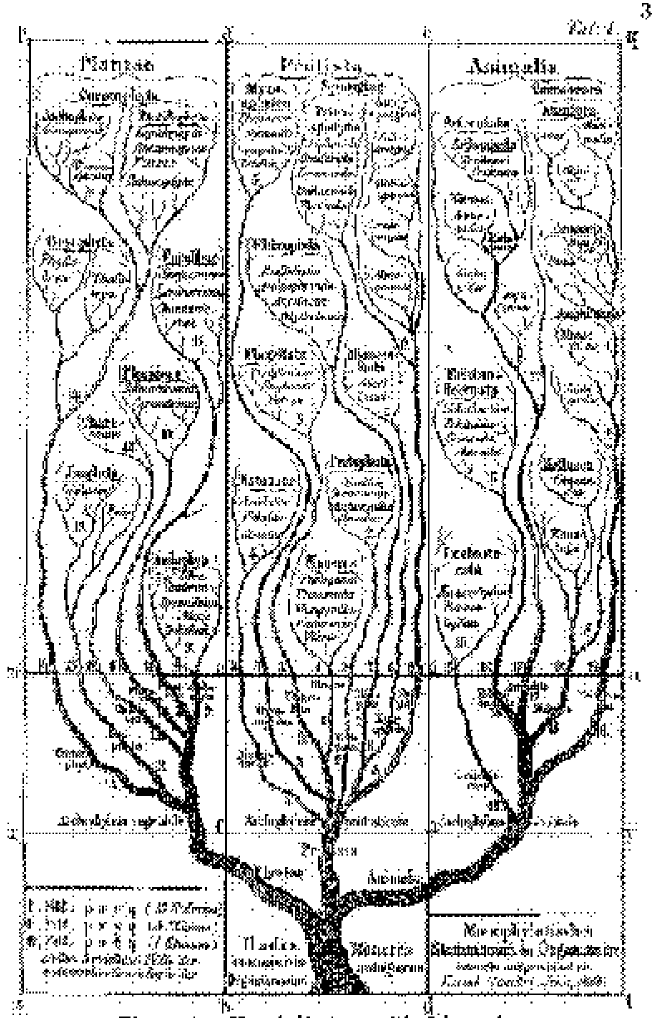
Algebraic varieties arising from phylogenetic trees

W. Buczynska, J.A. Wisniewski

Institute of Mathematics, Warsaw University, Poland

phylogenetics

Phylogenetics: reconstructing historical relation between species by analyzing their *present* features and putting their common ancestors in a diagram which forms a tree. [e.g. Häckel, 1866]



three (un?)related problems

- counting points in polyhedra

three (un?)related problems

- counting points in polyhedra
- networks of paths in a tree

three (un?)related problems

- counting points in polyhedra
- networks of paths in a tree
- Markov processes on a tree

⋆ product of functions

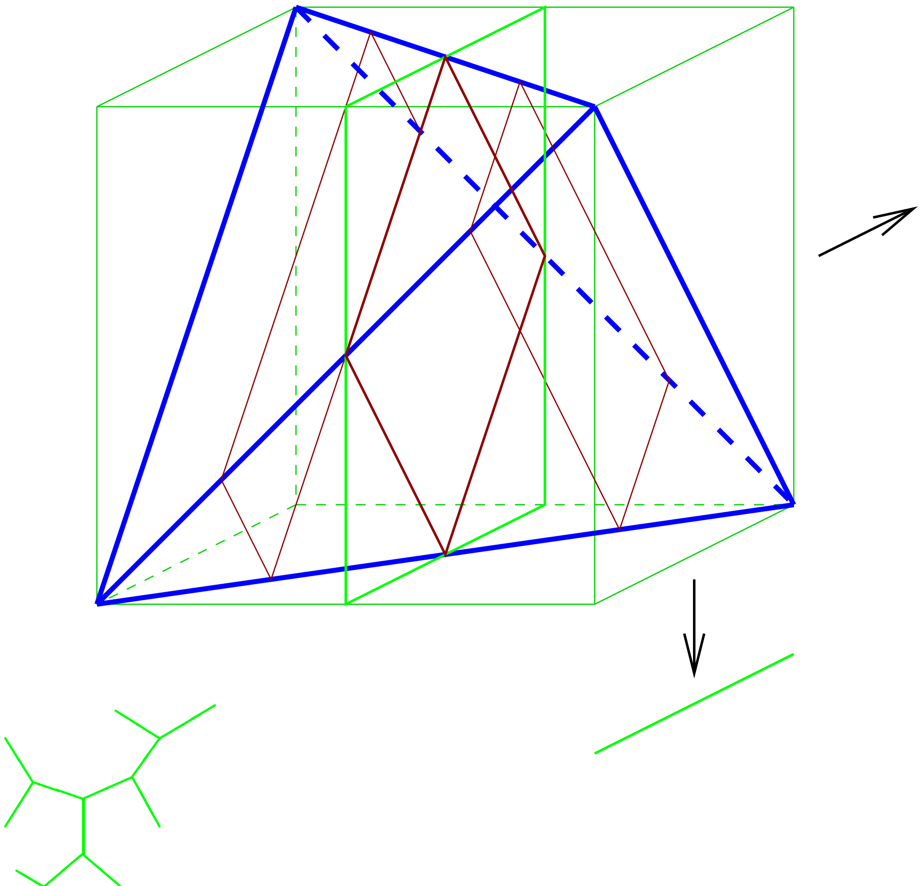
For a positive integer n let $[n] = \{0, \ldots n\}$.
Function $f : [n] \to \mathbb{Z}$ is symmetric if for every $k \in [n]$ it holds $f(k) = f(n-k)$.
By $\mathbf{1} : [n] \to \mathbb{Z}$ denote the unit function.

$\star$ product of functions

For a positive integer n let $[n] = \{0, \ldots n\}$.
Function $f : [n] \to \mathbb{Z}$ is symmetric if for every $k \in [n]$ it holds $f(k) = f(n-k)$.
By $\mathbf{1} : [n] \to \mathbb{Z}$ denote the unit function.
If $f_1\ f_2 : [n] \to \mathbb{Z}$ are symmetric functions then we define their symmetric product $f_1 \star f_2 : [n] \to \mathbb{Z}$ such that for $k \leq n/2$:

$$(f_1 \star f_2)(k) = 2 \cdot \left(\sum_{i=0}^{k-1} \sum_{j=0}^{i} f_1(i) f_2(k+i-2j) \right) + \left(\sum_{i=k}^{n-k} \sum_{j=0}^{k} f_1(i) f_2(k+i-2j) \right)$$

geometric interpretation of $\star$

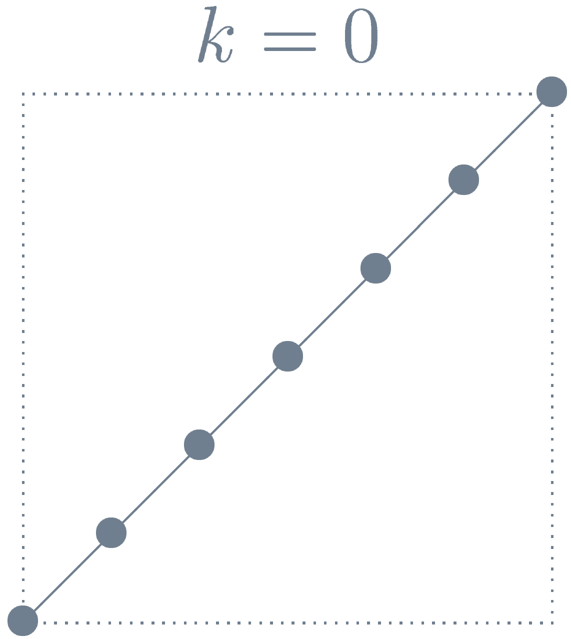


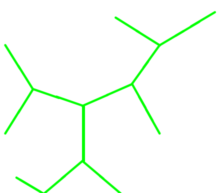
Consider the simplex Δ as in the picture

$(f_1 \star f_2)(k)$ is equal to the sum of products of f_1 and f_2 counted over points of lattice spanned by Δ in k-th slice of $n \cdot \Delta$

$(\mathbf{1} \star \mathbf{1})(k) = (k+1)(n-k+1)$ is the number of lattice points in k-th slice of $n \cdot \Delta$

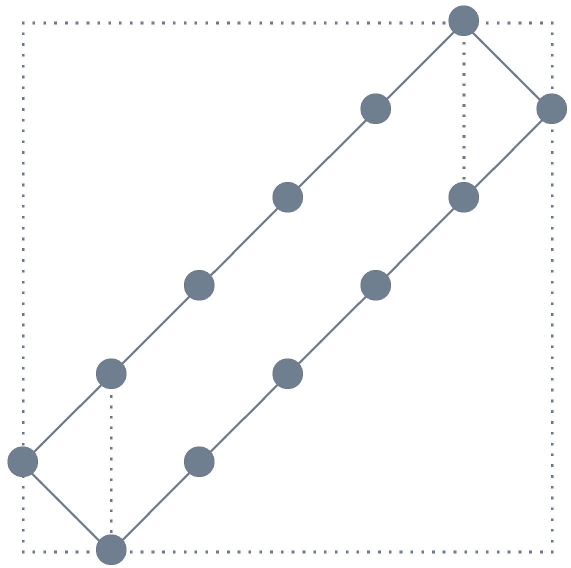
travel trough $6 \cdot \Delta$





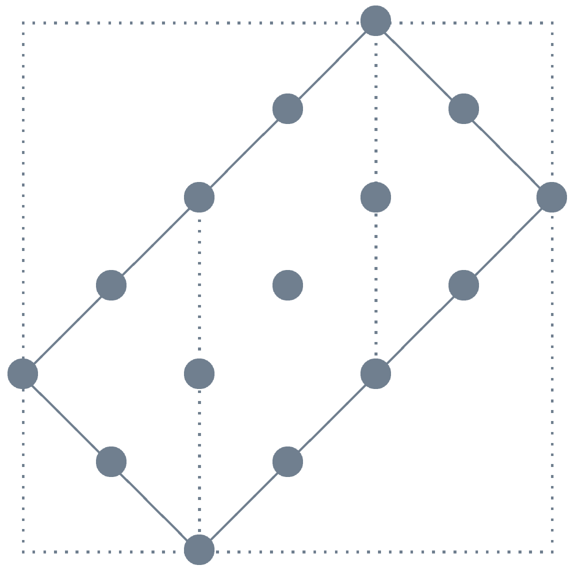
travel trough $6 \cdot \Delta$

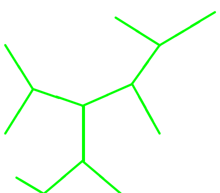
$k = 1$



travel trough $6 \cdot \Delta$

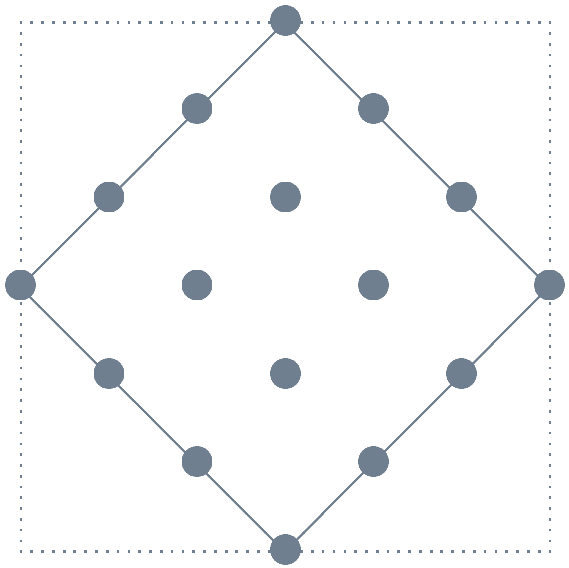
$k = 2$





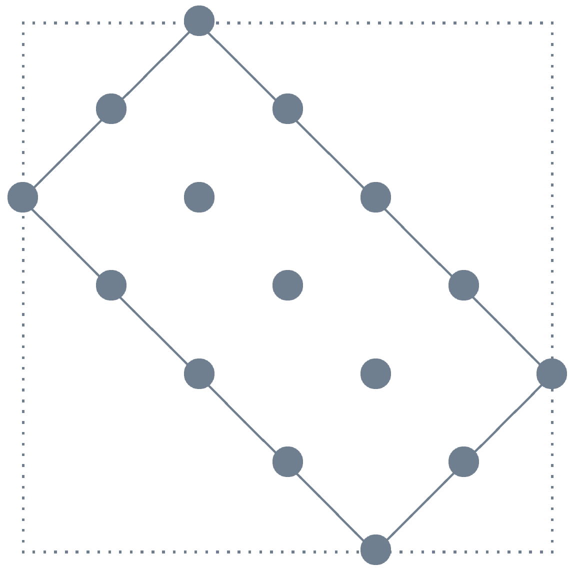
travel trough $6 \cdot \Delta$

$k = 3$

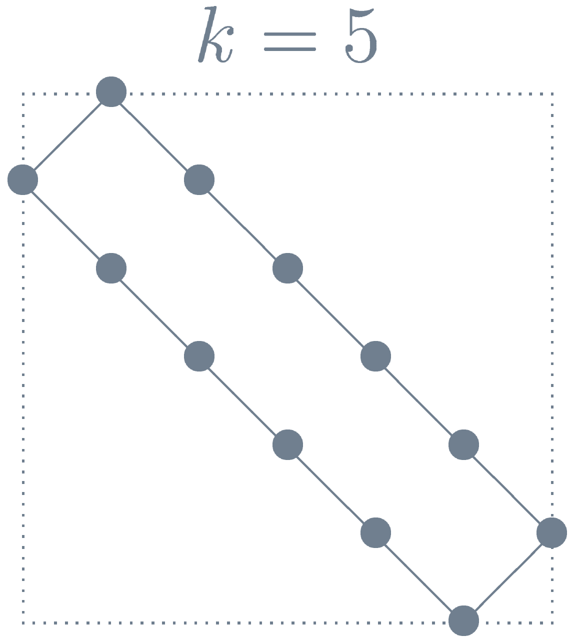


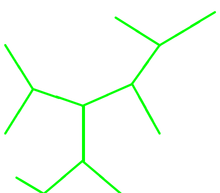
travel trough $6 \cdot \Delta$

$k = 4$



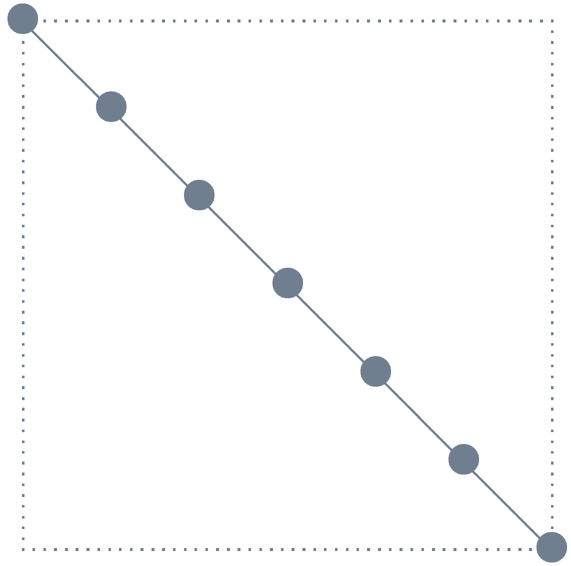
travel trough $6 \cdot \Delta$





travel trough $6 \cdot \Delta$

$k = 6$



properties of $\star$

- $\star$ is commutative, $f_1 \star f_2 = f_2 \star f_1$

properties of $\star$

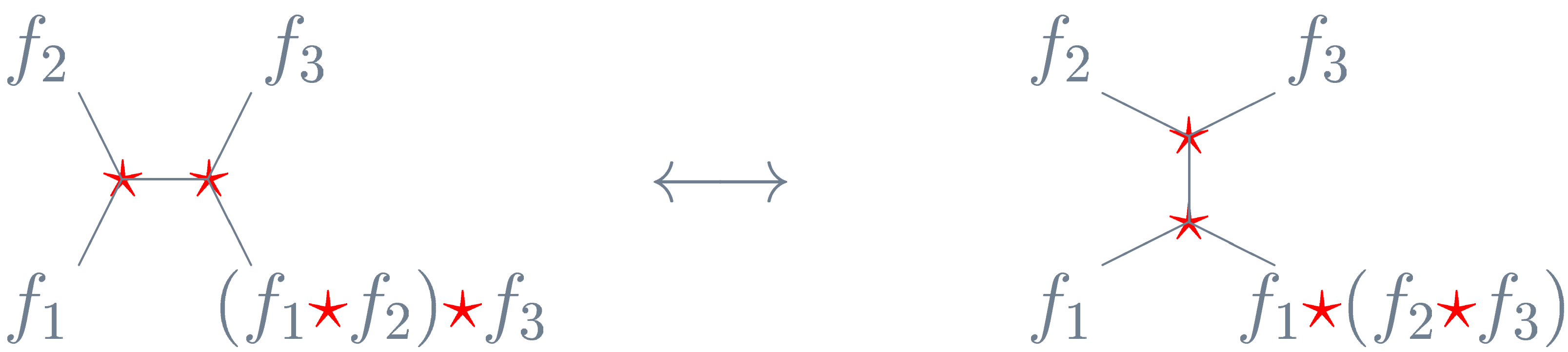
- $\star$ is commutative, $f_1 \star f_2 = f_2 \star f_1$
- $\star$ is usually nonassociative, i.e. $(f_1 \star f_2) \star f_3 \neq f_1 \star (f_2 \star f_3)$

properties of $\star$

- $\star$ is commutative, $f_1 \star f_2 = f_2 \star f_1$
- $\star$ is usually nonassociative, i.e. $(f_1 \star f_2) \star f_3 \neq f_1 \star (f_2 \star f_3)$
- however, [theorem 1] If Ω is the smallest set of functions closed under $\star$ and containing 1 then $\star$ is associative within Ω

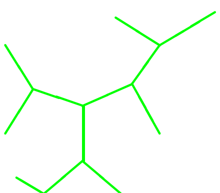
properties of $\star$

- $\star$ is commutative, $f_1 \star f_2 = f_2 \star f_1$
- $\star$ is usually nonassociative, i.e. $(f_1 \star f_2) \star f_3 \neq f_1 \star (f_2 \star f_3)$
- however, [theorem 1] If Ω is the smallest set of functions closed under $\star$ and containing 1 then $\star$ is associative within Ω
-



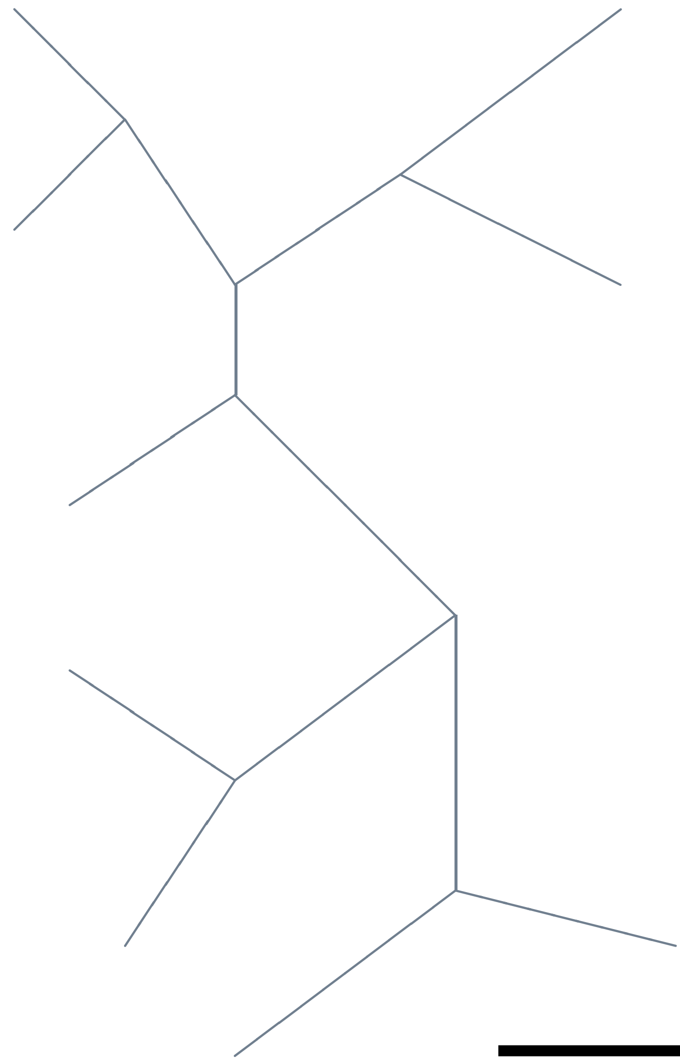
trees, sockets and networks

Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$



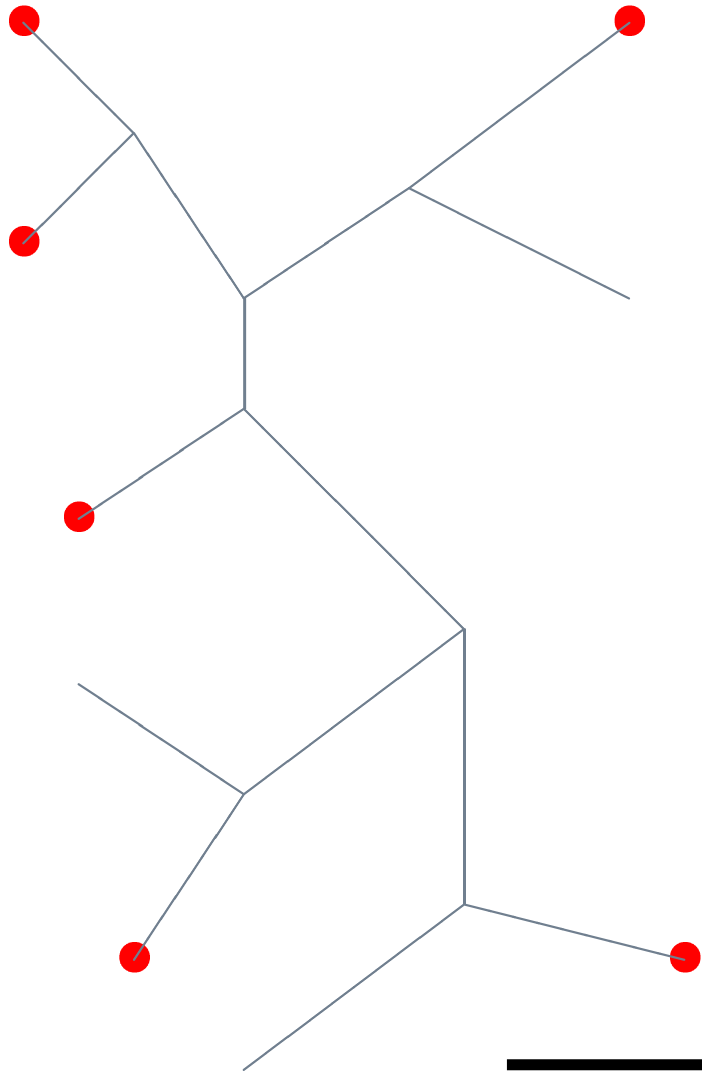
trees, sockets and networks

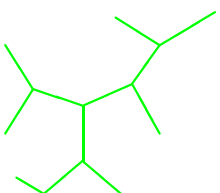
Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$



trees, sockets and networks

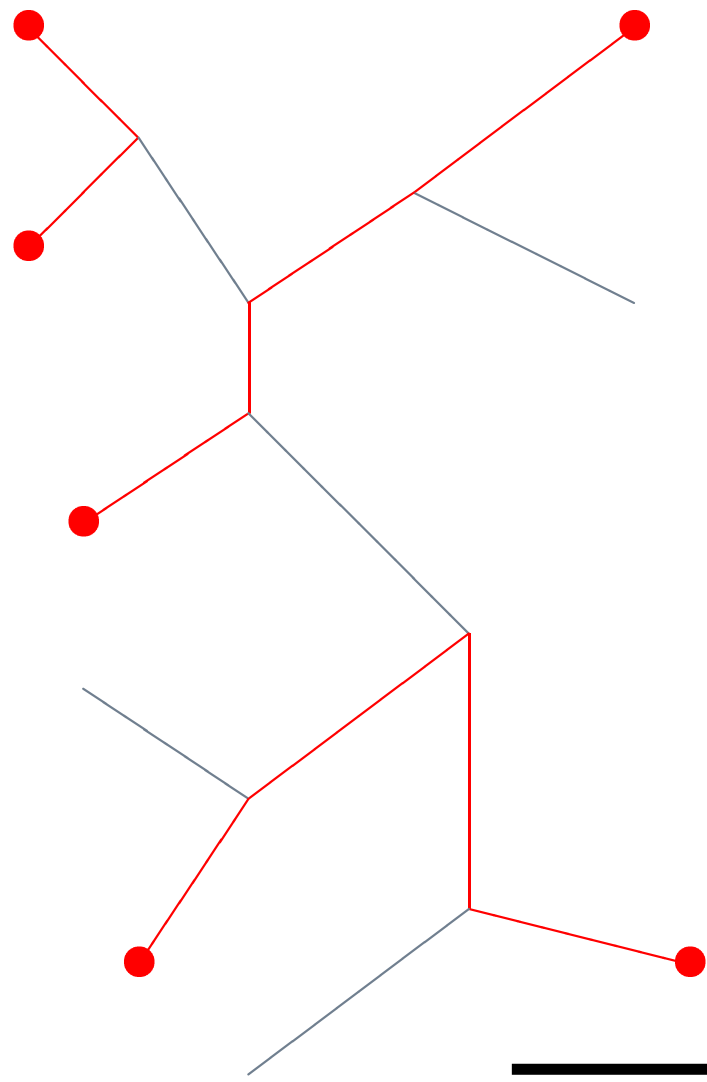
Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$





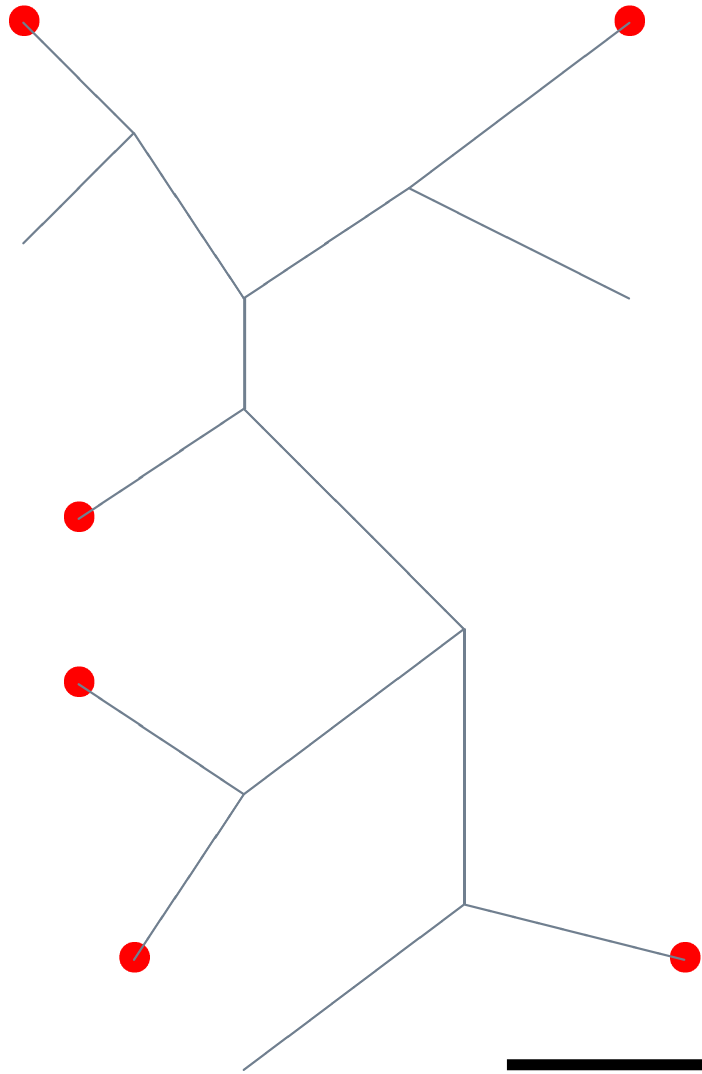
trees, sockets and networks

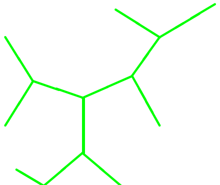
Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$



trees, sockets and networks

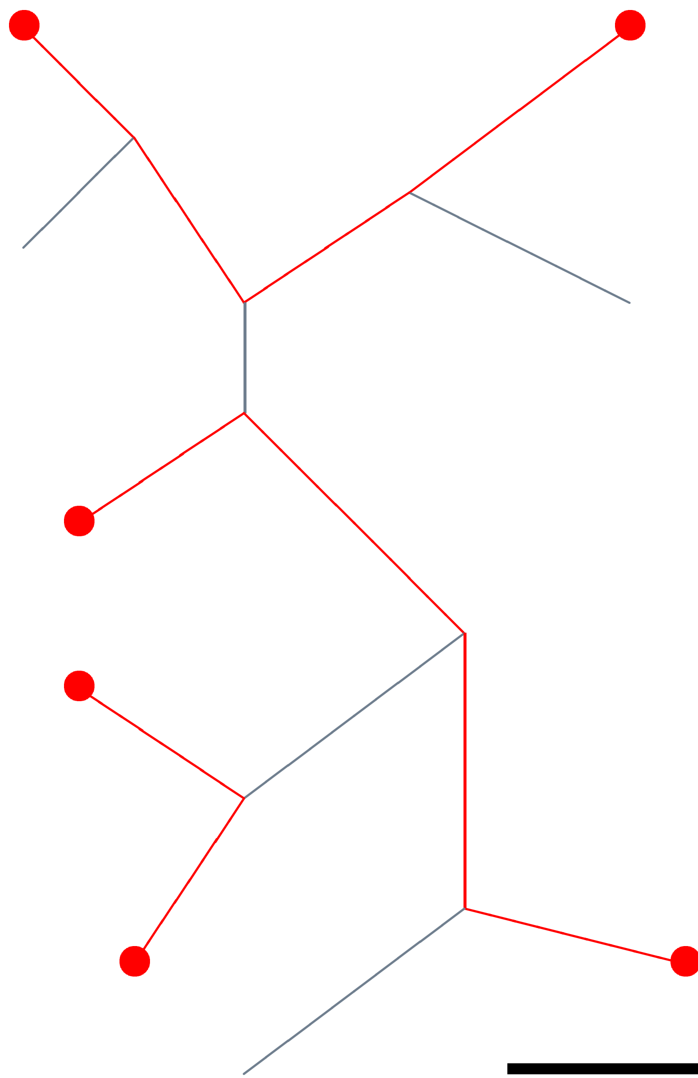
Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$





trees, sockets and networks

Consider a tree $\mathcal{T}$ which has $d+1$ leaves $\mathcal{L}$, $d-1$ inner trivalent nodes $\mathcal{N}$ and $2d-1$ edges $\mathcal{E}$; *socket* is a subset of $\mathcal{L}$ which has even number of elements; *path* in $\mathcal{T}$ is a connected union of edges, *network* is a set of non-meeting paths in $\mathcal{T}$ with ends in $\mathcal{L}$



varieties associated to trees

[lemma] There is a bijection between the set of sockets and networks, that is for every socket σ there exists a unique network $\mu(\sigma)$ whose end points are in σ

varieties associated to trees

[lemma] There is a bijection between the set of sockets and networks, that is for every socket σ there exists a unique network $\mu(\sigma)$ whose end points are in σ

For every edge $e \in \mathcal{E}$ we consider a $\mathbb{P}^1_e$ with homogeneous coordinates $[y_0^e, y_1^e]$. Moreover consider a projective space $\mathbb{P}_\Sigma$ of dimension $2^d - 1$ with homogeneous coordinates $[z_\sigma]$ indexed by sockets of $\mathcal{T}$.

varieties associated to trees

[lemma] There is a bijection between the set of sockets and networks, that is for every socket σ there exists a unique network $\mu(\sigma)$ whose end points are in σ

Define rational map $\prod_{e \in \mathcal{E}} \mathbb{P}^1_e \to \mathbb{P}_\Sigma$ such that

$$z_\sigma = \prod_{e \in \mu(\sigma)} y_1^e \cdot \prod_{e \notin \mu(\sigma)} y_0^e$$

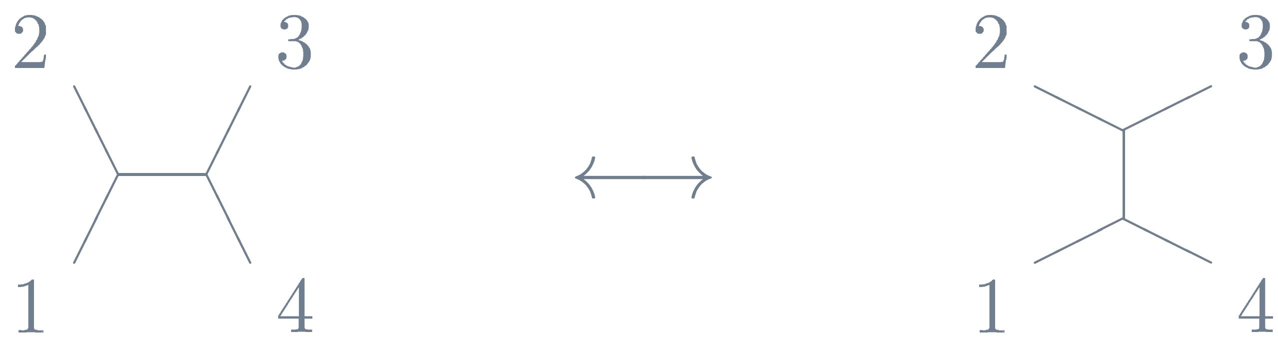
The model of the tree, $X(\mathcal{T}) \subset \mathbb{P}_\Sigma$, is the closure of the image of this map, $\dim X(\mathcal{T}) = 2d$.

deforming $X(\mathcal{T})$ within $\mathbb{P}_\Sigma$

Leaves of $\mathcal{T}$ can be labeled by numbers $1, \ldots, d+1$ or, equivalently, given $d+1$ points we can make them leaves of a (non-unique) tree $\mathcal{T}$.

deforming $X(\mathcal{T})$ within $\mathbb{P}_\Sigma$

Leaves of $\mathcal{T}$ can be labeled by numbers $1, \ldots, d+1$ or, equivalently, given $d+1$ points we can make them leaves of a (non-unique) tree $\mathcal{T}$.



deforming $X(\mathcal{T})$ within $\mathbb{P}_\Sigma$

Leaves of $\mathcal{T}$ can be labeled by numbers $1, \ldots, d+1$ or, equivalently, given $d+1$ points we can make them leaves of a (non-unique) tree $\mathcal{T}$. Thus, all the varieties representing different labeled trees can be embedded in a fixed $\mathbb{P}_\Sigma$

deforming $X(\mathcal{T})$ within $\mathbb{P}_\Sigma$

Leaves of $\mathcal{T}$ can be labeled by numbers $1, \ldots, d+1$ or, equivalently, given $d+1$ points we can make them leaves of a (non-unique) tree $\mathcal{T}$. Thus, all the varieties representing different labeled trees can be embedded in a fixed $\mathbb{P}_\Sigma$

These varieties can be non-isomorphic (one can check it), however [theorem 2] they are in the same connected component of the Hilbert scheme of $\mathbb{P}_\Sigma$, that is $X(\mathcal{T}_1)$ can be deformed to $X(\mathcal{T}_2)$ if only $\mathcal{T}_1$ and $\mathcal{T}_2$ have the same number of leaves.

binary Markov process on tree

Fix a root r in tree $\mathcal{T}$ - this implies an order $<$ on the set of vertexes $\mathcal{V} = \mathcal{L} \cup \mathcal{N}$. To each vertex $v \in \mathcal{V}$ assign a random variable ξ_v which takes value in $\{\alpha_1, \alpha_2\}$.

binary Markov process on tree

Fix a root r in tree $\mathcal{T}$ - this implies an order $<$ on the set of vertexes $\mathcal{V} = \mathcal{L} \cup \mathcal{N}$. To each vertex $v \in \mathcal{V}$ assign a random variable ξ_v which takes value in $\{\alpha_1, \alpha_2\}$.

Variables ξ_v determine a Markov process on $\mathcal{T}$ if (intuitively) the value of ξ_v depends only on the value of ξ_u, where u is the node immediately preceeding v.

binary Markov process on tree

Fix a root r in tree $\mathcal{T}$ - this implies an order $<$ on the set of vertexes $\mathcal{V} = \mathcal{L} \cup \mathcal{N}$. To each vertex $v \in \mathcal{V}$ assign a random variable ξ_v which takes value in $\{\alpha_1, \alpha_2\}$.

For each edge $e = \langle u, v \rangle$ bounded by vertexes $u < v$ define the transition matrix A^e:

$$A^e_{ij} = P(\xi_v = \alpha_j | \xi_u = \alpha_i)$$

binary Markov process on tree

Fix a root r in tree $\mathcal{T}$ - this implies an order $<$ on the set of vertexes $\mathcal{V} = \mathcal{L} \cup \mathcal{N}$. To each vertex $v \in \mathcal{V}$ assign a random variable ξ_v which takes value in $\{\alpha_1, \alpha_2\}$.
For each edge $e = \langle u, v \rangle$ bounded by vertexes $u < v$ define the transition matrix A^e:

$$A^e_{ij} = P(\xi_v = \alpha_j | \xi_u = \alpha_i)$$

and set the probability of the variable ξ_r at the root: $P^r_i = P(\xi_r = \alpha_i)$

from Markov to phylogenetics

For a Markov process on a rooted tree $\mathcal{T}$ as above

from Markov to phylogenetics

For a Markov process on a rooted tree $\mathcal{T}$ as above and any function $\mathcal{V} \ni v \to \rho(v) \in \{1,2\}$

$$P(\bigwedge_{v \in \mathcal{V}} \xi_v = \alpha_{\rho(v)}) = P^r_{\rho(r)} \cdot \prod_{e=\langle u,v \rangle \in \mathcal{E}} A^e_{\rho(u)\rho(v)}$$

from Markov to phylogenetics

For a Markov process on a rooted tree $\mathcal{T}$ as above and any function $\mathcal{L} \ni v \to \rho(v) \in \{1,2\}$

$$P(\bigwedge_{v\in\mathcal{L}} \xi_v = \alpha_{\rho(v)}) = \sum_{\widehat{\rho}} P^r_{\widehat{\rho}(r)} \cdot \prod_{e=\langle u,v\rangle\in\mathcal{E}} A^e_{\widehat{\rho}(u)\widehat{\rho}(v)}$$

where the sum is taken over all $\widehat{\rho} : \mathcal{V} \to \{1,2\}$ which extend ρ.

from Markov to phylogenetics

For a Markov process on a rooted tree $\mathcal{T}$ as above and any function $\mathcal{L} \ni v \to \rho(v) \in \{1, 2\}$

$$P(\bigwedge_{v \in \mathcal{L}} \xi_v = \alpha_{\rho(v)}) = \sum_{\widehat{\rho}} P^r_{\widehat{\rho}(r)} \cdot \prod_{e=\langle u,v \rangle \in \mathcal{E}} A^e_{\widehat{\rho}(u)\widehat{\rho}(v)}$$

where the sum is taken over all $\widehat{\rho} : \mathcal{V} \to \{1, 2\}$ which extend ρ.
Phylogenetics: understand the shape of $\mathcal{T}$ by looking at the distribution of $P(\bigwedge_{v \in \mathcal{L}} \xi_v = \alpha_{\rho(v)})$, that is

Fourier transformation

Phylogenetics wants to understand the locus of possible probability values of a Markov process on a fixed tree $\mathcal{T}$

$$\mathcal{X}(\mathcal{T}) :=$$
$$\{\zeta_\rho = P(\textstyle\bigwedge_{v\in\mathcal{L}} \xi_v = \alpha_{\rho(v)}) : A^e_{ij}, P^r_i \;\text{ are }\; \text{arbitrary}\}$$

in the simplex with coordinates ζ_ρ where $\zeta_\rho \geq 0$, $\sum_\rho \zeta_\rho = 1$.

Fourier transformation

Assume:

Fourier transformation

Assume:

- the root distribution is uniform, $P_1^r = P_2^r$

Fourier transformation

Assume:

- the root distribution is uniform, $P_1^r = P_2^r$
- the transition matrices are symmetric:

$$A_{12}^e = A_{21}^e, \quad A_{11}^e = A_{22}^e$$

Fourier transformation

Assume:

- the root distribution is uniform, $P_1^r = P_2^r$
- the transition matrices are symmetric:

$$A^e_{12} = A^e_{21}, \quad A^e_{11} = A^e_{22}$$

then [proposition] after suitable change of coordinates (and identifying spaces) the varieties $\mathcal{X}(\mathcal{T})$ and $X(\mathcal{T})$ coincide.

proof: working dictionary

Translate the original problem into toric geometry

proof: working dictionary

Translate the original problem into toric geometry

tree

proof: working dictionary

Translate the original problem into toric geometry

tree	variety

proof: working dictionary

Translate the original problem into toric geometry

tree | polytope | variety

proof: working dictionary

Translate the original problem into toric geometry

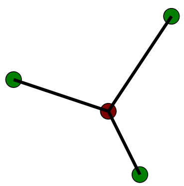
tree polytope variety

understand the basic objects

proof: working dictionary

Translate the original problem into toric geometry

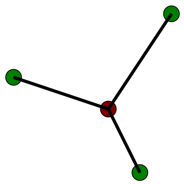
tree polytope variety



proof: working dictionary

Translate the original problem into toric geometry

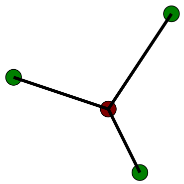
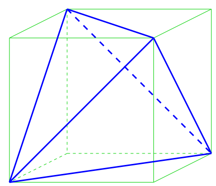
tree | polytope | variety



$\mathbb{P}^3$

proof: working dictionary

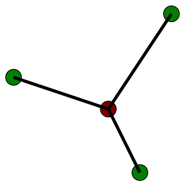
Translate the original problem into toric geometry

tree	polytope	variety
		$\mathbb{P}^3$

proof: working dictionary

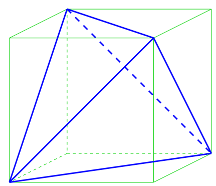
Translate the original problem into toric geometry

tree



a leaf

polytope



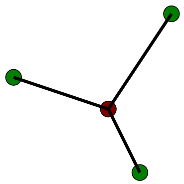
variety

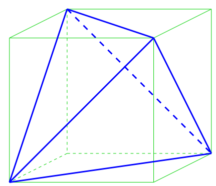
$\mathbb{P}^3$

proof: working dictionary

Translate the original problem into toric geometry

tree | polytope | variety



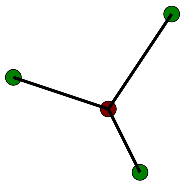
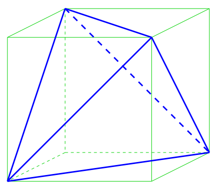


$\mathbb{P}^3$

a leaf | projection

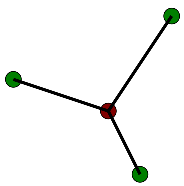
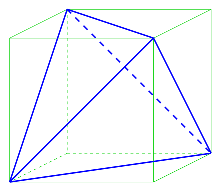
proof: working dictionary

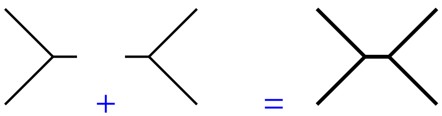
Translate the original problem into toric geometry

tree	polytope	variety
		$\mathbb{P}^3$
a leaf	projection	$\mathbb{C}^*$ action

proof: working dictionary

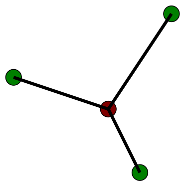
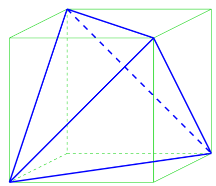
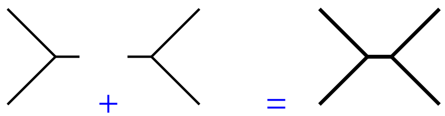
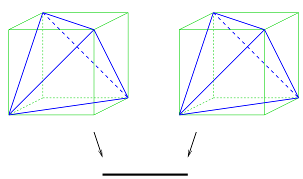
Translate the original problem into toric geometry

tree	polytope	variety
		$\mathbb{P}^3$
a leaf	projection	$\mathbb{C}^*$ action



proof: working dictionary

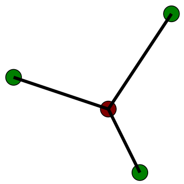
Translate the original problem into toric geometry

tree	polytope	variety
		$\mathbb{P}^3$
a leaf	projection	$\mathbb{C}^*$ action
		

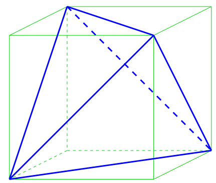
proof: working dictionary

Translate the original problem into toric geometry

tree



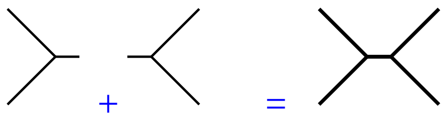
polytope



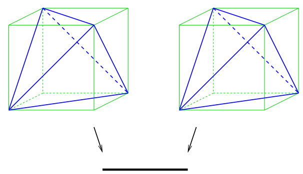
variety

$\mathbb{P}^3$

a leaf



projection

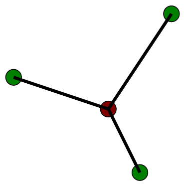
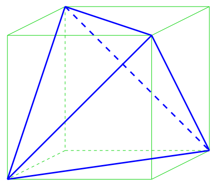
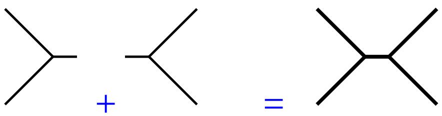
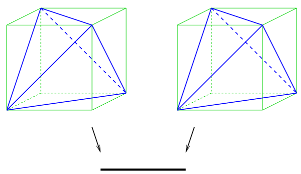


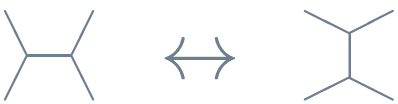
$\mathbb{C}^*$ action

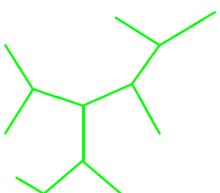
GIT quotient

proof: working dictionary

Translate the original problem into toric geometry

tree	polytope	variety
		$\mathbb{P}^3$
a leaf	projection	$\mathbb{C}^*$ action
		GIT quotient

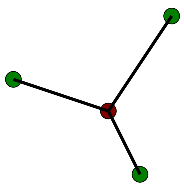




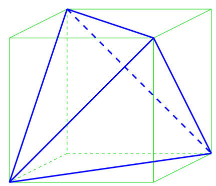
proof: working dictionary

Translate the original problem into toric geometry

tree



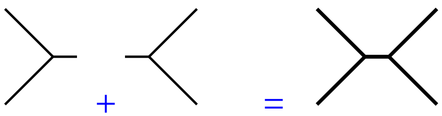
polytope

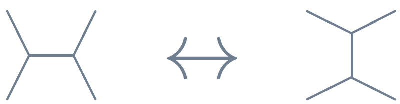


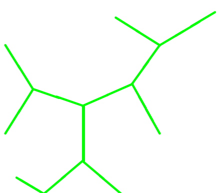
variety

$\mathbb{P}^3$

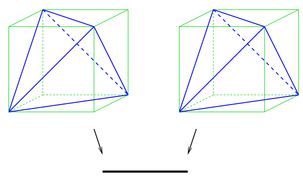
a leaf







projection



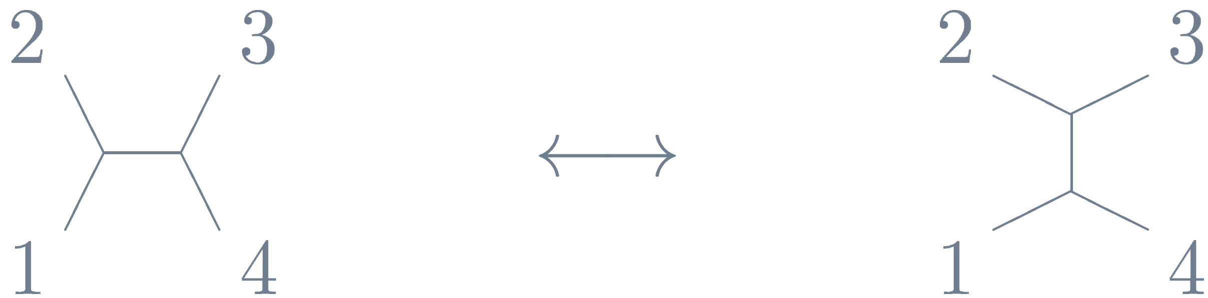
$\mathbb{C}^*$ action

GIT quotient

deformation

proof: the idea

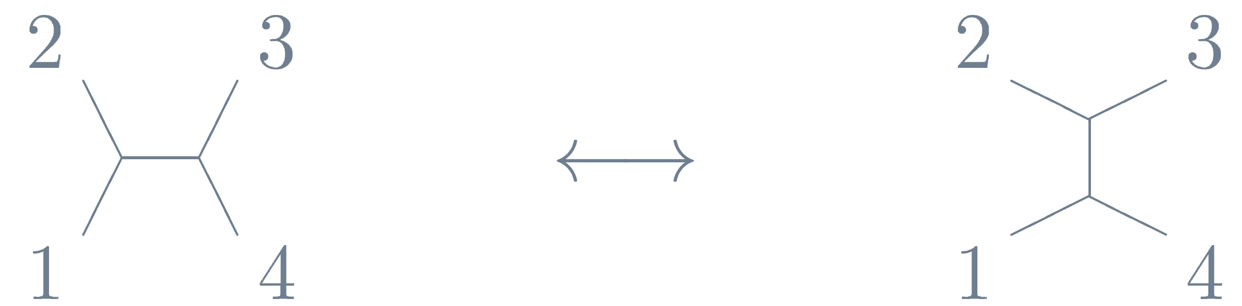
The mutation of a 4-leaf tree



can be explicitely written as deformation which preserves the action of $\mathbb{C}^*$ groups associated to leaves,

proof: the idea

The mutation of a 4-leaf tree



can be explicitely written as deformation which preserves the action of $\mathbb{C}^*$ groups associated to leaves, thus via GIT quotient it can be extended to a mutation of any tree along any inner edge