Contents

Abstract

The paper contains a revised, and extended by new results, part of the author's PhD thesis. The main objects that we study are toric varieties naturally associated to special Markov processes on trees. Such Markov processes can be defined by a tree T and a group G. They are called group-based models. The main, but not unique, motivation to consider these processes comes from phylogenetics. We study the geometry, defining equations and combinatorial description of the associated toric varieties. We obtain new results for a large class of not necessarily abelian group-based models, which we call G-models. We also prove that equations of degree 4 define the projective scheme representing the 3-Kimura model.

Acknowledgements. First and foremost I want to thank my supervisors Laurent Manivel and Jarosław Wiśniewski. It was a great pleasure and opportunity to learn from them. Throughout the work on my thesis they provided me with encouragement and ideas. I am sincerely grateful for their effort.

It is difficult to overstate my gratitude to Rosa-Maria Miró-Roig, who provided me with a lot of help. I am heartily thankful to Laura Costa, who together with Rosa-Maria, taught me a lot and presented inspiring and interesting problems.

I offer my sincerest gratitude to Laurent Bonavero, with whom I started the PhD studies. I thank also Sławomir Cynk, who with Laurent Bonavero introduced me to algebraic geometry.

I am very grateful to Elizabeth Allman, Marta Casanellas, Bernd Sturmfels and Seth Sullivant for their explanations, answers to my questions and important remarks. Their expert opinions have helped a lot to improve the results of the thesis.

I sincerely acknowledge the administrative help of Emmanuel Peyre and Anna Poczmańska, who made the PhD studies possible.

I am grateful to Jan Draisma and Markus Perling for their crucial remarks.

Writing this thesis would not be possible without the support of my family. I thank a lot Gabriela, Martyna, Małgorzata and Krzysztof Michałek for their love and support.

I thank a lot my friends: Roland Abuaf, Piotr Achinger, Agnieszka Bodzenta, Weronika Buczyńska, Jarosław Buczyński, Maria Donten-Bury, Nathan Ilten, Thomas Kahle, Grzegorz Kapustka, Michał Kapustka, Kaie Kubjas, Karine Kuyumzhiyan, Kevin Langlois, Michał Lasoń, Christopher Manon, Karol Pałka, Clélia Pech, Ronan Terpereau for interesting talks, questions, answers and a great working atmosphere.

The author is supported by DAAD PRIME program.

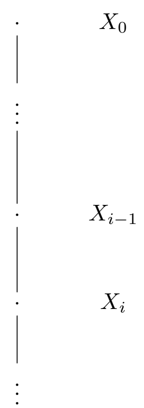
2010 *Mathematics Subject Classification*: Primary 14M25, 13P25; Secondary 52B20.
Key words and phrases: group-based model, algebraic phylogenetics, Kimura models.
Received 2.12.2013; revised version 30.7.2014.

1. Introduction

1.1. Short summary of results. To each Markov process on a tree one can associate an algebraic variety. Motivated by biology, we focus on Markov processes defined by a group action. We provide a precise description of the polytope representing the associated toric variety for a large class of models (Theorem 4.63). We provide conditions ensuring that the varieties obtained are normal (Proposition 4.73), as well as give examples when they are not (Proposition 4.74, Computation 4.75). One of the main tools we use is the generalization of the notions of sockets and networks introduced in [BW07] for the group $\mathbb{Z}_2$ to arbitrary abelian groups. In our setting the networks form a group (Definition 4.24) that acts on the variety. Moreover, the ambient space of the variety is the regular representation of this group.

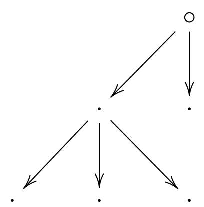
The main open problem that we address is a conjecture of Sturmfels and Sullivant [SS05, Conjecture 30] stating that the affine scheme associated to the 3-Kimura model is defined by an ideal generated in degree 4. Our strongest result states that the associated projective scheme can be defined by an ideal generated in degree 4 (Theorem 11.1). We also present a part of a joint work with Maria Donten-Bury: a method for generating phylogenetic invariants for any model. We prove that our method provides generators of the ideal for many models and trees if and only if the conjecture of Sturmfels and Sullivant holds (Proposition 6.8). We present some applications, for example to the identifiability problem in biology.

1.2. Motivations. Our motivations come from applied mathematics. Let us recall basic properties of Markov chains and Markov processes on trees. A Markov chain is a sequence of random variables $\{X_i\}$ that satisfy specific conditions. For a fixed state of a variable



X_{i-1} the variable X_i is independent of the set of all the variables X_{i-j} for $j > 1$. Typically, this chain is depicted vertically by associating a vertex to each variable and joining X_i to X_{i-1}. For a Markov chain we usually introduce conditional probabilities that specify all the properties of the chain. Suppose that each variable X_i can be in $a_i < \infty$ states. Then to each edge joining X_{i-1} and X_i we can associate an $a_{i-1} \times a_i$ matrix. The columns and rows of the matrix are indexed respectively by the states of X_{i-1} and X_i. The entries correspond to conditional probabilities. Namely, the entry indexed by a pair (p, q) of states equals the probability that X_i is in state q under the condition that X_{i-1} is in state p. These matrices are called *transition matrices*. If we know the distribution of X_0 and the transition matrices we can easily calculate the distributions of all other variables.

This construction can be directly generalized to rooted trees. By a *rooted tree* we will always mean a connected graph with one distinguished vertex and no cycles. By *leaves* we mean vertices of valency one. *Nodes* are vertices that are not leaves. We will sometimes identify leaves with edges adjacent to them. To simplify the language we assume that the tree is a directed graph and all the edges are directed away from the root. In the example below the root is denoted by $\circ$.

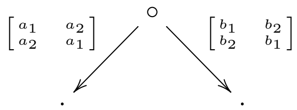


As before, to each vertex we associate a random variable. We say that a node v_1 is a *direct ancestor* of v_2 if there is an edge directed from v_1 to v_2. Note that there is always one direct ancestor, except for the root that does not have ancestors. The *descendants* of a vertex are all the vertices that can be reached from it by a directed path. The Markov property ensures that a variable X is independent of all other variables that are not its descendants once the state of the direct ancestor is fixed.

One of the possible approaches to problems in phylogenetics using algebraic geometry is as follows. We fix a rooted tree T that we suspect is a correct model of evolution. We consider any transition matrices with entries that are free parameters, which depend only on the biological model that we choose. To the space of parameters we add also possible distributions of the variable associated to the root. We calculate the distribution of random variables associated to leaves. More precisely, we get a map ([1]) $\pi \circ \widehat{\psi}$. Its domain parametrizes entries of transition matrices and possible distributions of the random variable associated to the root. Its image parametrizes all possible distributions of the random variables associated to leaves.

Example 1.1. In this example we suppose that each variable can be in two states, denoted by 0 and 1. There is one root with two descendants. The variable associated to the root attains the value 0 and 1 with probability respectively λ_0 and λ_1. The transition matrices are as follows:

([1]) The reason for choosing this notation will become clear in the following sections.



Hence there are six parameters. The leaves can be in four states. We order them as follows:

1) both leaves are in state 0,
2) the left leaf is in state 0 and the right one in state 1,
3) the left leaf is in state 1 and the right one in state 0,
4) both leaves are in state 1.

We obtain the map

$$\pi \circ \widehat{\psi} : (\lambda_0, \lambda_1, a_1, a_2, b_1, b_2) \to (\lambda_0 a_1 b_1 + \lambda_1 a_2 b_2, \lambda_0 a_1 b_2 + \lambda_1 a_2 b_1, \lambda_0 a_2 b_1 + \lambda_1 a_1 b_2, \lambda_0 a_2 b_2 + \lambda_1 a_1 b_1).$$

Let P be the point, established empirically, that represents the distribution of random variables associated to leaves. We would like to check if P belongs to the image of $\pi \circ \widehat{\psi}$. If it is not in the image, then we know that either the biological model we used is wrong, or the tree T is not the right one. If the point P is in the image, we can ask for a description of the fiber. However, determining whether P belongs to the image is hard in general. One of the methods bases on the fact that $\pi \circ \widehat{\psi}$ is an algebraic map. We can consider the Zariski closure of its image. This is an affine algebraic variety. One would like to describe its ideal and check whether the generators vanish at P. The elements of this ideal are called *phylogenetic invariants*.

This approach may not be very effective. The description of the ideal of a variety given by a parametrization is not an easy task. However, the maps we get are not arbitrary. As first observed by Evans and Speed [ES93] and Hendy and Penny [HP89] for certain models of evolution, the variety we consider is toric. More precisely, there are coordinates in which the parametrization map is given by monomials. This allows us to apply methods of toric geometry in order to determine the ideal of the variety.

Throughout the article we assume that *the random variable associated to the root has a uniform distribution.* This assumption is not motivated by biology. We use it only to obtain nicer results from the mathematical point of view. Hence, in our study the parameter space contains only the coefficients of the transition matrices.

One of the main aims of this article is to provide a general description of toric varieties appearing in phylogenetics. We present the most general known setting in Theorem 4.63. In particular, we believe that our approach covers all biological models of interest that are known to give rise to toric varieties. Further, we investigate the properties of the toric varieties obtained. We prove that varieties associated to certain biological models are normal (Proposition 4.73). However, we also give examples where the varieties are not normal (Computation 4.75). Next we address the question for which models the varieties associated to trivalent (2) trees belong to the same flat family. For the binary Jukes–

(2) The valency of all vertices is either one or three.

Cantor this fact was known to be true by [BW07], while for 3-Kimura it does not hold by [Kub12]. By calculating the Hilbert polynomials of many varieties we have found that most models do not have this property.

Another important task concerns phylogenetic invariants.

DEFINITION 1.2 (Claw tree). A *claw tree* $K_{n,1}$ is a tree with exactly one inner vertex and n leaves.

For many models, in particular those that are most important for us, the study of phylogenetic invariants of any tree reduces to the case of the claw tree [SS05], [AR08], [DK09]. However, establishing phylogenetic invariants in this special case has turned out to be difficult. We do not even know the degree in which the ideal of phylogenetic invariants is generated. There is a well-known conjecture by Sturmfels and Sullivant [SS05, Conjecture 1] that gives a precise upper bound for this degree. The conjecture is astonishingly similar to an old theorem of Noether. The theorem bounds the degree in which the ring of invariants of the group action on the polynomials is generated. However, as we will see in Section 5, it is hard to give a description of the whole algebra of the phylogenetic variety as a ring of invariants. Moreover, even if some description is possible, the order of the group is large (Corollary 5.6). One interesting observation is that the conjecture implies a description of the ideal as a sum of simpler ideals. In fact we propose a method for obtaining many phylogenetic invariants for any model for the claw tree (Section 6.2). We conjecture that our method gives a description of the whole ideal. We show that in many cases our conjecture is equivalent to the one made by Sturmfels and Sullivant (Proposition 6.8). Our strongest result, Theorem 11.1, proves a weaker, scheme-theoretic version of [SS05, Conjecture 2], sufficient for applications.

2. Toric varieties: the setting

Let us start with an introduction to toric geometry. The study of toric varieties is a relatively new subject. However, its origins can be traced back even to Newton who introduced the idea of representing a polynomial by lattice points. To a monomial in n variables $x_1^{a_1}\cdots x_n^{a_n} =: x^a$ one associates the point $(a_1,\dots,a_n)\in\mathbb{Z}^n$. The following definition will not be used in our article. However, we include it to give a reader not familiar with toric geometry first foundations.

DEFINITION 2.1 (Newton polytope). Let $f=\sum_{a\in\mathbb{N}^n}\alpha_a x^a$ be a polynomial in n variables. The *Newton polytope* of f is the convex hull of points associated to monomials x^a, such that $\alpha_a\neq 0$. The definition can be easily extended to Laurent polynomials.

To find information on Newton polytopes we advise the reader to consult [Stu98]. One of the first papers where toric varieties were studied in a systematic way is [KKMSD73]. The authors call toric varieties "toroidal embeddings" and view them as special compactifications of the algebraic torus $(\mathbb{C}^*)^n$. The classical references for toric varieties are [Oda87] and [Ful93]. The latter book focuses more on the torus action. Recently a new, very modern, user friendly book appeared [CLS11]. The point of view on toric varieties presented there is closest to ours. The reasons why toric varieties have recently become so popular are numerous. A few most important ones are:

(i) toric varieties are strongly related to combinatorial objects, which makes a lot of computations possible or at least easier,
(ii) toric varieties are simple, but fertile enough to provide a good testing ground for conjectures, proofs, theorems and examples,
(iii) toric varieties appear naturally as simplifications of other varieties,
(iv) toric varieties appear in applied mathematics.

This section contains well-known results. We present the proofs, trying to find the easiest and most direct ones. We hope that, with little effort, this section can be read by people not familiar with toric geometry. Details that are skipped can be considered as exercises. We avoid referring to any general theorems, as the theory is, at this level, easy enough to develop from scratch. Many ideas presented in this part come from [CLS11] and [Stu96]. Throughout the paper we will use the setting presented in this section. We encourage the reader familiar with toric geometry to take a look, because our approach is often different from the standard one.

In modern algebraic geometry, a variety is locally described as the spectrum of an algebra. Thus the most important object connected to an affine algebraic variety is its ideal containing all polynomials vanishing on it. Note however that many varieties can be constructed in a different way. Given k polynomials $f_1, \dots, f_k$ in n variables one can consider the map $(f_1, \dots, f_k) : \mathbb{C}^n \to \mathbb{C}^k$. The Zariski closure of the image is an algebraic variety. Furthermore, we can generalize this construction assuming that the f_i are Laurent polynomials. In this case the domain of the map is $(\mathbb{C}^*)^n$.

Let us start the discussion of toric geometry by introducing affine toric varieties. In simplest terms, the study of affine toric varieties is the study of the case where all f_i are monomials.

Definition 2.2 (Affine toric variety). Consider k Laurent monomials in n variables $f_i = x^{a_i}$, where $a_i \in \mathbb{Z}^n$. An *affine toric variety* is the Zariski closure of the image of the map $(f_1, \dots, f_k) : (\mathbb{C}^*)^n \to \mathbb{C}^k$.

Note that we do not require the affine toric variety to be normal. This issue will be addressed later. Moreover, affine toric varieties come with an embedding in the affine space. Recalling Newton's idea, the map (f_i) can be represented by k points $a_i \in \mathbb{Z}^n$. The geometry of these points is strongly related to the geometry of the affine toric variety. We will say that the variety is *associated* to the set of points $\{a_i\}$.

Proposition 2.3. *The ideal of the affine toric variety is generated by binomials. Suppose that the parametrization of the variety is given by k monomials f_i in n variables x_i. Let $P_i \in \mathbb{Z}^k$ be the point associated to f_i. A binomial $y_1^{b_1} \cdots y_k^{b_k} - y_1^{c_1} \cdots y_k^{c_k}$ for $b_i, c_i \in \mathbb{N}$ is in the ideal if and only if $\sum_i b_i P_i = \sum_i c_i P_i$.*

Proof. The binomials of the given form vanish on the image of the map $(f_1, \dots, f_k)$, hence also on its Zariski closure. We will prove that they not only generate the ideal, but span it as a vector space. Fix any order on the monomials. Suppose that the ideal is not spanned by the binomials of the given form. Let g be a polynomial in the variables y_i such that:

- g is in the ideal of the variety,
- g is not spanned by binomials of the given form,
- its leading coefficient is least possible.

Let $\alpha m(y_1, \dots, y_k)$ be the leading coefficient of g, where m is a monomial. As g is in the ideal, by replacing y_i by f_i we get a Laurent polynomial that is zero on $(\mathbb{C}^*)^n$. Hence it has to be zero. In particular the term $\alpha m(f_1, \dots, f_k)$ has to cancel with the term induced by some different monomial $\beta m'(f_1, \dots, f_k)$ appearing in g. Thus the monomials m and m' induce an integer relation between the points P_i. In particular $m - m'$ is a binomial of the chosen form. By subtracting $\alpha(m - m')$ from g we get a polynomial in the ideal with a strictly smaller leading coefficient, which gives a contradiction. ∎

The above proposition allows us to describe the algebra of an affine toric variety.

Definition 2.4 (Semigroup algebra). Let $(C, \oplus)$ be a monoid. As a vector space, the monoid algebra $\mathbb{C}[C]$ is spanned freely by the elements of C. Multiplication is defined as $c_1 c_2 := c_1 \oplus c_2$ for $c_1, c_2 \in C \subset \mathbb{C}[C]$ and extended to $\mathbb{C}[C]$ using the axioms of a $\mathbb{C}$-algebra.

Example 2.5. For the monoid $\mathbb{N}^n$ we obtain the algebra of polynomials in n variables. For the group $\mathbb{Z}^n$ we obtain the algebra of Laurent polynomials.

Corollary 2.6 (from Proposition 2.3). *Consider the affine toric variety parametrized by monomials f_i in n variables. Let $P_i \in \mathbb{Z}^n$ be the point representing f_i. Let C be the monoid generated by the points P_i. The algebra of the affine toric variety is $\mathbb{C}[C]$.* ∎

We will often be working with projective toric varieties.

Definition 2.7 (Projective toric variety). Consider $k+1$ Laurent monomials f_i in n variables. A *projective toric variety* is the Zariski closure of the image of the map $(f_1, \dots, f_{k+1}) : (\mathbb{C}^*)^n \to \mathbb{P}^k$.

If $P \subset \mathbb{Z}^n$ is the set of points representing the monomials f_i, we will say that the closure of the image of (f_i) in $\mathbb{P}^k$ is the projective toric variety *associated to* P, and we will denote it by $\mathbb{P}(X)_P$. We can adapt Proposition 2.3 and Corollary 2.6. First let us consider an affine cone over a projective toric variety. Its parametrization is as follows:

$$(\lambda f_1, \dots, \lambda f_{k+1}) : (\mathbb{C}^*)^{n+1} \to \mathbb{C}^{k+1}.$$

Notice that we have added a nonzero parameter λ, as we passed to the affine space. Of course λf_i is still a monomial. If f_i is represented by a point $P_i \in \mathbb{Z}^n$ then λf_i is represented by $P_i \times \{1\} \in \mathbb{Z}^{n+1}$. Thus in the projective case it is more natural to consider the points P_i in the lattice of dimension one greater, and set the last coordinate to 1. The monoid generated by $P_i \times \{1\}$ gives rise to a monoid algebra of the cone over the projective variety. Moreover, the last coordinate gives the grading of this algebra. The projective toric variety is the *Proj* of this graded algebra. Thus affine toric varieties correspond to finitely generated monoids in $\mathbb{Z}^n$. Projective toric varieties correspond to finitely generated monoids in $\mathbb{Z}^{n+1}$ with generators having last coefficient 1. As an exercise, the reader interested in this topic may extend these results to varieties embedded in weighted projective spaces.

Usually one assumes that a toric variety is normal. Let us explain why. We start by recalling basic definitions.

Definition 2.8 (Normal algebraic variety). An affine algebraic variety is *normal* if its algebra is integrally closed in its field of fractions. An abstract algebraic variety is *normal* if it can be covered by normal affine algebraic varieties.

The concept of normality is important for a number of reasons. Let us recall that smoothness implies normality. Moreover, the singular locus of a normal variety has codimension at least 2. Most toric geometers work with normal varieties, as they have a nice combinatorial description [Oda87, Theorem 1.4].

Definition 2.9 (Lattice). A *lattice* is a finitely generated abelian group with no torsion. In other words, a lattice is an abelian group isomorphic to $\mathbb{Z}^n$.

Consider a subset P of points in a lattice $M \simeq \mathbb{Z}^n$. As in Definition 2.7, the set P defines a projective toric variety $\mathbb{P}(X)_P$ together with an embedding. Let X be the affine cone over $\mathbb{P}(X)_P$. Let C be the monoid generated by the points of $P \times \{1\} \subset M \times \mathbb{Z}$. We know that $X = \operatorname{Spec} \mathbb{C}[C]$. Let $\widetilde{M} \subset M \times \mathbb{Z}$ be the sublattice generated by $P \times \{1\}$.

Definition 2.10 (Projective normality). We call the projective variety $\mathbb{P}(X)$ *projectively normal* if the affine cone X over this variety is normal.

Of course each projectively normal variety is normal. In the toric setting both normality and projective normality can be described in combinatorial language.

Definition 2.11 (Saturated monoid, saturation, saturated set of points). Let C be a monoid contained in a lattice $\widetilde{M}$. We say that C is *saturated* if for any $x \in \widetilde{M}$ and any positive integer k, we have $kx \in C$ if and only if $x \in C$.

For any monoid C one can define its *saturation* $\widetilde{C}$, the smallest saturated monoid containing C. In other words, $x \in \widetilde{C}$ if and only if $kx \in C$ for some positive integer k.

We say that a set of points is *saturated in a lattice M* if it generates a saturated monoid. We say that a set of points is *saturated* if it is saturated in the lattice it generates.

Definition 2.12 (Integral polytope). An *integral polytope* is a convex hull of a finite number of points in the lattice. As we will be dealing only with lattice polytopes, we will often call them just *polytopes*. We will also identify polytopes with the set of their lattice points.

Definition 2.13 (Normal polytope). We say that a polytope $P \subset M$ is *normal in the lattice M* if $P \times \{1\}$ is saturated in $M \times \mathbb{Z}$. We say that a polytope P is *normal* if it is normal in the lattice it generates.

In other words, a polytope P is normal in the lattice M if and only if for any $k \in \mathbb{N}$, any point $Q \in kP \cap M$ is a sum of k points from $P \cap M$.

Note that it is important to specify the lattice. Consider the polytope $P \subset M := \mathbb{Z}^3$. Suppose P has four integral points: $(0, 0, 0)$, $(1, 1, 0)$, $(0, 1, 1)$, $(1, 0, 1)$. This is a normal polytope. Note however that it is not normal in M. Indeed, $(1,1,1) \in 2P$ and $(1,1,1)$ is not the sum of two integral points of the polytope.

Remark 2.14. Some authors distinguish between normal polytopes, which is an intrinsic property of the polytope, and integrally closed polytopes, which in our setting corresponds to being normal in the ambient lattice; for details consult [BDGM15, Section 2].

Note that if the set $P \times \{1\}$ is saturated then P must be a polytope in the lattice it generates. Indeed, let M be the lattice spanned by P. Let $D \in M$ be a linear combination of points from P with positive coefficients summing to 1. By linear algebra, we can assume that the coefficients are rational. Hence some multiple of $D \times \{1\}$ is in the monoid generated by $P \times \{1\}$. As $P \times \{1\}$ is saturated, it must contain $D \times \{1\}$. Thus the convex hull of P intersected with M equals P. Hence P is a polytope.

Fact 2.15. *The variety $\mathbb{P}(X)_P$, defined by a set P of points, is projectively normal if and only if $P \times \{1\}$ is saturated. Equivalently, P must be a normal polytope.*

Fact 2.16. *Let $D \in P \times \{1\}$. Let $P_D = P \times \{1\} - D$, where the minus is the lattice operation. The variety $\mathbb{P}(X)_P$ associated to $P \times \{1\}$ is normal, if and only if for any $D \in P \times \{1\}$ the set P_D is saturated. In that case P need not be normal and is called* very ample.

Proof of Facts 2.15 and 2.16. Both facts are direct consequences of Proposition 2.23. For the first, the algebra of the cone over the variety equals the monoid algebra for the monoid C spanned by $P \times \{1\}$. The monoid C is saturated if and only if P is normal.

For the second, notice that points of $P \times \{1\}$ correspond to variables of the ambient projective space. Consider the affine subvariety of $\mathbb{P}(X)$ corresponding to setting one variable, corresponding to a point D, to 1. The algebra of this affine variety is the monoid algebra associated to the monoid spanned by P_D. ∎

Definition 2.17 (Cone, cone over a polytope). A *cone* is a finitely generated, saturated submonoid of a lattice.

In the literature it is often called a *convex polyhedral cone*. More precisely, we identify lattice points of the polyhedral cone with the cone.

Let P be a polytope that spans the lattice M. The *cone over* P is the saturation of the monoid spanned by $P \times \{1\} \subset M \times \mathbb{Z}$.

We will see in Proposition 2.23 that normal affine toric varieties are associated to finitely generated cones. Projectively normal projective toric varieties are associated to cones over normal polytopes.

There is one important case where even in the projective case one can consider the set P instead of $P \times \{1\}$. Suppose that P is contained in a hyperplane given by an equation $\sum a_i x_i = b$ for $b \neq 0$. In this case the monoid generated by P is isomorphic to the monoid generated by $P \times \{1\}$. In the first part of the article we will be considering such polytopes.

We now explain the name of toric variety. It is connected to the algebraic torus $\mathbb{T} = (\mathbb{C}^*)^n = \operatorname{Spec} \mathbb{C}[x_i^{\pm 1}]$. Under coordinatewise multiplication, $\mathbb{T}$ is an algebraic group. At the level of algebras the action is given by the morphism $\mathbb{C}[x_i^{\pm 1}] \to \mathbb{C}[x_i^{\pm 1}] \otimes \mathbb{C}[x_i^{\pm 1}]$

that associates to a generator x_i the tensor product $x_i \otimes x_i$. Note that an arbitrary Laurent polynomial f is not sent to $f \otimes f$: this is only true for monomials. Let us consider algebraic morphisms $\mathbb{T} \to \mathbb{C}^*$ that preserve the abelian group structure. These are called *characters*. Such a map is in particular a regular function on $\mathbb{T}$, hence must be given by a Laurent polynomial. As it must preserve the group structure, one can prove that it must be a monomial. By identifying a monomial with a lattice point we see that the characters form a lattice $\mathbb{Z}^n$. Intrinsically, one defines the sum of characters f and g by $(f+g)(x) = f(x)g(x)$.

DEFINITION 2.18 (Lattice of characters). The *lattice M of characters* of a torus $\mathbb{T}$ consists of all morphisms $\mathbb{T} \to \mathbb{C}^*$ of algebraic groups with addition defined by $(f+g)(x) = f(x)g(x)$.

Dually one defines one-parameter subgroups as morphisms of algebraic groups $\mathbb{C}^* \to \mathbb{T}$. By projecting onto coordinates we see that each such morphism is of the form $t \mapsto (t^{a_1}, \ldots, t^{a_n})$ for some $a_i \in \mathbb{Z}$. It can be identified with a point $(a_1, \ldots, a_n) \in \mathbb{Z}^n$. Hence the one-parameter subgroups also form a lattice.

DEFINITION 2.19 (Lattice of one-parameter subgroups). The lattice N of *one-parameter subgroups* of a torus $\mathbb{T}$ consists of all morphisms $\mathbb{C}^* \to \mathbb{T}$ of algebraic groups with addition defined by $(\lambda + \delta)(t) = \lambda(t)\delta(t)$.

It is well known that the lattices M and N are dual. The pairing can be described as follows. Fix $f \in M$ and $\lambda \in N$. The composition $f \circ \lambda$ is a morphism of one-dimensional tori. Hence it is of the form $t \mapsto t^a$. We define the product of f and λ to be equal to a. After using the identification of M and N with $\mathbb{Z}^n$ this is the standard scalar product.

As we have seen, the characters correspond exactly to monomials in the algebra of the torus. Hence, $\mathbb{T}$ is the spectrum of the monoid algebra $\mathbb{C}[M]$. Points of $\mathbb{T}$ correspond to maximal ideals of this algebra or to surjective morphisms of algebras $f : \mathbb{C}[M] \to \mathbb{C}$. Of course, to determine such a morphism it is enough to define it on M. As M is a group, its image has to be contained in $\mathbb{C}^*$. Moreover, since f is a map of algebras, the map $M \to \mathbb{C}^*$ must preserve the group structure. Hence the points of $\mathbb{T}$ correspond to maps $M \to \mathbb{C}^*$ that preserve the group structure. More precisely, for a point P we associate to a character χ its value at P.

DEFINITION 2.20 (Abstract toric variety). A *toric variety* X is an algebraic variety, finitely generated over $\mathbb{C}$, containing $\mathbb{T}$ as a dense open subset. Moreover, we require that the action of $\mathbb{T}$ on itself extends to an algebraic action on X.

A crucial fact is that an abstract toric variety that is affine is an affine toric variety in the sense of Definition 2.2. This is usually proved using the following important lemmas.

LEMMA 2.21. *Suppose that a torus $\mathbb{T}$ acts on a vector space V. Then there exists a basis of V such that the action is diagonal.*

Proof. For $t \in \mathbb{T}$ and $v \in V$ we have

$$tv = \sum \chi(t) A_\chi(v),$$

where the sum is over a finite collection of characters χ of $\mathbb{T}$. One can notice that A_χ are projections to subspaces on which $\mathbb{T}$ acts by multiplication by a value of the corresponding character. ∎

Lemma 2.22. *The algebra of an abstract toric variety X that is affine is a monoid algebra associated to a monoid contained in the character lattice of the torus associated to the variety.*

We propose an approach that proves this lemma directly.

Proof. As $\mathbb{T}$ is Zariski dense in X, the algebra A of X is a subalgebra of $\mathbb{C}[M]$. Fix $f \in A$. We know that $f = \sum_{i=1}^{k} a_i\chi_i$ for some $\chi_i \in M$ and $a_i \neq 0$. Let W be the vector space spanned by the characters χ_i for $i = 1, \ldots, k$. Consider the vector subspace $V := A \cap W$. Our first aim is to prove that $V = W$. Suppose that V is contained in a proper vector subspace. Let $(b_1, \ldots, b_k)$ be such that if $\sum_{i=1}^{k} d_i\chi_i \in V$, then $\sum_{i=1}^{k} d_ib_i = 0$. By the assumptions $\mathbb{T}$ acts on X, hence on A. The action of $c \in \mathbb{T}$ on χ_i is given by $\chi_i(c)\chi_i$. Hence the action of c on f gives $\sum_{i=1}^{k} a_i\chi_i(c)\chi_i \in V$. Thus for any $c \in \mathbb{T}$ we must have $\sum_{i=1}^{k} b_ia_i\chi_i(c) = 0$. Hence $\sum_{i=1}^{k} b_ia_i\chi_i$ must be identically zero on $\mathbb{T}$. This is possible only if all b_i are zero, which gives a contradiction.

Hence the algebra A is spanned as a vector space by characters of M. Obviously these characters must form a monoid. ∎

As we have seen, the algebra of an abstract toric variety X that is affine is equal to $\mathbb{C}[C]$ for a monoid $C \subset M$. As the algebra is finitely generated, so is the monoid C. Let $\chi_1, \ldots, \chi_k$ be generators of C. Consider the embedding of the torus acting on X by $(\chi_1, \ldots, \chi_k)$. By Corollary 2.6 its Zariski closure in $\mathbb{C}^k$ is isomorphic to X.

Proposition 2.23. *Let X be an affine toric variety. Let C be a monoid in the character lattice M of the torus acting on X. The variety X is normal if and only if C is a cone.*

Proof. First let us prove that if X is normal then C is saturated. Consider any point $kc \in C$ for $c \in M$. We want to prove that $c \in C$. For $m \in M$ let χ_m be the corresponding character. Consider a polynomial $f(X) = X^k - \chi_{kc}$ with coefficients in the algebra of X. Clearly χ_c satisfies the equation f. Moreover, as C spans M, the character χ_c is in the quotient field of the algebra of X. By the normality of X we know that χ_c is also in the algebra. Hence $c \in C$.

Now we want to prove that if C is saturated, then $\mathbb{C}[C]$ is normal. First note that the quotient field of $\mathbb{C}[C]$ is equal to the quotient field of $\mathbb{C}[M]$. As the torus is smooth, its algebra is normal. One can also prove this by noticing that the algebra is a UFD (being a localization of the polynomial ring). Consider any monic polynomial $f \in \mathbb{C}[C][x]$. Suppose that g is in the quotient field and satisfies $f(g) = 0$. From the normality of $\mathbb{C}[M]$ we know that $g \in \mathbb{C}[M]$. One can repeat the argument of Lemma 2.22. Namely we can act on the equation $f(g)$ by any point P of the torus. The action of P on f gives a monic polynomial with coefficients in $\mathbb{C}[C]$. Hence the action of P on g gives polynomials that are in the normalization of $\mathbb{C}[C]$. By the same arguments as in Lemma 2.22 we see that every character appearing in g with nonzero coefficient must be in the normalization of $\mathbb{C}[C]$. Thus we can assume that $g \in M$.

Suppose that f is of degree d. Notice that $f(g) = 0$ implies that $dg = d'g + c_0$ for some integer $0 \leq d' < d$ and $c_0 \in C$, as the character χ_{dg} must cancel with some other character. Thus $(d - d')g \in C$, and by normality $g \in C$. ∎

It is also worth mentioning how we can recover the torus of an affine toric variety given by a parametrization. There are a few equivalent ways to do this. Note that our construction of an affine or projective variety defines them with an embedding in an affine or projective space *with a distinguished system of coordinates*. These coordinates are in bijection with the points in the lattice that define the variety. The construction also distinguishes a dense torus in the embedding space. It contains all points with nonzero coordinates.

FACT 2.24. *Consider a parametrization $f = (f_1, \ldots, f_k) : \mathbb{T}' := (\mathbb{C}^*)^n \to \mathbb{C}^k$, where f_i are Laurent monomials in n variables. Let X be the Zariski closure of the image of this map. Let $\mathbb{T}'' = (\mathbb{C}^*)^k \subset \mathbb{C}^k$ be the torus containing all points with all coordinates different from zero, with the action given by coordinatewise multiplication. Let M' and M'' be the character lattices respectively of the tori $\mathbb{T}'$ and $\mathbb{T}''$. Then:*

- (i) *At the level of algebras the parametrization map f is induced by a group homomorphism $\tilde{f} : M'' \to M'$.*
- (ii) *The image $\mathbb{T}$ of $\mathbb{T}'$ in $\mathbb{T}''$ is Zariski closed, isomorphic to a torus, with the group action induced from $\mathbb{T}''$.*
- (iii) *The character lattice of $\mathbb{T}$ is equal to the image of $\tilde{f}$ or equivalently to the quotient of M'' by the kernel of $\tilde{f}$.*
- (iv) *The variety X contains $\mathbb{T}$ as a dense open subset and the action of $\mathbb{T}$ extends to X.*

One can identify the torus $\mathbb{T}$ that acts on the projective toric variety $\mathbb{P}(X)_P$. As in the affine case, it is the image of the parametrizing torus. It is also equal to the intersection of $\mathbb{P}(X)_P$ with a torus $\mathbb{T}''$ containing all points of the projective space with all coordinates different from zero. The action of $\mathbb{T}$ is induced from the action of $\mathbb{T}''$ on the projective space. Using the basis, the action is given by coordinatewise multiplication.

We will be often comparing a projective variety with its affine cone. The following discussion concerns the ambient spaces. There is a natural morphism $m : \mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{P}^n$. A system of coordinates determines a torus $\mathbb{T}'$ in $\mathbb{C}^{n+1}$ consisting of the points with all coordinates different from zero. Let M' be the character lattice of $\mathbb{T}'$. Choose a coordinate system on $\mathbb{P}^n$ compatible with the one on $\mathbb{C}^{n+1}$ under the morphism m. The image of $\mathbb{T}'$ is a torus $\mathbb{T}''$ consisting of the points with all coordinates different from zero. Let M'' be the character lattice of $\mathbb{T}''$. Note that $\mathbb{C}^{n+1}$ is a toric variety, with the action of $\mathbb{T}'$ given by coordinatewise multiplication. So is $\mathbb{P}^n$ with the action of $\mathbb{T}''$. Each coordinate of $\mathbb{C}^{n+1}$ is a character of M'. All coordinates determine a basis of M'. The morphism m can be restricted to $\mathbb{T}'$ and can be considered as a morphism of tori, preserving the group action. It induces a map of character lattices $\tilde{m} : M'' \to M'$. As m is a surjective morphism of tori, the morphism $\tilde{m}$ is injective. Hence M'' is a sublattice of M'. Using the basis of M' we can give a precise description of the elements that belong to M''. Namely, an element of M' belongs to M'' if and only if the sum of its coordinates in M' is zero.

Definition 2.25 (Face of a cone). Let C be any cone in a lattice M. Let $v \in M^* = \operatorname{Hom}(M, \mathbb{Z})$. Suppose that for any $c \in C$ we have $v(c) \geq 0$. Let $v^\perp$ be a hyperplane of M consisting of all elements x such that $v(x) = 0$. A *face* of the cone C is any subset of the form $v^\perp \cap C$ for some v satisfying the conditions above. Notice that a face of a cone is also a cone.

Equivalently, a face F of C can be defined as a submonoid satisfying:

- For any $c_1, c_2 \in C$ such that $c_1 + c_2 \in F$ we have $c_1, c_2 \in F$.

For an affine toric variety corresponding to a cone C, the faces of C correspond to orbits of the torus acting on it. Let us present this correspondence in detail. We fix a finitely generated monoid C in a lattice M, and its generators $\chi_1, \dots, \chi_k \in C$. As in Definition 2.2 the closure of the embedding of the torus $\operatorname{Spec} \mathbb{C}[M]$ in $\mathbb{C}^k$ by means of the characters χ_i is the affine toric variety $X := \operatorname{Spec} \mathbb{C}[C]$. Note that we distinguished a basis in $\mathbb{C}^k$, but not on the torus $\mathbb{C}[M]$. From Fact 2.24 we know that:

- the dense torus orbit of X contains precisely those points that have all coordinates different from zero,
- the character lattice of the torus acting on X is equal to the sublattice of M spanned by C.

We will generalize this to other orbits. Assume that C is a cone. Each orbit will be indexed by a face F of C. The face F determines a subset $I \subset \{1, \dots, k\}$ such that $i \in I$ if and only if $\chi_i \in F$. The orbit corresponding to F can be characterized as follows:

(1) the orbit contains precisely those points that have coordinates corresponding to $i \in I$ different from zero and all the other zero,
(2) the orbit is a torus with character lattice spanned by elements of F,
(3) the closure of the orbit is a toric variety given by the cone F,
(4) each point of the orbit is a projection of the dense torus orbit onto the subspace spanned by basis elements indexed by indices from I,
(5) the inclusion of the orbit in the variety is given by a morphism of algebras $\mathbb{C}[C] \to \mathbb{C}[F]$. This morphism is an identity on $F \subset \mathbb{C}[C]$ and zero on $C \setminus F$.

Note that each orbit will contain a unique distinguished point given by the projection of the point $(1, \dots, 1) \in \mathbb{C}^k$. We will only sketch the proof of these observations.

Proof. As in the case of the torus, we can identify points of X with monoid morphisms $C \to (\mathbb{C}, \cdot)$. Fix any $x \in X$. The characters $\chi \in C$ such that $\chi(x) \neq 0$ must form a face of F. Hence x determines a subset $I \subset \{1, \dots, k\}$. Of course the set of points in X with nonzero coordinates indexed by I and other coordinates zero is invariant under the action of the torus on X. So to prove (1) it is enough to prove that all these points are in one orbit. The point x represents a morphism $C \to (\mathbb{C}, \cdot)$ that is nonzero on F and zero on $C \setminus F$. Consider the restriction of this morphism to F. As it is nonzero, it can be extended to a morphism $M' \to \mathbb{C}^*$, where M' is the sublattice generated by F. Next we can extend this morphism to the lattice M'' generated by C. Thus we obtain a morphism $f : M'' \to \mathbb{C}^*$ that agrees with the one representing x on F. Note that f represents a point p in the torus acting on X. By the action of p^{-1} on x we obtain a point given by a morphism that associates 1 to elements from F and zero to elements from $C \setminus F$. Thus we have proved (1). Moreover, we showed that each orbit contains the distinguished point. Point (2) follows, as morphisms that are nonzero on F and zero on $C \setminus F$ are identified with morphisms from M' to $\mathbb{C}^*$. Point (3) is a consequence of (2) and the previous discussion on affine toric varieties. Indeed, we already know that the orbit is a torus with the lattice generated by F. This torus is the image of the torus $\operatorname{Spec} \mathbb{C}[M]$ in $\mathbb{C}^k$ by means of the characters from I and all other coordinates equal to zero. Let A be the affine space spanned by basis elements indexed by indices in I. The orbit corresponding to F is contained in A. In fact, by the construction it is the image of $\operatorname{Spec} \mathbb{C}[M]$ by the characters χ_i such that $i \in I$. The closure of this torus is exactly given by $\operatorname{Spec} \mathbb{C}[F]$, as generators of the monoid C contained in F are generators of F. Point (4) is obvious, as the point p constructed in the first part of the proof projects to x. ∎

We finish this section by stating some results about normal abstract toric varieties.

Definition 2.26 (Fan). A *fan* Σ is a finite collection of cones in a lattice that satisfy the following conditions:

(1) if a cone C is in the fan then all its faces are also in the fan,
(2) the intersection of any two cones from the fan is a face of both,
(3) for any cone $C \in \Sigma$, if $x \in C$ then $-x \notin C$.

A general, normal toric variety can be represented by a fan in the one-parameter subgroup lattice N.

Definition 2.27 (Dual cone). Let L and L' be dual lattices with the pairing given by $(\cdot,\cdot)$. Let $\delta \subset L$ be a cone in L. We define the *dual cone* $\delta^* \subset L'$ as

$$\delta^* = \{x \in L' : (x, y) \geq 0 \text{ for any } y \in \delta\}.$$

A toric variety X is constructed from a fan Σ by gluing together the affine schemes $\operatorname{Spec}(\mathbb{C}[\sigma_i^*])$, where $\sigma_i^* \subset M$ is a cone dual to $\sigma_i \in \Sigma$. One-dimensional cones in Σ are called *rays*. The generators of these monoids are called *ray generators*.

Many properties of the variety X can be described using the fan Σ. For example X is smooth if and only if for every cone σ_i the set of its ray generators can be extended to a basis of N. Moreover, to each ray generator v we may associate a unique T-invariant Weil divisor denoted by D_v. For fans containing maximal dimensional cones there is a well-known exact sequence

$$0 \to M \to \operatorname{Div}_T \to \operatorname{Cl}(X) \to 0, \tag{2.1}$$

where Div_T is the group of T-invariant Weil divisors and $\operatorname{Cl}(X)$ is the class group. The map $M \to \operatorname{Div}_T$ is given by

$$m \to \sum m(v_i) D_{v_i},$$

where the sum is taken over all ray generators v_i.

So far we have defined objects of the category of toric varieties. Not every algebraic morphism is a morphism in this category. Indeed, as toric varieties are endowed with the torus action, it is natural to distinguish those morphisms that respect this action.

Definition 2.28 (Toric morphism). Let $f : X \to Y$ be a morphism of toric varieties. Let $\mathbb{T}_X \subset X$ and $\mathbb{T}_Y \subset Y$ be the tori acting respectively on X and Y. We call f a *toric morphism* if $f(\mathbb{T}_X) \subset \mathbb{T}_Y$ and for any points $p, q \in \mathbb{T}_X$ we have

$$f(pq) = f(p)f(q).$$

Notice that, as the tori are Zariski dense in the varieties, this immediately implies that for any $p \in \mathbb{T}_X$ and $q \in X$ the same equality holds.

As the restriction of a toric morphism is a morphism of algebraic tori, it induces a map $\tilde{f} : M_Y \to M_X$ of character lattices. By dualizing, this gives a map $\tilde{f}^* : N_X \to N_Y$ of one-parameter subgroups. In fact, one can easily characterize which morphisms of one-parameter subgroups give rise to toric morphisms. For each cone δ in the fan representing X there must be a cone δ' in the fan representing Y such that $\tilde{f}^*(\delta) \subset \delta'$.

Much more information on this topic can be found in [CLS11], [Ful93].

3. Basic definitions

This section introduces objects that will be studied in the first part of the article. Subsection 3.1 is the most important. Other parts can be treated as motivations and examples.

We will be dealing with algebraic varieties associated to phylogenetic models. These varieties are always given as the closure of the image of a parametrization map—details will be presented in Section 3.1. A short algebraic introduction to the topic can be found in [ERSS05].

Let S be a finite set, called the *set of states*. In the biological setting, S is often supposed to have four elements. These elements correspond to four nucleobases. The set S is the codomain of random variables in the Markov process. Let $\Delta \subset \mathbb{R}^{|S|}$ be the probabilistic simplex that contains all the points with nonnegative coordinates summing to 1. The points of Δ parametrize all possible distributions of random variables with the set of states equal to S. In algebraic geometry, instead of considering the simplex Δ one considers the whole complex vector space $\mathbb{C}^{|S|}$.

Definition 3.1 (Space W). We define W to be the complex vector space spanned freely by elements of S. More precisely, $W = \bigoplus_{a \in S} \mathbb{C}_a$, where $\mathbb{C}_a$ is a field of complex numbers corresponding to the one-dimensional vector space spanned by $a \in S$.

Suppose that we are given a rooted tree T with edges directed from the root.

Definition 3.2 (Sets L, V, N and E). Let L, V, N and E be respectively the set of leaves, vertices, nodes and edges of the tree T. We have $V = L \cup N$ and $L \cap N = \emptyset$. We identify leaves with edges adjacent to them.

The objects we study are derived from Markov processes on a tree. To each vertex one can associate a random variable with the set of states equal to S. The Markov property ensures that the variable at a vertex depends only on the variable associated to its first ancestor. Formally, let X_i be a variable associated to a vertex v_i. Suppose that there is an edge directed from v_1 to v_2. Consider any set of vertices $v_3, \dots, v_j$ that are not descendants of v_2. Then $P(X_2 = x_2 \mid X_1 = x_1, X_3 = x_3, \dots, X_j = x_j) = P(X_2 = x_2 \mid X_1 = x_1)$, where $x_i \in S$. This mathematical model is applied for example in phylogenetics. The nodes of the tree correspond to species, and the Markov property describes the fact that evolutionary changes depend only on the direct ancestor. More information on Markov processes can be found for example in [Ibe09]. The reader interested in phylogenetics is advised to consult [PS05]. There one can also find a detailed explanation of the relationship between Markov processes on trees and models that we consider.

To define a model, we need to distinguish a subspace $\widehat{W} \subseteq \mathrm{End}(W)$.

Definition 3.3 (Transition matrix). Any element of the space $\widehat{W}$ represented as a matrix in the basis corresponding to S is called a *transition matrix*.

The entries of a transition matrix correspond in biology to probabilities of mutation. Most often, a model is distinguished by specifying the type of transition matrices.

Let us present some of the models.

(i) **The Cavender–Farris–Neyman model**, also called the **2-state Jukes–Cantor model** ([1]). This is the simplest model. It was first introduced in [Ney71]. In most biological articles it is called the Cavender–Farris–Neyman model or just the Neyman model. However, recently, especially in algebraic phylogenetics, it is called the 2-state Jukes–Cantor model or the binary model [SS05], [BW07], [ERSS05]. In this model, S has two elements and the transition matrices are of the type

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

This model has a lot of nice properties. One of the most interesting is that the algebraic varieties arising from trivalent trees with the same number of leaves are deformation equivalent—see [BW07] for the original, algebraic proof, and [Ilt10] for a combinatorial one. It is a general group-based model for the group $G = \mathbb{Z}_2$; the definition of general group-based models will be introduced in Subsection 4.1.

(ii) **3-Kimura model.** This is a four-state model introduced in [Kim81]. It is a general group-based model for the natural action of the group $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ on the nucleobases A, C, G, T [ES93]. The transition matrices are of the type

$$\begin{bmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{bmatrix}.$$

(iii) **2-Kimura model.** This is a model for four states introduced in [Kim80]. The transition matrices are of the type

$$\begin{bmatrix} a & b & c & b \\ b & a & b & c \\ c & b & a & b \\ b & c & b & a \end{bmatrix}.$$

([1]) We would like to thank Elizabeth Allman for the information on the ambiguity.

(iv) **Jukes–Cantor model.** This is the simplest model for four states, introduced in [JC69]. The transition matrices are of the type

$$\begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}.$$

(v) **General Markov model.** This model can be considered on any number of states, but for biological reasons it is typically considered for four states. The space $\widehat{W}$ is equal to the whole space of endomorphisms, $\operatorname{End} W$. Hence for four states the transition matrices are arbitrary:

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix}.$$

3.1. A variety associated to a model. We will associate an algebraic variety to a tree T and a space $\widehat{W} \subset \operatorname{End} W$. This is a standard construction. In the literature one can find a lot of generalizations of the approach presented here—see for example [DK09].

Definition 3.4 (Spaces W_v and $\widehat{W}_e$). To each vertex v of the tree we attach a complex vector space W_v with a fixed isomorphism $\mathrm{iso}_v : W \simeq W_v$. The images of the basis elements of W corresponding to states in S by iso_v give a basis of W_v. The elements of this basis are denoted by $\{\alpha_v\}$. We also consider a vector space $\widehat{W} \subset \operatorname{End}(W)$, determined by the model we choose. To each edge e of the given rooted tree T we associate a vector space $\widehat{W}_e$ isomorphic to $\widehat{W}$.

Remark 3.5. The natural basis of W induces an isomorphism $W \cong W^*$. Hence $\operatorname{End}(W) \cong W^* \otimes W \cong W \otimes W$. We may regard $\widehat{W}$ and each $\widehat{W}_e$ as subspaces of $W \otimes W$.

Definition 3.6 (Spaces W_V, $\widehat{W}_E$, W_L). We recall that V, L and E are respectively the set of vertices, leaves and edges of a tree. We define the following three spaces:

$$W_V = \bigotimes_{v \in V} W_v, \quad W_L = \bigotimes_{l \in L} W_l, \quad \widehat{W}_E = \bigoplus_{e \in E} \widehat{W}_e.$$

We call W_V the *space of all possible states* of the tree, W_L the *space of states of leaves* and $\widehat{W}_E$ the *parameter space*.

Definition 3.7 (The map $\widehat{\psi}$ [BW07, Construction 1.5]). Let $\widehat{\psi} : \widehat{W}_E \to W_V$ be a map whose dual is defined as

$$\widehat{\psi}^* \Big(\bigotimes_{v \in V} \alpha_v^* \Big) = \bigotimes_{e \in E} (\alpha_{v_1(e)} \otimes \alpha_{v_2(e)})^*_{|\widehat{W}_e}.$$

Here the edge e is directed from $v_1(e)$ to $v_2(e)$.

$\widehat{\psi}$ is just a map well known to biologists that to a given choice of matrices associates the probability distribution on the set of all possible states of vertices of the tree.

Example 3.8. Let us consider the binary Jukes–Cantor model. Fix the tree with one root r and two leaves l_1 and l_2. The spaces W and $\widehat{W}$ are two-dimensional. Hence the spaces W_V and $\widehat{W}_E$ are respectively 8- and 4-dimensional. The basis elements of W_V correspond to states of the variables associated to nodes of trees. Hence they can be indexed by triples (p, q, s) for $p, q, s = 0, 1$. Assume that the first element of the triple is associated to the state of r. The elements of $\widehat{W}$ are matrices of the type

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

Fix a simple tensor in $\widehat{W}_E$ represented by a pair of such matrices

$$\begin{bmatrix} a_1 & b_1 \\ b_1 & a_1 \end{bmatrix}, \quad \begin{bmatrix} a_2 & b_2 \\ b_2 & a_2 \end{bmatrix}.$$

To this element the morphism $\widehat{\psi}$ associates an element of W_V given as

$$a_1a_2(0,0,0) + a_1a_2(1,1,1) + a_1b_2(0,0,1) + a_1b_2(1,1,0) \\ + b_1a_2(0,1,0) + b_1a_2(1,0,1) + b_1b_2(0,1,1) + b_1b_2(1,0,0).$$

Thus $\widehat{\psi}$ associates to a given choice of transition matrices the "probability distribution" on the set of all possible states of the tree. This is up to a scalar, as we assume that the root has uniform distribution. Moreover, as we work over complex numbers and there are no probabilistic restrictions on elements of $\widehat{W}$, the map $\widehat{\psi}$ is obtained by the rule for Markov processes, but in general the elements of the image have no probabilistic meaning.

Recall $N = V \setminus L$ is the set of nodes of a tree. We consider the map $\delta = \sum \alpha_i^* \in W^*$ that sums all the coordinates.

Definition 3.9 (π). Let $\pi : W_V \to W_L$ be defined as $\pi = (\bigotimes_{v \in L} \mathrm{id}_{W_v}) \otimes (\bigotimes_{v \in N} \delta_{W_v})$. The map π sums the probabilities of all the states of vertices that differ only at nodes.

If we compose $\widehat{\psi}$ with π we obtain a map from $\widehat{W}_E$ to W_L. This induces a rational map

$$\check{\psi} : \prod_{e \in E} \mathbb{P}(\widehat{W}_e) \dashrightarrow \mathbb{P}(W_L).$$

The closure of the image of this map is denoted by $\mathbb{P}(X(T, W, \widehat{W}))$. This is the algebraic projective variety associated to the model that is the main object of our study. We will also consider the affine model $X(T, W, \widehat{W})$ that is the affine cone over this variety.

4. Group-based models

Some parts of this section were published separately in [Mic11]. Our aim is to investigate the properties of certain models. The space of transition matrices will be given as a subspace invariant under a group action. We will see under what conditions we obtain a toric variety. We will also study the properties of the varieties so obtained, and their connections with trees and groups. We have to point out that in this section we do not assume that a toric variety has to be normal. We only assume that a torus acts on a variety and one of the orbits is dense. This setting is most common when dealing with applications. Much information can be found in [Stu96]. The main drawback of this approach is that the varieties we consider will not be given by a fan. However, they can still be represented by polytopes, which do not have to be normal. For this reason we will often work with the character lattice M instead of the one-parameter subgroup lattice N.

We will be defining objects that will depend on a tree T and a group G. For any object O, if we want to stress its dependence on either T or G we indicate them with indices, O_G^T. For vector spaces on which a group G acts we use the standard notation for the subspace of invariants, by putting G in the superscript.

4.1. General group-based models. In our study we are mainly interested in specific models. We set the notation for general group-based models. We generalize the notions of "sockets" and "networks" introduced in [BW07]. This enables us to extend some of the results from $\mathbb{Z}_2$ to arbitrary abelian groups. We believe that these notions give a nice, unified description of the variety associated to the model.

The inspiration for this section comes from the work [ES93] of Evans and Speed who recognized a natural action of an abelian group G on S in a biological case. Namely, the group $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ acts on $\{A, C, G, T\}$ transitively and freely. Hence from now on we assume that we have a *transitive and free action of an abelian group G on S*. In such a situation, S is often called a *G-torsor*. The action of G on S extends naturally to the action of G on W. The fact that general group-based models give toric varieties was already observed in [ES93], [SSE93].

Definition 4.1 (A_g). For $g \in G$ let A_g be the transition matrix (equivalently the linear map) corresponding to the action of g on W.

By choosing one element of S and associating to it the neutral element of G we obtain an action-preserving bijection between the elements of S and G. The element associated to $a \in S$ will be denoted by g_a. Canonically the rows and columns of the transition matrix are labeled by elements of S. After fixing the bijection we can also label them with group elements, but this is not canonical. The choice of the bijection allows us also to find another basis of W, indexed by characters of G. This is done by the discrete Fourier transform.

Definition 4.2 (w_χ). Let $\chi \in G^*$ be any character of the group G. We define a vector $w_\chi \in W$ by

$$w_\chi = \sum_{a \in S} \chi(g_a)a.$$

By orthogonality of characters, the elements w_χ form a basis of W. Notice that although the choice of the bijection between S and G is not canonical, the one-dimensional spaces spanned by w_χ are. Changing the bijection just multiplies each vector w_χ by $\chi(g)$ for some $g \in G$. In the language of representation theory, W is the regular representation of G. The one-dimensional spaces spanned by w_χ are of course the unique irreducible one-dimensional representations corresponding to characters of G.

The group structure also naturally determines a specific model, a vector space $\widehat{W}$. This is done as follows. We have a natural action of G on $W \otimes W$—the action of g is just $g \otimes g$:

$$g\Big(\sum \lambda a_1 \otimes a_2\Big) = \sum \lambda g(a_1) \otimes g(a_2).$$

Definition 4.3 ($\widehat{W}$). Let G be an abelian group acting on the set S transitively and freely. By Remark 3.5 we identify $\mathrm{End}(W)$ with $W \otimes W$. For a general group-based model we define $\widehat{W}$ as the set of fixed points of the G action on $\mathrm{End}(W) \cong W \otimes W$.

Remark 4.4. In other words, we only take transition matrices that satisfy the following condition for any $g \in G$: If we permute the columns and rows of a matrix with a permutation corresponding to g, then we obtain the same matrix.

Hence the parameters in the transition matrices depend only on the difference of the group elements labelling the row and column of a given entry. In particular, the dimension of $\widehat{W}$ is equal to $|G|$.

In general we assume that the tree is rooted and directed away from the root. However, the construction from Subsection 3.1 can be easily generalized to other orientations of the edges of the tree. We make this assumption because it simplifies the formulations.

Remark 4.5. One can see that if $A \in \widehat{W}$, then $A^T \in \widehat{W}$. This means that if we consider a tree T with two different orientations, then the associated varieties are exactly the same. If a point is the image of some element of the parameter space with respect to a given orientation, then it is also the image of an element of the parameter space with respect to the other orientation. We just have to transpose matrices that are associated to edges with different orientation.

The following elements are invariant with respect to the G action, hence belong to $\widehat{W}$.

Definition 4.6 (Elements $l_\chi \in \widehat{W}$). Let χ be a character of G. We define

$$l_\chi(w_{\chi'}) := \begin{cases} w_\chi, & \chi = \chi', \\ 0, & \chi \neq \chi'. \end{cases}$$

It follows that $(l_\chi)_{\chi \in G^*}$ is a base of $\widehat{W}$. Moreover, $\widehat{W}$ is equal to the space of diagonal matrices in the basis $(w_\chi)_{\chi \in G^*}$. The following proposition gives the description of l_χ in terms of the basis associated to elements of S. We omit the proof, as it relies on basic computations.

Proposition 4.7.

$$l_\chi(a_0) = \frac{1}{|G|}\chi(g_{a_0}^{-1})w_\chi = \frac{1}{|G|}\sum_{a \in S}\chi(g_{a_0}^{-1}g_a)a.$$

The vectors l_χ are independent of the choice of the bijection between S and G. The element $g_{a_0}^{-1}g_a$ is a unique element of G that sends a_0 to a, hence does not depend on the bijection. The map l_χ is a projection onto the (canonical) one-dimensional subspace spanned by w_χ.

Using this basis we will see that the map $\widehat{\psi}$ is injective. Hence the induced algebraic map $\prod_{e \in E} \mathbb{P}(W_e) \to \mathbb{P}(W_V)$ is given by the full Segre system. The algebraic map $\pi \circ \widehat{\psi}$

will be given by a subsystem of the Segre system. We will describe it using the notions of "sockets" and "networks". Let us start with a few lemmas. The action of G on W extends to an action on W_V and W_L.

LEMMA 4.8. *The dimensions of G-fixed subspaces of W_V and W_L are as follows:*

$$\dim W_V^G = |G|^{|V|-1}, \qquad \dim W_L^G = |G|^{|L|-1}.$$

Proof. Let us consider the basis of W_V given by $(\bigotimes_{v\in V} w_{\chi_v})$. The action of g in this basis is diagonal, so the space of fixed vectors is spanned by fixed elements of this basis. As $g(w_\chi) = \chi(g^{-1})w_\chi$, we obtain

$$g\Big(\bigotimes_{v\in V} w_{\chi_v}\Big) = \bigotimes_{v\in V} \chi_v(g^{-1})w_{\chi_v} = \prod_{v\in W} \chi_v(g^{-1}) \bigotimes_{v\in V} w_{\chi_v},$$

so an element $\bigotimes_{v\in V} w_{\chi_v}$ is fixed if and only if for any $g \in G$ we have $\prod_{v\in V} \chi_v(g) = 1$. This is equivalent to $\sum_{v\in V} \chi_v$ being the trivial character (we use additive notation for the group G^* of characters). From this we see that $\dim W_V^G$ is equal to the number of sequences, indexed by vertices of the tree, of characters that sum to the neutral character. This gives us $|G^*|^{|V|-1}$ sequences and proves the first equality, as for abelian groups $|G^*| = |G|$. The proof of the second equality is the same. ∎

REMARK 4.9. The basis $\{\bigotimes_{v\in V} w_{\chi_v}\}$ of W_V depends on the choice of the bijection between S and G. However, the basis $\{\bigotimes_{v\in V} w_{\chi_v} : \sum_{v\in V} \chi_v = \chi_0\}$ of W_V^G is natural. Changing the bijection multiplies w_χ by $\chi(g)$ for a fixed $g \in G$. As $\sum_{v\in V} \chi_v = \chi_0$, we have $(\sum_{v\in V} \chi_v)(g) = 1$, and the vectors remain unchanged.

One can easily see that the image of $\widehat{W}_E$ in W_V is invariant with respect to the action of G.

PROPOSITION 4.10. *The map $\widehat{\psi}$ is a vector space isomorphism of $\widehat{W}_E$ and W_V^G. It takes the base $\{\bigotimes_{e\in E} |G| l_{\chi_e}\}$ bijectively onto the base $\{\bigotimes_{v\in V} w_{\chi_v} : \sum_{v\in V} \chi_v = \chi_0\}$, where χ_0 is the trivial character.*

Proof. Using Proposition 4.7 we can see that

$$\Big(\bigotimes_{v\in V} a_v\Big)^* \Big(\widehat{\psi}\Big(\bigotimes_{e\in E} |G| l_{\chi_e}\Big)\Big) = \prod_{e=(v_1,v_2)\in E} (-\chi_e)(g_{a_{v_1}})\chi_e(g_{a_{v_2}}).$$

For given characters χ_e, define characters χ_v for all v vertices of the tree by

$$\chi_v := \sum_{(v,w)\in E} \chi_{(v,w)} - \sum_{(w,v)\in E} \chi_{(w,v)}.$$

This corresponds to summing all characters on edges adjacent to v with appropriate signs, depending on the orientation of the edge. We consider an element $\bigotimes_{v\in V} w_{\chi_v}$, which is clearly in the chosen basis of W_V^G as each character χ_e is taken twice with different signs, so the sum of all χ_v is a trivial character. Moreover,

$$\bigotimes_{v\in V} w_{\chi_v} = \bigotimes_{v\in V} \Big(\sum_{a\in S} \chi_v(g_a)a\Big),$$

so $(\bigotimes_{v\in V} a_v)^*(\bigotimes_{v\in V} w_{\chi_v}) = \prod_{v\in V} \chi_v(g_{a_v})$, which proves the theorem. ∎

Corollary 4.11. *The morphism*

$$\psi : \prod_{e\in E} \mathbb{P}(\widehat{W}_e) \to \mathbb{P}(W_V^G)$$

is given by a full Segre system. In the basis from Proposition 4.10 *it is given by monomials.*

Our aim will be to obtain a result similar to Proposition 4.10 for the map $\pi\circ\widehat{\psi}$. Notice that apart from the action of G on $W\otimes W$ given by $g\otimes g$ that allowed us to define $\widehat{W}$, we have another action of G on $W\otimes W$ given by $g\otimes \mathrm{id}$, where id is the identity map.

Lemma 4.12. *The action* $g\otimes \mathrm{id}$ *restricts to* $\widehat{W}$.

Proof. It is enough to prove that the image of the action of $g\otimes \mathrm{id}$ on any element that is fixed for the action $g'\otimes g'$ is also fixed. Let $C\in\widehat{W}$. Then

$$\begin{aligned}(g'\otimes g')(g\otimes \mathrm{id})C &= (g'g\otimes g')(C) = (gg'\otimes g')(C) = (g\otimes \mathrm{id})(g'\otimes g')(C)\\ &= (g\otimes \mathrm{id})(C).\end{aligned}$$

Here we have used the fact that G is abelian. ∎

Definition 4.13 (The group G_N). For each $v\in N$, $g\in G$ and $e\in E$ we define an isomorphism $\rho^g_{v,e}$ of the space $\widehat{W}_e$. The action on $\widehat{W}_e$ depends on e and v. If e is not adjacent to v, it is the identity. If e is an outgoing edge from v, it is equal to $g\otimes \mathrm{id}$, and if e is an incoming edge, it is equal to $g^{-1}\otimes \mathrm{id}$.

For each $v\in N$ and $g\in G$ we define an isomorphism of $\widehat{W}_E$ by $\rho^g_v := \bigotimes_{e\in E}\rho^g_{v,e}$. We also define $G_N\subset \mathrm{End}(\widehat{W}_E)$ to be the group generated by all ρ^g_v.

Remark 4.14. It is crucial to realize how $g\otimes \mathrm{id}$ acts on elements of $\widehat{W}$ considered as morphisms. One can check that $g\otimes \mathrm{id}(A_{g'}) = A_{g'}\circ A_{g^{-1}}$, so the action of $g\otimes\mathrm{id}$ composes a given morphism with $A_{g^{-1}}$.

To obtain a nice description of the morphism $\pi\circ\widehat{\psi}$ we need a technical lemma.

Lemma 4.15. *We have* $G_N\cong G^{|N|}$. *There is a base in which* G_N *acts diagonally on* $\widehat{W}_E$.

Proof. Using 4.14 we obtain

$$\begin{aligned}(g\otimes \mathrm{id}(l_\chi))(w_{\chi'}) &= l_\chi A_{g^{-1}}(w_{\chi'}) = l_\chi A_{g^{-1}}\Big(\sum_{a\in A}\chi'(g_a)a\Big) = l_\chi\Big(\sum_{a\in S}\chi'(g_a)g^{-1}a\Big)\\ &= l_\chi\Big(\sum_{a\in S}\chi'(g_a g)a\Big) = \chi'(g)l_\chi(w_{\chi'}) = \chi(g)l_\chi(w_{\chi'}),\end{aligned}$$

where the last equality follows from the fact that $l_\chi(w_{\chi'})$ is nonzero only if $\chi=\chi'$. This proves that $g\otimes\mathrm{id}(l_\chi)=\chi(g)l_\chi$, which proves the theorem. ∎

Let F be any abelian group. In our examples $F=G$ or $F=G^*$. Consider the groups F^E and F^N. The elements of each are assignments of group elements respectively to edges and to nodes of the tree.

Definition 4.16 (Adding morphism *add*, projection p_v). We define a morphism add : $F^E\to F^N$. Let $m\in F^E$ and $p_v : F^N\to F$ be the projection onto the component indexed by the vertex $v\in N$. The element $p_v(add(m))$ is equal to the sum of the group elements associated by m to the edges incoming into v minus the sum of the group elements associated to the edges outgoing from v.

EXAMPLE 4.17. Consider $F = \mathbb{Z}_3$. Let T be a claw tree with three edges. We have

$$add : (\mathbb{Z}_3)^3 \to \mathbb{Z}_3,$$

where add is the usual sum in $\mathbb{Z}_3$.

DEFINITION 4.18 (Trivial signed sum). We say that an element $m \in F^E$ has *trivial signed sum* around a vertex v if $p_v(add(m))$ is the neutral element of F.

DEFINITION 4.19 (Map add'). We define a map $add' : F^L \to F$. This map sends a function to the sum of its values.

REMARK 4.20. As in Proposition 4.10, elements of the base of $\widehat{W_E}$ are in bijection with the sequences of characters indexed by edges of a tree. In other words, an element of the basis of $\widehat{W_E}$ can be described as assigning a character of G to each edge of a tree. Moreover, the elements of the basis of $\widehat{W_E}$ that are fixed under the action of G_N are exactly the assignments such that the signed sum of the characters around each inner vertex is the trivial character.

LEMMA 4.21. *The map $\pi : W_V \to W_L$ can be described as follows:*

$$\pi\Big(\bigotimes_{v\in V} w_{\chi_v}\Big) = |G|^{|N|} \bigotimes_{l\in L} w_{\chi_l}$$

if all the characters χ_v for the inner vertices are trivial, and zero otherwise.

Proof. First let us look at $\bigotimes_{v\in V} w_{\chi_v}$ in the old coordinates:

$$\bigotimes_{v\in V} w_{\chi_v} = \bigotimes_{v\in V}\Big(\sum_{a\in S} \chi_v(g_a)a\Big) = \sum_{(a_u)_{u\in V}\in S^V} \Big(\prod_{v\in V} \chi_v(g_{a_v})\Big)\Big(\bigotimes_{v\in V} a_v\Big),$$

where the sum $\sum_{(a_u)_{u\in V}\in S^V}$ is taken over all $|V|$-tuples (indexed by vertices) of basis vectors. In other words, this sum parametrizes the basis of W_V made of tensor products of base vectors corresponding to elements of G. This is equal to

$$\sum_{(a_u)_{u\in N}\in S^N} \sum_{(a_l)_{l\in L}\in S^L} \prod_{v\in N} \chi_v(g_{a_v}) \prod_{f\in L} \chi_f(g_{a_f}) \bigotimes_{v\in N} a_v \bigotimes_{f\in L} a_f.$$

We see that $\pi(\bigotimes_{v\in V} w_{\chi_v})$ is equal to

$$\begin{aligned}\sum_{(a_u)_{u\in N}\in S^N} \sum_{(a_l)_{l\in L}\in S^L} \prod_{v\in N} \chi_v(g_{a_v}) \prod_{f\in L} \chi_f(g_{a_f}) \bigotimes_{f\in L} a_f \\ = \Big(\prod_{v\in N}\Big(\sum_{g\in G} \chi_v(g)\Big)\Big) \sum_{(g_l)_{l\in L}\in G^N} \prod_{f\in L} \chi_f(g_l) \bigotimes_{f\in L} a_f.\end{aligned}$$

The product $\prod_{u\in N}(\sum_{g\in G} \chi_u(g))$ is zero unless all characters χ_u for $u \in N$ are trivial. In the latter case the product is equal to $|G|^{|N|}$. Of course

$$\sum_{(g_l)_{l\in L}\in G^N} \Big(\prod_{f\in L} \chi_f(g_l)\Big)\Big(\bigotimes_{l\in L} g_l\Big) = \bigotimes_{l\in L} w_{\chi_l},$$

which proves the proposition. ■

The following theorem is a direct generalization to arbitrary abelian groups of Theorem 2.12 from [BW07].

Theorem 4.22. *The spaces W_L^G and $(\widehat{W_E})^{G_N}$ are isomorphic.*

Proof. One can prove this using a dimension argument, but it is better to look how the bases are transformed. The base of $(\widehat{W_E})^{G_N}$ is given by $\bigotimes_{e\in E}|G|l_{\chi_e}$, where the signed sum of all characters at any vertex is trivial. This, thanks to Proposition 4.10, is transformed bijectively by the morphism $\widehat{\psi}:\widehat{W_E}\to W_V$ onto an independent set $\bigotimes_{v\in V}w_{\chi_v}$, where characters for inner vertices are trivial and the sum of all characters is trivial. By Lemma 4.21 the image of this set under π gives the set $|G|^{|N|}\bigotimes_{l\in L}w_{\chi_l}$, where the characters χ_l sum to the trivial character. The last set forms a base of W_L^G. ■

Corollary 4.23. *The morphism $\pi\circ\widehat{\psi}$ is a toric morphism.*

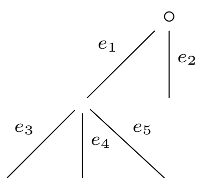
Proof. Follows from the proof of Theorem 4.22. ■

Our aim is to describe the monomials that define $\pi\circ\widehat{\psi}$. This motivates the following definitions of groups of sockets and networks.

Definition 4.24 (Groups $\mathfrak{S}$ and $\mathfrak{N}$). We fix an abelian group $F=G^*$. The *group of networks* $\mathfrak{N}$ is the kernel of the morphism *add*. The *group of sockets* $\mathfrak{S}$ is the kernel of the morphism *add′*.

Hence a socket is an assignment of characters from G^* to each leaf in such a way that the sum of all these characters is the trivial character. A network is an assignment of characters from G^* to each edge in such a way that the signed sum of characters at each inner vertex gives the trivial character.

Example 4.25. Consider the group $G\cong G^*=\mathbb{Z}_3$ and the following tree:



Here e_2, e_3, e_4 and e_5 are leaves. An example of a socket is the assignment $e_2\mapsto 1$, $e_3\mapsto 1$, $e_4\mapsto 2$, $e_5\mapsto 2$.

Example 4.26. We consider the same tree as in Example 4.25. We can make a network using the same assignment and extending it by $e_1\mapsto 2$.

Remark 4.27. Networks and sockets were introduced in [BW07]; see the discussion below. As the construction presented here directly generalizes the previous one, we decided to keep the name. However, networks could also be named *group based flows*. Indeed, the condition that at each vertex the sum of the elements assigned to the incoming edges equals the sum of the elements assigned to the outgoing edges is the well-known condition for a flow. The only difference is that we are assigning elements of an arbitrary group. As we will see in Proposition 4.30, there is a bijection between sockets and networks. This is similar to the theorem that for a flow the sum over all sources equals the sum over all sinks. One can also generalize the definition to arbitrary graphs—cf. [BBKM].

In [BW07], for the group $\mathbb{Z}_2$ the socket was defined as an even subset of leaves. That corresponds to assigning 1 to some leaves and 0 to the others. The condition that the subset has an even number of elements is just the condition that the elements from the group sum to the neutral element. We see that this definition is compatible with Definition 4.24. Networks were defined as subsets of edges such that there was an even number chosen around each inner vertex—this is also the condition of summing to the neutral element around each inner vertex.

Let us generalize the results on sockets and networks from [BW07].

LEMMA 4.28. *There are exact sequences of abelian groups*

$$0 \to \mathfrak{N} \to (G^*)^E \xrightarrow{add} (G^*)^N \to 0, \qquad 0 \to \mathfrak{S} \to (G^*)^L \xrightarrow{add'} G^* \to 0.$$

Proof. As add and add' are surjective, the lemma follows from Definition 4.24. ∎

DEFINITION 4.29 (Morphisms fo and bi). There is a group morphism $fo : (G^*)^E \to (G^*)^L$ that forgets all the components indexed by edges not adjacent to leaves. From the diagrams in Lemma 4.28 the image of $\mathfrak{N}$ by fo is contained in $\mathfrak{S}$. We define $bi : \mathfrak{N} \to \mathfrak{S}$ to be the restriction of fo.

We have the following diagram:

$$\begin{array}{ccccccccc} 0 & \to & \mathfrak{N} & \to & (G^*)^E & \xrightarrow{add} & (G^*)^N & \to & 0 \\ & & \downarrow^{bi} & & \downarrow^{fo} & & \downarrow^{-sum} & & \\ 0 & \to & \mathfrak{S} & \to & (G^*)^L & \xrightarrow{add'} & G^* & \to & 0 \end{array}$$

The map $-sum : (G^*)^N \to G^*$ associates to an $|N|$-tuple of characters minus their sum.

PROPOSITION 4.30. *For any tree and any abelian group G the morphism bi that assigns a socket to a network is a group isomorphism.*

Proof. Let n be a network. We know that the signed sum $p_v(add(n))$ around each inner vertex v is the neutral element. Hence $\sum_{v\in N} p_v(add(n)) = e$, where e is the neutral element. Consider an edge directed from v_1 to v_2, where $v_1, v_2 \in N$. Note that the group elements $n(v_1, v_2)$ and $n(v_1, v_2)^{-1}$ appear in $p_{v_1}(add(n))$ and $p_{v_2}(add(n))$. We see that $\sum_{v\in N} p_v(add(n)) = \sum_{l\in L} n(l)$. This means that the restriction of the network to leaves gives a socket.

Given a socket s we can define a function $n : E \to G$ inductively, starting from leaves, using the condition of summing to the neutral element around inner edges. The only nontrivial thing is to notice that the sum around the root also gives the neutral element. This follows from the previous equality $\sum_{v\in N} p_v(add(n)) = \sum_{l\in L} n(l)$ and the fact that $p_v(add(n)) = e$ for each node v different from the root. ∎

Each network naturally determines an element of the basis of $(\widehat{W_E})^{G_N}$, and each socket determines an element of the basis of W_L^G. The isomorphism in Theorem 4.22 just uses the natural bijection (Proposition 4.30). This motivates the following definition.

Definition 4.31 (Spaces $\widetilde{W_E}$, $\widetilde{W_L}$). We define the subspace $\widetilde{W_E} := (\widehat{W_E})^{G_N} \subset \widehat{W_E}$. Recall that basis elements of $\widehat{W_E}$ are indexed by elements of $(G^*)^E$ as in Remark 4.20. The basis elements of $\widetilde{W_E}$ correspond to elements of $\mathfrak{N}$.

We define the subspace $\widetilde{W_L} := W_L^G \subset W_L$. The basis elements of $\widetilde{W_L}$ correspond to assignments that form a socket—cf. proof of Lemma 4.8.

Using Theorem 4.22 we find that the variety $X(T, W, \widehat{W})$ is the closure of the image of the rational map induced by $\pi \circ \widehat{\psi}$:

$$\check{\psi} : \prod \widehat{W_e} = \mathbb{C}^{|G||E|} \to \widetilde{W_L},$$

where the coordinates of the domain are indexed by pairs (e, χ) for $e \in E$ and $\chi \in G^*$. The coordinates of the codomain are indexed by sockets (or equivalently networks). In fact the codomain is a regular representation of the group $\mathfrak{N}$. In forthcoming sections we will use the action of this group on the variety $X(T, W, \widehat{W})$.

Note that for a fixed basis of a vector space, the points with nonzero coordinates form an algebraic torus that acts on the space. Let us describe the affine map $\pi \circ \widehat{\psi}$ in toric terms.

Definition 4.32 (Lattices M_S, M_e, M_E). To each edge e we associated a vector space $\widehat{W_e}$ with the distinguished basis given by ω_χ. The points with nonzero coordinates in this basis form an algebraic torus with the action given by coordinatewise multiplication. We define M_e to be the character lattice of this torus.

The product vector space $\prod_{e\in E} \widehat{W_e}$ has a basis induced from each $\widehat{W_e}$. The points with nonzero coordinates form an algebraic torus with the character lattice given by M_E.

The vector spaces $\widetilde{W_E} \cong \widetilde{W_L}$ have a distinguished bases with elements corresponding to sockets. The points with nonzero coordinates form an algebraic torus with the character lattice given by M_S.

Let us note that the coordinate system on the vector space determines the basis of the lattice. The basis of each lattice M_e is indexed by characters. As $M_E = \bigoplus_{e\in E} M_e$, the basis of M_E is indexed by pairs (e, χ) where e is an edge and χ a character of G. The basis elements of M_S correspond to sockets or networks. The rational map $\check{\psi} : \prod_{e\in E} W_e \to \widetilde{W_E} \cong \widetilde{W_L}$ is an equivariant parametrization of a toric variety.

Definition 4.33 (Morphism $\widetilde{\psi}$). $\widetilde{\psi} : M_S \to M_E$ is the morphism of lattices induced by $\check{\psi}$.

In this setting the description of $\widetilde{\psi}$ is particularly simple. Let $f_n \in M_S$ be a basis vector corresponding to a network n. The element $\widetilde{\psi}(f_n)$ will be an element of the unit cube in M_E. Let $h_{(e,\chi)} \in M_E$ be the basis vector corresponding to a pair $(e, \chi) \in E \times G^*$ and let $h^*_{(e,\chi)}$ be its dual. We have

$$h^*_{(e,\chi)}(\widetilde{\psi}(f_n)) = \begin{cases} 1 & \text{if } n(e) = \chi, \\ 0 & \text{otherwise.} \end{cases}$$

We come to the most important definition of this section.

Definition 4.34 (Polytope P). We define the polytope $P \subset M_E$ to be the convex hull of the image of the basis of M_S by $\widetilde{\psi}$. In other words, the vertices of P correspond to networks. More precisely, each vertex has 1 on coordinates indexed by pairs that form a network, and 0 on other coordinates. Note that P is a subpolytope of the unit cube. Hence all its integer points are vertices.

Example 4.35. Consider a tree T with one inner vertex and three leaves l_1, l_2 and l_3. Let $G \cong G^* = \mathbb{Z}_2$. The lattice M_S is the 4-dimensional lattice generated freely by vectors $e_{(0,0,0)}, e_{(1,1,0)}, e_{(1,0,1)}, e_{(0,1,1)}$ that correspond to sockets/networks on T. The lattice M_E is a 6-dimensional lattice with basis vectors $f_{(l_i,g)}$ with $1 \le i \le 3$ and $g \in \mathbb{Z}_2$. We have

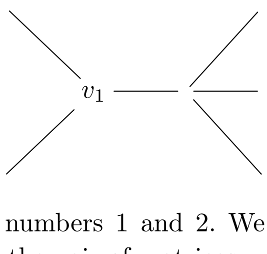
$$\widehat{\psi}(e_{(a,b,c)}) = f_{(l_1,a)} + f_{(l_2,b)} + f_{(l_3,c)}.$$

Hence each vertex of P will have three coordinates equal to zero and three to one. Consider the base of M_E in the order $f_{(l_1,0)}, f_{(l_1,1)}, \ldots, f_{(l_3,0)}, f_{(l_3,1)}$. The vertex corresponding to $e_{(0,0,0)}$ is $(1,0,1,0,1,0)$. In the same order $e_{(1,1,0)} \mapsto (0,1,0,1,1,0)$, $e_{(1,0,1)} \mapsto (0,1,1,0,0,1)$ and $e_{(0,1,1)} \mapsto (1,0,0,1,0,1)$. These are of course all the vertices of P.

Remark 4.36. Suppose that a tree T has a vertex v of degree two. Let $e_1 = (u,v)$ and $e_2 = (v,w)$ be respectively an incoming and outgoing edge. Consider any network n. We have $n(e_1) = n(e_2)$. Let T' be the tree obtained from T by removing the vertex v and the edges e_1, e_2, and adding an edge (u,w). We see that the polytope associated to T is isomorphic to the polytope associated to T'.

The polytope P is the polytope associated to the toric variety $X(T,G)$. The algebra of this variety is the algebra associated to the monoid generated by P in M_E. The generating binomials of a toric ideal associated to P correspond to integral relations between integer points of this polytope (Corollary 2.6). Hence in our situation phylogenetic invariants correspond to relations between networks. Each such relation can be described in the following way. We number all edges of a tree from 1 to e. The networks are specific e-tuples of group elements. For example, for the claw tree these are e-tuples of group elements summing to the neutral element. Each relation of degree d between the networks is encoded as a pair of matrices with d columns and e rows with entries that are group elements. We require that each column represents a network. Moreover, the rows of both matrices are the same up to permutation.

Example 4.37. Consider the binary Jukes–Cantor model and the tree



(4.1)

The leaves adjacent to v_1 have numbers 1 and 2. We assign 3 to the inner edge. An example of a relation is given by the pair of matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

The numbers 0 and 1 are treated as elements of $\mathbb{Z}_2$. By the definition of the socket the third row has to be the sum of both the first two and the last three rows.

Note that P does not have to generate the lattice M_E.

Definition 4.38 (Lattice $\widehat{M_E}$). We define the lattice $\widehat{M_E}$ as the sublattice of M_E generated by the vertices of P.

The lattices defined so far corresponded to affine objects. A rational map from a vector space to its projectivization is well defined at points with nonzero coordinates. Hence it induces a surjective morphism of tori, which corresponds to an injective morphism of character lattices.

Definition 4.39 (Degree functions $\deg_e$). Note that for a character lattice M with a distinguished basis we can define a function $\deg : M \to \mathbb{Z}$ that sums the coordinates. The degree of a lattice element is the degree of the monomial function associated to it. For the lattices M_e the corresponding degree functions are denoted by $\deg_e$.

Definition 4.40 (Lattices $M_{S,0}$, $M_{E,0}$ and $\widehat{M_{E,0}}$). For a lattice M_S we define $M_{S,0}$ as the sublattice of elements with coordinates summing to zero. In particular $M_{S,0}$ is the character lattice of the torus whose points are identified with points of $\mathbb{P}(\widetilde{W_E})$ with all coordinates different from zero.

We define $M_{E,0}$ to be the sublattice of M_E defined by the equalities $\deg_e = 0$ for each edge e. This is the character lattice of the torus whose points are identified with points of $\prod \mathbb{P}(W_e)$ with all coordinates different from zero.

We define $\widehat{M_{E,0}} := M_{E,0} \cap \widehat{M_E}$. This is the character lattice of the torus whose points are identified with points of the projective toric variety $\mathbb{P}(X(T))$ with all coordinates different from zero.

Recall that the basis of the lattice M_E is indexed by pairs (e, χ) where e is an edge and χ is a character of G. Also to each such pair we can associate a one-parameter subgroup in the dual of M_E. It is given as a morphism from M_E to $\mathbb{Z}$ that is the dual vector to the vector of the base of M_E that is indexed by the pair (e, χ). In particular for each leaf l and character $\chi \in G^*$ we obtain a one-parameter subgroup λ_l^χ. Using the morphism dual to $\tilde{\psi} : M_S \to M_E$, for each pair (e, χ) we obtain a one-parameter subgroup in the lattice dual to M_S. For each $t \in \mathbb{C}^*$ we have an action of $\lambda_l^\chi(t)$ on $\mathbb{A}^{(|L|-1)\times|G|} \supset X$. The weight of this action on the coordinate indexed by a socket s is either 1 or 0 depending on whether the socket s associates to the leaf l the character χ or not.

Remark 4.41. In [BW07] the authors considered only one one-parameter subgroup for each leaf although their group had two elements. Notice however that in our notation for the group $\mathbb{Z}_2$ the weights of the action of λ_l^0 are completely determined by the weights of the action of λ_l^1—one type weights are negations of the others. In our notation the authors of [BW07] considered only λ_l^1.

The setting presented here, where an abelian group G acts transitively and freely on the set of states, is best understood. The models obtained in this way are called *general group-based models*. Although this definition is quite clear, the question what is a group-based model is much less obvious. This motivates the discussion of the next section.

4.2. Notation. In Section 4.1 we have introduced the general group-based models. The key point of the definition was that the vector space $\widehat{W}$ was given as a subspace of $\operatorname{End} W$ invariant under the action of an abelian group that acts transitively and freely on the basis of W. This setting enabled us to apply the discrete Fourier transform and associate toric varieties with the models. There are a few possibilities to generalize this construction depending on the assumptions on the group, its action on the space W and properties of the associated variety.

The first idea would be to consider any action of any group on W. An even more general construction is presented in [DK09], where the vector space W may vary depending on the vertex of the tree. Such models are called *equivariant models*. Of course, in this case, in general one cannot apply the discrete Fourier transform, as the group G is not abelian. Moreover, if G is small the transition matrices may be too general and the associated variety will not be toric. For example, if G has only one element it is abelian. However, the corresponding model is just the general Markov model. The varieties associated to this model are an object of intensive study: see for example [AR08] and references therein. They are far from being toric, and establishing their properties even for the simplest tree is a great challenge. For example, it is an open problem to determine the ideal in the case of the tripod.

As we want to work with toric varieties, it is reasonable to make further assumptions. Let us notice that the adjective "general" indicates that other group-based models should be more specific. In other words, the subspace $\widehat{W}$ for a group-based model should contain specific transition matrices of a general group-based model. Thus we fix an abelian group H that acts on the space W transitively and freely. A group-based model will be obtained by requiring further conditions on the space of transition matrices.

Before stating some definitions, let us present the state of the art. In the literature one can find many references to group-based models [SS05], [APRS11], [PS05, p. 327]. In this setting one assumes that there is a bijection between elements of an abelian group and elements of S, as in general group-based models. One also requires that the entries of the transition matrices depend only on the difference of the group elements labelling the row and the column of the given entry. However, we allow the parameters for different differences to be the same—a formal definition is presented in Definition 4.43. This is a very general definition that covers many models, like Jukes–Cantor on any number of states, 2-Kimura or any general group-based model. However, for example in [APRS11], [SS05, p. 460], one can also find theorems, usually referring to [ES93], that group-based models are toric. This would require an additional assumption, e.g. that the invertible matrices in the model form a group—otherwise see Appendix 1, where after the

Fourier transform we do not get monomials but polynomials. The reason is that equality of variables before Fourier transform does not imply equality of parameters after the transform. We stress that the fact that Jukes–Cantor and 2-Kimura give rise to toric varieties was known before. To give a formal definition of group based-models we use a method of labellings by Sturmfels and Sullivant [SS05, Section 3].

Definition 4.42 (Labelling function). Let Lab be any finite set and H an abelian group. A *labelling function* is any function $f : H \to Lab$.

Later, we will consider special labellings, induced by group actions, which will turn out to have interesting properties.

Definition 4.43 (Group-based model). We define *group-based models* by specifying the space $\widehat{W}$ of transition matrices. Suppose that an abelian group H acts on the set S of states transitively and freely. For any $s_1, s_2 \in S$ we define a morphism $p_{s_1,s_2} : \operatorname{End} W \to \mathbb{C}$ by $p_{s_1,s_2}(M) = s_2^*(M(s_1))$ where $s_1 \in W$ is an element of the basis and s_2^* is an element of the dual basis. Let $g_{s_1,s_2} \in H$ be the unique element sending s_1 to s_2.

We fix any labelling function f on H. We define $\widehat{W}$ as the largest subspace of transition matrices M satisfying the following condition: For any $s_1, s_2, s_3, s_4 \in S$ such that $f(g_{s_1,s_2}) = f(g_{s_3,s_4})$ we have $p_{s_1,s_2}(M) = p_{s_3,s_4}(M)$.

Less formally but more intuitively, one labels the rows and columns of transition matrices with elements of H. The condition is that the entries labelled by (g_1, g_2) and (g_3, g_4) are equal if $(f(g_1), f(g_2)) = (f(g_3), f(g_4))$. Notice that the space $\widehat{W}$ is obtained from the space of transition matrices of a general group-based model by specific hyperplane sections. It is important to understand that in this setting the class of group-based models is much larger than the class of general group-based models. The latter are called "general" because the space $\widehat{W}$ is the most general. They correspond to labellings that are injective. This is a motivation for the next section. We will distinguish a class of group-based models, called G-models. For them, we will require that the labelling is given by a specific group action. In this setting the associated varieties will be toric, and we will provide an explicit description of the polytope.

4.3. G-models. This section contains results from [Mic11]. Our main aim is to introduce the general framework that would include all models of interest described as group-based, but still would give rise to toric varieties with an explicit construction of the associated polytope.

The setting of this section is sufficiently general to cover many Markov processes, in particular this will be a generalization of the results of Section 4.1. However, the inspiration is the 2-Kimura model, that is, the phylogenetic model in which the transition matrices are of the following type:

$$\begin{bmatrix} a & b & c & b \\ b & a & b & c \\ c & b & a & b \\ b & c & b & a \end{bmatrix}.$$

In this case, as in the previous section, we also have an abelian group, $H = \mathbb{Z}_2 \times \mathbb{Z}_2$, that acts on the basis (A, C, G, T) of a 4-dimensional vector space W. As we have seen, the fixed points of the action of H on $W \otimes W$ define the 3-Kimura model. We may however define a larger group G, namely the dihedral group of order 8, which contains H as a normal subgroup. The action of G on $W \otimes W$ defines the 2-Kimura model. Details of this construction can be found in [BDW09]. This motivates the following setting.

Let S be an n-element set of states. Let G be a subgroup (1) of $S_n = \mathrm{Sym}(S)$ acting on S. Suppose moreover that the group G contains a *normal, abelian* subgroup H and the action of H on S is transitive and free. Elements of S once again correspond to states of vertices of a phylogenetic tree T. We define W as in Definition 3.1.

The basic difference from the abelian case is that we define elements of $\widehat{W}$ as matrices fixed not only under the action of H, but under the whole action of G. We assume that $\mathrm{End}(W) \cong W \otimes W$ (cf. Remark 3.5).

Definition 4.44. Let
$$\widehat{W} = \Big\{ \sum_{a_i, a_j \in S} \lambda_{a_i,a_j} a_i \otimes a_j : \lambda_{a_i,a_j} = \lambda_{g(a_i),g(a_j)} \ \forall g \in G \Big\}.$$

Remark 4.45. The characterization of $\widehat{W}$ from Remark 4.4 is still valid. However, by additional symmetries the dimension is different.

Remark 4.46. The situation of the previous section corresponds to $G = H$.

Remark 4.47. As before, by choosing an element $e \in S$ we make a bijection between S and H. An element associated to $a \in S$ will be denoted by $h_a \in H$. The element e corresponds to the neutral element of H and is the index of the first row of transition matrices. Notice that the action of G on S (as permutations) will not generally be the same as the action of G on H (as a group).

We will often use the following easy observation.

Lemma 4.48. *Suppose that $h \in H$ as a permutation sends a to b, where $a, b \in S$. Then $h = h_b h_a^{-1}$.*

Proof. Both elements send a to b, so because H acts on S freely, they have to be equal. ∎

Definition 4.49 (G-model). Let G be a finite group acting on a finite set S. Suppose that G contains a normal, abelian subgroup H that acts on the set S transitively and freely. A *G-model* is an algebraic variety $X(T, W, \widehat{W})$ for W and $\widehat{W}$ as in Definitions 3.1 and 4.44.

Our aim is to provide the description of the associated toric varieties in this generalized setting. We will proceed in four steps.

(i) We introduce a general method for constructing endomorphisms of W from complex functions on H. We prove that under certain conditions (namely a function should be constant on orbits of the conjugation action of G on H), the resulting endomorphism

(1) Not necessarily abelian.

is in $\widehat{W}$. Such functions can be regarded as a generalization of class functions to pairs of groups.

(ii) We prove that some sums (over the orbits of the action of G on H^*) of characters of H are functions that can define elements of $\widehat{W}$. We also notice that we obtain a set of independent vectors of $\widehat{W}$.

(iii) Using dimension arguments we prove that the set defined in step (ii) is in fact a basis.

(iv) Finally, using theorems from Section 4.1, we provide, using the new coordinates, the toric description of the variety.

Definition 4.50. We define $\widehat{W}_H$ to be the vector space of matrices fixed under the action of H.

Remark 4.51. From the previous subsection we know that the closure of the image of the map

$$\psi : \prod_{e\in E} \mathbb{P}(\widehat{(W_H)_e}) \dashrightarrow \mathbb{P}(W_L)$$

is a toric variety. Moreover, we have found the base in which the described morphism is given by monomials. As $\widehat{W} \subset \widehat{W}_H$, our aim is to find a monomial description of the above map. We will use the base on $\widehat{W}_H$ to define the base of $\widehat{W}$.

Step 1: Correspondence between functions on H and endomorphisms of W. We are going to define some endomorphisms of W.

Definition 4.52. Let $f : H \to \mathbb{C}$ be *any* function. We define

$$l_f = \frac{1}{|H|} \sum_{a,b\in S} f(h_a^{-1}h_b)a \otimes b.$$

Remark 4.53. Notice that by Proposition 4.7 this definition is consistent with the definition of l_χ for $\chi \in H^*$. Moreover, the vector l_f depends only on the function f and not on the bijection between S and H, as $h_a^{-1}h_b$ is the only element from H that sends a to b.

Proposition 4.54. *Consider the conjugation action of G on H:*

$$(g, h) \mapsto ghg^{-1}.$$

If f is constant on orbits of this action then $l_f \in \widehat{W}$.

Proof. Fix $g \in G$. We focus on two entries of the matrix l_f, namely (a_1, b_1) and (a_2, b_2), where

$$g(a_1) = a_2 \quad \text{and} \quad g(b_1) = b_2.$$

From the definition of l_f these entries are respectively $f(h_{a_1}^{-1}h_{b_1})$ and $f(h_{a_2}^{-1}h_{b_2})$. By Remark 4.4 we want to prove that $f(h_{a_1}^{-1}h_{b_1}) = f(h_{a_2}^{-1}h_{b_2})$. Consider the element $gh_{b_1}h_{a_1}^{-1}g^{-1}$. Clearly it is in H (because H is a normal subgroup of G) and sends a_2 to b_2. From Lemma 4.48 we obtain

$$gh_{b_1}h_{a_1}^{-1}g^{-1} = h_{b_2}h_{a_2}^{-1}.$$

This completes the proof, as f is constant on orbits of the conjugation action. ■

Step 2: Appropriate functions on H. In the abelian case we considered the characters of H. As G was equal to H, these functions were of course constant on (one-element) orbits of the action of G on H. In the general case it may happen that we do not have the equality

$$\chi(ghg^{-1}) = \chi(h).$$

Of course this equality holds if the character of H extends to a character of G, but this is not always the case. If we define the vectors l_χ for $\chi \in H^*$ they may not be in $\widehat{W}$. To obtain the vectors in $\widehat{W}$ we will sum some characters to obtain functions that satisfy the condition of Proposition 4.54. Consider the action of G on H^* given by

$$\chi^g(h) = \chi(ghg^{-1}).$$

Let O be the set of orbits of this action. The elements of O give a *partition* of H^*. Let us define for each $o \in O$ a function $f_o : H \to \mathbb{C}$.

DEFINITION 4.55 (Function f_o). Let $f_o = \sum_{\chi \in o} \chi$. Here we are summing characters as complex-valued functions, not as characters, so this is the usual sum, not the product. We obtain $l_{f_o} = \sum_{\chi \in o} l_\chi$.

PROPOSITION 4.56. *The function f_o satisfies the conditions of Proposition 4.54, that is, it is constant on orbits of the conjugation action of G on H.*

Proof. As the action of g' is a permutation of the orbit o, we have

$$f_o(g'hg'^{-1}) = \sum_{\chi \in o} \chi(g'hg'^{-1}) = \sum_{\chi \in o} (g', \chi)(h) = \sum_{\chi \in o} \chi(h) = f_o(h). \ \blacksquare$$

COROLLARY 4.57. *The vectors l_{f_o} for $o \in O$ are in $\widehat{W}$. Moreover, as l_χ forms a basis of $\widehat{W}_H$, and l_{f_o} are sums over a partition of this basis, they are independent.*

PROPOSITION 4.58. *Any complex function constant on orbits of O is a linear combination of the functions f_o.*

Proof. Let f be a function constant on orbits. As the characters of H span the space of all functions, we know that $f = \sum_{\chi \in H^*} a_\chi \chi$. We have to prove that the coefficients of χ in the same orbit are the same. Let $\chi_1^g = \chi_2$. We know that for any $h \in H$ we have

$$\sum_{\chi \in H^*} a_\chi \chi(h) = f(h) = f(ghg^{-1}) = \sum_{\chi \in H^*} a_\chi \chi(ghg^{-1}) = \sum_{\chi \in H^*} a_\chi \chi^g(h).$$

From the linear independence of characters we see that $a_{\chi_1} = a_{\chi_2}$, which completes the proof. ■

COROLLARY 4.59. *The number of orbits in O (and so the number of vectors l_{f_o}) is equal to the number of orbits of the conjugation action of G on H.*

Proof. This follows by comparing the dimensions of the spaces of complex functions on H that are constant on orbits. ■

Step 3: Dimension of $\widehat{W}$. We are going to prove that the dimension of $\widehat{W}$ is equal to the number of orbits, $|O|$. First note that all the entries of any matrix in $\widehat{W}$ (in the basis S) are determined by the entries in the first row. This follows from Section 4.1. We see that $\dim \widehat{W}$ is equal to the number of independent parameters in the first row, which is indexed by e. The action of G imposes some conditions, namely the entry in the eth row and ath column and the entry in the eth row and bth column for $a,b\in S$ have to be equal if and only if there exists $g\in G$ such that

$$g(e)=e \quad\text{and}\quad g(a)=b.$$

LEMMA 4.60. *The following conditions are equivalent:*

(i) *there exists $g\in G$ that sends e to e and a to b,*
(ii) *h_a and h_b are in the same orbit of the action $(g,h)=ghg^{-1}$.*

Proof. Of course h_a and h_b are in the same orbit if and only if h_a^{-1} and h_b^{-1} are in the same orbit. For the proof we use the latter variant.

(i)⇒(ii): From Lemma 4.48 we know that $gh_a^{-1}g^{-1}=h_b^{-1}$, because both sides send b to e.

(i)⇐(ii): Suppose that $gh_a^{-1}g^{-1}=h_b^{-1}$. Let $g'=h_b^{-1}gh_{g^{-1}(b)}$. The element g' sends e to e, but $g'=gh_a^{-1}h_{g^{-1}(b)}$, hence it also sends a to b. ■

PROPOSITION 4.61. *The dimension of $\widehat{W}$ is equal to $|O|$.*

Proof. Classes of equal parameters in the first row of matrices in $\widehat{W}$ correspond bijectively to orbits of the action of G on H, from Lemma 4.60 and remarks at the beginning of this subsection. By Corollary 4.59 this finishes the proof. ■

COROLLARY 4.62. *The elements l_{f_o} for $o\in O$ form a basis of $\widehat{W}$.*

Proof. The vectors l_{f_o} are independent by Corollary 4.57. Their number equals the dimension of the space by Proposition 4.61. ■

Step 4: Toric description of G-models. Consider the basis of $\widehat{W}_e$ consisting of the vectors l_{f_o}. We consider the inclusion map $i:\widehat{W}_e\to\widehat{(W_H)}_e$ in the bases consisting respectively of l_{f_o} and l_χ. We know that $l_{f_o}=\sum_{\chi\in o}l_\chi$. Let us describe the morphism i in the coordinates corresponding to the basis l_{f_o} of $\widehat{W}_e$ and to the basis l_χ of $\widehat{(W_H)}_e$. Fix $\chi\in o$. We have $l_\chi^*(i(x))=l_{f_o}^*(x)$.

This shows that the map from $\prod_{e\in E}\mathbb{P}(\widehat{W}_e)$ to $\mathbb{P}(W_L)$ that parametrizes the model is also given by monomials—these are exactly the monomials from Section 4.1, where we just set some variables equal to each other. Let us describe which variables are identified. We recall that variables in the abelian case correspond to networks. Fix two networks n_1 and n_2. We identify them if and only if for each edge e the characters $n_1(e)$ and $n_2(e)$ are in the same orbit of the adjoint G action.

We have the following commutative diagram:

$$\begin{array}{ccccc}
\prod_{e\in E}\mathbb{P}(\widehat{W}_e) & \to & \mathbb{P}(\widehat{W}_E) & \dashrightarrow & \mathbb{P}(W_L)\\
\downarrow & & \downarrow & & \updownarrow\\
\prod_{e\in E}\mathbb{P}(\widehat{W_H}_e) & \to & \mathbb{P}(\widehat{W_H}_E) & \dashrightarrow & \mathbb{P}(W_L)
\end{array}$$

This proves the main theorem of this section.

THEOREM 4.63. *Let G be a finite group that acts faithfully on a finite set S. Let H be a normal, abelian subgroup of G. Suppose that the action of H on S is transitive and free. Let $\widehat{W}$ be the space of matrices invariant with respect to the action of G and let W be* *the vector space spanned freely by the elements of S. Then the G-model $X(T, W, \widehat{W})$ is toric for any tree T. Its polytope is a projection of the polytope associated to the general group-based model for H. The projection is determined by the adjoint action of G on H.*

We will now describe the lattices of characters of the tori that appear in the construction. As in Section 4.1, there is a lattice M_S with basis elements corresponding to sockets, and two lattices $\widehat{M}_{E,H} \subset M_{E,H}$. The letter has basis elements indexed by pairs (e, χ) where $e \in E$ is an edge of the tree and $\chi \in H^*$ is a character.

Definition 4.64 (Lattice $M_{E,G}$). Let $M_{E,G}$ be the lattice with basis elements indexed by pairs (e, o), where $e \in E$ and o is an orbit of the adjoint action of G on H^*.

Let $f_{e,\chi} \in M_{E,H}$ be the basis element indexed by the pair (e, χ). Let $f_{e,o} \in M_{E,G}$ be the basis element indexed by (e, o). There is a natural projection $M_{E,H} \to M_{E,G}$; to $f_{e,\chi}$ we associate $f_{e,o}$, where $\chi \in o$. The image of a polytope $P \subset M_{E,H}$ for the general group-based model is a polytope $\tilde{P}$ that is associated to the variety representing the G-model. Hence $\tilde{P}$ is a subpolytope of a unit cube. An element $\sum_{e \in E} f_{e,o_e}$ is a vertex of $\tilde{P}$ if and only if there exist characters $\chi_{o_e} \in o_e$ such that $\sum_{e \in E} f_{e,\chi_{o_e}}$ is a vertex of P. The lattice spanned by $\tilde{P}$ will be denoted by $\widehat{M}_{E,G}$. The following diagram commutes:

$$\begin{tikzcd} M_S \arrow[r] \arrow[dr] & M_{E,H} \arrow[d] \\ & M_{E,G} \end{tikzcd}$$

The morphisms from M_S correspond to embeddings of both models in an affine space. The vertical surjective morphism corresponds to inclusion of models. Indeed, by introducing new conditions on transition matrices for a G-model we restrict the image, hence there is a natural inclusion in a general group-based model.

We finish this section by presenting relations of G-models to labellings of Definition 4.42. From Lemma 4.60 it follows that the entries of the transition matrix labelled respectively by $(h_1, h_2) \in H^2$ and $(h_3, h_4) \in H^2$ are equal if $h_1^{-1}h_2$ and $h_3^{-1}h_4$ are in the same orbit of the adjoint action of G on H. Let Lab be the set of orbits of the adjoint action of G on H. The labelling function $f : H \to Lab$ associates to an element its orbit.

Definition 4.65 (m-friendly labelling, friendly labelling [SS05, Definition 8]). Let H be any abelian group and Lab any finite set. Fix a labelling function $f : H \to Lab$. For $m \geq 3$ consider the set

$$Z = \Big\{(g_1, \ldots, g_m) \in H^m : \sum_{i=1}^{m-1} g_i = g_m\Big\}.$$

Consider the induced map $\tilde{f} : Z \subset H^m \to Lab^m$ and denote by π_i the projection $\pi_i : H^m \to H$ onto the ith coordinate. The function f is called *m-friendly* if, for every $l = (l_1, \ldots, l_m) \in \tilde{f}(Z) \subset Lab^m$,

$$\pi_i(\tilde{f}^{-1}(l)) = f^{-1}(l_i) \quad \text{for all } i = 1, \ldots, m.$$

A labelling is *friendly* if it is m-friendly for all $m \geq 3$.

Lemma 4.66. *The labellings for G-models are friendly.*

Proof. Fix an m-tuple of orbits $(o_1, \dots, o_m)$ for the adjoint action of G on an abelian normal subgroup H. Suppose that there exist $h_i \in o_i$ such that $\prod_{i=1}^{m-1} h_i = h_m$. Fix any $\tilde{h}_{i_0} \in o_{i_0}$. There is $g \in G$ such that $\tilde{h}_{i_0} = g h_{i_0} g^{-1}$. Consider the element $(gh_1g^{-1}, \dots, gh_mg^{-1})$. Let $\tilde{f}$ and π_i be as in Definition 4.65. Of course $\tilde{f}(gh_1g^{-1}, \dots, gh_mg^{-1}) = (o_1, \dots, o_m)$. Moreover, $\pi_{i_0}(gh_1g^{-1}, \dots, gh_mg^{-1}) = \tilde{h}_{i_0}$, which proves that the labelling is friendly. ∎

The main reason for introducing friendly labellings is that they allow one to apply an important inductive procedure. Assuming that we are dealing with a model given by friendly labelling, the variety associated to any tree T can be described in terms of the varieties associated to claw trees. The polytope associated to a tree T is a fiber product of polytopes associated to claw trees. More information can be found in Section 4.5 and [Sul07], [SS05, Lemma 12].

At this point we should make a remark about the difference between group elements and characters. To define the space of transition matrices for a G-model we used a G action on $\mathrm{End}(W)$. We considered the basis of W that corresponded to states, or by choosing a bijection to elements of an abelian group. The adjunction action of G on H allowed us to define the labelling that described the G-model. Note however that this is *not* the labelling that identifies the coordinates of the parametrization of the variety. In the latter case the variables correspond to pairs (e, χ) where $\chi \in H^*$. The labelling identifies the variables corresponding to pairs with characters on the second coordinate that are in the same orbit. Hence the set of labels is the set of orbits of the adjoint action of G on H^*. The labelling associates to a character its orbit in the adjoint action. The same proof as for Lemma 4.66 shows that this is also a friendly labelling.

4.4. Example of a 2-Kimura model. In this subsection we will show how the construction from the previous subsection works on Kimura models. We will also present an algorithm for constructing a polytope of a model for a given group G with a normal subgroup H. The method was described in a different language in [SS05]. The main difference (apart from the notation) is that the authors assumed the existence of a friendly labelling function that described which characters are identified. In the case of G-models we know this function exactly: it associates to a given character its orbit under the G action. This is a friendly labelling.

If $G = H$ the construction is particularly easy. The polytope has $|G|^{|E|-|N|}$ vertices and the algorithm works in time $O(|N|(|G|^{|E|-|N|}))$ assuming that we can perform group operations in unit time.

Algorithm 1. INPUT: A rooted tree T and an abelian group G.

OUTPUT: Vertices of the polytope associated to the toric variety representing the model for the tree T and the group G.

(i) Orient the edges of the tree from the root.
(ii) For each inner vertex choose one outgoing edge.
(iii) Fix a bijection $b : G \to B \subset \mathbb{Z}^{|G|}$, where B is the standard basis of $\mathbb{Z}^{|G|}$.

(iv) Consider all possible assignments of elements of G to nonchosen edges (there are $|G|^{|E|-|N|}$ such assignments).
(v) For each such assignment, complete the assignment by assigning an element of G to each chosen edge in such a way that the signed sum of elements around each inner vertex gives the neutral element in G.
(vi) For each complete assignment, output the vertex of the polytope: $(b(g_e)_{e\in E})$, where g_e is the element of the group assigned to the edge e.

EXAMPLE 4.67. For the 3-Kimura model corresponding to the group $\mathbb{Z}_2\times\mathbb{Z}_2$ on a tree with one inner vertex and three leaves, the vertices of P correspond to triples of characters of the group that sum to the neutral character:

1) $(0,0),(0,0),(0,0)$, 2) $(0,0),(1,0),(1,0)$, 3) $(1,0),(0,0),(1,0)$,
4) $(1,0),(1,0),(0,0)$, 5) $(0,0),(0,1),(0,1)$, 6) $(0,1),(0,0),(0,1)$,
7) $(0,1),(0,1),(0,0)$, 8) $(0,0),(1,1),(1,1)$, 9) $(1,1),(0,0),(1,1)$,
10) $(1,1),(1,1),(0,0)$, 11) $(0,1),(1,0),(1,1)$, 12) $(0,1),(1,1),(1,0)$,
13) $(1,0),(1,1),(0,1)$, 14) $(1,0),(0,1),(1,1)$, 15) $(1,1),(0,1),(1,0)$,
16) $(1,1),(1,0),(0,1)$.

In the coordinates of the lattice, this gives us vertices of the polytope:

1) $(1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0)$, 2) $(1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0)$,
3) $(0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0)$, 4) $(0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0)$,
5) $(1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0)$, 6) $(0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0)$,
7) $(0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0)$, 8) $(1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1)$,
9) $(0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1)$, 10) $(0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0)$,
11) $(0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1)$, 12) $(0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0)$,
13) $(0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0)$, 14) $(0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1)$,
15) $(0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0)$, 16) $(0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0)$.

The basis for $\widehat{W}$ for 3-Kimura (in previous notation, the vectors $l_\chi=\sum\chi(h_a^{-1}h_b)a\otimes b)$ is the following:

$$l_1=\begin{bmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{bmatrix},\qquad l_2=\begin{bmatrix}1&-1&1&-1\\-1&1&-1&1\\1&-1&1&-1\\-1&1&-1&1\end{bmatrix},$$
$$l_3=\begin{bmatrix}1&-1&-1&1\\-1&1&1&-1\\-1&1&1&-1\\1&-1&-1&1\end{bmatrix},\quad l_4=\begin{bmatrix}1&1&-1&-1\\1&1&-1&-1\\-1&-1&1&1\\-1&-1&1&1\end{bmatrix}.$$

For the 2-Kimura model the four elements of H, treated as permutations decomposed into cycles, are:

$$(1)(2)(3)(4);\ (1,2)(3,4);\ (1,3)(2,4);\ (1,4)(2,3).$$

The group G is spanned by H and the transposition $(3,4)$.

If we consider the action of G on H^* we obtain the following three orbits:

(i) The orbit of the trivial character. It contains only the trivial character, so the vector

$$f_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

is in $\widehat{W}_G$ and will be considered as the first basis vector.

(ii) The orbit of the character that assigns -1 to $(1,3)(2,4)$ and $(1,4)(2,3)$ and 1 to other elements. It also has only one element. For example,

$$\chi((3,4)(1,3)(2,4)(3,4)) = \chi((1,4)(2,3)) = -1 = \chi((1,3)(2,4)).$$

This means that the vector

$$f_2 = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

is a basis vector of $\widehat{W}_G$.

(iii) The orbit that contains the remaining two characters. If we take their sum (as functions, not characters) we obtain a function that associates 2 to $(1)(2)(3)(4)$, -2 to $(1,2)(3,4)$ and 0 to the other two elements. This gives

$$f_3 = \begin{bmatrix} 2 & -2 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & 2 \end{bmatrix}.$$

This is the sum of the other two l_χ.

We obtain $f_1 = l_1$, $f_2 = l_4$, $f_3 = l_2 + l_3$. Let $F = \{f_1, f_2, f_3\}$ and $L = \{l_1, \dots, l_4\}$. From the previous section we know that F is a basis of $\widehat{W}_G$ and L is a basis of $\widehat{W}_H$. This can be checked directly in this example. Let us now look at the map for the tripod tree $\perp$. Elements of $\widehat{W}_G$ are special elements of $\widehat{W}_H$. We have a map

$$(f_j^{e_i})_{j=1,2,3,\, i=1,2,3} \to (l_j^{e_i})_{j=1,\dots,4,\, i=1,2,3}.$$

Here j parametrizes base vectors and i parametrizes edges. Our model is the composition of this map and a model map for H. The image of the first map is given by the condition that the coordinates corresponding to $l_2^{e_i}$ and $l_3^{e_i}$ are equal for $i = 1,2,3$. Let us see this directly.

The fixed bijection b from Algorithm 1 is the following:

$$b(e) = (1,0,0,0), \quad b(\chi_3) = (0,1,0,0),$$
$$b(\chi_1) = (0,0,1,0), \quad b(\chi_2) = (0,0,0,1),$$

where χ_1 and χ_3 are in the same orbit. The domain of $\widehat{\psi}$ for the group H is $\{(x_1, \dots, x_{12}) : x_i \in \mathbb{C}\}$ in the order as in Example 4.67 (we fix an isomorphism with $\chi_1 = (1,0)$ and

$\chi_3 = (0,1)$). Hence the subspace $\prod_{e\in E}(\widehat{W_G})_e$ is given by the conditions $x_2 = x_3$ (the coordinates of l_2 and l_3 for $\widehat{W}_H^{e_1}$), $x_6 = x_7$, $x_{10} = x_{11}$.

This procedure works generally. After having fixed the polytope for a subgroup H that is in the lattice M (whose coordinates are indexed by edges and characters of H), we consider a morphism from M onto the lattice M' (whose coordinates are indexed by edges and orbits of characters of H) that just assigns a character to a given orbit. This morphism sums coordinates that are in the same orbit of the action of G on H^*. The image of the polytope P is the polytope of our model. For 3-Kimura we sum coordinates ordered as in Example 4.67 obtaining a polytope for the 2-Kimura model:

1) (1, 0, 0, 1, 0, 0, 1, 0, 0), 2) (1, 0, 0, 0, 1, 0, 0, 1, 0),
3) (0, 1, 0, 1, 0, 0, 0, 1, 0), 4) (0, 1, 0, 0, 1, 0, 1, 0, 0),
5) (1, 0, 0, 0, 1, 0, 0, 1, 0), 6) (0, 1, 0, 1, 0, 0, 0, 1, 0),
7) (0, 1, 0, 0, 1, 0, 1, 0, 0), 8) (1, 0, 0, 0, 0, 1, 0, 0, 1),
9) (0, 0, 1, 1, 0, 0, 0, 0, 1), 10) (0, 0, 1, 0, 0, 1, 1, 0, 0),
11) (0, 1, 0, 0, 1, 0, 0, 0, 1), 12) (0, 1, 0, 0, 0, 1, 0, 1, 0),
13) (0, 1, 0, 0, 0, 1, 0, 1, 0), 14) (0, 1, 0, 0, 1, 0, 0, 0, 1),
15) (0, 0, 1, 0, 1, 0, 0, 1, 0), 16) (0, 0, 1, 0, 1, 0, 0, 1, 0).

After removing double entries we get the following vertices:

1) (1, 0, 0, 1, 0, 0, 1, 0, 0), 2) (1, 0, 0, 0, 1, 0, 0, 1, 0),
3) (0, 1, 0, 1, 0, 0, 0, 1, 0), 4) (0, 1, 0, 0, 1, 0, 1, 0, 0),
5) (1, 0, 0, 0, 0, 1, 0, 0, 1), 6) (0, 0, 1, 1, 0, 0, 0, 0, 1),
7) (0, 0, 1, 0, 0, 1, 1, 0, 0), 8) (0, 1, 0, 0, 1, 0, 0, 0, 1),
9) (0, 1, 0, 0, 0, 1, 0, 1, 0), 10) (0, 0, 1, 0, 1, 0, 0, 1, 0).

4.5. Further notation and applications. In this section we will introduce notation concerning specific group-based models. We start by introducing the so called "time-reversibility" condition. This condition forces the transition matrices to be symmetric [PS05, Lemma 17.2]. It is satisfied for many models considered in applications, for example for the 3-Kimura model. One can notice that a general group-based model gives rise to symmetric transition matrices if and only if all nonneutral group elements are of order two. We have to point out that in the literature one often adds to the definition of group-based models the requirement that matrices are symmetric [BDW09], [PS05, p. 328]. We do not use this convention. This leads to the following definition.

Definition 4.68 (general symmetric group-based model, symmetric group-based model). Let H be an abelian group acting transitively and freely on the set S of states. We define the *general symmetric group-based model* as the model associated to the vector space $\widehat{W}$ given as the maximal space of *symmetric* matrices invariant with respect to the H action.

Analogously we define a *symmetric group-based model* as a model associated to a subspace of $\widehat{W}$ given by hyperplane sections that make some parameters of the transition matrices equal.

Symmetric group-based models do not have to be toric. For a counterexample, one can consider the general group-based model for $\mathbb{Z}_6$. The transition matrices are of the following type:

$$\begin{bmatrix} a & b & c & d & e & f \\ f & a & b & c & d & e \\ e & f & a & b & c & d \\ d & e & f & a & b & c \\ c & d & e & f & a & b \\ b & c & d & e & f & a \end{bmatrix}$$

Let us consider a symmetric submodel with transition matrices of the type

$$\begin{bmatrix} a & a & c & d & c & a \\ a & a & a & c & d & c \\ c & a & a & a & c & d \\ d & c & a & a & a & c \\ c & d & c & a & a & a \\ a & c & d & c & a & a \end{bmatrix}.$$

After the Fourier transform we do not get a map given by monomials—see Appendix 1. However, the general symmetric group-based models belong to the class of G-models. In particular, we can provide an explicit toric description.

Proposition 4.69. *General symmetric group-based models are special G-models.*

Proof. Suppose that H is any abelian group. Let G be a semidirect product of H by $\mathbb{Z}_2$ where the action of $1 \in \mathbb{Z}_2$ on h gives h^{-1}. In this case the subspace invariant with respect to the G action gives the general symmetric group-based model. ∎

There are two abelian groups of order 4. For $\mathbb{Z}_2 \times \mathbb{Z}_2$ the general symmetric group-based model is the same as the general group-based model, and is the 3-Kimura model. For $\mathbb{Z}_4$ the general symmetric group-based model is the 2-Kimura model. Notice however that the class of general symmetric group-based models does not include Jukes–Cantor on four states, which is a G-model. It can be obtained for example by an embedding of $\mathbb{Z}_2 \times \mathbb{Z}_2$ in S_4 as a normal subgroup; more precisely, as $\{\mathrm{id}; (12)(34); (13)(24); (14)(23)\}$.

We would like to finish this subsection by restating the results of Sturmfels and Sullivant obtained for group-based models, in the case of G-models. We have seen that to each tree T and a G-model we can associate a polytope P. Fix a group G with a normal abelian subgroup H. The polytope P defines a projective toric variety as described in Section 2, and this is the variety representing the model. For the general group-based model the points of P correspond to networks (Definition 4.24), that is, special assignments of characters of a group to edges of the tree. Using the labelling method we identify two networks if for each edge the associated characters are in the same orbit of the adjoint action of G on H^*.

Definition 4.70 (Join of two trees, split of a tree into two subtrees). Fix a tree T with an inner edge $e = (v_1, v_2)$. We distinguish two subsets S_1 and S_2 of vertices of T. The set S_1 contains all descendants of v_1, including v_1. The set S_2 contains all vertices that are not descendants of v_2, including v_2. Let T_1 and T_2 be induced subtrees of T with vertices given respectively by S_1 and S_2. Note that the edge e is a distinguished leaf both in T_1 and T_2. One can specify the roots of T_1 and T_2 arbitrarily. The canonical choice is to take respectively v_1 and v_2.

We call the trees T_1 and T_2 the *split* of T. The tree T is a *join* of T_1 and T_2 (with a distinguished edge e).

Friendly labellings allow describing the polytope associated to T as a fiber product of the polytopes associated to T_1 and T_2. In particular we can give a description of the polytope of any tree knowing just the polytopes associated to claw trees [Sul07], [KR14].

Recall that the polytope associated to the tree T is contained in the lattice $M_{E,G}$ with the basis given by pairs (k, o), where k is an edge of T and o is an orbit of the adjoint G action on H.

FACT 4.71 ([Sul07, Theorem 12], [SS05, Theorem 23]). *Let T be a join of two trees T_1 and T_2 with a distinguished edge e. Let M be the lattice associated to the tree T. Consider a G-model associated to a group G with a normal abelian subgroup H. Let M_1 and M_2 be the corresponding lattices for the trees T_1 and T_2. Let M_e be the lattice generated by the basis elements (e, o), where o is any orbit of the adjoint G action on H and e is a fixed edge. There are natural projections $p_1 : M_1 \to M_e$ and $p_2 : M_2 \to M_e$.*

The polytope associated to the tree T is a fiber product over the projections p_1 and p_2 of the polytopes associated to the trees T_1 and T_2.

4.6. Normality of G-models. We have seen that the models associated to a group containing a normal, abelian subgroup are toric. The monomial parametrization map is sufficient for applications. However, for an algebraic geometer this would not be enough, as one would also need to prove the normality of these varieties. We will now address this problem. By normality we will mean projective normality, that is, normality of the affine cone equivalent to normality of polytopes. We will see that in general one cannot expect a G-model to be normal, but in many cases it is. First let us start with a technical lemma. Different versions of it that worked only for polytopes with a unimodular cover were presented in [BW07] and [Zwi]. Recently these results were generalized in [EKS14].

LEMMA 4.72. *Let P_1 and P_2 be two normal polytopes contained respectively in lattices L_1 and L_2 spanned by the points of the polytopes. Suppose that we have morphisms $p_i : L_i \to L$ of lattices for $i = 1, 2$ such that $p_i(P_i) \subset S$, where S is a standard simplex (the convex hull of the standard basis). Then the fiber product $P_1 \times_L P_2$ is normal in the lattice spanned by its points.*

Proof. Let $q \in n(P_1 \times_L P_2)$ for some positive integer n. Let q_i be the projection of q to L_i. Suppose q is in the lattice spanned by the points of $P_1 \times_L P_2$. Hence q is equal to the sum of points that belong to $P_1 \times_L P_2$ with integral coefficients summing to n. We know that it is in the convex hull of $n(P_1 \times_L P_2)$. Hence each q_i is the sum of points that belong to P_i with coefficients summing to n, and is in the convex hull of nP_i. This means that $q_i \in nP_i \cap L_i$. From the assumptions we obtain

$$q_i = \sum_{j=1}^{n} v_j^i,$$

with each $v_j^i \in P_i$. We also know that $p_1(q_1) = p_2(q_2)$, and this is an element of nS. Moreover, $p_i(v_j^i) \in S$. Notice that each element of nS can be *uniquely* written as the sum of n elements of S. This means that the collections $(p_1(v_1^1), \dots, p_1(v_n^1))$ and $(p_2(v_1^2), \dots, p_2(v_n^2))$ are the same up to permutation, so we can assume that $p_1(v_j^1) = p_2(v_j^2)$. Thus we can lift each pair (v_j^1, v_j^2) to a point $v_j \in P_1 \times_L P_2$ that projects respectively to v_j^1 and v_j^2. One obtains $q = \sum_{j=1}^n v_j$, which completes the proof. ∎

By Fact 4.71 the polytope associated to a tree with more than one inner vertex is the fiber product of polytopes associated to trees with a strictly smaller number of inner vertices. By Lemma 4.72, if we want to prove normality of a polytope associated to any trivalent tree we only have to consider normality of a polytope for a tripod. More generally, if we want to prove normality of a polytope associated to a tree with vertices of valency less than or equal to m we have to check the normality of polytopes associated to claw trees with at most m leaves.

PROPOSITION 4.73. *Let us consider a trivalent tree. The G-models for the abelian groups $\mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2$, $\mathbb{Z}_3$ and $\mathbb{Z}_4$ are normal.*

Proof. One can find the polytopes for the tripod and check their normality using the Macaulay computer program [GS]. The proposition then follows from Lemma 4.72. ∎

PROPOSITION 4.74. *The polytope of the 2-Kimura model for the tripod is not normal. Moreover, the projective variety associated to the model is not normal.*

Proof. As the second part of the statement is stronger, we prove only that part. The polytope P of the 2-Kimura model has vertices

$$\begin{array}{ll}
1)\ (1,0,0,1,0,0,1,0,0), & 2)\ (1,0,0,0,1,0,0,1,0),\\
3)\ (0,1,0,1,0,0,0,1,0), & 4)\ (0,1,0,0,1,0,1,0,0),\\
5)\ (1,0,0,0,0,1,0,0,1), & 6)\ (0,0,1,1,0,0,0,0,1),\\
7)\ (0,0,1,0,0,1,1,0,0), & 8)\ (0,1,0,0,1,0,0,0,1),\\
9)\ (0,1,0,0,0,1,0,1,0), & 10)\ (0,0,1,0,1,0,0,1,0).
\end{array}$$

Let $Q = (1,0,0,1,0,0,1,0,0)$ be a vertex of P. By Fact 2.16 it is enough to prove that the monoid C generated by the integral points of $P - Q$ is not saturated. Let us consider the cone $\tilde{C}$ that is the saturation of C. The point $L = (-1,0,1,-1,0,1,-1,0,1)$ is in C, as $2L$ is equal to

$$(-1,0,1,-1,0,1,0,0,0) + (-1,0,1,0,0,0,-1,0,1) + (0,0,0,-1,0,1,-1,0,1).$$

The point L is also in the lattice spanned by the vertices as

$$\begin{aligned}L = {} & (0,1,0,0,1,0,0,0,1) - (0,1,0,0,1,0,1,0,0) + (0,1,0,0,0,1,0,1,0)\\ & - (0,1,0,0,0,1,0,1,0) + (0,0,1,0,1,0,0,1,0) - (0,0,1,0,1,0,0,1,0).\end{aligned}$$

However, it is not an integral sum with positive coefficients of the vertices of $P - Q$. Indeed, each vertex of $P - Q$ with 0 on the second, fifth and eighth coordinate has an even sum of third, sixth and ninth coordinates. However, the sum of these coordinates for L is odd. ∎

In a joint work with Maria Donten-Bury [DBM12] we managed to get further results. Using the implementation of Algorithm 1 one can obtain the set of vertices of the polytope related to the group under study and the tripod. We applied Polymake [GJ00] to check the normality of this polytope (in the lattice generated by its vertices). We obtained:

COMPUTATION 4.75. The polytope associated with the G-model for the tripod and the group $G = H = \mathbb{Z}_6$ is not normal. Hence the affine algebraic variety representing this model is not normal.

In particular, the class of abelian models contains nonnormal models. We believe it can be difficult to characterize the class of groups for which G-models are normal, or even to determine a big (infinite) class of normal, toric G-models. On the other hand, one has the following result:

PROPOSITION 4.76. *Let T be a phylogenetic tree and let G_1 be a subgroup of an abelian group G_2. If the variety corresponding to the tree T and the group G_1 is not normal, then the variety corresponding to the tree T and the group G_2 is not normal either.*

Proof. Let M_i be a lattice whose basis is indexed by pairs of an edge of the tree and an element of the group G_i. The inclusion $G_1 \subseteq G_2$ gives a natural injective morphism $f : M_1 \to M_2$. Let $P_i \subset M_i$ be the polytope associated to the model for the tree T and the group G_i. Let $\tilde{M}_i \subset M_i$ be a sublattice spanned by vertices of the polytope P_i.

As P_1 is not normal in the lattice spanned by its vertices, there exists a point $x \in nP_1 \cap \tilde{M}_1$ that is not a sum of n vertices of P_1. Let $y = f(x)$. The vertices of P_1 are mapped to vertices of P_2. We see that $y \in nP_2 \cap \tilde{M}_2$. If P_2 were normal in $\tilde{M}_2$, we would be able to write $y = \sum_{i=1}^{n} q_i$ with $q_i \in P_2$.

Notice that each point in $f(M_1)$ has a zero entry for each coordinate indexed by any edge and any element $g \in G_2 \setminus G_1$. In particular y has zero entries for these coordinates. As all entries of all vertices of P_2 are nonnegative, this proves that all entries indexed by any edge and any $g \in G_2 \setminus G_1$ are zero for q_i. However, vertices of P_2 that have nonzero entries for all coordinates indexed by an edge and $g \in G_1$ are in the image of P_1. Hence $q_i = f(p_i)$ for some $p_i \in P_1$. We see that $x = \sum p_i$, which is impossible. ∎

In particular, all abelian groups G such that $|G|$ is divisible by 6 give rise to non-normal models.

5. Description of the variety using the group action

Let us describe precisely the characters of the torus that is the dense orbit of the variety associated to a model. Let us fix a tree T and an abelian group H. We have the following diagram:

$$\begin{tikzcd} \widehat{\psi} : M_S \arrow[r, two heads] & \widehat{M_E} \subset M_E \\ M_{S,0} \arrow[u, hook] \arrow[ur] & \end{tikzcd}$$

Let us define a sublattice of M_E.

Definition 5.1 (M_{deg}).

$$M_{\mathrm{deg}} = \{m \in M_E : \deg_{e_1}(m) = \deg_{e_2}(m) \text{ for all } e_1, e_2 \in E\}.$$

Proposition 5.2. *The lattice $\widehat{M_E}$ is contained in the sublattice M_{deg}.*

Proof. For any basis element $b \in M_S$ corresponding to a socket and for any edge $e \in E$ we have $\deg_e(\widehat{\psi}(b)) = 1$. Hence the image of any element of M_S satisfies the relations in the definition of M_{deg}. ■

Of course the elements of $\widehat{M_E}$ satisfy more relations. We will describe them now.

Definition 5.3 (Morphism $\mathfrak{add}$). There is a natural surjective group morphism $\mathfrak{add} : M_E \to (H^*)^N$. For a node $n \in N$ let $p_n : (H^*)^N \to H^*$ be the projection onto the corresponding factor. Let $f_{e,\chi} \in M_E$ be a basis element corresponding to an edge e and a character $\chi \in H^*$. We define

$$p_n(\mathfrak{add}(f_{e,\chi})) = \begin{cases} \chi_0 & \text{if } n \text{ is not adjacent to } e, \\ \chi & \text{if } e \text{ is an edge incoming to } n, \\ -\chi & \text{if } e \text{ is an edge outgoing from } n, \end{cases}$$

where χ_0 is the neutral character.

We say that an element $m \in \widehat{M_E}$ *has a trivial sum around a node* n if $p_n(\mathfrak{add}(m)) = \chi_0$.

Consider the composition $\mathfrak{add} \circ \widehat{\psi}$. Let $s \in M_S$ be a basis element corresponding to a network $\tilde{s} \in \mathfrak{N} \subset (H^*)^E$. We have $\mathfrak{add} \circ \widehat{\psi}(s) = add(\tilde{s})$. However, by Definition 4.24 we have $add(\tilde{s}) = 0$, hence $\mathfrak{add} \circ \widehat{\psi}(s) : M_S \to (H^*)^N$ is zero. This means that $\widehat{M_E}$ is contained in the kernel of the morphism $\mathfrak{add}$.

We will prove that there is an exact sequence

$$0 \to \widehat{M_E} \to M_{\mathrm{deg}} \to (H^*)^N \to 0,$$

where the last morphism is the restriction of $\mathfrak{add}$ to M_{deg}. In particular the ranks of $\widehat{M_E}$ and M_{deg} are equal.

Corollary 5.4. *The dimension of the affine variety associated to the model is equal to the dimension of the dense torus orbit, that is,*

$$\dim \widehat{M_E} = \dim M_{\mathrm{deg}} = (|H| - 1)|E| + 1.$$

The dimension of the projective variety equals $(|H| - 1)|E|$.

We have to prove the following lemma.

Lemma 5.5. *Every element of M_{deg} that is in the kernel of $\mathfrak{add}$ belongs to $\widehat{M_E}$.*

Proof. We proceed by induction on the number of inner vertices of the tree. First let us assume that the tree T is a claw tree with l leaves. The elements of M_{deg} can be described by sequences of length l given by elements $(\sum a_\chi^1 \chi, \dots, \sum a_\chi^l \chi)$ with $\sum a_\chi^1 = \dots = \sum a_\chi^l$.

We prove that elements of the form $(g_1 + g_2 - g_1 g_2 - \chi_0, 0, \dots, 0)$, where $g_1, g_2 \in H^*$ are any characters, are in $\widehat{M_E}$. Such an element is equal to $(g_1, g_1^{-1}, \chi_0, \dots, \chi_0) + (g_2, \chi_0, g_2^{-1}, \chi_0, \dots, \chi_0) - (g_1 g_2, g_1^{-1}, g_2^{-1}, \chi_0, \dots, \chi_0) - (\chi_0, \dots, \chi_0)$. Each element of the sum is given by a socket, hence it is in $\widehat{M_E}$.

We now fix any element $(\sum a^1_\chi \chi, \dots, \sum a^l_\chi \chi) = m \in M_{\deg}$ that is in the kernel of $\mathfrak{add}$. We will reduce it modulo the image of M_S to zero. Let us assume that $\sum a^1_\chi = \dots = \sum a^l_\chi = d$.

Using elements as above we can reduce m and assume that for $\chi \neq \chi_0$ the coefficient a^j_χ for each $1 \le j \le l$ is zero apart from one character for each j for which the coefficient can be one. More precisely, if there are two characters with a positive (resp. negative) coefficient, we can replace them with their sum plus (resp. minus) the trivial character. If one entry is equal to $g_1 - g_2$ we add $g_2 + g_1 g_2^{-1} - g_1 - \chi_0$. If there is one negative g on an entry we add $g + g^{-1} - 2\chi_0$.

In other words, m is equal to $(\chi_1, \dots, \chi_l) + (d-1)(\chi_0, \dots, \chi_0)$ modulo the image of M_S. As $\sum \chi_j = \chi_0$ in H^*, this element is in the image of M_S.

Now we will prove the induction step. Let us fix a tree T with at least two inner vertices. We may choose an inner edge e of T such that cutting along e we obtain two trees T_1 and T_2 (the tree T is a join of T_1 and T_2) with strictly lower number of inner vertices. In one of the trees, say T_2, we have to choose a root—this will be a vertex belonging to the edge e. In this way all edges of T_2 are oriented as in T apart from e which has an opposite direction. An element $m \in M_{\deg}$ gives two elements $m_i \in M^i_{\deg}$ for $i = 1, 2$ that are also in the kernels of $\mathfrak{add}$ for both trees. By induction hypothesis we can find two elements $s_i \in M^i_S$ whose images give m_i. Let $s_i = \sum c^i_j b^i_j$ where b^i_j is the basis of M^i_S corresponding to sockets on T_i. Let us consider the multisets Z_i that are the projections of $\sum c_j b^i_j$ onto the edge e—each b_j distinguishes an element on e. The multiset Z_i has c_j elements distinguished by b^i_j with a minus sign if $c_j < 0$. Z_i is a signed multiset of characters. Let Z'_i be a multiset obtained by reductions cancelling χ with $-\chi$ in Z_i. The multiset Z'_1 is just the signed multiset of characters corresponding to m_e. The multiset Z'_2 gives the same multiset as Z'_1 if we inverse all characters. This means that we can pair together elements from Z'_1 and Z'_2 so that each pair gives rise to a socket on the tree T. The image of the sum of these sockets does not have to give m yet. We have to lift also the sockets that we cancelled by passing from Z_i to Z'_i. This is done as follows. Suppose that two sockets b_1 and b'_1 give χ on the edge e, and so b_1 and $-b'_1$ cancel each other in Z_1. We choose any socket s on T_2 that gives χ^{-1} on the edge e. We can glue together b_1 and s obtaining a socket (b_1, s) of the tree T, and analogously (b'_1, s). The image of the difference of the sockets $(b_1, s) - (b'_1, s)$ on the edges of the tree T_1 is the same as the difference $b_1 - b'_1$ and zero on the edges belonging to T_2. In this way we obtain the socket of T whose image agrees with $\sum c_j b^i_j$ on T_i, hence is equal to m. ∎

Corollary 5.6. *For the tree T and the group H, the dense torus orbit of the affine variety representing the model has a natural description as a quotient of the dense orbit of the torus of the parameter space by the $H^N \times (\mathbb{C}^*)^{|E|-1}$ action.*

Proof. The characters of the dense orbit of the parameter space are given by the lattice M_E. Its algebra is $\mathbb{C}[M_E] = \mathbb{C}[x^{\pm 1}_{(e,\chi)}]_{e \in E,\, \chi \in H^*}$. First let us describe the action of $Gr = (\mathbb{C}^*)^{|E|-1}$. We regard this torus as a subtorus of $(\mathbb{C}^*)^{|E|}$ with an additional condition that the product of all coordinates is one. Hence an element of Gr is just an assignment of a nonzero complex number to each edge of the tree T so that the product of all these numbers is one. The action of Gr just multiplies $x_{(e,\chi)}$ by the complex number associated to e. In this way the invariant monomials are those whose degree with respect to each edge is the same, hence $M_E^{Gr} = M_{\deg}$.

The coordinates of the group H^N are indexed by nodes. There is a natural diagonal action of H^N on the algebra $\mathbb{C}[M_E]$. Fix a node $v \in N$. The action of $h \in H$ considered as an element of H^N, equal to h on the coordinate indexed by v and the neutral element on the other coordinates, is as follows:

- for an edge e incoming to v we have $h(x_{(e,\chi)}) = \chi(h)x_{(e,\chi)}$,
- for an edge e outgoing from v we have $h(x_{(e,\chi)}) = (\chi(h))^{-1}x_{(e,\chi)}$,
- for the other edges $h(x_{(e,\chi)}) = x_{(e,\chi)}$.

First notice that elements of $\widehat{M_E}$ are invariant by the action of H^N. They are in the kernel of $\mathfrak{add}$, so the signed sum of characters around each inner vertex gives a trivial character. But the action of $h \in H \subset H^N$ just multiplies the monomial by the value at h of the character that is a signed sum of characters associated to edges adjacent to v, hence by 1. Conversely, if the signed sum of characters at any $h \in H$ is 1, the sum has to be a trivial character. So an element of $M_{\deg}$ is invariant with respect to the H^N action if and only if it is in the kernel of $\mathfrak{add}$, so by Lemma 5.5 if and only if it belongs to $\widehat{M_E}$. ∎

Note that the group $H^N \times (\mathbb{C}^*)^{|E|-1}$ acts also on the algebra of the parameter space $\mathbb{C}[x_{(e,\chi)}]_{e\in E,\,\chi\in H^*}$. However, the quotient is not equal to the variety representing the model, contrary to what is stated in [CFS08, Theorem 3.6]. Indeed, the algebra of the variety is generated by the polytope (contained in the positive quadrant of $M_{\deg}$) and is invariant by the action of $H^N \times (\mathbb{C}^*)^{|E|-1}$. However, the invariant monomials of $\mathbb{C}[x_{(e,\chi)}]_{e\in E,\,\chi\in H^*}$ correspond to all the monomials of $\widehat{M_E}$ that are in the positive quadrant of M_E. Not all such monomials are generated by the polytope. For example, for the 3-Kimura model the monomial $x_{e_0,\chi}^2 \prod_{e_i\in E} x_{e_i,e}^2$, where e is the trivial character, is invariant for any χ and any distinguished edge e_0 (because $\chi + \chi = e$). This is not, however, the sum of any two vertices of the polytope associated to the variety.

Let us present some applications.

Corollary 5.7. *There is an exact sequence of groups*

$$M_{S,0} \to M_{E,0} \to (H^*)^{|N|} \to 0,$$

where the first map is given by $\widehat{\psi}$, and the second is the restriction of $\mathfrak{add}$ to $M_{E,0}$. ∎

This corollary can be applied in the identifiability problem to determine the parameters of transition matrices. We will do this in Section 10.4.1.

Fix an abelian group H and a tree T. We will prove that the group $\mathfrak{N}$ of networks acts on the variety $X(T,G)$. Recall that the ambient space $\widetilde{W_L}$ is a regular representation of $\mathfrak{N}$.

Proposition 5.8. *The action of the group $\mathfrak{N}$ of networks on $\widetilde{W_L}$ restricts to the variety $X(T,G)$.*

Proof. Consider the parametrization morphism $\pi_L \circ \widehat{\psi} : \mathbb{C}^{|E|\,|H^*|} \to \widetilde{W_L}$. The basis vectors of the affine space $\mathbb{C}^{|E|\,|H^*|}$ are indexed by pairs $(e,\chi) \in E \times H^*$. We denote the corresponding basis elements by $b_{(e,\chi)}$. For $t \in \mathbb{C}^{|E|\,|H^*|}$ we define $t_{(e,\chi)} := b^*_{(e,\chi)}(t)$. The basis elements of $\widetilde{W}_L$ are indexed by networks $n \in (H^*)^E$. We identify a network with a sequence of characters $n = (n_e := \chi_e)_{e \in E}$ indexed by edges. Note that the group of networks acts also on the domain $\mathbb{C}^{|E|\,|H^*|}$ by

$$(n(t))_{(e,\chi)} := t_{(e,n_e^{-1}\chi)}.$$

It is easy to check that the morphism $\pi_L \circ \widehat{\psi}$ is equivariant. ■

6. Phylogenetic invariants

This section contains results of joint work with Maria Donten-Bury [DBM12]. We investigate the most important objects of phylogenetic algebraic geometry—ideals of phylogenetic invariants. The main problem in this area is to give an effective description of the whole ideal of the variety associated to a given model on a tree. Our task is to find an efficient way to compute generators of these ideals.

We suggest a way of obtaining all phylogenetic invariants of a claw tree of a G-model—more precisely, we conjecture that our invariants generate the whole ideal of the variety. These, together with Fact 4.71, could provide an algorithm listing all generators of the ideal of phylogenetic invariants for any tree and for any G-model (so in particular for a general group-based model).

6.1. Inspirations. The inspirations for our method were the conjectures made by Sturmfels and Sullivant [SS05]. They are still open but, as we will see, they strongly support our ideas. In particular, we will prove later that our algorithm listing the generators of the ideal works for the 3-Kimura model if we assume that the weaker conjecture made in [SS05] holds.

First we introduce some notation. As before, let $K_{n,1}$ be a claw tree with n leaves. Let $\phi(G, n) = d$ be the least natural number such that the ideal associated to $K_{n,1}$ for the group-based model G is generated in degree d. The phylogenetic complexity of the group G is defined as $\phi(G) = \sup_n \phi(G, n)$. Note that by [SS05, Theorem 23] (see also [Sul07, Theorem 12]) the number $\phi(G, n)$ bounds the degree in which the ideal associated to any tree of valency at most n is generated. Based on numerical results, Sturmfels and Sullivant suggested the following conjecture:

Conjecture 6.1. *For any abelian group G we have $\phi(G) \le |G|$.*

This conjecture was separately stated for the 3-Kimura model, that is, for $G = \mathbb{Z}_2 \times \mathbb{Z}_2$.

Still very little is known about the function ϕ apart from the case of the binary Jukes–Cantor model (see also [CP07]):

Proposition 6.2 (Sturmfels and Sullivant [SS05]). *For the binary Jukes–Cantor model, $\phi(\mathbb{Z}_2) = 2$.* ■

There are also some computational results—to the table in [SS05] presenting the computations made by Sturmfels and Sullivant a few cases can be added.

Using `4ti2` software [tt] we obtained the following:

COMPUTATION 6.3.

- $\phi(\mathbb{Z}_3, 6) = 3$,
- $\phi(\mathbb{Z}_5, 4) = 4$,
- $\phi(\mathbb{Z}_8, 3) = 8$,
- $\phi(\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2, 3) = 8$.

For the 3-Kimura model we do not even know whether the function ϕ is bounded. As we will see later, this conjecture is strongly related to the one stated in the next section.

6.2. A method for obtaining phylogenetic invariants. We propose a method that is inspired by the geometry of the varieties we consider. First we have to introduce some notation.

DEFINITION 6.4. Let V_i be the set of vertices of a tree T_i for $i = 1, 2$. Let e be an inner edge of T_2 joining $v_1, v_2 \in V_2$. We say that the tree T_1 is obtained from T_2 by *contraction of an edge* e if:

- $V_1 = \{v\} \cup (V_2 \setminus \{v_1, v_2\})$,
- for $w \in V_1 \setminus \{v\}$ a pair (v, w) is an edge of T_1 if and only if (v_1, w) or (v_2, w) is an edge of T_2,
- for $w \in V_1 \setminus \{v\}$ a pair (w, v) is an edge of T_1 if and only if (w, v_1) or (w, v_2) is an edge of T_2,
- for $w, u \in V_1 \setminus \{v\}$ a pair (w, u) is an edge of T_1 if and only if (w, u) is an edge of T_2.

In such a situation we say that T_2 is a *prolongation* of T_1.

REMARK 6.5. Note that these definitions are not the same as the definitions of flattenings introduced in [AR08] and further studied in [DK09].

Assume that we are in an abelian case, that is, we are dealing with a general group-based model. Using Algorithm 1 one can see that vertices of the polytope correspond to sockets. As explained in Section 2, vertices of the polytope correspond to coordinates of the ambient space of the variety. In this setting the variety $X(T_1)$ associated to the tree T_1 is in a natural way a subvariety of $X(T_2)$. Notice that we can identify sockets of both varieties, as we may identify their leaves, so both varieties are contained in $\mathbb{P}^s$, where s is the number of sockets. The natural inclusion corresponds to a projection of the character lattices: we forget all the coordinates corresponding to the edge joining the vertices v_1 and v_2. The details are presented in Proposition 7.1. In this setting the following conjecture is natural:

CONJECTURE 6.6. *The variety $X(K_{n,1})$ is equal to the (scheme-theoretic) intersection of all the varieties $X(T_i)$, where T_i is a prolongation of $K_{n,1}$ that has only two inner vertices, both of valency at least three.*

As $X(K_{n,1})$ is a subvariety of $X(T_i)$ for any prolongation T_i, one inclusion is obvious. Note also that the valency assumption is made, because otherwise the conjecture would be obvious: one of the varieties that we intersect would be equal to $X(K_{n,1})$ by Remark 4.36.

All T_i have a strictly smaller maximal valency than $K_{n,1}$, so if the conjecture holds then we can inductively use Theorem 23 of Sturmfels and Sullivant [SS05] (see also [Sul07, Theorem 12]) to obtain all phylogenetic invariants for a given model for any tree of any valency, knowing just the ideal of the tripod. In such a case the ideal of $X(K_{n,1})$ is just the sum of ideals of trees with smaller valency. More precisely, if Conjecture 6.6 holds then the degree in which the ideals of claw trees are generated cannot grow when the number of leaves increases. This means that $\phi(G) = \phi(G, 3)$, which can be computed in many cases. In particular, Conjecture 6.6 implies all cases of Conjecture 6.1 in which we can compute $\phi(G, 3)$—this includes the most interesting 3-Kimura model.

Remark 6.7. Note that the varieties $X(T_1)$ and $X(T_2)$ are naturally contained in the same ambient space for any model, even if it does not give rise to toric varieties. Indeed, using the construction of the variety presented in Section 3 one can see that the ambient space depends only on the leaves of the tree. Hence if we can identify the leaves of trees, we can identify the ambient spaces of the associated varieties. Thus Conjecture 6.6 can help to compute the ideals of claw trees for a large class of phylogenetic models.

Of course one may argue that Conjecture 6.6 above is too strong to be true. Later we will prove it for the binary Jukes–Cantor model. We will also consider two weakenings of this conjecture, which can still have a lot of applications. The first modification just states that Conjecture 6.6 holds for n large enough.

Proposition 6.8. *For any G-model Conjecture* 6.6 *holds for n large enough if and only if the function ϕ is bounded.*

Proof. One implication is obvious. Suppose that Conjecture 6.6 holds for $n > n_0$. We choose d such that the ideals associated to $K_{l,1}$ are generated in degree m for $l \leq n_0$. Using Conjecture 6.6 and the results of [SS05], we can describe the ideal associated to $K_{n,1}$ as the sum of ideals generated in degree m. It follows that this ideal is also generated in degree m, so the function ϕ is bounded by m.

For the other implication assume that $\phi(n) \leq m$. Let us consider any binomial B that is in the ideal of the claw tree and is of degree less than or equal to m. We prove that B belongs to the ideal of some prolongation of a tree T, which is in fact more than stated in Conjecture 6.6.

Such a binomial can be described as a linear relation between (at most m) vertices of the polytope of the variety. Each vertex is given by an assignment of orbits of characters to edges such that there exist representatives of orbits that sum to a trivial character. Let us fix such representatives, so that each vertex is given by n characters summing to a trivial character.

Now the binomial B can be presented as a pair of matrices A_1 and A_2 with characters as entries. Each column of the matrices is a vertex of the polytope. The matrices have at most m columns and exactly n rows. Consider the matrix $A = A_1 - A_2$, that is, the entries of the matrix A are the characters that are differences of entries of A_1 and A_2. We can subdivide the first column of A into groups of at most $|H|$ elements summing to a trivial character. Then inductively we can subdivide the rows into groups of at most $|H|^i$ elements summing to a trivial character in each column up to the ith one.

For $n > |H|^m + 1$ we can find a set S of rows of A such that the characters sum to a trivial character in each column restricted to S, and both the cardinality of S and of its complement are greater than 1. Note that the sums of the entries lying in a chosen column and in the rows in S are the same in A_1 and A_2. Therefore, adding to both matrices an extra row whose entries are equal to the sum of the entries in S gives a representation of the binomial B on a prolongation of T. ∎

In particular, this proof shows that if Conjecture 6.1 of Sturmfels and Sullivant holds for the 3-Kimura model, then Conjecture 6.6 also holds for this model for $n > 257$. Later we will significantly improve this estimation.

For the second modification of Conjecture 6.6 let us recall a few facts on toric varieties. Let T_1 and T_2 be two tori with lattices of characters respectively M_1 and M_2. Assume that both tori are contained in a third torus T with character lattice M. The inclusions give natural isomorphisms $M_1 \simeq M/K_1$ and $M_2 \simeq M/K_2$, where K_1 and K_2 are torsion free lattices corresponding to the characters that are trivial when restricted respectively to T_1 and T_2. The ideal of each torus (inside the algebra of the big torus) is generated by binomials corresponding to such trivial characters. The points of T are given by monoid morphisms $M \to \mathbb{C}^*$. The points of T_i are those morphisms that associate 1 to each character from K_i. We see that the points of $T_1 \cap T_2$ are those morphisms $M \to \mathbb{C}^*$ that associate 1 to each character from the lattice $K_1 + K_2$. Of course the (possibly reducible) intersection Y is generated by the ideal corresponding to $K_1 + K_2$. This lattice may not be saturated, but Y contains a distinguished torus T', namely one of its connected components. If K' is the saturation of the lattice $K_1 + K_2$, then the characters of T' are given by the lattice M/K'. Suppose that X is a toric variety that contains the dense torus orbit equal to T. Let X_i be the toric variety that is the closure of T_i, and X' be the closure of T' in X. We call the toric variety X' the *toric intersection* of X_1 and X_2. The definition extends to a greater number of toric varieties embedded equivariantly in one toric variety. The most important case that we will use is when X is the affine space and the X_i are affine toric varieties.

In the setting of Conjecture 6.6 we conjecture the following:

Conjecture 6.9. *The toric variety $X(T)$ is the toric intersection of all the toric varieties $X(T_i)$.*

This conjecture differs from the previous one by the fact that we allow the intersection to be reducible, with one distinguished irreducible component equal to $X(T)$. We state this conjecture because it can be checked using only the tori. As the points important from the biological point of view are contained in the torus (see [CFS08, Definition 2.13]), this conjecture is a weaker version of Conjecture 6.6 which is still suitable for applications. Moreover, it is quite easy to check it for trees with small enough number of leaves using computer programs. To explain this properly, let us consider the following general setting.

Assume that the tori T_i are associated to polytopes P_i, and that T is just the torus of the projective space $\mathbb{P}^n \supseteq T_i$ consisting of the points with all coordinates different from zero. Let A_i be a matrix whose columns represent the vertices of the polytope P_i. The characters trivial on T_i, or respectively binomials generating the ideal of T_i, are exactly represented by integer vectors in the kernel of A_i. The characters trivial on the intersection are given by integer vectors in the sum of lattices $\ker A_1 + \ker A_2$.

Note that the ideal of the toric intersection T' of the tori T_i in T is generated by binomials corresponding to characters trivial on T', that is, by the saturation of $\ker A_1 + \ker A_2$. These binomials define a toric variety in $\mathbb{P}^n$. This variety is contained in the intersection (in fact it is a toric component) of the toric varieties that are the closures of T_i. Equality may not hold however, as the intersection might be reducible.

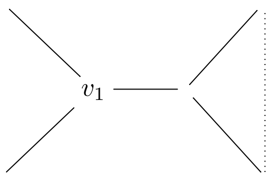
In Conjecture 6.9 we have to compare two tori, one contained in the other. To do this, it is enough to compare their dimensions, that is, the ranks of the character lattices. Note that the dimension of $T_1 \cap T_2$ is n minus the dimension (as a vector space) of $\ker A_1 + \ker A_2$, as it is equal to the rank of the lattice $\mathbb{Z}^n \cap (\ker A_1 + \ker A_2)$. To compute this dimension it is enough to compute the ranks of the matrices A_1, A_2 and B, where B is the matrix obtained by putting A_1 under A_2 (that is, $\ker B = \ker A_1 \cap \ker A_2$). This can be done very easily using GAP [GAP]. The results obtained for small trees will be used in the following section.

6.3. Main results. To support Conjecture 6.6 let us consider the case of the binary Jukes–Cantor model. This model is well understood [BW07], [CP07], [SS05].

PROPOSITION 6.10. *Conjecture* 6.6 *holds for the binary Jukes–Cantor model.*

Proof. We use the same notation as in the proof of Proposition 6.8.

Let us fix the number of leaves l. We claim that we can find two special trees T_1 and T_2 for which the scheme-theoretic intersection $X(T1, \mathbb{Z}_2) \cap X(T_2, \mathbb{Z}_2)$ equals $X(K_{l,1}, \mathbb{Z}_2)$. We number the leaves from 1 to l. The trees T_1 and T_2 are isomorphic as graphs but have different leaf labelling. The topology of the trees is as follows:



For the tree T_1 the leaves adjacent to v_1 have numbers 1 and 2. For the tree T_2 they are numbered 1 and 3. The ideal of the variety associated to a tree for the group $\mathbb{Z}_2$ is always generated in degree 2 by Proposition 6.2. Hence the generators of the ideals are of the form $n_1 n_2 = n_3 n_4$ where n_i for $1 \le i \le 4$ are the coordinates corresponding to networks. Each binomial equality corresponds to a pair of matrices (M_0, M_1), with entries that are group elements, whose columns represent networks and rows are the same up to permutation. Hence each generator of the ideal of $X(K_{l,1}, \mathbb{Z}_2)$ is represented by a pair of $2 \times l$ matrices with entries from $\mathbb{Z}_2$. Moreover, the sum in each column is the neutral element, and the rows of both matrices are the same up to permutation. As we can permute the columns of each matrix, we may assume that the first rows of both matrices coincide. Consider any such generator (M_0, M_1) in the ideal of $X(K_{l,1}, \mathbb{Z}_2)$. First suppose that the entries in the first row are the same, that is, either 00 or 11.

Then the relation holds for both $X(T_1)$ and $X(T_2)$. Hence we may suppose that the first row is 01 or 10. If the second row were 00 or 11, then the relation would hold for $X(T_1)$. The same reasoning holds for the third row and $X(T_2)$. Hence all three rows in both matrices are either 01 or 10. If the second (resp. third) rows are the same in both matrices then the relation holds for $X(T_1)$ (resp. $X(T_2)$). So the only possibility left is that the second and third rows of M_1 are respectively the negations of the second and third rows of M_0. In this case the relation does not hold in any $X(T_i)$, but we can generate it. We consider a matrix M that is equal to M_0 with the first two rows permuted. The pair (M_0, M) represents a relation in $X(T_1)$. Moreover, the pair (M, M_1) represents a relation in $X(T_2)$. ∎

From the proof above it follows that in fact to obtain the variety of the claw tree for the binary Jukes–Cantor model it is enough to intersect two varieties corresponding just to three subdivisions. These subdivisions correspond to S containing exactly the first and second rows or the first and third rows. Note that it is not enough to intersect two varieties corresponding to *any* prolongations—see Section 7.

Now we prove the following conditional result for the 3-Kimura model:

PROPOSITION 6.11. *If Conjecture* 6.1 *of Sturmfels and Sullivant holds then Conjecture* 6.6 *holds for* $n > 8$.

Proof. We use the same notation as in Proposition 6.8. Consider any binomial of degree k represented by a pair of matrices (M_1, M_2) with entries given by group elements. Let $A = M_1 - M_2$, where minus is the group substraction. The matrix A has k columns with entries from $\mathbb{Z}_2 \times \mathbb{Z}_2$. Let A' be the matrix with $2k$ columns and entries from $\mathbb{Z}_2$, obtained from A by applying the two projections $\mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2$ to each entry. Recall that the matrices M_1 and M_2 had the same rows up to permutation. This means that also after each projection the rows were the same up to permutation. Note that the difference of two vectors with entries from $\mathbb{Z}_2$ that are the same up to permutation always has an even number of 1's. Thus if we consider any row of A' and either odd or even entries of this row, the number of 1's is always even.

Once again we may assume that the entries in the first row of A' are neutral elements, that is, they are zero. Let A'' be the matrix obtained by deleting the first row of A'. For each subset of rows of A'' we may consider a vector of length equal to the number of columns of A'', whose entries are given by sums of group elements from the subset. Note that this vector always has an even number of 1's both in even and odd columns. Because we assume Conjecture 6.1, the matrix A'' has at most eight columns. By the pigeonhole principle, if $n > 8$ then we can find two subsets of rows of A'' that are not complements of each other, such that their sum vector is the same. If we take the symmetric difference of these subsets, we obtain a proper, nonempty subset S of rows of A'', summing in each column to the neutral element. We add the first row of A' to S or its complement, so that both sets have more than one element. Thus we obtain a subdivision of the set of rows of A such that the given binomial is in the ideal of the tree corresponding to this division. ∎

For $n \leq 8$ we checked, using Polymake, 4ti2, Macaulay2 and GAP, that the toric intersection of the tori of the subdivisions gives the torus of the claw tree. We used the linear algebra described in the previous section. This proves that if Conjecture 6.1 holds for the 3-Kimura model, then Conjecture 6.9 holds. Moreover, in all the checked cases it was enough to consider just two subdivisions. This is not a coincidence, as we will prove in Section 10.

To summarize, we know that for the 3-Kimura model Conjecture 6.6 implies both Conjectures 6.9 and 6.1, and moreover Conjecture 6.1 implies Conjecture 6.9, and for $n > 8$ also Conjecture 6.6.

7. Interactions between trees and varieties

The ideas from the preceding sections are general. We can define an order on trees with l leaves as follows. We say that $T_1 \leq T_2$ if T_1 can be obtained from T_2 by a series of contractions of inner edges. Here by an *edge contraction* we mean identifying two vertices of a given edge as in Definition 6.4. The smallest tree with l leaves is the claw tree $K_{l,1}$ with one inner vertex. This is a part of a construction of the tree space [BHV01]. We fix an abelian group G.

PROPOSITION 7.1. *If $T_1 \leq T_2$ then $X(T_1, G) \subset X(T_2, G)$.*

Proof. Although the statement is very easy, we believe that the following discussion may be helpful to better understand the forthcoming sections. Both trees have the same number of leaves, so we can make a natural bijection between their sockets. This gives an isomorphism of the ambient spaces $\widetilde{W_E}$. As $T_1 \leq T_2$, we can make an injection from the edges of T_1 to the edges of T_2. Note that a network on T_2 restricted to the edges of T_1 is a network on T_1. This gives a projection $\pi : M_E^{T_1} \twoheadrightarrow M_E^{T_2}$. The map π simply forgets the coordinates indexed by (e, g), where e is an edge of T_2 not corresponding to an edge of T_1. Moreover, the projection of P^{T_2} is equal to P^{T_1}. The following diagram commutes:

$$\begin{tikzcd} & M_E^{T_2} \arrow[d, two heads] \\ M_S \arrow[ur] \arrow[dr] & \\ & M_E^{T_1} \end{tikzcd}$$

Any relation between the vertices of P^{T_2} is also a relation between the vertices of P^{T_1}. Hence any polynomial in the ideal of $X(T_2, G)$ is also in the ideal of $X(T_1, G)$. ■

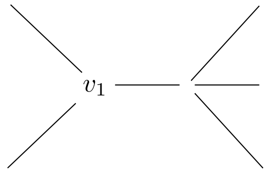
The surjective morphism of algebras corresponding to the inclusion of varieties is given by the restriction of the surjective morphism between $M_E^{T_2}$ and $M_E^{T_1}$ to the cones spanned by the polytopes P^{T_2} and P^{T_1}.

It is natural to ask what is the relation between $X(T_0, G)$ and the scheme-theoretic intersection of all $X(T, G)$ for $T_0 < T$. Conjecture 6.6 states that if there exists at least one $T > T_0$, then they are equal. So far we only know that the answer is positive for $G = \mathbb{Z}_2$ [CP07], [SS05], [DBM12].

Conjecture 6.6 can be stated for *any phylogenetic model*, not necessarily given by a group ([1]), in particular for a general Markov model. One would also be interested to know

([1]) I would like to particularly thank Elizabeth Allman for discussions on this topic.

exactly what is the intersection of a few varieties associated to different trees. In particular how many ideals do we have to sum to obtain the ideal associated to the claw tree? One could also hope that the intersection of $X(T_1, G)$ and $X(T_2, G)$ is equal to $X(T, G)$ where T is the largest tree smaller than T_1 and T_2. Here we present a counterexample. We will prove that the scheme-theoretic intersection $X(T_1, \mathbb{Z}_2) \cap X(T_2, \mathbb{Z}_2)$ does not have to be equal to $X(K_{l,1}, \mathbb{Z}_2)$ even if $K_{l,1}$ is the only tree smaller than T_1 and T_2. We consider the case of five leaves, $l = 5$. The trees T_1 and T_2 are isomorphic as graphs but have different leaf labelling. Their topology is as follows:



(7.1)

For the tree T_1 the leaves adjacent to v_1 have numbers 1 and 2. The tree T_2 is isomorphic, with two distinguished leaves labelled with 4 and 5. We consider the relation given by the pair of matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}.$$

This corresponds to a generator of the ideal of $X(K_{5,1}, \mathbb{Z}_2)$. Consider any relation involving the first matrix and some other matrix M for $X(T_1)$ or $X(T_2)$. One can see that the first two rows of M must be negations of each other, and the third one is 00. Hence it is impossible to generate the relation above.

8. Computational results

This section contains results of the joint work with Maria Donten-Bury [DBM12]. We used the implementation of Algorithm 1.

8.1. Hilbert–Ehrhart polynomials. The binary Jukes–Cantor model (for trivalent trees) has an interesting property, stated and proved in [BW07]: an elementary mutation of a tree gives a deformation of the associated varieties (see Construction 3.23). This implies that the binary Jukes–Cantor models of trivalent trees with the same number of leaves are deformation equivalent [BW07, Theorem 3.26]. As it was not obvious what to expect for other G-models, we computed the Hilbert–Ehrhart polynomials, which are invariants of deformation, in some simple cases.

Let us recall basic facts about the Hilbert polynomials for projective toric varieties. Suppose that our variety corresponds to a polytope $P \times \{1\}$ contained in the lattice M spanned by its integral points. There are two functions that one can associate to the polytope P.

(i) Let $h : \mathbb{N} \to \mathbb{N}$ where $h(n)$ is the number of points in the monoid generated by $P \times \{1\}$ with the last coordinate equal to n. We call h the *Hilbert function*.
(ii) Let $e : \mathbb{N} \to \mathbb{N}$ where $e(n)$ is the number of integral points in nP, or equivalently in $n(P \times \{1\})$. We call e the *Ehrhart function*.

The function e is a polynomial function, thus we call it the *Ehrhart polynomial*. The function h is a polynomial function for large enough values. The polynomial $\tilde{h}$ such that for n large enough $\tilde{h}(n) = h(n)$ is called the *Hilbert polynomial*. From the definition of normal polytope (Definition 2.13), we see that the Hilbert function equals the Ehrhart polynomial if and only if P is normal, that is, if and only if the associated variety is projectively normal. The associated variety is normal if and only if the Hilbert polynomial equals the Ehrhart polynomial [Stu96, Theorem 13.11]. In this case we call it the *Hilbert–Ehrhart polynomial*.

8.1.1. Numerical results. We checked models for two different trees with six leaves (this is the least number of leaves for which there are nonisomorphic trees, exactly two), the *snowflake* and the 3-*caterpillar*. The most interesting ones were the cases of the biologically meaningful 2-Kimura and 3-Kimura models.

To determine the Hilbert–Ehrhart polynomial of a G-model we compute the number of lattice points in multiples of its polytope. Even if it is not possible to get enough data to determine the polynomials (e.g. because numbers are too big), sometimes we can establish that the polynomials for two models are not equal, because their values for some n are different.

Before we completed our computations, Kubjas computed the numbers of lattice points in the third dilations of the polytopes for the 3-Kimura model on the snowflake and the 3-caterpillar with six leaves, and got respectively 69248000 and 69324800 points [Kub12]. Thus she proved that the varieties associated with these models are not deformation equivalent.

Our computations confirm her results for the 3-Kimura model, and also give the following

Computation 8.1. The varieties associated with 2-Kimura models for the snowflake and the 3-caterpillar trees have different Ehrhart polynomials. In the second dilations of the polytopes there are 56992 lattice points for the snowflake and 57024 for the 3-caterpillar.

Also, the pairs of varieties associated with G-models for the snowflake and the 3-caterpillar trees and

(i) $G = H = \mathbb{Z}_3$,
(ii) $G = H = \mathbb{Z}_4$,
(iii) $G = H = \mathbb{Z}_5$,
(iv) $G = H = \mathbb{Z}_7$

have different Hilbert–Ehrhart polynomials, and therefore are not deformation equivalent. (For these pairs the G-models are normal, which can be checked using Polymake.) The precise results of the computations are presented in Appendix 2.

In the cases of

(i) $G = H = \mathbb{Z}_8$,
(ii) $G = H = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$,
(iii) $G = H = \mathbb{Z}_9$

the varieties have different Hilbert functions. We have not been able to check if they are normal, but if they are then their Hilbert–Ehrhart polynomials are different.

8.2. Some technical details. The first attempt to compute the numbers of lattice points in dilations of a polytope was the direct method: constructing the list of lattice points in nP by adding the vertices of P to the lattice points in $(n-1)P$ and reducing repeated entries. This algorithm is not very efficient, but (after adding a few technical upgrades to the implementation) we were able to confirm Kubjas' results [Kub12]. However, this method does not work for nonnormal polytopes. As we planned to investigate the 2-Kimura model, we had to implement another algorithm.

The second idea is to compute inductively the relative Hilbert polynomials, i.e. the number of points in the nth dilation of the polytope intersected with the fiber of the projection onto the group of coordinates that correspond to a given leaf. Our approach is quite similar to the methods used in [Kub12] and [Sul07].

First we compute two functions for the tripod. Let $P \subset \mathbb{Z}^{3m} \cong \mathbb{Z}^m \times \mathbb{Z}^m \times \mathbb{Z}^m$ be the polytope associated to a tripod. Let $pr_i : \mathbb{Z}^{3m} \cong \mathbb{Z}^m \times \mathbb{Z}^m \times \mathbb{Z}^m \to \mathbb{Z}^m$ be the projection onto the ith group of coordinates. We distinguish one edge of the tripod corresponding to the third group of coordinates in the lattice. Let f be a function such that $f(a)$ for $a = (a_1, \ldots, a_m) \in \mathbb{Z}^m$ is the number of lattice points in $(a_1 + \cdots + a_m)P$ that project to a under pr_3. We compute $f(a)$ for sufficiently many values of a to proceed with the algorithm.

Example 8.2. The polytope P for the binary Jukes–Cantor model has the following vertices:

$$v_1 = (0,1,0,1,0,1), \qquad v_3 = (1,0,0,1,1,0),$$
$$v_2 = (0,1,1,0,1,0), \qquad v_4 = (1,0,1,0,0,1).$$

These are the only integral points in P. In this case $f(1,0) = 2$, because there are exactly two points, $(1,0,0,1,1,0)$ and $(0,1,1,0,1,0)$, that are in $1P = P$ and project to $(1,0)$ via the third projection.

The function f will be our base for induction. Next, we need to compute the number of points in the fiber of the projection onto two distinguished leaves. Let g be a function such that $g(a,b)$ for $(a,b) = (a_1, \ldots, a_m, b_1, \ldots, b_m) \in \mathbb{Z}^m \times \mathbb{Z}^m$ is the number of lattice points in $(a_1 + \cdots + a_m)P$ that project to a under pr_3 and to b under pr_2. We compute $g(a,b)$ for sufficiently many pairs (a,b) to proceed with the algorithm.

Let T be a tree with a corresponding polytope P and a distinguished leaf l. Let h be a function such that $h(a)$ for $a = (a_1, \ldots, a_m) \in \mathbb{Z}^m$ is the number of points in the fiber of the projection corresponding to the leaf l of $(a_1 + \cdots + a_m)P$ onto a. We construct a new tree T' by attaching a tripod to the chosen leaf l of T. We call T' the join of T and the tripod. The chosen leaf of T' will be one of the leaves of the attached tripod. As proved in [BW07], [SS05], [Mic11], [Sul07] (depending on the model), the polytope associated to a join of two trees is a fiber product of the polytopes associated to these trees. Thus we can calculate the function h' for T' by the following rule: $h'(a) = \sum_b g(a,b)h(b)$, where the sum is taken over all $b \in \mathbb{Z}^m$ such that $g(a,b) \neq 0$.

This allows us to compute inductively the relative Hilbert polynomial. The last tripod could be attached in the same way. Then one obtains the Hilbert function from the relative Hilbert functions simply by summing over all possible projections. However, it is better to do the last step in a different way.

Suppose that as before we are given a tree T with a distinguished leaf l and a corresponding relative Hilbert function h. We compute the Hilbert function of the tree T' that is a join of the tree T and a tripod using the equality $h'(n) = \sum_a f(a)h(a)$, where $a = (a_1, \ldots, a_m)$ and $\sum a_i = n$. The function f is the basis for induction introduced above.

Thus decomposing the snowflake and the 3-caterpillar trees into joins of tripods, we can inductively compute (a few small values of) the corresponding Hilbert functions. This method also works for nonnormal models, if only the Hilbert function for the tripod can be computed. In particular, for the 2-Kimura model the computations turned out to be possible, because its polytope for the tripod is quite well understood at least to describe fully its second dilation. More precisely, the points of the polytope and the point constructed in the proof of Proposition 4.74 generate the cone over the polytope. This way we obtained the results of Computation 8.1.

9. Categorical setting

The aim of this section is to present a category GM of G-models and its connections with other categories. As an application of the theory we will present a proof of Conjecture 6.9 for the 3-Kimura model.

9.1. Category of G-models. A G-model is the following set of data:

- a tree T,
- a group G,
- a normal, abelian subgroup $H \lhd G$.

Recall that the group G acts on the characters H^* by adjunction, $\chi^g(h) = \chi(ghg^{-1})$. This motivates the following definition.

Definition 9.1 (Compatible morphism of subgroups). Fix two pairs (H_i, G_i) where H_i is an abelian, normal subgroup of G_i for $i = 1, 2$. We say that a morphism $f : H_1 \to H_2$ is *compatible* if the dual morphism $f^* : H_2^* \to H_1^*$ preserves the orbits of the groups G_i. That is, for any pair of characters $\chi, \chi' \in H_2^*$ in the same orbit of the G_2 action the images $f^*(\chi)$ and $f^*(\chi')$ are in the same orbit of the G_1 action.

Remark 9.2. Note that in the abelian case, that is, $G_i = H_i$, all morphisms are compatible. Note also that compatibility does not mean that the orbits of the adjoint action of G_i on H_i are preserved by f.

Now we are ready to state the definition of the category GM.

Definition 9.3 (Category GM of G-models). Let GM be the category whose objects are all triples (T, G, H) as described above. A morphism in GM between (T_1, G_1, H_1) and (T_2, G_2, H_2) will be a pair of maps $f : T_1 \to T_2$ and $g : H_1 \to H_2$, where g is a compatible group morphism and f is a morphism of graphs, that is, an isomorphism onto its image.

We define the category of polytopes *Poly*.

Definition 9.4 (Category *Poly* of polytopes). Let *Poly* be the category whose objects are pairs $(P, \widehat{M})$, where $\widehat{M}$ is a lattice and P a lattice polytope that spans the whole lattice. A morphism from $(P_1, \widehat{M_1})$ to $(P_2, \widehat{M_2})$ is a lattice morphism from $\widehat{M_1}$ to $\widehat{M_2}$ that takes points of P_1 to points of P_2.

9.1.1. Construction of the functor F. Our aim is to define a contravariant functor F from GM to *Poly*. We have already done this on objects; to a tree T and a group $G \rhd H$ we associate a pair $(\tilde{P}, \widehat{M_{E,G}})$ as in the discussion after Definition 4.64. Let us define the functor F on morphisms. Suppose that we have a morphism in GM, that is, a pair of morphisms $f : T_1 \to T_2$ and $g : H_1 \to H_2$. Let $P_i \subset \widehat{M_i}$ be the polytope and the lattice corresponding to the tree T_i with the group $G_i \rhd H_i$. Let also M_i be the lattice with the basis elements indexed by (e, o) (cf. Definition 4.64) where e is an edge of T_i and o an orbit in H_i^*. The lattice M_i contains the lattice $\widehat{M_i}$. The morphism g gives a morphism of characters $g^* : H_2^* \to H_1^*$. We proceed in two steps.

Step 1: The group morphism. We consider a polytope $\widetilde{P}$ associated to the tree T_2 with the group $G_1 \rhd H_1$. Let M' be the lattice associated to this tree. The basis of M' is indexed by pairs (e, o), where e is an edge of T_2 and o is an orbit in H_1^*. Using the morphism g^*, we can define a morphism $m : M_2 \to M'$ by sending a character over an appropriate edge to its image by g^*. Of course, the points of P_2 are mapped to the points of $\widetilde{P}$, because the condition of summing to the trivial character is preserved by the action of the morphism and so are the orbits. This means that we can restrict m to a morphism $m' : \widehat{M_2} \to \widehat{M'}$, where $\widehat{M'}$ is a sublattice of M' spanned by points of $\widetilde{P}$. This gives a morphism in *Poly* from $(P_2, \widehat{M_2})$ to $(\widetilde{P}, \widehat{M'})$.

Step 2: The tree morphism. Here we forget the coordinates corresponding to edges that are not in the image. Of course the condition of summing to the trivial character around vertices that are in the image is preserved.

Remark 9.5. In the "big" lattice M_i our morphism always consists in:

- first, summing up coordinates (that correspond to the orbits of characters in the inverse image of a given orbit),
- second, forgetting coordinates indexed by pairs (e, o) where e is an edge not in the image of the morphism of trees.

However, we have to remember about smaller lattices and the fact that the image of our polytope may not span the whole "small" lattice $\widehat{M_i}$ (if g^* is not surjective).

Next we consider a covariant functor from *Poly* to the category of algebras. We associate to a polytope $P \subset M$ the monoid algebra for the submonoid of $\mathbb{Z} \times M$ spanned by

$\{1\} \times P$. We have a well-known contravariant functor from the category of algebras to the category of varieties. In the toric case it was described in Section 2. Composing all, we obtain a covariant functor from GM to the category of toric varieties.

REMARK 9.6. Note that first we associate to a polytope $P \subset M$ the algebra associated to the submonoid of $\mathbb{Z} \times M$ spanned by $\{1\} \times P$. This is not necessarily a cone, as P does not have to be normal. Then we associate to this algebra a variety. This does *not* have to be a toric variety associated to a polytope in the sense of [Ful93], [CLS11]—that construction always gives a normal variety.

9.2. Morphisms of groups and rational maps of varieties. The motivation for this subsection is the following observation: if we look at graded algebras (or respectively projective varieties), then the map of graded algebras obtained from the map of polytopes in general gives only a rational map of varieties. However, we obtain a morphism, for example if the map of graded algebras is surjective.

This observation allows us to define a functor G from GM to *Proj*, where *Proj* is the category of embedded projective varieties with rational morphisms. The functor G is a composition of the functor F from the previous section, a natural functor that associates to a polytope a graded algebra generated in degree one (cf. Remark 9.6), and a well-known functor that associates to a graded algebra a projective variety [Har77, p. 76].

In particular, consider the abelian case, that is, a full subcategory $GM^{\mathrm{ab}} \subset GM$ containing all objects for which $G = H$. Then to each morphism of groups $G_1 \to G_2$ we can associate a rational morphism of projective varieties. Note that this is a well-defined morphism of affine cones over the projective varieties. More information on the abelian case can be found in Section 9.3.

Consider a G-model (T_1, G_1, H_1). The affine variety associated to this model can be realized as a subvariety of $\mathbb{A}^s$, where s is the number of vertices of the associated polytope. Notice that a morphism between two G-models that is an identity on trees induces an equivariant morphism of ambient spaces.

The following description of a morphism between varieties will be useful in the following sections. Consider two G-models (T, G_1, H_1) and (T, G_2, H_2). Let $f : H_1 \to H_2$ be a compatible morphism that, together with an identity on T, induces a morphism of G-models. Let P_1 and P_2 be the polytopes associated to the two models. As in Definition 4.64 the polytope P_i is contained in the lattice M_{E,G_i} with basis elements indexed by pairs (e, o) for e an edge of T and o an orbit of the G_i action on H_i^*. The vertices of P_i correspond also to coordinates of the affine space containing the affine variety associated to the model. Note that f^* induces a morphism $m : M_{E,G_2} \to M_{E,G_1}$. Each vertex of P_2 can be *represented* by an assignment of characters from H_2^* to edges. The morphism m is simply an application of f^* to the representatives.

PROPOSITION 9.7. *Consider the setting described above. Let s_i be the number of vertices of P_i, and let $\mathbb{A}^{s_i}$ be the affine space containing the affine variety associated to (T, G_i, H_i). Every morphism m of the G-models induces a morphism $\tilde{m} : \mathbb{A}^{s_1} \to \mathbb{A}^{s_2}$ of affine spaces. This is an equivariant morphism induced by the restriction of m to positive* *quadrants. More precisely, let e_v^* be the coordinate corresponding to the vertex $v \in P_2$. Then $e_v^*(\tilde{m}(x)) = e_{m(v)}^*(x)$.* ∎

Now fix morphisms from (T, G_i, H_i) to (T, G_0, H_0) that are identities on trees and are given by compatible group morphisms $f_i : H_i \to H_0$. Let P_i be the polytope associated to the model (T, G_i, H_i). Let M_{S_i} be the lattice with basis elements indexed by vertices of P_i. We obtain a morphism of lattices $m : M_{S_0} \to \prod M_{S_i}$. Let s_i be the dimension of M_{S_i}. Let $p_j : \prod M_{S_i} \to M_{S_j}$ be the projection to the jth factor.

REMARK 9.8. The morphism of lattices described above corresponds to the morphism of ambient spaces $\prod \mathbb{A}^{s_i} \to \mathbb{A}^{s_0}$ that can be described in coordinates as follows:

A coordinate corresponding to a vertex $v_0 \in P_0$ is the product of all coordinates corresponding to the vertices $p_j(m(v_0)) \in P_j$.

9.3. Abelian case. In this section we will establish connections between morphisms of abelian groups and morphisms of the corresponding varieties. Once again, our main aim is an application in geometry. We are building a set up for the next section. That is why we restrict to special cases. This reduces the complexity of the language but still gives a geometric insight. Let us fix a tree T.

Let $f : G_1 \to G_2$ be a morphism of abelian groups. It induces morphisms of groups of sockets $\mathfrak{S}^{G_2} \to \mathfrak{S}^{G_1}$. This gives the following commutative diagram:

$$\begin{tikzcd} M_{S,G_1} \arrow[r] & \widehat{M}_{E,G_1} \\ M_{S,G_2} \arrow[r] \arrow[u] & \widehat{M}_{E,G_2} \arrow[u] \end{tikzcd}$$

Hence the morphism $\widehat{M}_{E,G_1} \to \widehat{M}_{E,G_2}$ of the character lattices restricts to cones over polytopes. This gives a morphism of the algebras of the associated varieties. The morphism $M_{S,G_2} \to M_{S,G_1}$ restricts to positive quadrants of both lattices. Hence we get a morphism of ambient spaces $\widehat{f} : \widehat{W}_{L,G_1} \to \widehat{W}_{L,G_1}$ compatible with a morphism of varieties $\widehat{f'} : X(T, G_1) \to X(T, G_2)$. This gives a covariant functor from the category of abelian groups to the category of embedded affine toric varieties. Moreover, if f^* is injective (resp. surjective) then $\widehat{f'}$ is dominant (resp. injective). The second assertion is an easy exercise. We also need the following setting. Suppose that we have morphisms $\phi_i : G_i \to G$ for $i = 1, \ldots, m$. Just as above, this gives us a morphism $f_i : X(T, G_i) \to X(T, G)$ of embedded varieties. Let P be the polytope associated to $X(T, G)$ and let P_i be the polytope associated to $X(T, G_i)$. Consider the induced morphism $\tilde{f} : \widehat{M}_{E,G} \to \prod \widehat{M}_{E,G_i}$. If the product $f_1^* \times \cdots \times f_m^* : G^* \to \prod G_i^*$ is surjective, then $\tilde{f}$ restricted to the monoid spanned by P is surjective onto the monoid spanned by $\prod P_i$. However, in general, if the product $f_1^* \times \cdots \times f_m^*$ is injective then the restriction of $\tilde{f}$ to the monoid generated by P does not have to be injective. If $\tilde{f}$ is injective, then it induces a dominant map from the product $\prod X(T, G_i)$ to $X(T, G)$.

10. Applications to the 3-Kimura model, part 1

Our aim is to prove Conjecture 6.9 for $G = \mathbb{Z}_2 \times \mathbb{Z}_2$. The results of this section were also presented in [Mic].

Conjecture 10.1. *The dense torus orbit of the toric variety $X(K_{l,1}, \mathbb{Z}_2 \times \mathbb{Z}_2)$ is the intersection of the dense torus orbits of the varieties $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$, where T is any tree with l leaves different from the claw tree.*

Note that all dense torus orbits are contained in the dense torus orbit O of the projective (or affine) ambient space. In the algebraic set O all the orbits under consideration are closed subschemes. Hence Conjecture 10.1 can be regarded in a set-theoretic or in a scheme-theoretic version. Both are equivalent. This follows for example from a more general statement of [ES96, Corollary 2.2], and is particularly simple in the toric case. However, because the proofs of both versions are basically the same for $G = \mathbb{Z}_2 \times \mathbb{Z}_2$, we have decided to include both. Moreover, this also gives an idea of how the elements of the ideal of $X(K_{l,1}, \mathbb{Z}_2 \times \mathbb{Z}_2)$ can be generated by elements of the ideals of $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$.

The main idea of the proof is to extend the results known for binary models to the 3-Kimura model. Binary models are very well understood and have a lot of special properties [BW07]. In particular, from Proposition 6.10 we know that Conjecture 6.6 holds for $G = \mathbb{Z}_2$. As G is abelian we will be identifying G with G^*. In particular, in this subsection we assume that networks and sockets assign to edges group elements, not characters. This convention does not change anything, but simplifies the language.

We have three natural projections $f_i : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2$ for $i = 1, \dots, 3$. The map $f_1 \times f_2 \times f_3 : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ is injective. Moreover, it induces a dominant map from the product of three binary models onto the 3-Kimura model. This map is the key tool that will allow us to transfer some of the properties from the binary model to the 3-Kimura model. Unfortunately the map is not surjective, but just dominant. We can projectivise the varieties, but then we get a rational map. It turns out that a combined use of both maps allows us to derive the main theorem.

Let $f_i^* : M_{S,\mathbb{Z}_2\times\mathbb{Z}_2} \to M_{S,\mathbb{Z}_2}$ be the morphism of lattices induced by f_i. More precisely, a socket that assigns to an edge e a group element $g \in \mathbb{Z}_2 \times \mathbb{Z}_2$ is sent to a socket that assigns to e the element $f_i(g) \in \mathbb{Z}_2$. Let $i : M_{E,\mathbb{Z}_2\times\mathbb{Z}_2} \to M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2}$ be the morphism of lattices induced by $f_1 \times f_2 \times f_3$. A basis vector indexed by (e, g) is sent to the product of three basis vectors indexed respectively by $(e, f_1(g))$, $(e, f_2(g))$ and $(e, f_3(g))$. For sublattices spanned by basis vectors indexed by a fixed edge, the morphism i can be described in coordinates as

$$(a, b, c, d) \to (a + c, b + d, a + b, c + d, a + d, b + c).$$

In particular, we see that i is indeed injective. Let $g : M_{S,\mathbb{Z}_2} \to M_{E,\mathbb{Z}_2}$ be the morphism of lattices that corresponds to the parametrization map of the binary model (cf. Definition 4.33). Let $g_0 : M_{S,\mathbb{Z}_2\times\mathbb{Z}_2} \to M_{E,\mathbb{Z}_2\times\mathbb{Z}_2}$ be the morphism of lattices that corresponds to the parametrization map of the 3-Kimura model.

We have the following commutative diagram:

$$\begin{tikzcd}
M_{S,\mathbb{Z}_2} \times M_{S,\mathbb{Z}_2} \times M_{S,\mathbb{Z}_2} \arrow[r, "g\times g\times g"] & M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2} \\
M_{S,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[u, "f_1^*\times f_2^*\times f_3^*"] \arrow[r, "g_0"] & M_{E,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[u, "i"']
\end{tikzcd}$$

The following fact follows from Corollary 5.4.

Fact 10.2. *The dimension of the affine* 3*-Kimura model is* $3|E|+1$*. The dimension of the product of three affine binary models is* $3(|E|+1)$*. The dimension of the projective* 3*-Kimura model is* $3|E|$*. The dimension of the product of three projective binary models is* $3|E|$*.* ∎

It follows that if we consider projective varieties representing the models, the dominant morphism from the product of three binary models to the 3-Kimura model described above becomes a rational, generically finite map. As a map between projective varieties is not a morphism, we will restrict our attention to dense orbits of tori. On these tori orbits all maps are well defined and are represented by morphisms of lattices.

10.1. Maps of dense torus orbits. Consider the following diagram:

$$\begin{tikzcd}
M_{S,\mathbb{Z}_2} \times M_{S,\mathbb{Z}_2} \times M_{S,\mathbb{Z}_2} \arrow[rr, "g\times g\times g"] & & M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2} \times M_{E,\mathbb{Z}_2} & \\
& M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2} \arrow[ul, hook'] \arrow[rr] & & \widehat{M}_{E,0,\mathbb{Z}_2} \times \widehat{M}_{E,0,\mathbb{Z}_2} \times \widehat{M}_{E,0,\mathbb{Z}_2} \arrow[ul, hook'] \\
M_{S,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[uu, "f_1^*\times f_2^*\times f_3^*"] \arrow[rr, "g_0"] & & M_{E,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[uu, hook, "i"] & \\
& M_{S,0,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[uu, "f"] \arrow[ul, hook'] \arrow[rr, "h"] & & \widehat{M}_{E,0,\mathbb{Z}_2\times\mathbb{Z}_2} \arrow[ul, hook'] \arrow[uu, hook, "j"']
\end{tikzcd} \tag{10.1}$$

The back rectangle is just the previous diagram. The front rectangle is induced from it by taking sublattices (cf. Definition 4.40). At the level of varieties the back is the affine picture, while the front is the projective picture. The left square with lattices of type M_S corresponds to morphisms of ambient spaces. The square on the right describes the maps between varieties, or parametrizing spaces. The upper square corresponds to the product of three binary models, while the bottom square to the 3-Kimura model.

Let us explain the morphism j. It is injective, as it is a restriction of i. The lattice $\widehat{M}_{E,0}$ is the character lattice of the torus acting on the projective toric variety representing the model. The morphism j is induced by the rational finite map from the product of three $\mathbb{P}(X(T,\mathbb{Z}_2))$ to $\mathbb{P}(X(T,\mathbb{Z}_2\times\mathbb{Z}_2))$. Via the coordinate system we can identify dense torus orbits with the tori.

Definition 10.3 (The torus $\mathbb{T}_X$). Let X be any toric variety in an affine or projective space with a distinguished coordinate system. Suppose that X is embedded equivariantly, as in Section 2. The dense torus orbit of X will be denoted by $\mathbb{T}_X \subset X$. Recall that $\mathbb{T}_X$ consists precisely of those points of X that have all coordinates different from 0.

The morphism j of character lattices is induced by a finite morphism from $\mathbb{T}_{(\mathbb{P}(X(T,\mathbb{Z}_2)))^3} = (\mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2))})^3$ to $\mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2\times\mathbb{Z}_2))}$. From the discussion in the proof of Proposition 7.1 we also know that the morphism of ambient spaces does not depend on the tree, but only on the number l of leaves. Hence the vertical morphisms of lattices in the left part of Diagram 10.1 are the same for all trees with l leaves.

10.2. Idea of the proof. The main reason for passing to tori is that we want to have a well-defined dominant finite map. This allows us to take advantage of toric geometry. For example, we know that the number of points in the fiber of the morphism of tori $(\mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2))})^3 \to \mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2\times\mathbb{Z}_2))}$ is equal to the index I_1 of the image of j in $(\widehat{M}_{E,0,\mathbb{Z}_2})^3$.

For the projective ambient spaces the situation is a little different. The morphism $f : M_{S,0,\mathbb{Z}_2\times\mathbb{Z}_2} \to (M_{S,0,\mathbb{Z}_2})^3$ is not injective, so the corresponding morphism of tori is not surjective. We will show that the image of f in $(M_{S,0,\mathbb{Z}_2})^3$ is of finite index, say I_2. This means that the corresponding morphism of tori is finite with each fiber having I_2 elements. Moreover, we will show that $I_2 = I_1$. Hence we get the diagram

$$\begin{tikzcd}
\mathbb{T}_{(\mathbb{P}(\widetilde{W}_{E,\mathbb{Z}_2})^3)} \arrow[r] & \mathbb{T}_{\mathbb{P}(\widetilde{W}_{E,\mathbb{Z}_2\times\mathbb{Z}_2})} \\
\mathbb{T}_{(\mathbb{P}(X(T,\mathbb{Z}_2)))^3} \arrow[u, hook] \arrow[r, two heads] & \mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2\times\mathbb{Z}_2))} \arrow[u, hook]
\end{tikzcd}$$

where the horizontal maps are finite, étale of the same degree.

This means that if we consider the morphism of projective ambient spaces, then the preimage of $\mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2\times\mathbb{Z}_2))}$ is precisely $\mathbb{T}_{(\mathbb{P}(X(T,\mathbb{Z}_2)))^3}$. Hence any intersection results that hold for the binary model must also hold for the 3-Kimura model. In particular, since Conjecture 6.6 holds for the binary model, we obtain a set-theoretic version of Conjecture 10.1 for the 3-Kimura model. By easy algebraic arguments we will also prove Conjecture 10.1 scheme-theoretically for the 3-Kimura model.

10.3. Proof. Our first step will be to understand the morphism of projective ambient spaces $(\mathbb{P}(\widetilde{W}_{E,\mathbb{Z}_2}))^3 \dashrightarrow \mathbb{P}(\widetilde{W}_{E,\mathbb{Z}_2\times\mathbb{Z}_2})$. This is a well-defined map on dense tori orbits. The map of tori corresponds to the morphism $f : M_{S,0,\mathbb{Z}_2\times\mathbb{Z}_2} \to (M_{S,0,\mathbb{Z}_2})^3$ of lattices. This morphism depends only on the number of leaves, not on the tree.

By the definition we can embed the group $\mathfrak{S}$ of sockets in G^l. We can also view the group $\mathfrak{S}$ as a $\mathbb{Z}$-module. This gives group morphisms $M_S \to \mathfrak{S} \to G^l$. The element of the basis of M_S indexed by a socket s is mapped to the socket s.

Example 10.4 (The case of the binary model and trivalent claw tree). Consider the tree $K_{3,1}$ and the group $\mathbb{Z}_2$. We have four sockets: $(0,0,0), (1,1,0), (1,0,1), (0,1,1)$. By coordinatewise action they form a subgroup of $(\mathbb{Z}_2)^3$. The lattice M_S is freely generated by four basis vectors $e_{(0,0,0)}, e_{(1,1,0)}, e_{(1,0,1)}, e_{(0,1,1)}$. The morphism $M_S \to \mathfrak{S}$ maps $e_{(a,b,c)}$

to (a,b,c). Of course $ke_{(a,b,c)}$ is mapped to $k(a,b,c)$. For example, $3e_{(1,1,0)}$ is mapped to $(1,1,0)+(1,1,0)+(1,1,0)=(1,1,0)$.

Lemma 10.5. *We have an exact sequence of groups*

$$M_{S,0,\mathbb{Z}_2\times\mathbb{Z}_2} \to (M_{S,0,\mathbb{Z}_2})^3 \to (\mathbb{Z}_2)^l,$$

where the first morphism is given by f, and the second is the sum of three morphisms $M_{S,0,\mathbb{Z}_2} \to (\mathbb{Z}_2)^l$ described above (1).

Proof. It is clear that this is a complex. Let $(b'_i)_{i\geq 0}$ be the basis of $M_S^{\mathbb{Z}_2}$ corresponding to sockets. Let s_i be the socket corresponding to b'_i. Moreover, suppose that b'_0 corresponds to the trivial socket, that is, the neutral element of $\mathfrak{S}$. Let b_i be the basis of $M_{S,0,\mathbb{Z}_2}$ defined as $b_i = b'_i - b'_0$ for $i>0$. Note that an element (b'_i, b'_j, b'_k) is in the image of $f_1^* \times f_2^* \times f_3^*$ if and only if the corresponding three sockets s_i, s_j, s_k sum to the neutral element of $\mathfrak{S}$. Hence the elements of the form $(b_i,b_i,0)=(b'_i,b'_i,b'_0)-(b'_0,b'_0,b'_0)$ are in the image of f. We see that $(2b_i,0,0)=(b_i,b_i,0)+(b_i,0,b_i)-(0,b_i,b_i)$ is also in the image. Furthermore, for any two sockets s_i and s_j there exists a socket $s_k := s_i + s_j$ such that (b_i,b_j,b_k) is in the image of f. This reduces any element from $(M_{S,0,\mathbb{Z}_2})^3$ to an element $(b_i,0,0)$ modulo the image of f or to 0. Hence any element is in the image if the XOR of all its coordinates is zero. ∎

Definition 10.6 (The kernel K). For any tree T let $K^T = K_1^T \times K_2^T \times K_3^T \subset M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2}$ be the restriction of the kernel of the morphism $g\times g\times g$ to $M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2} \times M_{S,0,\mathbb{Z}_2}$.

Each character in K^T is a character of $(\mathbb{T}_{\mathbb{P}(\widetilde{W}_E^{\mathbb{Z}_2})})^3$ that is the trivial character when restricted to $(\mathbb{T}_{\mathbb{P}(X(\mathbb{Z}_2))})^3$. Each such character is a triple of characters of $\mathbb{T}_{\mathbb{P}(\widetilde{W}_E^{\mathbb{Z}_2})}$. Each character of the triple is a quotient m_1/m_2 of monomials of the same degree on the projective space $\mathbb{P}(\widetilde{W}_E^{\mathbb{Z}_2})$. The polynomials $m_1 - m_2$ span (2) the ideal of the toric variety $\mathbb{P}(X(\mathbb{Z}_2))$. We want to view characters as functions. Hence we restrict our attention to $(\mathbb{T}_{\mathbb{P}(\widetilde{W}_E^{\mathbb{Z}_2})})^3$. In the algebra of this torus the ideal of $(\mathbb{T}_{\mathbb{P}(X(\mathbb{Z}_2))})^3$ is generated by the elements $k-1$, where $k\in K^T$.

Definition 10.7 (The kernel D). For any tree T let D^T be the kernel of the map h defined in Diagram 10.1.

The elements of D represent characters trivial on the projective 3-Kimura variety. In the setting described at the end of Section 6.2 we want to prove that the sublattices D^T for different trees T with l leaves generate the sublattice $D^{K_{l,1}}$. The idea is to push the lattices D to $(M_{S,0,\mathbb{Z}_2})^3$ using the morphism f. Next we use the results on binary models to obtain the generation for $f(D)$. Using properties of the image of f we are able to conclude the generation in $M_{S,0,\mathbb{Z}_2\times\mathbb{Z}_2}$. The following lemma enables us to restrict to the image of f instead of regarding the whole lattice $(M_{S,0,\mathbb{Z}_2})^3$.

Lemma 10.8. *For any tree T the kernel K^T is a sublattice of the image of f.*

(1) In this case the second operation is often called XOR.

(2) They do not only generate the ideal, but even span it as a vector space.

Proof. It is enough to show that $K_1^T \times \{0\} \times \{0\} \subset \operatorname{Im} f$. Suppose that $m = \sum_i a_i b_i \in K_1^T$, where each b_i is as in the proof of Lemma 10.5. Hence $b_i = (g_1^i - e, \dots, g_l^i - e)$, where e is the neutral element of $\mathbb{Z}_2$ and the $g_j^i \in \mathbb{Z}_2$ are elements forming a socket. We know that $g(m) = 0$. In particular, the coordinates of M_E indexed by leaves are equal to zero. Fix k that is a number of a leaf, $1 \le k \le l$. Let us look at all coordinates indexed by (k, q) where $q \in \mathbb{Z}_2$. The restriction of M_E to these coordinates is a free abelian group spanned by elements of $\mathbb{Z}_2$. Hence $\sum_i a_i(g_k - e) = 0$ in the free abelian group generated formally by elements of $\mathbb{Z}_2$. Hence, a fortiori, $\sum_i a_i(g_k - e) = e$ where now the sum is taken in $\mathbb{Z}_2$. As the action in $\mathfrak{G}$ is coordinatewise, we see that the image of m in $\mathfrak{G}$, and hence in $\mathbb{Z}_2^l$, is the neutral element. Using Lemma 10.5 we see that $m \in \operatorname{Im} f$. ∎

Proposition 10.9. *The index of the image of f in $(M_{S,0,\mathbb{Z}_2})^3$ is equal to the index of the image of j in $(\widehat{M}_{E,0,\mathbb{Z}_2})^3$.*

Proof. This is a consequence of Lemma 10.8. ∎

Corollary 10.10. *Conjecture* 10.1 *holds set-theoretically.*

Proof. The index of the image of f equals the degree of a finite map of tori. In particular, we are in the situation of Diagram 10.2. The corollary follows from the discussion at the beginning of Section 10.2. ∎

Now we will prove Conjecture 10.1 scheme-theoretically. Let $T_0 = K_{l,1}$. We consider trees T_i such that the ideal of $\mathbb{T}_{\mathbb{P}(X(T,\mathbb{Z}_2))}$ is the sum of the ideals $\mathbb{T}_{\mathbb{P}(X(T_i,\mathbb{Z}_2))}$. Let K^{T_i} be the kernel of $g \times g \times g$ for the tree T_i. Let D^{T_i} be the kernel of h for the tree T_i. We know from Proposition 6.10 that the lattices K^{T_i} for $i > 0$ span K^{T_0}.

Theorem 10.11. *The lattices D^{T_i} for $i > 0$ span D^{T_0}. Conjecture* 10.1 *holds scheme-theoretically.*

Proof. Let $a \in D^{T_0}$. We know that $f(a) \in K_{\mathbb{Z}_2}^{T_0}$, so $f(a) = \sum k_i$ where $k_i \in K_{\mathbb{Z}_2}^{T_i}$. Using Lemma 10.8 we can find $k_i' \in D^{T_i}$ such that $f(k_i') = k_i$. This means that $a - \sum k_i'$ is in the kernel of f. In particular, as j is injective, $a - \sum k_i'$ belongs to every D^{T_i}, hence we obtain the desired decomposition. ∎

Remark 10.12. From Proposition 6.10 it is enough to take two (particular) different $i > 0$ to span D^{T_0}, as it was in the case of the binary model.

10.4. Applications to phylogenetics. In this section we present a few applications. The basic result that we use is by Marta Casanellas and Jesús Fernández-Sánchez [CFS08]. It states that all points important for biologists are contained in the dense torus orbit of $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$. Thus, following [CFS08], we call points of the dense torus orbit *biologically meaningful.* In Section 10 we gave a precise description of this orbit for any tree. This is sufficient for biologists.

People dealing with applications are usually interested in trivalent trees. Let us motivate the use of other trees. The first, obvious reason is that they can appear (at least hypothetically) as right models of evolution. This however is a degenerate situation that is often neglected. The next subsection presents another reason.

10.4.1. Identifiability. Dealing with applications we are given a point P in the space of all possible probabilities $\widetilde{W_L}$. The first question is for which trees this point can be realized. More precisely, for which trees T do we have $P \in X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$? We are interested in knowing if there is only one such tree T, or there are several possibilities. This is a first part of the *identifiability* problem. Hence Conjecture 6.6 is a question about the locus of points for which the identifiability problem cannot be resolved at all. Of course, a generic point that belongs to any of the varieties belongs to exactly one $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$ with T trivalent. Much more is known about the identifiability of different models. For the precise results the reader is advised to consult [AR06] or [APRS11] and the references therein.

In particular we see that points that belong to some $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$ where T is not trivalent cannot identify the tree topology. Hence the question about the locus of these points, or equivalently about the polynomials defining such varieties, may give some results for trivalent trees. However, as the situation in Section 7 shows, the phylogenetic invariants of two varieties $X(T, \mathbb{Z}_2)$ for two different trees do not generate the ideal of the variety associated to their degeneration.

The second, but equally important question about the identifiability is to give the description of the fiber of the parametrization map of the model $\check{\psi}^{-1}(P)$. A biologist aims at distinguishing one point in the fiber. This would enable to identify not only the tree topology, but also the corresponding probabilities of mutation. The algebraic setting allows us to give a description of this fiber. We assume that P is biologically meaningful, that is, contained in the dense torus orbit. Equivalently, all coordinates of P after the Fourier transform are different from zero. We prefer to work up to multiplicity, that is, regard the projectivization of $\check{\psi}$ denoted by $\check{\psi}_{\mathbb{P}}$. The fiber $\check{\psi}_{\mathbb{P}}^{-1}(P)$ is contained in the dense torus orbit of $\prod \mathbb{P}(W_e)$. As this parameter space is of the same dimension as the image, we know that $\check{\psi}_{\mathbb{P}}$ is a generically finite map. Moreover, when restricted to dense torus orbits it is étale and finite. Hence each fiber is finite and contains the same number of points, independent of P. This number is the index of the lattice $\widehat{M_E}$ in a saturated sublattice of M_E. Of course we do not claim that all the points in the fiber have a probabilistic meaning. We just prove that from the algebraic point of view there is always a fixed, finite number of possible candidates for transition matrices.

We will now give a precise description of a general fiber for a general group-based model corresponding to an abelian group H. From Corollary 5.4 we know that the map of projective tori parametrizing the model is a finite map. By dualizing the exact sequence in Corollary 5.7, we see that the kernel has a group structure isomorphic to $H^{|N|}$. By [CFS08] the only biologically meaningful points are contained in the dense torus orbit.

COROLLARY 10.13. *Let T be any tree and H any abelian group. Let $\mathbb{P}(X)$ be the projective variety associated to the model. Let $x \in \mathbb{P}(X)$ be a biologically meaningful point. Up to multiplication by a constant, there are exactly $|H|^{|N|}$ parameters in the fiber of x. In other words, there are exactly $|H|^{|N|}$ possible transition matrices.* ■

Note that we do not use further restrictions on the parameters of transition matrices. For example, we do not assume that the parameters are real. This condition for sure further decreases the number of possible transition matrices. However, we see that when we consider complex parameters, the number of possible parameters is already finite and moreover independent of the point considered.

10.4.2. Phylogenetic invariants. The main theorem gives an inductive way of obtaining phylogenetic invariants of any tree. It is an open problem if these invariants generate the whole ideal. It is proved however that they give a description of all biologically meaningful points in the case of the 3-Kimura model. The method is very simple. Suppose that we know the phylogenetic invariants for all trees with vertices of degree less than or equal to d. By the results of [SS05] it is enough to describe the phylogenetic invariants for the claw tree $K_{d+1,1}$. For 3-Kimura, to obtain the description of the dense torus orbit we just take the sum of two ideals (cf. Remark 10.12). They are both associated to trees with the same topology. The tree has two inner vertices v_1 and v_2 of degrees 3 and d respectively. The difference between the ideals is a consequence of different labelling of leaves. For one tree the leaves adjacent to v_1 are labelled 1 and 2. For the second tree they are labelled 1 and 3. Notice that in fact we have to compute just one ideal. The second one can be obtained by permuting the variables.

11. Applications to the 3-Kimura model, part 2

The aim of this section is to further investigate Conjecture 6.1 for the 3-Kimura model. These results were also presented in [Mic13].

Let I_n be the ideal of the variety $X(T, \mathbb{Z}_2 \times \mathbb{Z}_2)$ where T is a claw tree with n leaves. Let I'_n be the subideal of I_n generated in degree 4. The conjecture of Sturmfels and Sullivant states that $I_n = I'_n$ for any n. In this section we will prove that I_n and I'_n define the same projective scheme. This is equivalent to the fact that their saturations are equal [Har77, Exercise 5.10 b)]. In particular, it follows that they define the same affine set. One concludes that in order to check if any point belongs to the variety, it is enough to consider the phylogenetic invariants of degree 4. By [SS05, Theorem 23] the result will follow for any tree. Let us state the main theorem of this section.

Theorem 11.1. *Consider any tree T and the 3-Kimura model. The ideal of the variety associated to it and the subideal generated by polynomials of degree at most 4 define the same projective scheme.*

We hope that the method presented in this section can be applied to other problems of the type, "prove that a toric projective scheme can be defined by an ideal generated in degree d". Indeed, recently the method was applied in [LM14, Bru13]. In general, let I be the ideal of a projective toric variety. Let I' be the subideal generated in degree d. The aim is to prove that the saturation of I' with respect to the irrelevant ideal equals I.

Suppose that the variety is given by a polytope P, with points corresponding to coordinates of the ambient projective space—as in Section 2. Proving that the saturation of I' equals I is equivalent to proving that I' and I are equal in each localization with respect to any coordinate, represented by a point $Q \in P$. Thus we have to prove that any generator of I multiplied by a sufficiently high power of the variable corresponding to Q belongs to I'.

Let us translate this condition into the combinatorial language. The generators of I correspond to relations between points of $P \times \{1\}$. Fix a relation $\sum A_i = \sum B_j$, where $A_i, B_j \in P \times \{1\}$. Multiplying the corresponding element of the ideal by the variable corresponding to Q is equivalent to adding Q to both sides of the relation. Thus we have to prove that the binomial corresponding to the relation $\sum A_i + mQ = \sum B_j + mQ$ is generated by binomials from I of degree at most d for m sufficiently large.

A binomial corresponding to a relation $\sum R_i = \sum S_i$ between points of a polytope is generated in degree d if and only if one can transform $\sum R_i$ to $\sum S_i$ using a sequence of simple steps. In each single transformation one can replace points $R_1, \dots, R_k$ for $k \leq d$ by $R'_1, \dots, R'_k$ if they satisfy the relation $\sum_{i=1}^k R_i = \sum_{i=1}^k R'_i$. In that case we say that the relation is generated in degree d.

The proof scheme is very simple:

(∗) (i) Using degree d relations reduce A_i and B_i to some simple, special points of $P \times \{1\}$ contained in a subset $L_Q \subset P$.
(ii) Show that any relation between points of L_Q is generated in degree d.

In general, any of these two points can be very difficult.

Remark 11.2. It is well-known that the projective toric variety defined by a polytope P is covered by affine subsets given by localizations by coordinates corresponding to vertices. Thus one may be tempted to prove that $I = I'$ only in the localizations by vertices. Note however that in general we do not know if the scheme defined by I' is also covered by localizations by coordinates corresponding to vertices. Indeed, I' and I may be different at the set-theoretical level. For example if *Proj* I' contains a point that is zero at the coordinates corresponding to vertices and nonzero at some other coordinates, then such a point will not belong to any localization with respect to vertices. However, if $\operatorname{rad} I' = I$ then of course it is enough to consider localizations with respect to vertices.

As our polytopes have only vertices, the problem described in Remark 11.2 does not concern us.

Remark 11.3. We have the following equivalences for a toric ideal I given by a polytope $P \times \{1\}$.

- All relations between vertices of $P \times \{1\}$ are generated in degree $d \Leftrightarrow$ the ideal I is generated in degree d.
- For any point $Q \in P \times \{1\}$ and any relation there is an integer m such that after adding mQ to both sides of the relation, it is generated in degree $d \Leftrightarrow$ the projective scheme defined by I can be defined by an ideal generated in degree d.
- For any relation there are [1] points $Q_i \in P \times \{1\}$ such that after adding $\sum Q_i$ to both sides, it is generated in degree $d \Leftrightarrow$ the dense torus orbit of the variety is defined by an ideal generated in degree d in the algebra of the ambient torus.

In order to prove Theorem 11.1, following the arguments of [SS05, Chapter 5] one immediately reduces to the case when $T = K_{1,l}$ for some $l \in \mathbb{N}$. Let us restate the

[1] Not necessarily different.

general definitions in the case of the group $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ (the 3-Kimura model) on a claw tree $K_{1,l}$.

DEFINITION 11.4 (Group-based flow). A *group-based flow* is an assignment of elements of the group $\mathbb{Z}_2 \times \mathbb{Z}_2$ to edges of $K_{1,l}$ such that the sum of all elements is the natural element.

We will identify a group-based flow with an l-tuple of group elements summing to zero. The sum of such l-tuples will be a coordinatewise sum, where each entry is treated as an element of the free abelian group generated by elements of $\mathbb{Z}_2 \times \mathbb{Z}_2$. Each group-based flow represents a vertex of a polytope P. The addition described above is the addition in the lattice generated by the vertices of the polytope.

EXAMPLE 11.5. For $l = 4$,

$$\begin{aligned}\big((0,0)+(0,1),(1,0)+(1,1),2(0,1),2(0,0)\big)&+\big((0,1),(1,0),(1,1),(0,0)\big)\\ &=\big((0,0)+2(0,1),2(1,0)+(1,1),2(0,1)+(1,1),3(0,0)\big).\end{aligned}$$

The neutral group-based flow is $\mathfrak{nt} = ((0,0),(0,0),(0,0),(0,0))$.

DEFINITION 11.6 (Pair, triple). We say that a group-based flow is a *pair* if the cardinality of its support is 2. We say that a group-based flow is a *triple* if the cardinality of its support is 3.

LEMMA 11.7. *Suppose that Q is a group-based polytope with a group G acting transitively on its integral points. The projective scheme $Proj\,\mathbb{C}[Q]$ can be represented by an ideal generated in degree at most d if and only if there exists a point $R \in Q$ such that for any relation $\sum A_i = \sum B_i$ between points of Q, for m sufficiently large $\sum A_i + mR = \sum B_i + mR$ is generated in degree d.*

Proof. As G acts transitively, the assumption that there exists a point R with the given property is equivalent to the fact that the property holds for all integral points of Q. ■

LEMMA 11.8. *Consider a claw tree $K_{1,n}$ and a finite abelian group G. Consider any relation $\sum A_i = \sum B_i$ of integral points of P. There exists m such that the relation $m\mathfrak{nt} + \sum A_i = m\mathfrak{nt} + \sum B_i$ can be transformed, using only quadrics, to a relation among group-based flows with support of cardinality at most $D(G)$, where $D(G)$ is the Davenport constant of the group G. In particular, when G is not a cyclic group then the supports are of cardinality at most $|G| - 1$.*

Proof. Consider any group-based flow A. Suppose its support is of cardinality greater than $D(G)$. Then we can find a proper subset S of edges in the support such that $\sum_{e \in S} A(e)$ is the neutral element of G, where the addition is taken in G. Thus $\mathfrak{nt} + A$ equals the sum of two group-based flows with strictly smaller support. The lemma follows easily. ■

By Lemmas 11.7 and 11.8 we have to generate relations only between group-based flows that are pairs and triples. This completes the first step of the method $(*)$ presented above. The set L consists of pairs and triples. Note that this part of the proof can be adjusted to other groups G for L consisting of group-based flows with support of cardinality at most $|G|$.

Fix any relation $\sum A_i = \sum B_i$, where A_i and B_i are either pairs or triples. Our aim is to transform $\sum A_i$ to $\sum B_i$ in a series of steps, each time replacing at most four A_i by group-based flows with the same sum ([2]). We assume that among A_i there are more or the same number of triples as among B_i. We first try to reduce the relation, so that consecutively:

(i) among A_i there are as few triples as possible,
(ii) among B_i there are as few triples as possible,
(iii) the degree of the relation is as small as possible.

More precisely, let t and t' be the number of triples among respectively A_i and B_i. Let d' be the degree of the relation. Our proof will be inductive on (t, t', d') with lexicographic order.

To prove Theorem 11.1 we consider separately three cases depending on the number of triples among A_i. The cases are:

(a) there are no triples,
(b) there is exactly one triple,
(c) there are at least two triples.

We say that a family of group-based flows *agrees on an index* j of an edge if they all associate the same element to j, and j belongs to their *support*. We will denote by g_1, g_2 and g_3 the three nonneutral elements of $\mathbb{Z}_2 \times \mathbb{Z}_2$. A triple that associates g_1 to index a, g_2 to b, and g_3 to c is denoted by (a, b, c). A pair that associates an element g_i to indices d and e will be denoted by $(d, e)_{g_i}$ and called a g_i *pair*. We say that g_i is *contained in the group-based flow* if there exists an index j such that the group-based flow associates g_i to j. We believe that the following proofs are impossible to follow without a piece of paper. We strongly encourage the reader to note what group-based flows appear on both sides of the relation at each step of the proof.

11.1. The case with no triples. First note that there are no triples among B_i. Without loss of generality we may assume that A_1 is a pair equal to $(a, b)_{g_1}$. Hence there exists $(b, c)_{g_1}$ among B_i for some index c. If $c = a$ we can reduce this pair, hence we assume $c \neq a$. There exists a group-based flow, say A_2, that is $(c, d)_{g_1}$. If $d = b$ we can reduce this pair. We consider two cases:

1) $d \neq a$. Then we use the degree two relation $(a, b)_{g_1} + (c, d)_{g_1} = (a, d)_{g_1} + (b, c)_{g_1}$ and we can reduce $(b, c)_{g_1}$.

2) $d = a$. Then there is a group-based flow, say B_1, given by $(a, e)_{g_1}$. If $e = b$ or $e = c$, we can reduce this pair. In the other cases we use the relation $(a, e)_{g_1} + (b, c)_{g_1} = (a, b)_{g_1} + (e, c)_{g_1}$ and we reduce $(a, b)_{g_1}$.

Notice that in this very easy case we have only used degree two relations.

([2]) We are also allowed to add the group-based flow $\mathfrak{nt}$ to both sides.

11.2. The case with one triple. Let A_1 be the only triple among A_i.

LEMMA 11.9. *There is exactly one triple among B_i.*

Proof. By the assumptions we know that there is at most one triple among B_i. We exclude the case when there are no triples by comparing the parity of the number of times the element g_1 appears on both sides of the relation. ∎

By the previous lemma we may assume that B_1 is the only triple among B_i. Without loss of generality, assume $A_1 = (1, 2, 3)$.

11.2.1. The triples agree on at least two elements in their support. Suppose that $A_1 = (1, 2, 3)$ and $B_1 = (1, 2, c)$. Of course if $c = 3$ we can make a reduction. Otherwise we must have a pair $(c, d)_{g_3}$ among A_i. If $d \neq 3$ then we use the relation $(c, d)_{g_3} + (1, 2, 3) = (1, 2, c) + (3, d)_{g_3}$ and reduce the triples. Assume $d = 3$. Analogously, we can assume there is a pair $(3, c)_{g_3}$ among B_i, hence we can reduce this pair.

11.2.2. The triples agree on exactly one element in their support. Consider the case when the triples agree on at least one element, say 1, in their common support. By the previous case we may assume that they agree on exactly one element.

As before let $A_1 = (1, 2, 3)$ and $B_1 = (1, b, c)$. We consider three cases.

1) $b \neq 3$. There must be a pair $(b, d)_{g_2}$ among A_i. If $d \neq 2$ then we can apply the relation $(b, d)_{g_2} + A_1 = (1, b, 3) + (d, 2)_{g_2}$. This reduces to the case 11.2.1. So we assume $d = 2$. There must be a pair $(2, e)_{g_2}$ among B_i. We may assume $e \neq b$, as otherwise we would be able to make a reduction. Hence there must also be a pair $(e, f)_{g_2}$ among A_i. If $f \neq b$ we can use the relation $(e, f)_{g_2} + (2, b)_{g_2} = (e, 2)_{g_2} + (f, b)_{g_2}$ and reduce $(e, 2)_{g_2}$. For $f = b$ we must have a pair $(b, g)_{g_2}$ among B_i. If $g = 2$ or $g = e$ then this pair can be reduced. In the other case we use the relation $(e, 2)_{g_2} + (b, g)_{g_2} = (e, g)_{g_2} + (b, 2)_{g_2}$ and reduce $(b, 2)_{g_2}$.

2) $c \neq 2$. This case is analogous to 1).

3) $b = 3$ and $c = 2$.

LEMMA 11.10. *If there is a pair $(p, q)_{g_2}$ among A_i such that $p, q \neq 2$ then we may assume that it is equal to* $(1, 3)$.

Proof. Suppose that $p \neq 1, 2, 3$ and $q \neq 2$. We apply the relation $(p, q)_{g_2} + A_1 = (1, p, 3) + (q, 2)_{g_2}$ and reduce to case 2) $c \neq 2$. ∎

Analogously, if there is a pair $(p, q)_{g_2}$ among B_i such that $p, q \neq 3$ then we can assume it is equal to $(1, 2)_{g_2}$.

Notice that there must be a pair $(3, d)_{g_2}$ among A_i and a pair $(2, e)_{g_2}$ among B_i. From Lemma 11.10, d is either 2 or 1, and e is either 3 or 1. We will consider several subcases.

3.1) $d = 2$. If $e = 3$ then we can make a reduction of pairs. If $e = 1$ we must have a pair $(1, f)_{g_2}$ among A_i. If $f = 2$ we make a reduction, hence we assume $f = 3$. This means that there must be a pair $(3, g)_{g_2}$ among B_i. If $g = 2$ or $g = 1$ we can make a reduction. Otherwise we apply the relation $(1, 2)_{g_2} + (3, g)_{g_2} = (1, 3)_{g_2} + (2, g)_{g_2}$ and reduce the pair $(1, 3)_{g_2}$.

3.2) $e = 3$. This case is similar to 3.1).

3.3) $d = 1$ and $e = 1$. As this is the only case left, we may repeat the same reasoning for g_3. In particular, we can assume there is a pair $(1,2)_{g_3}$ among A_i. We see that we can reduce the triples by applying the following relation:

$$(1,2,3) + (1,3)_{g_2} + (1,2)_{g_3} = (1,3,2) + (1,2)_{g_2} + (1,3)_{g_3}.$$

This is a degree three relation.

11.2.3. The triples do not agree on any element of the support. We want to reduce to one of the previous cases. Consider the following two cases.

1) *The triples A_1 and B_1 have different supports.* Once again let $(1,2,3) = A_1$ and let $(a,b,c) = B_1$. We may assume that a is not in the support of A_1. We see that there must be a pair $(a,f)_{g_1}$ among A_i. If $f \neq 1$ we can use the relation $(a,f)_{g_1} + A_1 = (a,2,3) + (f,1)_{g_1}$. This reduces to the case 11.2.2, hence we assume that $f = 1$. There must be a pair $(g,1)_{g_1}$ among B_i. If $g = a$ we can reduce this pair, so we assume $g \neq a$. Notice that there must be a pair $(g,h)_{g_1}$ among A_i. If $h \neq a$ then we can use the relation $(1,a)_{g_1} + (g,h)_{g_1} = (g,1)_{g_1} + (h,a)_{g_1}$ and reduce the pair $(g,1)_{g_1}$. So we can assume $h = a$. Then there must be a pair $(a,i)_{g_1}$ among B_i. If $i = 1$ then we can reduce it. Otherwise we can use the relation $(g,1)_{g_1} + (a,i)_{g_1} = (g,a)_{g_1} + (1,i)_{g_1}$ and reduce the pair $(g,a)_{g_1}$.

2) *The set $\{1,2,3\}$ is the support of B_1 and A_1.* Remember that by the assumption 11.2.3 the triples A_1 and B_1 do not agree on any element from their support. Without loss of generality we may assume $A_1 = (1,2,3)$ and $B_1 = (2,3,1)$. Hence there must be a pair $(2,a)_{g_1}$ among A_i, and $(1,b)_{g_1}$ among B_i. If $a = 1$ and $b = 2$ then the pairs are the same and can be reduced. As both cases are symmetric, we can assume that $a \neq 1$.

If $a \neq 3$ we can use the relation $(2,a)_{g_1} + (1,2,3) = (a,2,3) + (2,1)_{g_1}$. This reduces to the case with different supports. We are left with the case $a = 3$. There must be a pair $(3,z)_{g_1}$ among B_i. If $z \neq 1$ we can use the relation $(3,z)_{g_1} + B_1 = (z,3,1) + (2,3)_{g_1}$. This would enable us to reduce the $(2,3)_{g_1}$ pair and decrease the degree. So we can assume that $z = 1$. So far we have shown that there must be pairs $(2,3)_{g_1}$ among A_i and $(3,1)_{g_1}$ among B_i [3]. By the same reasoning for g_2 and g_3 we see that we can use the following relation:

$$(1,2,3) + (2,3)_{g_1} + (1,3)_{g_2} + (1,2)_{g_3} = (2,3,1) + (2,3)_{g_3} + (1,3)_{g_1} + (1,2)_{g_2}.$$

Notice that this is a degree 4 relation. It enables us to reduce triples.

11.3. The case with at least two triples. We suppose that there are at least two triples among A_i.

LEMMA 11.11. *If there are two triples A_1, A_2 among A_i that do not agree on any element of their supports then we can make a reduction. Thus we can assume that any two triples among A_i agree on at least one index.*

[3] Notice that we have made the symmetry assumption $a \neq 1$. The symmetric assumption would be $b \neq 2$. However, as the result we got was symmetric, also for $b \neq 2$ we prove the existence of the same pairs.

Proof. The assumptions are equivalent to $A_1 = (a,b,c)$, $A_2 = (d,e,f)$ with $a \neq d$, $b \neq e$, $c \neq f$. We apply the relation $A_1 + A_2 + \mathfrak{nt} = (a,d)_{g_1} + (b,e)_{g_2} + (c,f)_{g_3}$ that reduces the number of triples. ∎

Lemma 11.12. *If there is no index on which all triples among A_i agree then we can make a reduction.*

Proof. Suppose there is no index on which all triples among A_i agree. We may consider only two cases by Lemma 11.11.

1) *Any two triples from A_i agree on at least two elements.* Consider any triple $A_1 = (1,2,3)$. Since not all triples from A_i associate g_1 to 1, there is a triple $(a,2,3)$ with $a \neq 1$ among A_i. There must also be a triple that does not associate g_2 to 2. But this cannot happen, as the triple must agree with both $(1,2,3)$ and $(a,2,3)$ on two indices, which gives a contradiction.

2) *There exist two triples that agree only on one index.* Let $A_1 = (1,2,3)$ and $A_2 = (1,b,c)$ with $b \neq 2$ and $c \neq 3$. By the case assumption there is a triple $A_3 = (d,e,f)$ with $d \neq 1$. Any two triples have to agree on at least one element by Lemma 11.11. Hence without loss of generality we can assume $e = b$ and $f = 3$. We can apply the relation

$$A_1 + A_2 + A_3 + \mathfrak{nt} = (d,1)_{g_1} + (2,b)_{g_2} + (3,c)_{g_3} + (1,b,3).$$

This relation reduces the number of triples. ∎

By the previous lemma we may assume that there exists an index, say 1*, such that all triples A_i associate to it the same nonneutral element, say g_1.*

Definition 11.13 (k). Let k be the number of indices on which all triples among A_i agree. We know that $1 \leq k \leq 3$.

We proceed inductively on k, as for $k = 0$ we already know from Lemma 11.12 how to reduce the relation. Hence from now on decreasing k is also a reduction.

Lemma 11.14. *Suppose that all triples A_i associate g_j to an index l. If there is a pair $(x,y)_{g_j}$ among A_i with $l \neq x, y$, then either $\{l,x,y\}$ is the support of all triples among A_i or we can make a reduction.*

Proof. Suppose that there is a triple A_i with support $\{l,b,c\}$ different from $\{l,x,y\}$. We can assume $x \neq b,c$. We apply the relation $A_i + (x,y)_{g_j} = \tilde{A}_i + (l,y)_{g_j}$, where $\tilde{A}_i$ associates g_j to x and agrees with A_i on b and c. This relation reduces k. ∎

Lemma 11.15. *Suppose that all triples from A_i associate g_j to an index l. If all pairs $(x,y)_{g_j}$ among A_i have l in the support then we can reduce all such pairs.*

Proof. Recall that t is the number of triples among A_i. Let p be the number of g_j pairs among A_i. Let t_1' (resp. t_2') be the number of triples in B_i that assign (resp. do not assign) g_j to l. Let p_1' (resp. p_2') be the number of g_j pairs among B_i that have (resp. do not have) l in the support. We know that $t \geq t_1' + t_2'$. Comparing the number of times g_j appears in A_i and B_i we get

$$t + 2p = t_1' + t_2' + 2(p_1' + p_2').$$

Comparing the number of times g_j appears on index l we get

$$t + p = t'_1 + p'_1.$$

This forces $t'_2 = p'_2 = 0$, $t = t'_1$ and $p = p'_1$. Hence all g_j pairs and triples among A_i and B_i must assign g_j to l. Hence the multisets of pairs must be the same for A_i and B_i. ∎

LEMMA 11.16. *Suppose that all triples from A_i associate g_j to an index l. If there are g_l pairs among A_i, then we can make a reduction.*

Proof. Without loss of generality we assume $g_l = g_1$. By Lemma 11.15, it is enough to prove that if there are pairs $(a, b)_{g_1}$ among A_i with $a, b \neq 1$ then we can make a reduction. Suppose that there is such a pair. By Lemma 11.14 all the triples among A_i must have the support $\{1, a, b\}$. So either $k = 1$ or $k = 3$. If $k = 1$ we can apply the relation

$$(1, a, b) + (1, b, a) + (a, b)_{g_1} + \mathfrak{nt} = (1, a)_{g_1} + (1, b)_{g_1} + (a, b)_{g_2} + (a, b)_{g_3}.$$

This reduces the number of triples. Thus we can assume that all triples among A_i are equal to $(1, a, b)$.

CLAIM. *Consider any pair $(c, d)_{g_2}$ among A_i. We can assume that its support is contained in $\{1, a, b\}$.*

Proof of the Claim. Suppose this is not the case, that is, $c \notin \{1, a, b\}$. By Lemma 11.14 we can assume $d = a$. There are two cases to consider:

1) There is a g_2 pair among A_i that does not contain a in the support. It must be equal to $(1, b)_{g_2}$ by Lemma 11.14. We can apply the relation $(1, b)_{g_2} + (a, c)_{g_2} = (c, 1)_{g_2} + (a, b)_{g_2}$. Applying once again Lemma 11.14 to the pair $(c, 1)_{g_2}$ we can make a reduction.

2) All g_2 pairs among A_i contain a in the support. By Lemma 11.15 we can make a reduction. ∎

Thus the support of all g_2 pairs among A_i is contained in $\{1, a, b\}$. The same holds for g_1 and g_3 pairs. Thus all group-based flows among A_i have support contained in $\{1, a, b\}$. Hence the same must hold for B_i. So our relation is a relation only on three indices. It is well-known [SS05] that the ideal for a tree with three edges is generated in degree 4, so in particular the relation in question is generated in degree 4. ∎

COROLLARY 11.17. *If all triples among A_i associate g_j to an index l, then there are no g_j pairs among A_i. Consequently, there are no g_j pairs among B_i and all triples among B_i associate g_j to l. Moreover, the number of triples among A_i equals the number of triples among B_i.* ∎

By the previous corollary *we assume* that there are no g_1 pairs among A_i or B_i. Moreover, there are the same number of triples among A_i and B_i and they all associate g_1 to 1.

LEMMA 11.18. *If all the triples among A_i and B_i have support contained in $\{1, 2, 3\}$ then we can make a reduction.*

Proof. Suppose all triples have support contained in $\{1, 2, 3\}$. In this case $k = 1$ or $k = 3$. If $k = 1$ then among A_i there is a triple $(1, 2, 3)$ and $(1, 3, 2)$. Any triple among B_i is equal to one of these. In particular, one of these triples can be reduced. If $k = 3$ there are no pairs. All triples among A_i and B_i are equal, thus the relation is trivial. ∎

11.3.1. Case $k = 1$. We first consider the most difficult case $k = 1$. As always let $A_1 = (1, 2, 3)$ and $B_1 = (1, b, c)$. As the proof is quite complicated, we include a diagram that describes the most important cases. While reading the proof, the reader is encouraged to follow at which node we are. The proof is "depth-first, left-first".

$$\begin{tikzcd}
 & & k=1 \arrow[dll] \arrow[d] \arrow[dr] & \\
b=2 \arrow[d] \arrow[drr] \arrow[drrr] & & c=3 & \text{any triples agree on exactly one index} \\
\text{no } (3,l)_{g_3}, (c,w)_{g_3} \arrow[d] \arrow[dr] & & (3,l)_{g_3} \arrow[d] \arrow[dr] & (c,w)_{g_3} \\
(c,f)_{g_2} & (1,c,g) & (1,3,p) & (3,o)_{g_2}
\end{tikzcd}$$

We start with the left node in the second row—assume $b = 2$. Then we may assume $c \neq 3$, or else we can reduce.

We move to the leftmost node in the third row: suppose that there is no g_3 pair among A_i that has c in the support, and symmetrically there is no g_3 pair among B_i that has 3 in the support. There must be a triple $(1, e, c)$ among A_i. If $e \neq 3$ then we apply the relation $(1, 2, 3) + (1, e, c) = (1, 2, c) + (1, e, 3)$ and reduce the triple $(1, 2, c)$. We may assume $e = 3$. Analogously, we may assume that there is a triple $(1, c, 3)$ among B_i. Hence there must be either a pair $(c, f)_{g_2}$ or a triple $(1, c, g)$ among A_i.

We continue to the leftmost node in the fourth row; suppose there is a pair $(c, f)_{g_2}$. If $f \neq 2$ we apply the relation $(1, 2, 3) + (c, f)_{g_2} = (1, c, 3) + (f, 2)_{g_2}$ and reduce the triple $(1, c, 3)$. If $f = 2$ we apply the relation $(1, 3, c) + (c, 2)_{g_2} = (1, 2, c) + (3, c)_{g_2}$ and reduce the triple $(1, 2, c)$.

Hence we can assume that there is a triple $(1, c, g)$ among A_i, the second node in the fourth row. If $g \neq 2$ then we apply the relation $(1, c, g) + (1, 2, 3) = (1, 2, g) + (1, c, 3)$ and reduce the triple $(1, c, 3)$. For $g = 2$ we apply the relation $(1, 2, 3) + (1, 3, c) + (1, c, 2) = (1, 2, c) + (1, 3, 2) + (1, c, 3)$ and reduce the triple $(1, 2, c)$.

We continue to the second node in the third row. We assume that there is a pair $(3, l)_{g_3}$ among B_i. If $l \neq c$ we apply the relation $(1, 2, c) + (3, l)_{g_3} = (1, 2, 3) + (c, l)_{g_3}$ and reduce the triple $(1, 2, 3)$. If there were a pair $(c, m)_{g_3}$ among A_i then analogously we could assume $m = 3$ and we would be able to reduce this pair. So there must be a triple $(1, n, c)$ among A_i. If $n \neq 3$ then we apply the relation $(1, 2, 3) + (1, n, c) = (1, n, 3) + (1, 2, c)$ and reduce the triple $(1, 2, c)$. So we assume $A_2 = (1, 3, c)$. Hence there is either a pair $(3, o)_{g_2}$ or a triple $(1, 3, p)$ among B_i.

We move to the third node in the fourth row: suppose that there is a triple $(1, 3, p)$ among B_i. If $p \neq 2$ we apply the relation $(1, 2, c) + (1, 3, p) = (1, 2, p) + (1, 3, c)$ and we reduce $(1, 3, c)$. So we assume $p = 2$. We apply the relation $(1, 3, 2) + (3, c)_{g_3} = (1, 3, c) + (2, 3)_{g_3}$ and reduce the triple $(1, 3, c)$.

We pass to the fourth node in the fourth row: we assume that there is a pair $(3,o)_{g_2}$ and there is no triple $(1,3,p)$ among B_i. If $o \neq 2$ then we apply the relation $(1,2,c)+(3,o)_{g_2} = (1,3,c)+(2,o)_{g_2}$ and reduce $(1,3,c)$. So we assume there is a pair $(2,3)_{g_2}$ among B_i. Suppose that this pair appears $r > 0$ times among B_i. Note that we may assume that there are no pairs $(2,s)_{g_2}$ among A_i. Indeed, suppose that there is such a pair. If $s \neq 3$ then we apply the relation $(1,3,c)+(2,s)_{g_2} = (1,2,c)+(3,s)_{g_2}$ and reduce the triple $(1,2,c)$. If $s = 3$ we reduce the pair $(2,3)_{g_2}$. Hence we assume there are at least $r+1$ triples of the type $(1,2,t)$ among A_i. If there is a triple with $t \neq 3$ then we apply the relation $(1,3,c)+(1,2,t) = (1,3,t)+(1,2,c)$ and reduce the triple $(1,2,c)$. Hence we assume there are at least $r+1$ triples $(1,2,3)$ among A_i. Notice that we may assume there are no triples of the type $(1,y,3)$ among B_i. Indeed, in that case we could apply the relation $(1,y,3)+(2,3)_{g_2} = (1,2,3)+(y,3)_{g_2}$ and reduce $(1,2,3)$. Hence we assume there are at least $r+1$ pairs of the type $(3,u)_{g_3}$ among B_i. If $u \neq c$ then we apply the relation $(1,2,c)+(3,u)_{g_3} = (1,2,3)+(c,u)_{g_3}$ and reduce the triple $(1,2,3)$. Hence we assume there are at least $r+1$ pairs $(3,c)_{g_3}$ among B_i. Note that we can assume there are no pairs of the type $(c,v)_{g_3}$ among A_i. Indeed, if $v = 3$ we could reduce this pair. If $v \neq 3$ we apply the relation $(1,2,3)+(c,v)_{g_3} = (1,2,c)+(3,v)_{g_3}$ and reduce the triple $(1,2,c)$. Hence we must have at least $r+1$ triples of the type $(1,z,c)$ among A_i. If $z \neq 3$ we apply the relation $(1,2,3)+(1,z,c) = (1,2,c)+(1,z,3)$ and reduce the triple $(1,2,c)$. So we may assume there are at least $r+1$ triples $(1,3,c)$ among A_i. Note that the elements g_2 on 3 cannot be reduced: among B_i there are only r pairs containing them and no triples. The contradiction finishes this case.

Consider the third node in the third row: there is a pair $(c,w)_{g_3}$ among A_i. This is completely analogous to the second node in this row, which was already considered.

Also the second node in the second row, $c = 3$, is analogous to the first node in the second row.

We are left with the last, third node in the second column: any two triples A_i and B_j agree on exactly one index, that is, on 1. By Lemma 11.18, there is a triple among B_i, say B_1, with support different from some triple in A_i, say A_1. Exchanging g_2 and g_3 if necessary, we can assume $b \neq 2$ and $b \neq 3$. By the case assumption there must be a pair $(b,d)_{g_2}$ among A_i. If $d \neq 2$ then we apply the relation $(1,2,3)+(b,d)_{g_2} = (1,b,3)+(d,2)_{g_2}$ and reduce to the case $b = 2$ ([4]). Analogously, we must have the same pair among B_i and it can be reduced.

11.3.2. Case $k = 2$ or $k = 3$. Suppose now that $k = 2$. Let $A_1 = (1,2,3)$ and $B_1 = (1,2,c)$. If we cannot reduce B_1 then there must be a pair $(c,d)_{g_3}$ among A_i and a pair $(3,e)_{g_3}$ among B_i. If $d = 3$ and $e = c$ we can reduce the pairs. Thus we can assume that $d \neq 3$. We apply the relation $(1,2,3)+(c,d)_{g_3} = (1,2,c)+(3,d)_{g_3}$ and reduce the triple $(1,2,c)$.

The last, easiest case is $k = 3$. Then all triples are equal to $(1,2,3)$, and there are no pairs by Corollary 11.17. Hence we can reduce the triples. This finishes the proof of Theorem 11.1.

([4]) Notice that we do not reduce to the case $k = 2$, as if this were true we would have already been in the first node in the second column $b = 2$.

12. Open problems

We have already presented a few conjectures. Here we would like to give a list of problems that should be much easier, but still we find them interesting.

We start with the questions concerning normality. We already know that many general group-based models give rise to projectively normal varieties for trivalent trees. However, not much is known about trees of higher valency. Of course, by Proposition 4.72, it is enough to consider claw trees. The normality questions are important, as many toric methods work only for normal polytopes. We have already applied some of them to compute Hilbert functions. Further applications to the conjecture of Sturmfels and Sullivant could be possible by the methods of "finite generation in rings with infinitely many variables"; for more details see [HS12], [DK14].

CONJECTURE 12.1. *Let T be any tree. The polytope representing the binary 3-Kimura model on T is normal.*

In [DBM12] it was proved that the *projective* variety representing the 2-Kimura model is not normal. We also know that the affine variety representing the general group-based model for $\mathbb{Z}_6$ is not normal.

Another question is to what extent the methods of Section 11 can be applied to other abelian groups. The following conjecture appearing in the author's PhD was recently proved in [DB14].

CONJECTURE 12.2. *The projective scheme associated to the group-based model for $\mathbb{Z}_3$ and any tree can be represented by an ideal generated in degree 3.*

We finish by restating, in our opinion, the most interesting, important and difficult Conjecture 6.6.

CONJECTURE 12.3. *The variety $X(K_{n,1})$ is equal to the (scheme-theoretic) intersection of all the varieties $X(T_i)$, where T_i is a prolongation of $K_{n,1}$ that has only two inner vertices, both of them of valency at least 3.*

Appendix 1

Here we give an explicit example when the equality of parameters before the Fourier transform does not imply the equality after it.

Let $G = \mathbb{Z}_6$. The transition matrices are of the form

$$\begin{bmatrix} a & b & c & d & e & f \\ f & a & b & c & d & e \\ e & f & a & b & c & d \\ d & e & f & a & b & c \\ c & d & e & f & a & b \\ b & c & d & e & f & a \end{bmatrix}.$$

The matrix of the type above corresponds to a function $g : G \to \mathbb{C}$ such that $g(0) = a$, $g(1) = b$, $g(2) = c$, $g(3) = d$, $g(4) = e$ and $g(5) = f$. The Fourier transform of g

gives $\widehat{g}(\chi_0) = a+b+c+d+e+f$, $\widehat{f}(\chi_1) = a+jb+j^2c+j^3d+j^4e+j^5f$, $\widehat{f}(\chi_2) = a+j^2b+j^4c+d+j^2e+j^4f$ etc. where j is a primitive sixth root of unity. We consider a submodel defined by $g(0)=g(1)=g(5)$ and $g(2)=g(4)$. This corresponds to $a=b=f$ and $c=e$. The Fourier transform gives $(x_0,x_1,x_2,x_3,x_4,x_5) = (3a+2c+d, 2a-c-d, -c+d, -a+2c-d, -c+d, 2a-c-d)$. This defines a linear subspace given by $x_4=x_2$, $x_5=x_1$ and $x_1+3x_2+2x_3=0$. The latter is not an equality of distinct variables.

Appendix 2

Here we present the precise results of computing the Hilbert–Ehrhart polynomials for a few G-models. The results are taken from a joint work with Maria Donten-Bury [DBM12].

For the groups $\mathbb{Z}_8$, $\mathbb{Z}_2\times\mathbb{Z}_2\times\mathbb{Z}_2$ and $\mathbb{Z}_9$ we computed only the Hilbert function and, as we could not check normality, we do not know if it is equal to the Hilbert–Ehrhart polynomial.

Models for $G=H=\mathbb{Z}_3$.

dilation	snowflake	3-caterpillar
1	243	243
2	21627	21627
3	903187	904069
4	21451311	21496023
5	330935625	331976637
6	3647265274	3662146270
7	30770591364	30920349834
8	209116329075	210269891871
9	1189466778457	1196661601837
10	5831112858273	5868930577941
11	25205348411361	25377886917819

Models for $G=H=\mathbb{Z}_2\times\mathbb{Z}_2$ (3-Kimura).

dilation	snowflake	3-caterpillar
1	1024	1024
2	396928	396928
3	69248000	69324800
4	5977866515	5990170739
5	291069470720	291864710144
6	8967198289920	8995715702784

Models for $G=H=\mathbb{Z}_4$.

dilation	snowflake	3-caterpillar
1	1024	1024
2	396928	396928
3	69248000	69324800
4	6122557220	6138552524
5	310273545216	311525688320
6	10009786400352	10062179606880

Models for $G = H = \mathbb{Z}_5$.

dilation	snowflake	3-caterpillar
1	3125	3125
2	3834375	3834375
3	2229584375	2230596875
4	640338121875	642089603125

Models for $G = H = \mathbb{Z}_7$. In this case the first three dilations of the polytopes have the same number of points. The numbers of points in the fourth dilations were too large to obtain precise results. Hence we computed only the numbers of points mod 64, which is sufficient to prove that the Hilbert–Ehrhart polynomials are different.

dilation	snowflake	3-caterpillar
1	16807	16807
2	117195211	117195211
3	423913952448	423913952448
4	$\equiv 54 \bmod 64$	$\equiv 14 \bmod 64$

Models for $G = H = \mathbb{Z}_8$.

dilation	snowflake	3-caterpillar
1	32768	32768
2	454397952	454397952
3	3375180251136	3375013036032

Models for $G = H = \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$.

dilation	snowflake	3-caterpillar
1	32768	32768
2	454397952	454397952
3	3375180251136	3375013036032

Models for $G = H = \mathbb{Z}_9$.

dilation	snowflake	3-caterpillar
1	59049	59049
2	1499667453	1499667453
3	20938605820263	20937202945056

References

[APRS11] E. S. Allman, S. Petrović, J. A. Rhodes, and S. Sullivant, *Identifiability of two-tree mixtures for group-based models*, IEEE/ACM Trans. Comput. Biology and Bioinformatics 8 (2011), 710–722.

[AR06] E. S. Allman and J. A. Rhodes, *The identifiability of tree topology for phylogenetic models, including covarion and mixture models*, J. Comput. Biol. 13 (2006), 1101–1113.

[AR08] E. S. Allman and J. A. Rhodes, *Phylogenetic ideals and varieties for the general Markov model*, Adv. Appl. Math. 40 (2008), 127–148.

[BDGM15] M. Beck, J. Delgado, J. Gubeladze, and M. Michałek, *Very ample and Koszul segmental fibrations*, J. Algebraic Combin. 42 (2015), 165–182.

[BHV01] L. J. Billera, S. P. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. Appl. Math. 27 (2001), 733–767.

[Bru13] W. Bruns, *The quest for counterexamples in toric geometry*, in: Commutative Algebra and Algebraic Geometry, Ramanujan Math. Soc., 2013, 45–61.

[BBKM] W. Buczyńska, J. Buczyński, K. Kubjas, and M. Michałek, *On the graph labellings arising from phylogenetics*, Central Eur. J. Math. 11 (2013), 1577–1592.

[BDW09] W. Buczyńska, M. Donten, and J. A. Wiśniewski, *Isotropic models of evolution with symmetries*, in: Contemp. Math. 496, Amer. Math. Soc., 2009, 111–132.

[BW07] W. Buczyńska and J. A. Wiśniewski, *On geometry of binary symmetric models of phylogenetic trees*, J. Eur. Math. Soc. 9 (2007), 609–635.

[CFS08] M. Casanellas and J. Fernández-Sánchez, *Geometry of the Kimura 3-parameter model*, Adv. Appl. Math. 41 (2008), 265–292.

[CLS11] D. A. Cox, J. B. Little, and H. K. Schenck, *Toric Varieties*, Amer. Math. Soc., 2011.

[CP07] J. Chifman and S. Petrović, *Toric ideals of phylogenetic invariants for the general group-based model on claw trees $k_{1,n}$*, in: Algebraic Biology, Lecture Notes in Computer Sci. 4545, Springer, 2007, 307–321.

[DB14] M. Donten-Bury, *Phylogenetic invariants for $\mathbb{Z}_3$ scheme-theoretically*, arXiv:1405.4169 (2014).

[DBM12] M. Donten-Bury and M. Michałek, *Phylogenetic invariants for group-based models*, J. Algebr. Statist. 3 (2012), 44–63.

[DK09] J. Draisma and J. Kuttler, *On the ideals of equivariant tree models*, Math. Ann. 344 (2009), 619–644.

[DK14] J. Draisma and J. Kuttler, *Bounded-rank tensors are defined in bounded degree*, Duke Math. J. 163 (2014), 35–63.

[ES96] D. Eisenbud and B. Sturmfels, *Binomial ideals*, Duke Math. J. 84 (1996), 1–45.

[EKS14] A. Engström, T. Kahle, and S. Sullivant, *Multigraded commutative algebra of graph decompositions*, J. Algebraic Combin. 39 (2014), 335–372.

[ERSS05] N. Eriksson, K. Ranestad, B. Sturmfels, and S. Sullivant, *Phylogenetic algebraic geometry*, in: Projective Varieties with Unexpected Properties (Siena, 2004), de Gruyter, 2005, 237–255.

[ES93] S. N. Evans and T. P. Speed, *Invariants of some probability models used in phylogenetic inference*, Ann. Statist. 21 (1993), 355–377.

[Ful93] W. Fulton, *Introduction to Toric Varieties*, Ann. of Math. Stud. 131, Princeton Univ. Press, Princeton, NJ, 1993.

[GAP] *GAP—Groups, Algorithms, and Programming, Version 4.4.10*, The GAP Group, http://www.gap-system.org, 2007.

[GJ00] E. Gawrilow and M. Joswig, *Polymake: a framework for analyzing convex polytopes*, in: Polytopes—Combinatorics and Computation, G. Kalai and G. M. Ziegler (eds.), Birkhäuser, 2000, 43–73.

[GS] D. R. Grayson and M. E. Stillman, *Macaulay2, a software system for research in algebraic geometry*, http://www.math.uiuc.edu/Macaulay2/.

[Har77] R. Hartshorne, *Algebraic Geometry*, Grad. Texts in Math. 52, Springer, New York, 1977.

[HP89] M. Hendy and D. Penny, *A framework for the quantitative study of evolutionary trees*, Systematic Zoology 38 (1989), 297–309.

[HS12] C. J. Hillar and S. Sullivant, *Finite Gröbner bases in infinite dimensional polynomial rings and applications*, Adv. Math. 229 (2012), 1–25.

[Ibe09] O. C. Ibe, *Markov Processes for Stochastic Modeling*, Elsevier/Academic Press, Amsterdam, 2009.

[Ilt10] N. Ilten, *Deformations of rational varieties with codimension-one torus action*, Doctoral thesis, FU Berlin, 2010.

[JC69] T. Jukes and C. Cantor, *Evolution of protein molecules*, in: Mammalian Protein Metabolism, Academic Press, 1969, 21–132.

[KR14] T. Kahle and J. Rauh, *Toric fiber products versus Segre products*, Abh. Math. Sem. Univ. Hamburg 84 (2014), 187–201.

[KKMSD73] G. Kempf, F. F. Knudsen, D. Mumford, and B. Saint-Donat, *Toroidal Embeddings. I*, Lecture Notes in Math. 339, Springer, Berlin, 1973.

[Kim80] M. Kimura, *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences*, J. Molecular Evolution 16 (1980), 111–120.

[Kim81] M. Kimura, *Estimation of evolutionary distances between homologous nucleotide sequences*, Proc. Nat. Acad. Sci. USA 78 (1981), 454–458.

[Kub12] K. Kubjas, *Hilbert polynomial of the Kimura 3-parameter model*, J. Algebr. Statist. 3 (2012), 64–69.

[LM14] M. Lasoń and M. Michałek, *On the toric ideal of a matroid*, Adv. Math. 259 (2014), 1–12.

[Mic] M. Michałek, *Toric geometry of the 3-Kimura model for any tree*, Adv. Geom. 14 (2014), 11–30.

[Mic11] M. Michałek, *Geometry of phylogenetic group-based models*, J. Algebra 339 (2011), 339–356.

[Mic13] M. Michałek, *Constructive degree bounds for group-based models*, J. Combin. Theory Ser. A 120 (2013), 1672–1694.

[Ney71] J. Neyman, *Molecular studies of evolution: A source of novel statistical problems*, in: Statistical Decision Theory and Related Topics (Lafayette, IN, 1970), Academic Press, New York, 1971, 1–27.

[Oda87] T. Oda, *Convex Bodies and Algebraic Geometry*, Springer, 1987.

[PS05] L. Pachter and B. Sturmfels (eds.), *Algebraic Statistics for Computational Biology*, Cambridge Univ. Press, New York, 2005.

[Stu96] B. Sturmfels, *Gröbner Bases and Convex Polytopes*, University Lecture Series, vol. 8, Amer. Math. Soc., 1996.

[Stu98] B. Sturmfels, *Polynomial equations and convex polytopes*, Amer. Math. Monthly 105 (1998), 907–922.

[SS05] B. Sturmfels and S. Sullivant, *Toric ideals of phylogenetic invariants*, J. Comput. Biol. 12 (2005), 204–228.

[Sul07] S. Sullivant, *Toric fiber products*, J. Algebra 316 (2007), 560–577.

[SSE93] L. A. Székely, M. A. Steel, and P. L. Erdős, *Fourier calculus on evolutionary trees*, Appl. Math. 14 (1993), 200–210.

[tt] 4ti2 team, *4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces*, www.4ti2.de.

[Zwi] P. Zwiernik, *Toric geometry of statistical binary graph models*, MSc thesis, Warsaw Univ., 2006.