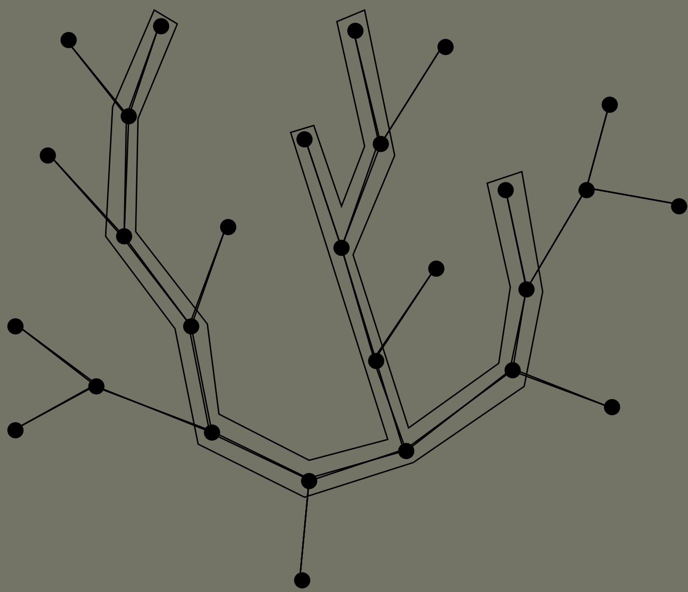
The space of phylogenetic trees and the tropical geometry of flag varieties



FPSAC, Nagoya, August 3, 2012

Christopher Manon
www.math.berkeley.edu/∼manonc
supported by NSF fellowship DMS-0902710

Reminder: some representation theory

G a reductive group over $\mathbb{C}$ (e.g. $SL_n(\mathbb{C})$, $SP_{2n}(\mathbb{C})$)

Finite dimensional irreducible representations are indexed by the lattice points λ in a cone Δ_G.

Example: $\Delta_{SL_n(\mathbb{C})} = \{(a_1, \ldots, a_{n-1}) | a_i \geq a_j \geq 0,\ i < j\}$

$\omega_m = (1, \ldots, 1, 0, \ldots, 0)$

$V(\omega_m) = \bigwedge^m(\mathbb{C}^n)$

Reminder: flag varieties

A flag variety G/P is a complete quotient of G by a parabolic subgroup $P \subset G$.

Any flag variety is the orbit through the highest weight vector in $\mathbb{P}(V(\lambda))$ for some $\lambda \in \Delta$.

$$G/P \cong G \circ [v_\lambda] \in \mathbb{P}(V(\lambda))$$

In particular G/P is projective, cut out by the homogeneous ideal I_λ, with projective coordinate ring

$$R_\lambda = \bigoplus_{N \geq 0} H^0(G/P, L_\lambda^{\otimes N}) = \bigoplus_{N \geq 0} V(N\lambda^*)$$

-Borel-Bott-Weil Theorem.

Example: $SL_n(\mathbb{C})$

Let $P_{m,n}$ be the parabolic subgroup of $SL_n(\mathbb{C})$ of the form

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

where A is $m \times m$, and C is $n - m \times n - m$.

$$SL_n(\mathbb{C})/P_{m,n} \cong Gr_m(\mathbb{C}^n) = SL_n(\mathbb{C}) \circ [z_1 \wedge \ldots \wedge z_m] \subset \mathbb{P}(\bigwedge^m(\mathbb{C}^n))$$

$$R_{\omega_m} = \bigoplus_{N \geq 0} V(N\omega_m^*) = \mathbb{C}[\ldots z_{i_1 \ldots i_m} \ldots]/I_{m,n}$$

This algebra is known as the Plücker algebra, and $I_{m,n}$ is the Plücker ideal.

Reminder: tropical varieties

$\mathbb{T} = \mathbb{R} \cup \{-\infty\}$

$a \oplus b = max\{a, b\}$
$a \otimes b = a + b.$

$X = \{x_1, \ldots, x_n\}, f(X) = \sum C_{\vec{m}} \vec{x}^{\vec{m}} \in \mathbb{C}[X]$

$T(f) = \bigoplus_{C_{\vec{m}} \neq 0} (\bigotimes_{i=1}^{n} m_i x_i) = max\{\ldots, \sum_{i=1}^{n} m_i x_i, \ldots\}$

$tr(f) = \{\vec{p} \in \mathbb{T}^n | T(f)$ has two maxima at $\vec{p}\}$

$I \subset \mathbb{C}[X],\ tr(I) = \bigcap_{f \in I} tr(f)$

Reminder: tropical varieties

There is always a finite set $M \subset I$ such that $tr(I) = \bigcap_{f \in M} tr(f)$.

-Bogart, Jensen, Speyer, Sturmfels, Thomas

weighted (phylogenetic) trees

Tree T with n leaves labeled $\{1, \ldots, n\}$
$P_T = \{w : Edge(T) \to \mathbb{R} | \geq 0$ on internal edges$\}$

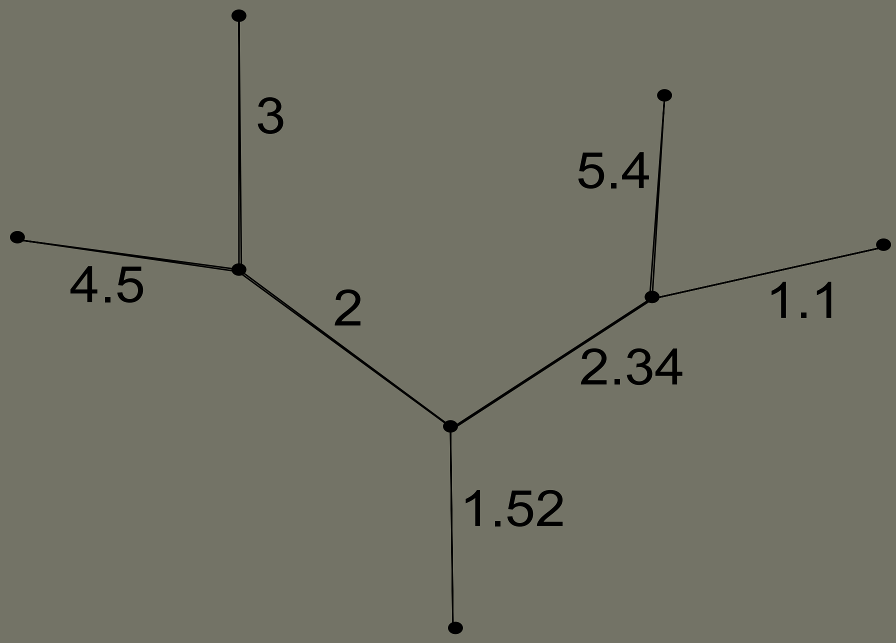


Figure 1:

$\cong \mathbb{R}_{\geq 0}^{|Edge(T)|-|Leaf(T)|} \times \mathbb{R}^{|Leaf(T)|}$

The space of trees $\mathcal{T}^n$

For $\psi : T' \to T$ a map of n-trees there is a map $\psi^* : P_T \to P_{T'}$.

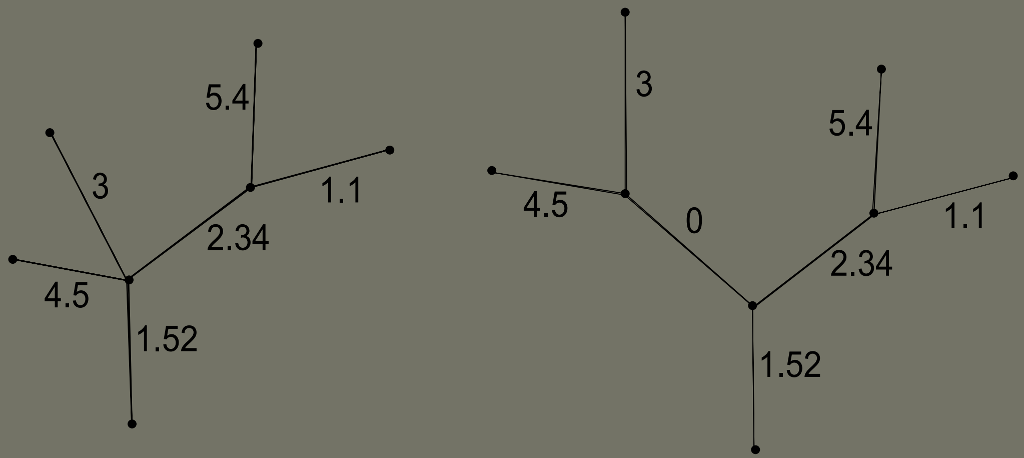


Figure 2:

We define $\mathcal{T}^n = \coprod_{|Leaf(T)|=n} P_T / \sim$

This space was studied by Billera, Holmes and Vogtman.

Dissimilarity vectors

For $1 < m < n$, and $\{i_1, \ldots, i_m\} \subset \{1, \ldots, n\}$ define

$$d_{i_1,\ldots,i_m}(T, w) = \sum_{e \in C(i_1,\ldots,i_m)} w(e).$$

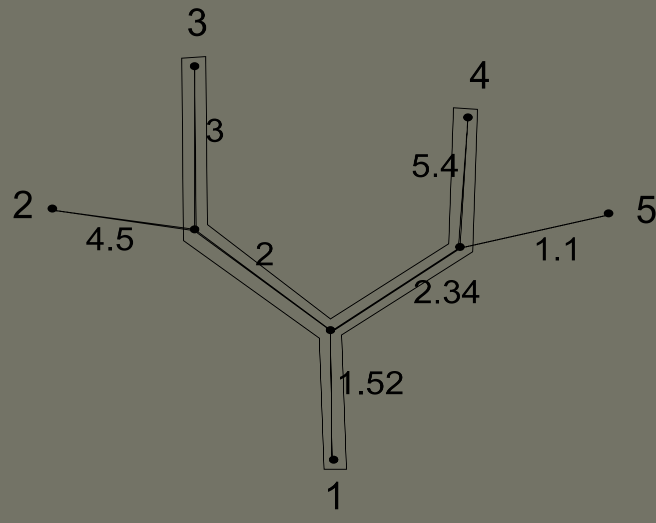


Figure 3: $d_{134}(T, w) = 14.26$

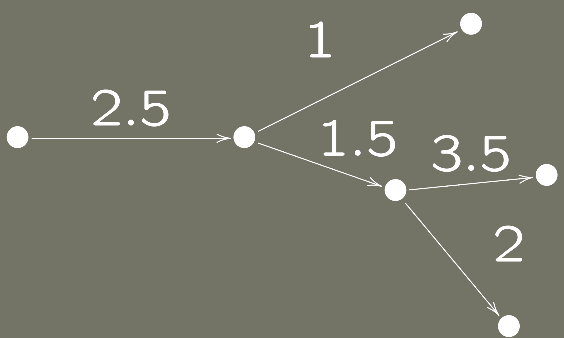
Dissimilarity vectors

Definition:

-We call $d^m(T,w) = (\ldots d_{i_1,\ldots,i_m}(T,w)\ldots)$ the m-dissimilarity vector of (T,w).

-We call $d^m : \mathcal{T}^n \to \mathbb{R}^{\binom{n}{m}}$ the m-dissimilarity map.

$d^3($



$) = (8.5, 7, 9.5, 8)$

2-dissimilarity vectors

The map $d^2 : \mathcal{T}^n \to \mathbb{R}^{\binom{n}{2}}$ is $1-1$.

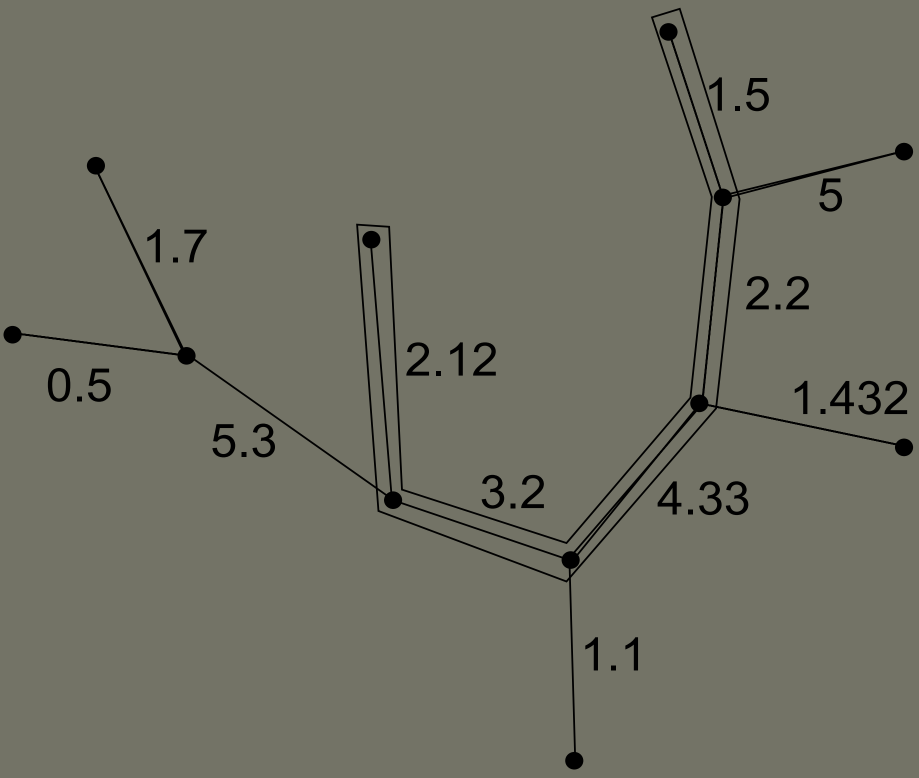


Figure 4: $d_{34} = 13.35$

d^2 and the tropical Grassmannian

$I_{2,n} =< \{z_{ij}z_{kl} - z_{ik}z_{jl} + z_{il}z_{jk} | 1 \leq i < j < k < l \leq n\} >$
$\subset \mathbb{C}[\ldots z_{ij} \ldots],$

$\mathbb{C}[\ldots z_{ij} \ldots]/I_{2,n} = R_{\omega_2}$

Theorem [Speyer, Sturmfels]:
The image of d^2 coincides with $tr(I_{2,n})$.

Plücker relations form a tropical basis of $I_{2,n}$,
so the $d^2(\mathcal{T}) \in \mathbb{R}^{\binom{n}{2}}$ satisfy

$$max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\}.$$

Conjecture of Cools, Pachter, Speyer

Conjecture [Cools, Pachter, Speyer]:

Theorem [Iriarte-Giraldo, M]:

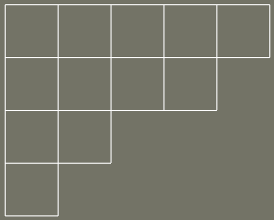
$d^m(\mathcal{T}^n) \subset tr(I_{m,n}) \subset \mathbb{R}^{\binom{n}{m}}$

This implies $d^m(T, w)$ satisfies $T(f)$ for all $f \in I_{m,n}$.

Plücker relations no longer make a tropical basis when $m > 2$.

back to representation theory

An irreducible representation of $GL_n(\mathbb{C})$ is given by a Young diagram.



The representation $V(\lambda)$ has a basis given by semi-standard fillings τ of λ by the indices $\{1, \ldots, n\}$.

1	1	2	4	5
3	3	3	6	
4	5			
5				

Example: $V(\omega_m) = \bigwedge^m(\mathbb{C}^n)$

$\bigwedge^m(\mathbb{C}^n)$ has a basis of elements $z_{i_1} \wedge \ldots \wedge z_{i_m}$, with $i_1 < \ldots < i_m$. This determines a semi-standard filling of a column of m boxes.

1
3
4
5
6

$$z_1 \wedge z_3 \wedge z_4 \wedge z_5 \wedge z_6$$

back to dissimilarity functions

We have seen that $z_{i_1} \wedge \ldots \wedge z_{i_m}$ tropicalizes to the dissimilarity function $d_{i_1,\ldots,i_m} : \mathcal{T}^n \to \mathbb{R}$. What about other semi-standard tableaux?

Definition:
We define $d_\tau : \mathcal{T}^n \to \mathbb{R}$ to be $\sum d_{I_k}$, where I_k are the columns of τ.

This construction gives a function on $\mathcal{T}^n$ for any basis member of a representation of $GL_n(\mathbb{C})$.

Yes.

Theorem [M]:

Let I_λ be the ideal which cuts the flag variety $GL_n(\mathbb{C})/P$ out of $\mathbb{P}(V(\lambda))$.
There is a map of complexes $d^\lambda : \mathcal{T}^n \to tr(I_\lambda)$ where $d^\lambda = (d_{\tau_1}, \ldots, d_{\tau_t})$

Tropical Theory: valuations

Let A be an algebra over $\mathbb{C}$.
A valuation $v : A \to \mathbb{T}$ is a function which satisfies the following.

$$v(ab) = v(a) + v(b) = v(a) \otimes v(b)$$

$$v(a + b) \leq max\{v(a), v(b)\} = v(a) \oplus v(b)$$

$$v(C) = 0 \text{ for } 0 \neq C \in \mathbb{C}$$

$$v(0) = -\infty.$$

Tropical theory: lifting

For A an algebra over $\mathbb{C}$, let $\mathbb{V}_{\mathbb{T}}(A)$ be the set of valuations of A into $\mathbb{T}$ over $\mathbb{C}$.

Theorem [Payne]:
For any presentation

$$0 \longrightarrow I \longrightarrow \mathbb{C}[X] \longrightarrow A \longrightarrow 0$$

there is a surjective map $\pi_X : \mathbb{V}_{\mathbb{T}}(A) \to tr(I)$ given by $\pi_X(v) = (\ldots v(x_i) \ldots)$.

Instead of the ideal I_λ, we consider the algebra R_λ

Valuations from chains of groups

Theorem[M]:
For any commutative $G-$algebra A, and every chain of subgroups

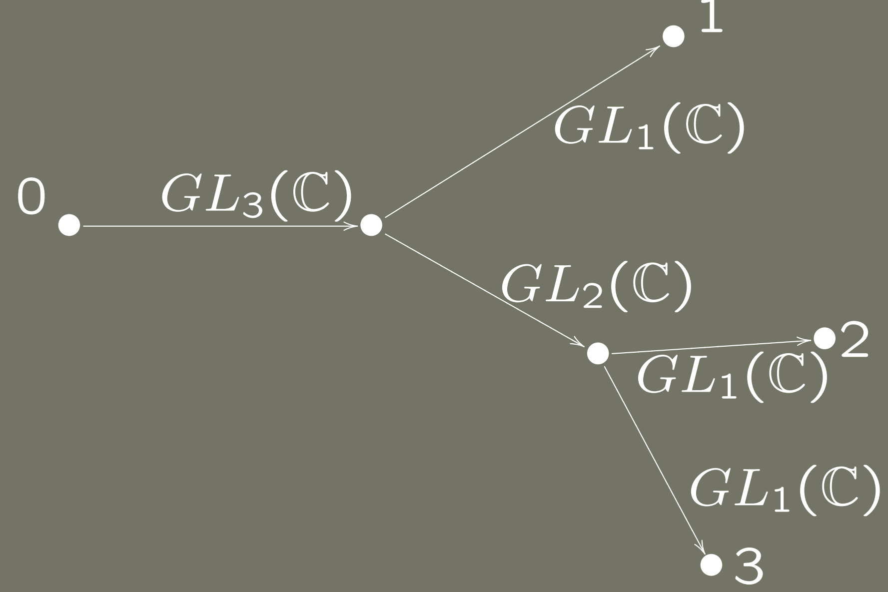
$$G_1 \xrightarrow{\phi_1} \dots \xrightarrow{\phi_{k-1}} G_{k-1} \xrightarrow{\phi_k} G,$$

there is a cone of valuations $D_{\vec{\phi}}$ in $\mathbb{V}_{\mathbb{T}}(A)$.

These cones glue together into a polyhedral complex $\mathcal{D}(G)$ of valuations on A.

Strategy: Find a way to turn trees into chains of subgroups of $GL_n(\mathbb{C})$.

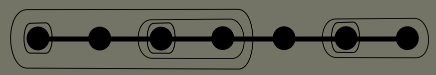
Trees and subgroups of $GL_n(\mathbb{C})$

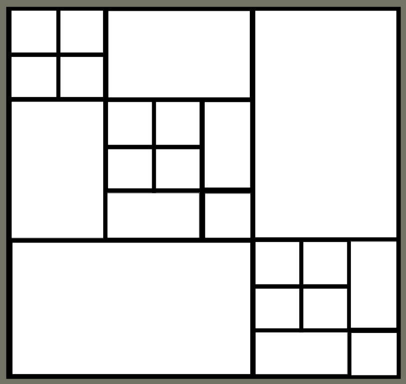


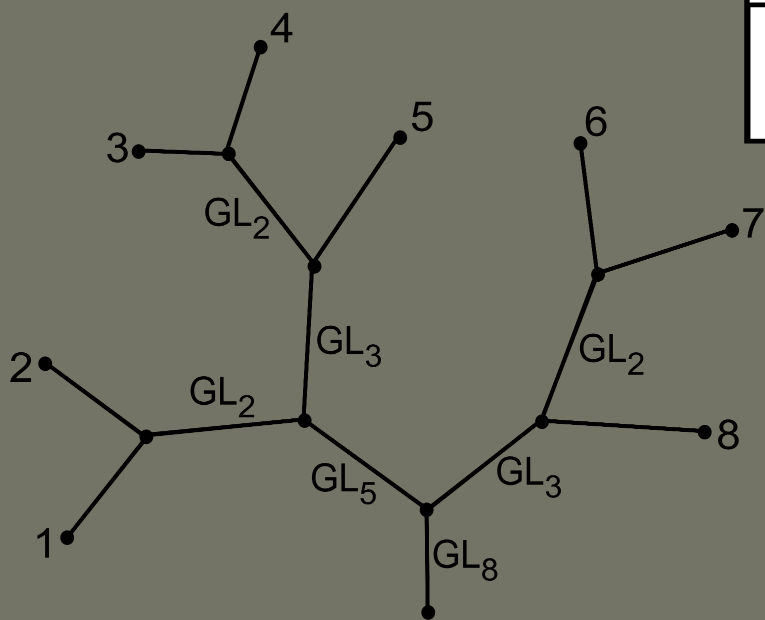
$$1 \to GL_1(\mathbb{C}) \times GL_1(\mathbb{C}) \times GL_1(\mathbb{C}) \to GL_1(\mathbb{C}) \times GL_2(\mathbb{C}) \to GL_3(\mathbb{C})$$

Rule: to an edge $e \in E(T)$ assign the group $GL_k(\mathbb{C})$ where k is the number of leaves "above" e.

Trees and subgroups of $GL_n(\mathbb{C})$







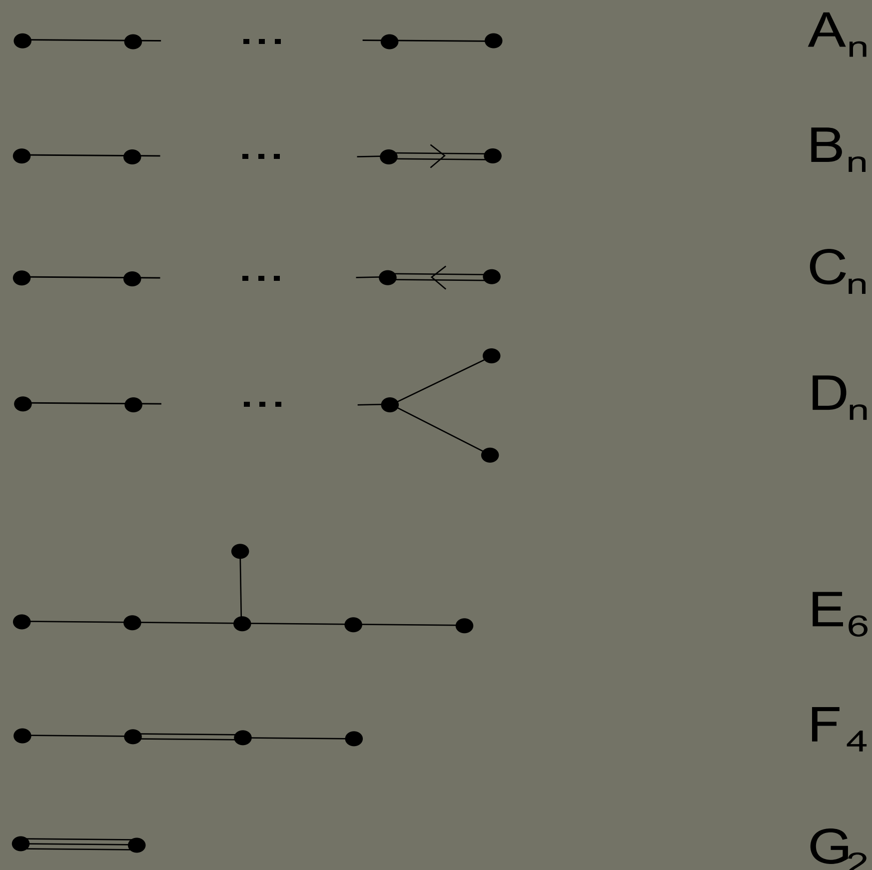
More general groups

For any Dynkin Diagram Γ, let $G(\Gamma)$ be the associated simply connected, semi-simple group over $\mathbb{C}$.

The diagram Γ also gives a hyperplane arrangement, we let B_Γ denote the Bergman fan of the associated matroid.

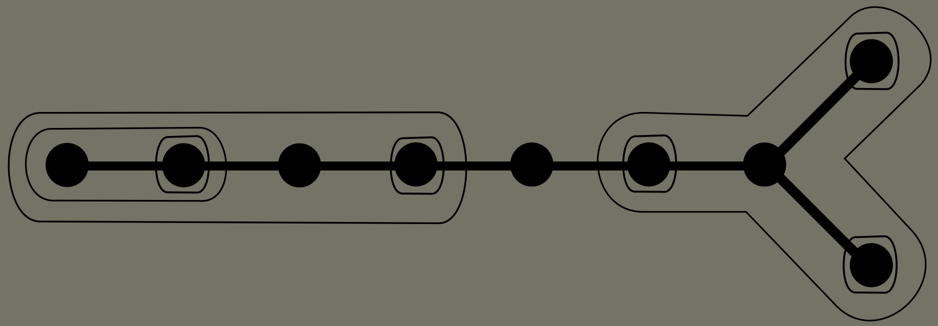
This space was studied by Ardilla, Klivans and Williams.

Dynkin Diagrams



Bergman Fan B_Γ

Faces of B_Γ are indexed by "tubings" of Γ. This is related to the graph associahedron of the Dynkin diagram.



Tubings also correspond to chains of Levi subgroups in $G(\Gamma)$.

Other groups

For Γ a Dynkin diagram, B_Γ the Bergman Fan of its associated hyperplane arrangement, and $tr(I_\lambda)$ the tropical variety associated to an ideal I_λ which cuts out a flag variety $G(\Gamma)/P$, we have the following:

Theorem [M]:
There is a map of complexes

$$\pi_\lambda : B_\Gamma \to tr(I_\lambda)$$

Other groups

Just as semi-standard tableaux give functions on $\mathcal{T}^n$, basis members (ie canonical basis, standard monomials, etc) of $V(\lambda)$ give functions on B_Γ.

$GL_n(\mathbb{C})$	$G(\Gamma)$
$\mathcal{T}^n$	B_Γ
$GL_{k_1}(\mathbb{C}) \times \ldots \times GL_{k_m}(\mathbb{C}) \subset GL_n(\mathbb{C})$	$L \subset G(\Gamma)$
semi-standard tableaux	standard monomials

Other Generalizations: Buildings

The complex $\mathcal{D}(G)$ has a $G-$action. An element $g \in G$ takes a cone $D_\phi \subset \mathcal{D}(G)$ to $D_{g\circ[\phi]}$

$$1 \longrightarrow H \xrightarrow{\phi} G \xrightarrow{Ad_g} G$$

Applying this to the cone defined by $1 \subset T \subset G$ for a maximal torus T yields a copy of (a cone over) the spherical building of G inside $\mathcal{D}(G)$. For A a $G-$algebra we have,

$$\mathcal{B}(\mathbb{C}, G) \subset \mathcal{D}(G) \to \mathbb{V}_{\mathbb{C}}(A)$$

Results like this have been obtained by Berkovich, and Remy, Thuillier, Werner in the context of flag varieties.

Other Generalizations: Buildings

In fact, the product complex $\mathcal{B}(\mathbb{C}, G(\Gamma)) \times B(\Gamma)$ is a subcomplex of $\mathcal{D}(G(\Gamma))$.

References

L.J. Billera, S. Holmes and K. Vogtmann: Geometry of the space of phylogenetic trees, Advances in Applied Mathematics 27 (2001) 733767.

F. Cools: On the relation between weighted trees and the tropical Grassmannian, Journal of Symbolic Compututation, Vol 44, Issue 8, (2009), 1079 -1086

B. Giraldo: Dissimilarity vectors of trees are contained in the tropical Grassmannian, The Electronic Journal of Combinatorics 17, no. 1, 2010

C. Manon: Dissimilarity maps on trees and the representation theory of $SL_m(\mathbb{C})$, to appear J. Alg. Combinatorics.

L. Pachter, B. Sturmfels: Algebraic statistics for computational biology, Cambridge University Press, New York 2005

D. Speyer, B.Sturmfels: The Tropical Grassmannian, Adv. Geom. 4. no. 3, (2004), 389-411.

Thankyou!