$  is rather difficult, and we do not know its growth rate. ◇

*Remark 5.7.* For this section, the same references as in [Remark 5.6](#) are relevant. Indeed, we will comment on one momentarily.  $\diamond$

One way to address [Speculation 3.3](#) or [Speculation 5.2](#) could be the following. [[BB03](#), Theorem 1] implies that, for fixed  $m \in \mathbb{Z}_{\geq 0}$ , the set

$$\{w \in S_n | P_w(1) \leq m\}$$

is characterized by *pattern avoidance*. For example,  $m = 1$  corresponds to the avoidance of



as nicely explained in [[Woo09](#), Appendix]. In turn, the avoidance of these patterns is well studied and surprisingly strong results are known. For example, for the two patterns 4231 and 3412, we have

$$\#\{w \in S_n | w \text{ avoids } 4231\} \sim (1.15\dots) \cdot (0.95\dots)^n \cdot n!, \quad \#\{w \in S_n | w \text{ avoids } 3412\} \sim (1.14\dots) \cdot (0.95\dots)^n \cdot n!,$$

with precise values for the numbers, see [[OEI23](#), A113228 and A113229] and, for example, [[DK13](#), Section 4.2.4]. Using results of this form systematically might lead to some insight regarding [Speculation 3.3](#) or [Speculation 5.2](#).

## 6. UNIMODALITY OF KL POLYNOMIALS

Recall that a *multimodal distribution*, say with  $k$  modes, is a probability distribution with  $k$  local peaks of the distribution. The case  $k = 1$  is called *unimodal*,  $k = 2$  is *bimodal*,  $k = 3$  is called *trimodal* etc.

Applying the same terminology to the vector  $P_w = [1, a_1, \dots, a_d] \in \mathbb{Z}_{\geq 0}^{d+1}$  for  $w \in S_n$  raises the following questions:

(6.i) What is the limit  $n \rightarrow \infty$  of

$$\text{uni}_n = \#\{P_w | w \in S_n, P_w \text{ is unimodal}\} / \#\{P_w | w \in S_n\}?$$

In other words, we are interested in the percentage of unimodal KL polynomials. We can ask the same question for the multiset: What is the limit  $n \rightarrow \infty$  of

$$\text{uni}_n^{\text{mult}} = \#\{P_w | w \in S_n, P_w \text{ is unimodal}\} / \#\{P_w | w \in S_n\}?$$

(6.ii) Similarly, for  $k \in \mathbb{Z}_{\geq 3}$ , let  $\text{multi}_n^k$  denote the percentage of  $k$  modal KL polynomials that are not  $j$  modal for  $j < k$ , taken among the KL polynomials that are not  $j$  modal for  $j < k$ . What is

$$\lim_{n \rightarrow \infty} \text{multi}_n^k?$$

For example,  $\text{multi}_n^2$  is the percentage of bimodal KL polynomials among all nonunimodal KL polynomials. We could track the same statistics for the multiset version.

Let us first point out that there are essentially no unimodal polynomials (similarly for bimodal etc., but we did not find any more explicit data):

**Lemma 6.1.** *We have*

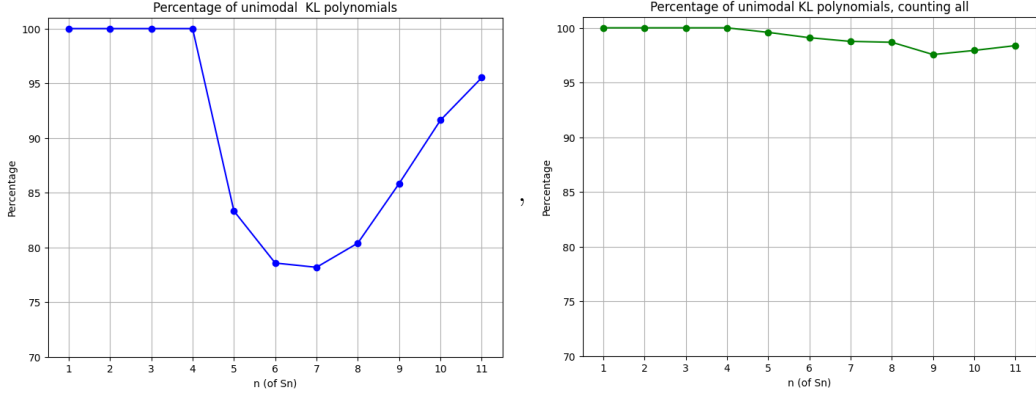
$$\#\{P | P \in \mathbb{Z}_{\geq 0}^{\infty}, \sum P = n, P_w \text{ is unimodal}\} / \#\{P | P \in \mathbb{Z}_{\geq 0}^{\infty}, \sum P = n\} \in O(n^{-a}) \text{ for all } a \in \mathbb{R}_{>1}.$$

*That is, the percentage of unimodal  $[a_0, \dots, a_d] \in \mathbb{Z}_{\geq 0}^{\infty}$  goes to zero superpolynomially fast.*

*Proof.* The number of compositions of  $n$  is  $2^{n-1}$ . Combine this with [[Aul51](#), Theorem 2].  $\square$

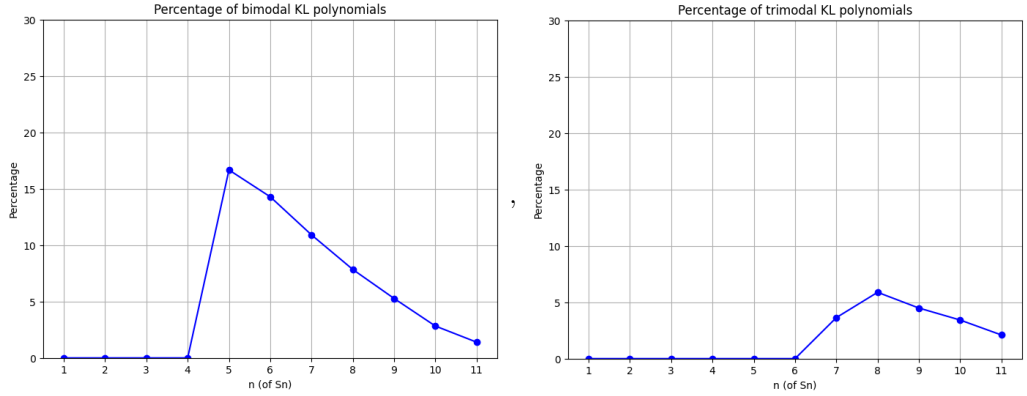
*Remark 6.2.* We expect something similar to be true for bimodal etc., but we did not find any more explicit statement. However, we will use a different way to generate polynomials below, but we have not found a percentage count for these. We still see [Lemma 6.1](#) and the data below as an indication that the KL polynomials have a remarkable distribution.  $\diamond$

The data is as follows:



$n$	1	2	3	4	5	6	7	8	9	10	11
% uni(set)	100	100	100	100	83.33	78.57	78.18	80.39	85.84	91.62	95.51
% uni(multiset)	100	100	100	100	99.59	99.10	98.76	98.68	97.56	97.93	98.38

This shows the percentage of unimodal KL polynomials (note that the bottom value is 70% not 0%), either for the set of polynomials or the multiset. For multimodal, the plots are:



$n$	1	2	3	4	5	6	7	8	9	10	11
% bi	0	0	0	0	16.67	14.29	10.91	7.84	5.27	2.83	1.40
% tri	0	0	0	0	0	0	3.64	5.88	4.49	3.42	2.10

(Note that the top value is 30% not 100%.) Similarly for the multiset versions. This is not spelled out because the data set is too small and the percentages are essentially all zero for  $S_{11}$ .

**Conjecture 6.3.** Almost all KL polynomials are unimodal, i.e.:

$$\lim_{n \rightarrow \infty} \text{uni}_n = \lim_{n \rightarrow \infty} \text{uni}_n^{\text{mult}} = 1.$$

*Remark 6.4.* Given that the fire seems to go out as in [Speculation 3.3](#), one could expect most KL polynomials to be rather small and thus to have a higher percentage of unimodal. In particular, one could say that  $\lim_{n \rightarrow \infty} \text{uni}_n = 1$  is more surprising than its multiset version.  $\diamond$

The following is not fully justified by the above data, as e.g. there are more trimodal KL polynomials for  $S_{11}$  than bimodal ones. However, we think that the pattern is that they all peak shortly after they appear, which skews the picture we get for the data we have. So:

**Speculation 6.5.** Almost all KL polynomials that are not unimodal are bimodal, almost all KL polynomials that are not unimodal or bimodal are trimodal, etc., i.e.:

$$\lim_{n \rightarrow \infty} \text{multi}_n^k = 1.$$

Similarly for the multiset versions.

We sadly do not see any way to prove [Conjecture 6.3](#) or [Speculation 6.5](#) at the time of writing this paper. However, although we do not know any connection, related effects are known in the literature. Most notable what one could call log concavity of R polynomials, see e.g. [\[Bre24, Conjecture 2.4\]](#) for details and a precise formulation, and, for example, [\[BB05, Corollary 5.5.3\]](#) for the relation to KL polynomials.

## 7. ROOTS OF KL POLYNOMIALS: ALL ROOTS

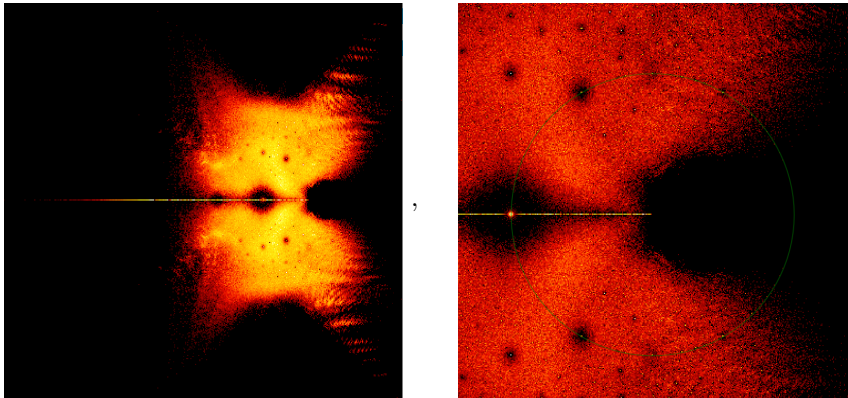
We now compare the following two pictures, following the ideas in [BCD23] and the references to various blogs therein. For other root plots of ‘famous’ polynomials see e.g. [WW01] (albeit this reference takes a quite different perspective).

The first picture is the multiset of roots of the set of KL polynomials. In formulas, we plot

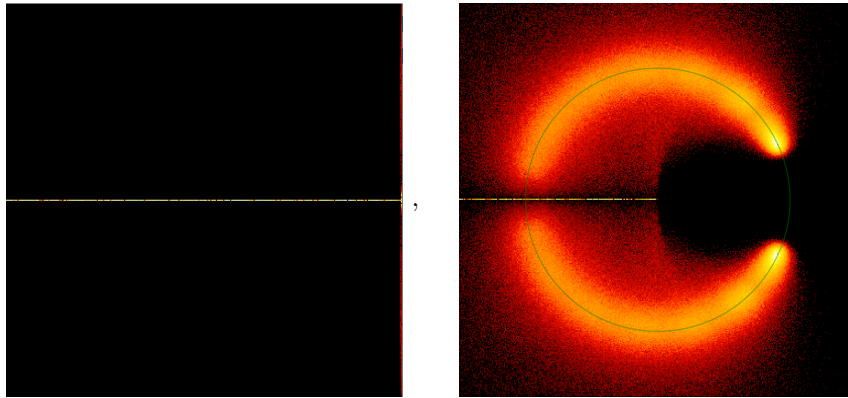
$$(\text{roots}(p) | p \in \{P_w | w \in S_{11}\})$$

with the brightness indicating the density (bright=high density). The second plot is of the same type, but with a different set of polynomials in  $1 + \mathbb{Z}_{\geq 0}[v]$ . For that plot, it was created using 293189 polynomials in  $1 + \mathbb{Z}_{\geq 0}[v]$ , with a randomly chosen degree in  $\{1, \dots, 15\}$  (the maximal degree of a KL polynomial for  $S_{11}$  is 15), and then randomly chosen coefficients  $\{0, \dots, 61582\}$ . These numbers are chosen in this way because the total number of distinct KL polynomials for  $S_{11}$  is 293189, the maximal degree is 15, and the maximal coefficient is 61582.

First, the KL polynomials with the leftmost points being  $\approx -6.6532$  and  $-1.5$ :



Next we have the random polynomials, and the leftmost points are now at  $\approx -61581.9988$  and  $-1.5$ :



The percentage of real roots is  $\approx 24.12\%$ ,  
the percentage of roots with  $|\text{root}|$  in  $[0.9, 1.1]$  is  $\approx 44.15\%$ ,  
the average real value is  $\approx -3849.98$ .

In both cases, the right plot is a zoomed plot with a marked unit circle.

The plots are very different. To make this into a clearer statement, let us first recall a few facts about the distribution of roots of random polynomials. Notable are:

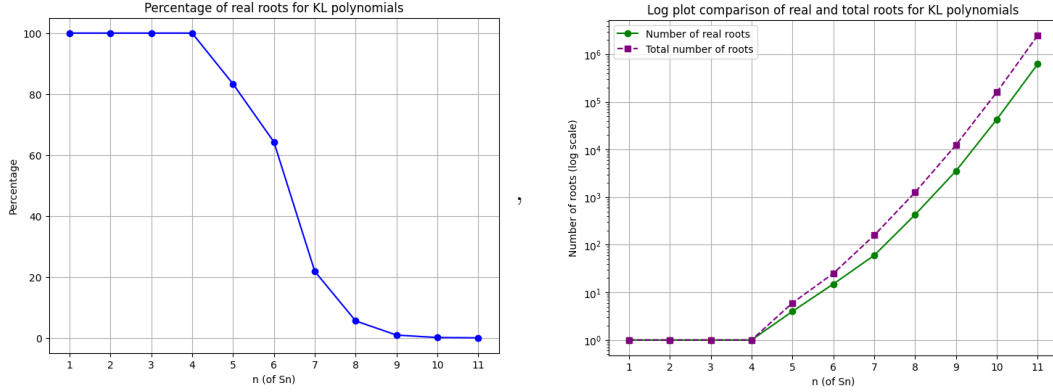
- (7.a) The tail of negative real numbers. The tail is explained in [Section 8](#) below.
- (7.b) The high number of real roots. By [Kac49], the number of expected real roots of a polynomial of degree  $k$  is  $2/\pi \ln k$ . To simplify our calculation, assume that the average degree of our sample is  $e^2$  for the usual  $e \approx 2.71\dots$ . Then the expected percentage of real roots is  $4\pi/e^2 \approx 17.23\%$ .
- (7.c) The reason is that they cluster near the unit circle. This is expected for random polynomials; see, for example, [SV95]. Fairly explicit formulas for the distribution are known, for example, see [MBF<sup>+</sup>97], but the only thing we notice here is that, in the limit, almost all roots will have an absolute value in  $[1 - \epsilon, 1 + \epsilon]$  for a fixed  $\epsilon \in \mathbb{R}_{>0}$ .



(7.d) The many holes in the KL plot. We do not know a general statement, but see [BCD23] for similar patterns (this should be true for other integer valued polynomials as well). The case of coefficients in  $\{-1, 0, 1\}$  is addressed, for example, in [CKW17]. The only exception is ‘forbidden region’ (the black pizza slice around zero for real value  $> 0$ ): this is true in general, and appears in both pictures, and is a classical result; see e.g. [Fil88, Lemma 2].

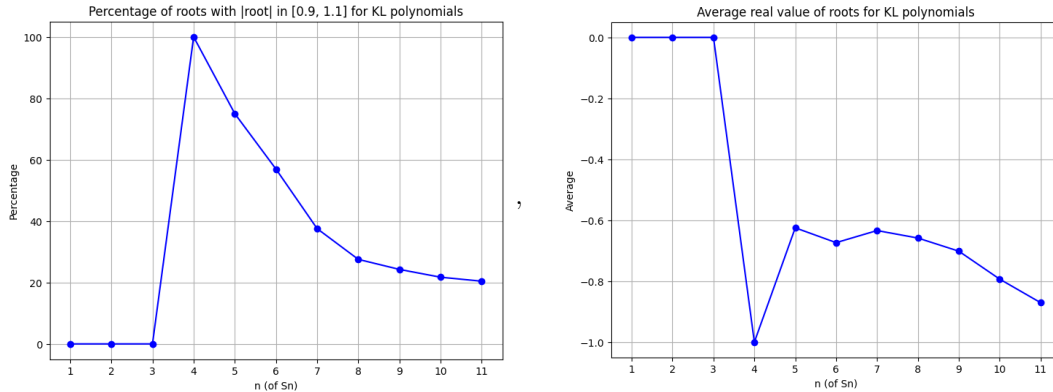
*Remark 7.1.* The references [Kac49, SV95] are for polynomials with complex coefficients since we have not found any references for integer-valued polynomials.  $\diamond$

Here are a few statistics for the roots of KL polynomials. First, the percentage of real roots:



$n$	1	2	3	4	5	6	7	8	9	10	11
%	100	100	100	100	83.33	64.29	21.82	5.56	0.92	0.11	0.01

Second, how far they are, on average, from the origin:



$n$	1	2	3	4	5	6	7	8	9	10	11
%	0	0	0	100	75	56.94	37.55	27.53	24.26	21.72	20.42
av. real	0	0	0	-1	-0.625	-0.6736	-0.6339	-0.6585	-0.7016	-0.7933	-0.8706

As one can see, these are very different from the random distribution. So we ask:

**Question 7.2.** *What is a random KL polynomial (defined so that one gets values closer to the above patterns)?*

We do not know how to attack this problem beyond studying and adjusting the above references.

### 8. ROOTS OF KL POLYNOMIALS: PERRON–FROBENIUS ROOT

We now address the tail in Section 7.

A polynomial  $f = [a_0, \dots, a_d] \in 1 + \mathfrak{v}\mathbb{Z}[\mathfrak{v}]$  satisfies the (*negative*) **Perron–Frobenius (PF for short) property**, or simply **is PF**, if there exists a root  $\lambda \in \mathbb{R}_{\leq 0}$  of  $f$  with  $-\lambda > \max\{|\mu| \mid \mu \text{ is a root of } f, \mu \neq \lambda\}$ . Such a  $\lambda$  is called the **PF root** of  $f$ . The name is explained by the proof of the following statement.

**Lemma 8.1.** *Almost all  $f \in 1 + \mathfrak{v}\mathbb{Z}[\mathfrak{v}]$  satisfy the PF property, i.e.:*

$$\lim_{M \rightarrow \infty} \{f \in 1 + \mathfrak{v}\mathbb{Z}[\mathfrak{v}] \mid \max[a_0, \dots, a_d] \leq M, d \leq M, f \text{ is PF}\} / \{f \in 1 + \mathfrak{v}\mathbb{Z}[\mathfrak{v}] \mid \max[a_0, \dots, a_d] \leq M, d \leq M\} = 1.$$

*Proof.* Let  $\bar{f}$  be  $f$  normalized, i.e. if  $f = [1, a_1, \dots, a_d]$ ,  $a_d \neq 0$ , then  $\bar{f} = [1/a_d, a_1/a_d, \dots, 1]$ . Denote its coefficients by  $b_i$ . The companion matrix of  $f$ , having  $\bar{f}$  as its characteristic polynomial, is

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & -b_0 \\ 1 & 0 & 0 & \dots & 0 & -b_1 \\ 0 & 1 & 0 & \dots & 0 & -b_2 \\ 0 & 0 & 1 & \dots & 0 & -b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -b_{d-1} \end{pmatrix}.$$

Ignoring the sign in the final column, the associated graph is strongly connected, unless some  $b_i = 0$  (and this is an if and only if), and the Perron–Frobenius theorem applies. Taking the sign then back into account gives the negative Perron–Frobenius eigenvalue, and therefore root. Finally, vanishing of a coefficient occurs with probability  $1 - (\frac{M-1}{M})^M$ , but the companion matrix will still satisfy the PF property with one exception: if  $b_{d-1} = 0$ . As this happens with probability  $\frac{1}{M}$ , the limit  $M \rightarrow \infty$  of the converse condition, as in the statement, is 1.  $\square$

Since KL polynomials are either zero or in  $1 + \sqrt[n]{\mathbb{Z}[v]}$ , we can ask:

(8.i) What is the limit  $n \rightarrow \infty$  of

$$\text{PF}_n = \#\{P_w | w \in S_n, P_w \text{ is PF}\} / \#\{P_w | w \in S_n\}?$$

Similarly, but for the multiset, we would like to know the limit  $n \rightarrow \infty$  of

$$\text{PF}_n^{\text{mult}} = \#(P_w | w \in S_n, P_w \text{ is PF}) / \#(P_w | w \in S_n).$$

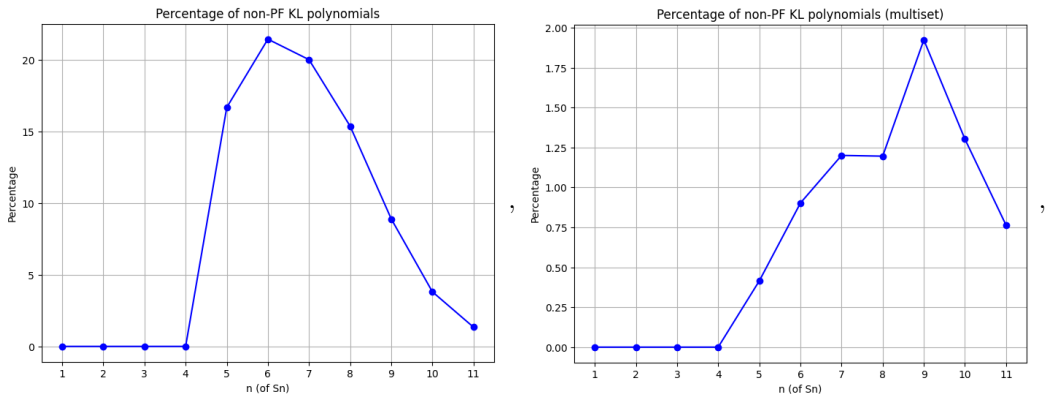
(8.ii) Also, what is the asymptotic behavior  $n \rightarrow \infty$  of

$$\text{PFmax}_n = \max\{-\lambda | \lambda \text{ is PF root of } P_w, w \in S_n, P_w \text{ is PF}\}?$$

(The sign ensures that this is a nonnegative real number.)

For random polynomials e.g. [Gem86, Bai97] implies that the PF root gets very large. In particular, almost surely we have  $\text{PFmax}(\text{random}) \geq \sqrt{M}$  where  $M$  is the bound for the coefficients. In Section 7 we have  $M = 61582$ , so  $\text{PFmax}(\text{random}) \geq 248.1572$ . Moreover, Conjecture 5.4 would imply that  $\text{PFmax}(\text{random})$  grows superexponentially if the KL polynomials were randomly distributed.

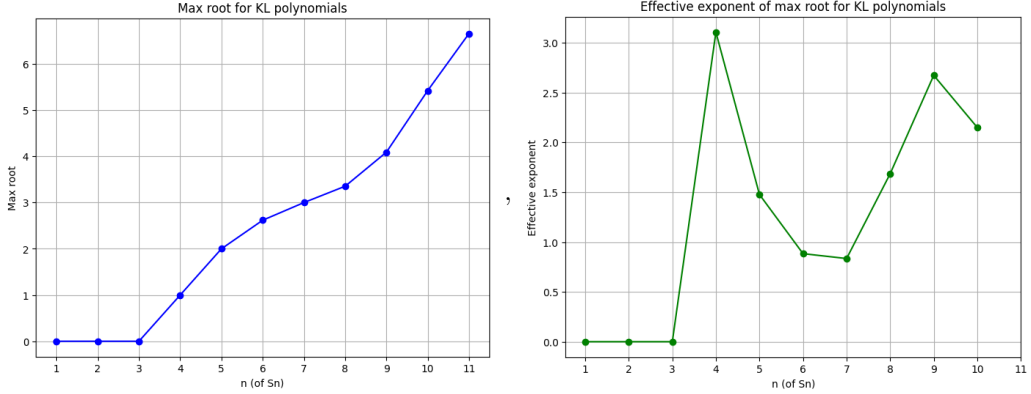
Now to the KL polynomials, the left side for the set of polynomials and the right for the multiset:



$n$	1	2	3	4	5	6	7	8	9	10	11
% set	0	0	0	0	16.67	21.43	20	15.36	8.89	3.81	1.34
% multiset	0	0	0	0	0.41	0.90	1.2	1.20	1.92	1.30	0.76

(Again, let us point out that the top of the diagrams is not at 100%.)

Regarding the size of the PF roots, we have the following:



$n$	1	2	3	4	5	6	7	8	9	10	11
max	0	0	0	1	2	2.618	3	3.354	4.090	5.4206	6.6532

One pattern seems to be clear, so:

**Conjecture 8.2.** *Almost all KL polynomials satisfy the PF property, i.e.:*

$$\lim_{n \rightarrow \infty} \text{PF}_n = 1, \quad \lim_{n \rightarrow \infty} \text{PF}_n^{\text{mult}} = 1.$$

Let us motivate [Speculation 8.3](#), which is up next. Assuming [Conjecture 6.3](#), almost all KL polynomials are unimodal. In fact, the ones with the largest coefficients tend to be unimodal as well. However, it seems that the roots of unimodal polynomials grow slowly. The most famous example of this is  $(1 + v + \dots + v^l)^{k-1}$ , which we have seen in [Section 5](#): Their roots are roots of unity. A nontrivial example is *chromatic roots*, the roots of chromatic polynomials (that these polynomials are unimodal is a nontrivial theorem of [[Huh12](#)]). Since the complete graph with  $k$  vertices has a chromatic root at  $k - 1$ , the rate of growth of the chromatic roots is linear in the number of vertices.

**Speculation 8.3.** *We have subexponential growth, i.e.:*

$$\text{PFmax}_n \in O(\gamma^n) \text{ for all } \gamma \in \mathbb{R}_{>1}.$$

Being probably related to [Section 6](#), we do not know how to prove [Conjecture 8.2](#) or [Speculation 8.3](#).

## 9. THE KL BALLMAPPER

Mapper algorithms are classical tools in TDA, originally introduced in [[SMC07](#)], designed to explore and visualize data. These algorithms combine dimensionality reduction, clustering, and graph network techniques to create a graph out of data. We will use a mild modification that is a bit more suitable for our purpose, as it gives nicer visualizations, *ball mapper* from [[Dło19](#)]. These types of algorithms have found important applications, even in the real-world sciences; see, for example, [[Dło24](#)] for references and examples of applications. Moreover, the current section is inspired by [[DGS24](#)], which uses the ball mapper on data from knot theory, Tic-Tac-Toe games, and classical cancer databases. We now explain the results if one applies these methods to the KL polynomials.

In fact, following [[BBD<sup>+</sup>22](#)], we also used the so-called *H polynomials* together with the KL polynomials. See [[BBD<sup>+</sup>22](#), Section 3.1] for the definition. The point is that knowing the tuple  $(P_w, \ell(w))$  is equivalent to knowing the H polynomial for  $w$  used in the ball mapper.

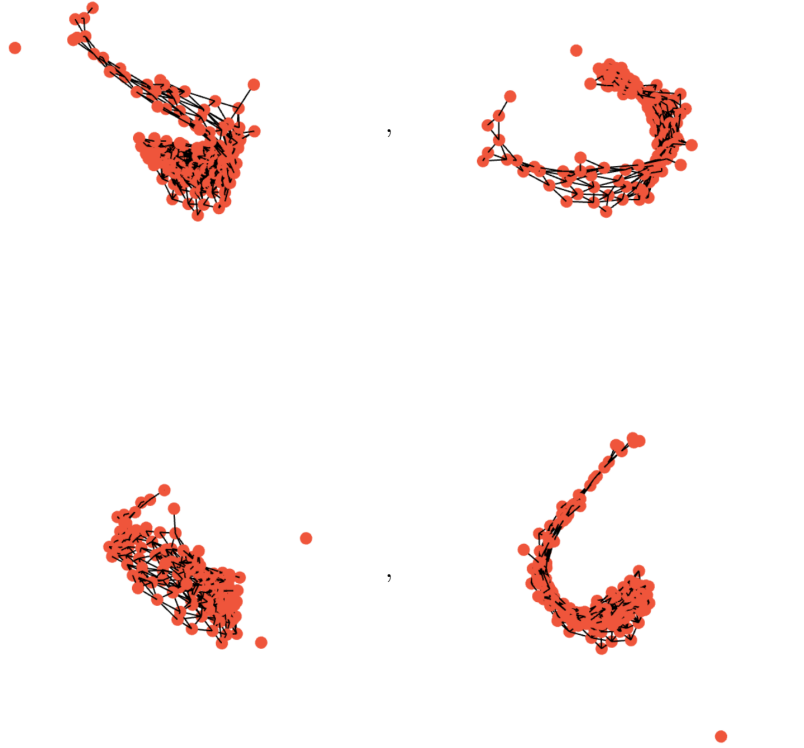
*Remark 9.1.* The data sets for the H polynomials can also be found in [[LTV24a](#)]. ◇

We will not recall the definition of the mapper and ballmapper algorithms. Here is a very brief summary of what the reader needs to know; see the above references for details:

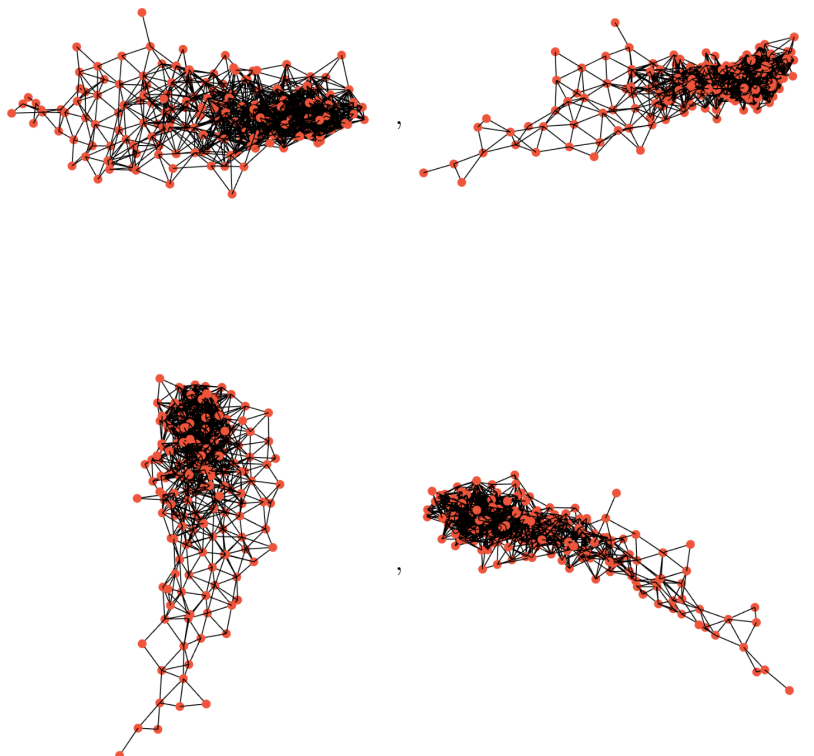
- (a) The algorithm we used creates a graph  $G = G(\epsilon)$  from a point cloud using balls of a prefixed radius  $\epsilon \in \mathbb{R}_{>0}$ . All points in a ball with this radius are collapsed into a vertex of  $G$ , i.e.  $G$  which has more vertices the smaller  $\epsilon$ . The main point then is the study of the family of graphs when one varies  $\epsilon$ .
- (b) In order to improve the graph plots, we rescaled the KL and H polynomials, followed by a principal component analysis (PCA).
- (c) For visualization in 3D, we used force-directed drawing algorithms of (coil-)spring-type and Kamada–Kawai-type as provided by the Python library Bokeh. Details about these graph visualization algorithms can be found, for example, in [[Tam14](#), Chapter 12].

(d) More details are explained in [LTV24a]. We give only a sample of what can be found in [LTV24a]; in particular, the graphs are much more impressive in the 3D plot in [LTV24a].

We now present the KL ball mappers for  $S_{11}$ . First, the spring embedding for  $\epsilon = 4 \cdot 10^{-9}$  and the KL polynomials, rotated to four different angles:

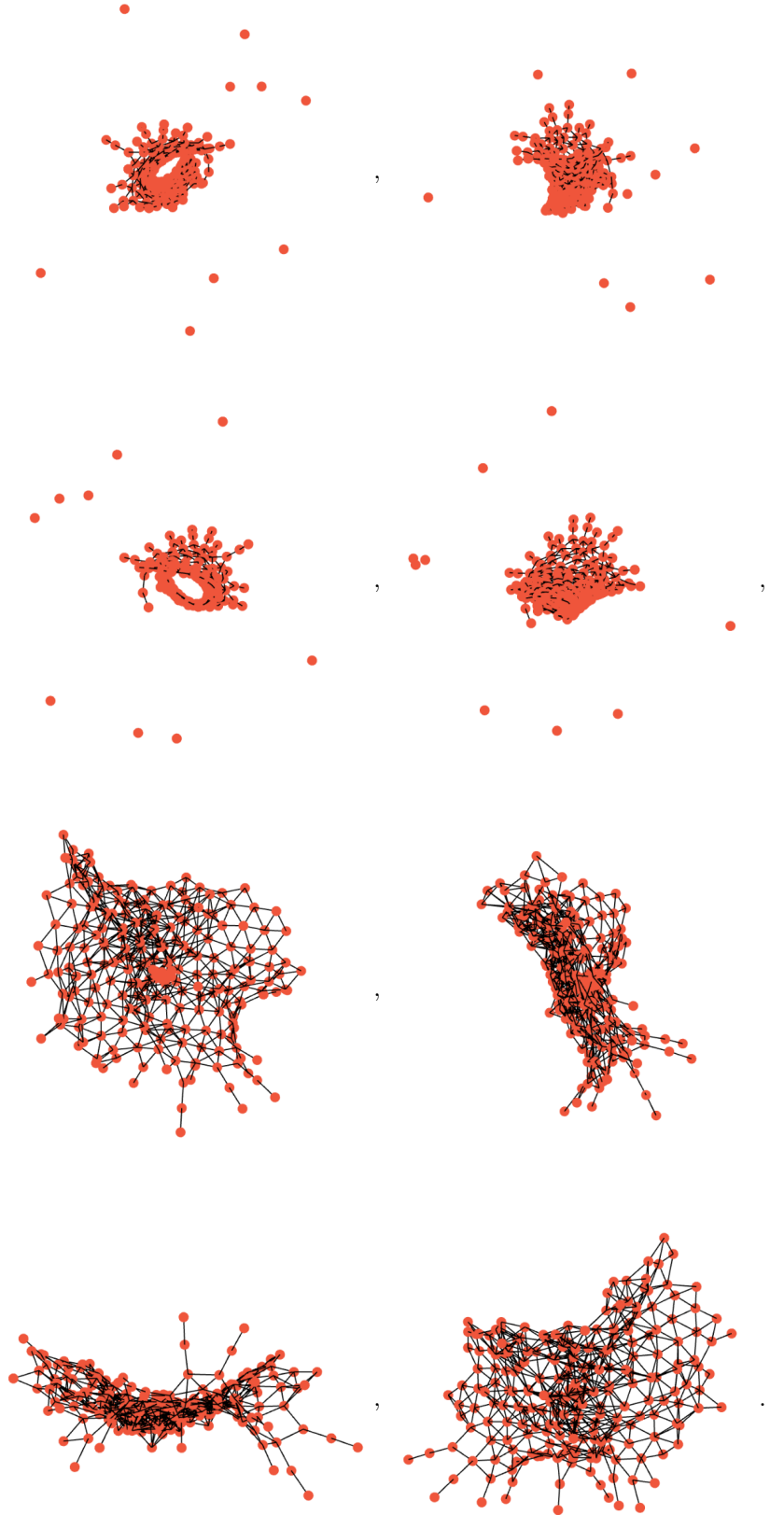


Next, the Kamada–Kawai embedding for the same  $\epsilon$ , the KL polynomials, and again in four different angles:



We point out, as the reader is invited to check using the code in [LTV24a], that these arc-like (top) or fish (bottom) shapes are quite stable when varying  $\epsilon$ , but also when varying  $n$  from roughly  $n = 8$  onward. This might indicate some stabilization process that happens for  $n \rightarrow \infty$  but, at the same time, it underlines that the small  $n$  does not provide a sample representative of KL polynomials.

Similarly, the H polynomials in spring and the Kamada–Kawai embedding, both for the same  $\epsilon = 10^{-9}$ , are:

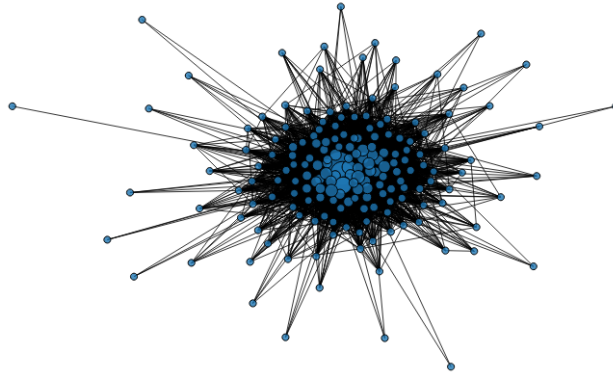


The same comments as for the KL polynomials apply, but this time for this web-type shape.

ChatGPT says about the KL polynomial ball mapper and the H polynomial ball mapper, both for the spring embedding:

- ▶ “The overall shape resembles a ‘fish’ or a ‘stream,’ with the dense cluster forming a ‘head’ and the sparser connections on the left resembling a ‘tail.’ It gives the impression of directional flow or gradient density across the network.”
- ▶ “The overall structure looks like a dense, irregular web, with some areas more interconnected than others. The shape does not conform to a specific geometric form, but appears somewhat sprawling and organic, indicative of a complex network with a dense central region and sparser peripheral connections.”

Finally, again there is a huge difference between randomly generated polynomials as in Section 7 and KL polynomials. The ball mapper for the former is a random graph:



Thus, in analogy to Question 7.2, we ask:

**Question 9.2.** *What is a random KL polynomial (defined so that one gets BM closer to the above)?*

*Remark 9.3.* We also tried one of the most classical tools from TDA on the KL polynomials: persistent homology. However, the data set of KL polynomials starts to be interesting very late and is then too large: we were not able to extract meaningful results.  $\diamond$

## REFERENCES

- [AK92] H.H. Andersen and W.X. Kexin. Representations of quantum algebras. The mixed case. *J. Reine Angew. Math.*, 427:35–50, 1992. doi:10.1515/crll.1992.427.35.
- [Aul51] F.C. Auluck. On some new types of partitions associated with generalized Ferrers graphs. *Proc. Cambridge Philos. Soc.*, 47:679–686, 1951.
- [BCD23] J.C. Baez, J.D. Christensen, and S. Derbyshire. The beauty of roots. *Notices Amer. Math. Soc.*, 70(9):1495–1497, 2023. URL: <https://arxiv.org/abs/2310.00326>, doi:10.1090/noti2789.
- [Bai97] Z.D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997. doi:10.1214/aop/1024404298.
- [Bai24] J. Baine. (co)Minuscule Hecke categories. 2024. URL: <https://arxiv.org/abs/2409.16131>.
- [BB03] S.C. Billey and T. Braden. Lower bounds for Kazhdan–Lusztig polynomials from patterns. *Transform. Groups*, 8(4):321–332, 2003. URL: <https://arxiv.org/abs/math/0202252>, doi:10.1007/s00031-003-0629-x.
- [BB05] A. Björner and F. Brenti. *Combinatorics of Coxeter groups*, volume 231 of *Graduate Texts in Mathematics*. Springer, New York, 2005.
- [BBD<sup>+</sup>22] C. Blundell, L. Buesing, A. Davies, P. Veličković, and G. Williamson. Towards combinatorial invariance for Kazhdan–Lusztig polynomials. *Represent. Theory*, 26:1145–1191, 2022. URL: <https://arxiv.org/abs/2111.15161>, doi:10.1090/ert/624.
- [BM01] T. Braden and R. MacPherson. From moment graphs to intersection cohomology. *Math. Ann.*, 321(3):533–551, 2001. URL: <https://arxiv.org/abs/math/0008200>, doi:10.1007/s002080100232.
- [Bre24] F. Brenti. Some open problems on Coxeter groups and unimodality. In *Open problems in algebraic combinatorics*, volume 110 of *Proc. Sympos. Pure Math.*, pages 23–37. Amer. Math. Soc., Providence, RI, [2024] ©2024. URL: <https://arxiv.org/abs/2410.09897>.
- [BK09] J. Brundan and A. Kleshchev. Blocks of cyclotomic Hecke algebras and Khovanov–Lauda algebras. *Invent. Math.*, 178(3):451–484, 2009. URL: <https://arxiv.org/abs/0808.2032>, doi:10.1007/s00222-009-0204-8.
- [CKW17] D. Calegari, S. Koch, and A. Walker. Roots, Schottky semigroups, and a proof of Bandt’s conjecture. *Ergodic Theory Dynam. Systems*, 37(8):2487–2555, 2017. URL: <https://arxiv.org/abs/1410.8542>, doi:10.1017/etds.2016.17.
- [dC06] F. du Cloux. Positivity results for the Hecke algebras of noncrystallographic finite Coxeter groups. *J. Algebra*, 303(2):731–741, 2006. URL: <https://arxiv.org/abs/math/0506448>, doi:10.1016/j.jalgebra.2005.10.004.
- [Cvi08] P. Cvitanović. *Group theory*. Princeton University Press, Princeton, NJ, 2008. Birdtracks, Lie’s, and exceptional groups. doi:10.1515/9781400837670.
- [DVB<sup>+</sup>21] A. Davies, P. Veličković, L. Buesing, S. Blackwell, D. Zheng, N. Tomašev, R. Tanburn, P. Battaglia, C. Blundell, A. Juhász, M. Lackenby, G. Williamson, D. Hassabis, and K. Pushmeet. Advancing mathematics by guiding human intuition with AI. *Nature*, 600:70–74, 2021. doi:10.1038/s41586-021-04086-x.



- [Dło19] P. Dłotko. Ball mapper: a shape summary for topological data analysis. 2019. URL: <https://arxiv.org/abs/1901.07410>.
- [Dło24] P. Dłotko. On the shape that matters—topology and geometry in data science. *Eur. Math. Soc. Mag.*, (132):5–13, 2024. doi:10.4171/MAG/190.
- [DGS24] P. Dłotko, D. Gurnari, and R. Sazdanovic. Mapper–Type Algorithms for Complex Data and Relations. *J. Comput. Graph. Statist.*, 33(4):1383–1396, 2024. URL: <https://arxiv.org/abs/2109.00831>, doi:10.1080/10618600.2024.2343321.
- [DK13] V. Dotsenko and A. Khoroshkin. Shuffle algebras, homology, and consecutive pattern avoidance. *Algebra Number Theory*, 7(3):673–700, 2013. URL: <https://arxiv.org/abs/1109.2690>, doi:10.2140/ant.2013.7.673.
- [EGNO15] P. Etingof, S. Gelaki, D. Nikshych, and V. Ostrik. *Tensor categories*, volume 205 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2015. doi:10.1090/surv/205.
- [EM24] A. Evseev and A. Mathas. Content systems and deformations of cyclotomic KLR algebras of type  $a$  and  $c$ . *Annals of Representation Theory*, 1(2):193–297, 2024. URL: <https://arxiv.org/abs/2209.00134>, doi:10.5802/art.8.
- [FK10] P. Fendley and V. Krushkal. Link invariants, the chromatic polynomial and the Potts model. *Adv. Theor. Math. Phys.*, 14(2):507–540, 2010. URL: <https://arxiv.org/abs/0806.3484>.
- [Fil88] M. Filaseta. Irreducibility criteria for polynomials with nonnegative coefficients. *Canad. J. Math.*, 40(2):339–351, 1988. doi:10.4153/CJM-1988-013-6.
- [FL21] J. Flake and R. Laugwitz. On the monoidal center of Deligne’s category  $\text{Rep}(S_t)$ . *J. Lond. Math. Soc. (2)*, 103(3):1153–1185, 2021. URL: <https://arxiv.org/abs/1901.08657>, doi:10.1112/jlms.12403.
- [FKK98] I.B. Frenkel, M.G. Khovanov, and A.A. Kirillov, Jr. Kazhdan–Lusztig polynomials and canonical basis. *Transform. Groups*, 3(4):321–336, 1998. URL: <https://arxiv.org/abs/q-alg/9709042>, doi:10.1007/BF01234531.
- [Gem86] S. Geman. The spectral radius of large random matrices. *Ann. Probab.*, 14(4):1318–1328, 1986. doi:10.1214/aop/1176992372.
- [Gib21] J. Gibson. IHecke and ASLoc for Magma. 2021. <https://github.com/joelgibson/IHecke>, <https://github.com/joelgibson/ASLoc>.
- [HP08] A. Hammett and B. Pittel. How often are two permutations comparable? *Trans. Amer. Math. Soc.*, 360(9):4541–4568, 2008. URL: <https://arxiv.org/abs/math/0605490>, doi:10.1090/S0002-9947-08-04478-4.
- [Huh12] J. Huh. Milnor numbers of projective hypersurfaces and the chromatic polynomial of graphs. *J. Amer. Math. Soc.*, 25(3):907–927, 2012. URL: <http://arxiv.org/abs/1008.4749>, doi:10.1090/S0894-0347-2012-00731-0.
- [JW17] L.T. Jensen and G. Williamson. The  $p$ -canonical basis for Hecke algebras. In *Categorification and higher representation theory*, volume 683 of *Contemp. Math.*, pages 333–361. Amer. Math. Soc., Providence, RI, 2017. URL: <https://arxiv.org/abs/1510.01556>, doi:10.1090/conm/683.
- [Jon86] V.F.R. Jones. A new knot polynomial and von Neumann algebras. *Notices Amer. Math. Soc.*, 33(2):219–225, 1986.
- [Kac49] M. Kac. On the average number of real roots of a random algebraic equation. II. *Proc. London Math. Soc. (2)*, 50:390–408, 1949. doi:10.1112/plms/s2-50.5.390.
- [KL79] D. Kazhdan and G. Lusztig. Representations of Coxeter groups and Hecke algebras. *Invent. Math.*, 53(2):165–184, 1979. doi:10.1007/BF01390031.
- [KL09] M. Khovanov and A.D. Lauda. A diagrammatic approach to categorification of quantum groups. I. *Represent. Theory*, 13:309–347, 2009. URL: <https://arxiv.org/abs/0803.4121>, doi:10.1090/S1088-4165-09-00346-X.
- [KK99] M. Khovanov and G. Kuperberg. Web bases for  $\text{sl}(3)$  are not dual canonical. *Pacific J. Math.*, 188(1):129–153, 1999. URL: <https://arxiv.org/abs/q-alg/9712046>, doi:10.2140/pjm.1999.188.129.
- [Kup96] G. Kuperberg. Spiders for rank 2 Lie algebras. *Comm. Math. Phys.*, 180(1):109–151, 1996. URL: <https://arxiv.org/abs/q-alg/9712003>.
- [LTV24b] A. Lacabanne, D. Tubbenhauer, and P. Vaz. A formula to evaluate type  $a$  webs and link polynomials. *Ark. Mat.*, 62(1):83–101, 2024. URL: <https://arxiv.org/abs/2209.12169>, doi:10.4310/ARKIV.2024.v62.n1.a5.
- [LTV24a] A. Lacabanne, D. Tubbenhauer, and P. Vaz. Code and more for the paper Big data approaches to Kazhdan–Lusztig polynomials. 2024. <https://github.com/dtubbenhauer/kldata>.
- [LHS22] J.S.F. Levitt, M. Hajij, and R. Sazdanovic. Big data approaches to knot theory: understanding the structure of the Jones polynomial. *J. Knot Theory Ramifications*, 31(13):Paper No. 2250095, 20, 2022. URL: <https://arxiv.org/abs/1912.10086>, doi:10.1142/s021821652250095x.
- [Lus03] G. Lusztig. *Hecke algebras with unequal parameters*, volume 18 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 2003. URL: <https://arxiv.org/abs/math/0208154>.
- [MMM<sup>+</sup>23] M. Mackaay, V. Mazorchuk, V. Miemietz, D. Tubbenhauer, and X. Zhang. Simple transitive 2-representations of Soergel bimodules for finite Coxeter types. *Proc. Lond. Math. Soc. (3)*, 126(5):1585–1655, 2023. URL: <https://arxiv.org/abs/1906.11468>, doi:10.1112/plms.12515.
- [MPT14] M. Mackaay, W. Pan, and D. Tubbenhauer. The  $\mathfrak{sl}_3$ -web algebra. *Math. Z.*, 277(1-2):401–479, 2014. URL: <https://arxiv.org/abs/1206.2118>, doi:10.1007/s00209-013-1262-6.
- [MT23] A. Mathas and D. Tubbenhauer. Cellularity of  $\text{klr}$  and weighted  $\text{klrw}$  algebras via crystals. 2023. URL: <https://arxiv.org/abs/2309.13867>.
- [MBF<sup>+</sup>97] G. Mezincescu, D. Bessis, J.-D. Fournier, G. Mantica, and F.D. Aaron. Distribution of roots of random real generalized polynomials. *J. Statist. Phys.*, 86(3-4):675–705, 1997. URL: <https://arxiv.org/pdf/chao-dyn/9606012>, doi:10.1007/BF02199115.
- [MPS17] S. Morrison, E. Peters, and N. Snyder. Categories generated by a trivalent vertex. *Selecta Math. (N.S.)*, 23(2):817–868, 2017. URL: <https://arxiv.org/abs/1501.06869>, doi:10.1007/s00029-016-0240-3.
- [OEI23] OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences, 2023. Published electronically at <http://oeis.org>.
- [PRH14] D. Plaza and S. Ryom-Hansen. Graded cellular bases for Temperley–Lieb algebras of type  $A$  and  $B$ . *J. Algebraic Combin.*, 40(1):137–177, 2014. URL: <https://arxiv.org/abs/1203.2592>, doi:10.1007/s10801-013-0481-6.
- [Rou08] R. Rouquier. 2-Kac–Moody algebras. 2008. URL: <http://arxiv.org/abs/0812.5023>.
- [RTW32] G. Rumer, E. Teller, and H. Weyl. Eine für die Valenztheorie geeignete Basis der binären Vektorinvarianten. *Nachrichten von der Ges. der Wiss. Zu Göttingen. Math.-Phys. Klasse*, pages 498–504, 1932. In German.



- [RT22] H.M. Russell and J. Tymoczko. The transition matrix between the Specht and  $\mathfrak{sl}_3$  web bases is unitriangular with respect to shadow containment. *Int. Math. Res. Not. IMRN*, (5):3371–3416, 2022. doi:10.1093/imrn/rnaa290.
- [RT19] H.M. Russell and J.S. Tymoczko. The transition matrix between the Specht and web bases is unipotent with additional vanishing entries. *Int. Math. Res. Not. IMRN*, (5):1479–1502, 2019. URL: <https://arxiv.org/abs/1701.01868>, doi:10.1093/imrn/rnx164.
- [SS24] R. Sazdanovic and D. Scofield. Structure of the chromatic polynomial. 2024. URL: <https://arxiv.org/abs/2411.15088>.
- [SSV98] B. Shapiro, M. Shapiro, and A. Vainshtein. Kazhdan–Lusztig polynomials for certain varieties of incomplete flags. In *Proceedings of the 7th Conference on Formal Power Series and Algebraic Combinatorics (Noisy-le-Grand, 1995)*, volume 180, pages 345–355, 1998. doi:10.1016/S0012-365X(97)00124-6.
- [SV95] L.A. Shepp and R.J. Vanderbei. The complex zeros of random polynomials. *Trans. Amer. Math. Soc.*, 347(11):4365–4384, 1995. doi:10.2307/2155041.
- [SMC07] G. Singh, F. Memoli, and G. Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007. doi:10.2312/SPBG/SPBG07/091-100.
- [Soe97] W. Soergel. Kazhdan–Lusztig polynomials and a combinatoric[s] for tilting modules. *Represent. Theory*, 1:83–114, 1997. doi:10.1090/S1088-4165-97-00021-6.
- [STWZ23] L. Sutton, D. Tubbenhauer, P. Wedrich, and J. Zhu. SL2 tilting modules in the mixed case. *Selecta Math. (N.S.)*, 29(3):39, 2023. URL: <https://arxiv.org/abs/2105.07724>, doi:10.1007/s00029-023-00835-0.
- [Tam14] R. Tamassia, editor. *Handbook of graph drawing and visualization*. Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, 2014. URL: <https://cs.brown.edu/people/rtamassi/gdhandbook/>.
- [Ten07] B.E. Tenner. Pattern avoidance and the Bruhat order. *J. Combin. Theory Ser. A*, 114(5):888–905, 2007. URL: <https://arxiv.org/abs/math/0604322>, doi:10.1016/j.jcta.2006.10.003.
- [Tub23] D. Tubbenhauer. On rank one 2-representations of web categories. 2023. To appear in *Algebr. Comb.* URL: <https://arxiv.org/abs/2307.00785>.
- [War11] G.S. Warrington. Equivalence classes for the  $\mu$ -coefficient of Kazhdan–Lusztig polynomials in  $S_n$ . *Exp. Math.*, 20(4):457–466, 2011. URL: <https://arxiv.org/abs/1010.3961>, doi:10.1080/10586458.2011.565260.
- [Web19] B. Webster. Weighted Khovanov–Lauda–Rouquier algebras. *Doc. Math.*, 24:209–250, 2019. URL: <https://arxiv.org/abs/1209.2463>, doi:10.25537/dm.2019v24.209-250.
- [Wil24] G. Williamson. Is deep learning a useful tool for the pure mathematician? *Bull. Amer. Math. Soc. (N.S.)*, 61(2):271–286, 2024. URL: <https://arxiv.org/abs/2304.12602>, doi:10.1090/bull/1829.
- [Woo09] A. Woo. Permutations with Kazhdan–Lusztig polynomial  $P_{i,d,w}(q) = 1 + q^h$ . *Electron. J. Combin.*, 16(2):Research Paper 10, 32, 2009. With an appendix by S. Billey and J. Weed. URL: <https://arxiv.org/abs/0809.2374>, doi:10.37236/76.
- [WW01] F.Y. Wu and J. Wang. Zeroes of the Jones polynomial. *Phys. A*, 296(3-4):483–494, 2001. URL: <https://arxiv.org/pdf/cond-mat/0105013>, doi:10.1016/S0378-4371(01)00189-3.
- [Yam89] S. Yamada. An invariant of spatial graphs. *J. Graph Theory*, 13(5):537–551, 1989. doi:10.1002/jgt.3190130503.

A.L.: LABORATOIRE DE MATHÉMATIQUES BLAISE PASCAL (UMR 6620), UNIVERSITÉ CLERMONT AUVERGNE, CAMPUS UNIVERSITAIRE DES CÉZEAUX, 3 PLACE VASARELY, 63178 AUBIÈRE CEDEX, FRANCE, [WWW.NORMALESUP.ORG/~LACABANNE](http://WWW.NORMALESUP.ORG/~LACABANNE), ORCID 0000-0001-8691-3270

Email address: [abel.lacabanne@uca.fr](mailto:abel.lacabanne@uca.fr)

D.T.: THE UNIVERSITY OF SYDNEY, SCHOOL OF MATHEMATICS AND STATISTICS F07, OFFICE CARSLAW 827, NSW 2006, AUSTRALIA, [WWW.DTUBBENHAUER.COM](http://WWW.DTUBBENHAUER.COM), ORCID 0000-0001-7265-5047

Email address: [daniel.tubbenhauer@sydney.edu.au](mailto:daniel.tubbenhauer@sydney.edu.au)

P.V.: INSTITUT DE RECHERCHE EN MATHÉMATIQUE ET PHYSIQUE, UNIVERSITÉ CATHOLIQUE DE LOUVAIN, CHEMIN DU CYCLOTRON 2, 1348 LOUVAIN-LA-NEUVE, BELGIUM, [HTTPS://PERSO.UCLOUVAIN.BE/PEDRO.VAZ](https://perso.uclouvain.be/pedro.vaz), ORCID 0000-0001-9422-4707

Email address: [pedro.vaz@uclouvain.be](mailto:pedro.vaz@uclouvain.be)