

# On the combinatorial structure of intervals in groups

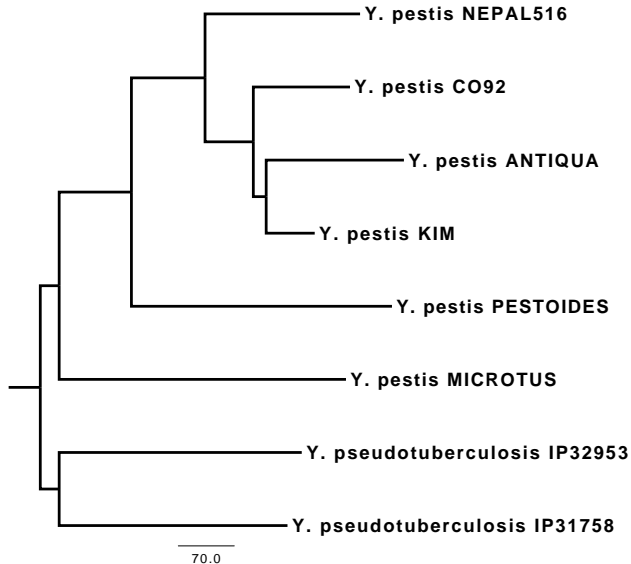
Attila Egri-Nagy, joint work with Andrew Francis and Volker Gebhardt

Centre for Mathematics Research,  
School of Computing, Engineering and Mathematics  
University of Western Sydney

General Algebra and Its Applications 2013

# Motivation

Reconstructing phylogeny trees.



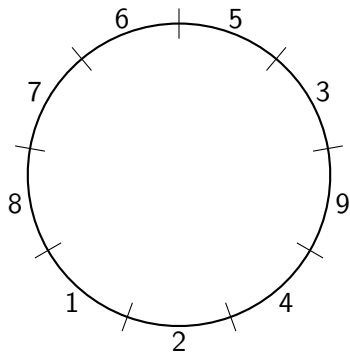
# Motivation

Calculating the evolutionary distance between genomes.

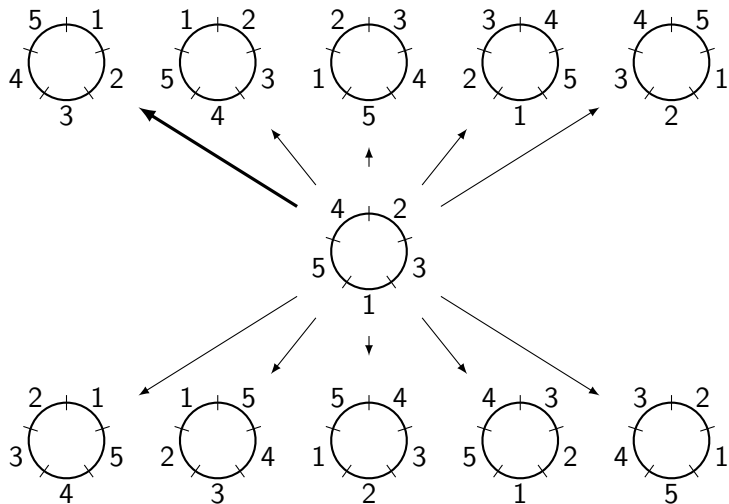
	KIM	ANTIQUA	MICROTUS	CO92	NEPAL516	PESTOIDES	Yp_IP31758	Yp_IP32953
KIM	0	233	738	188	334	515	758	738
ANTIQUA	233	0	750	319	449	664	719	712
MICROTUS	738	750	0	745	659	809	695	706
CO92	188	319	745	0	366	595	697	760
NEPAL516	334	449	659	366	0	659	641	759
PESTOIDES	515	664	809	595	659	0	753	695
Yp_IP31758	758	719	695	697	641	753	0	589
Yp_IP32953	738	712	706	760	759	695	589	0

# Biology $\rightarrow$ Math

Genome  $\rightarrow$  permutations



# The dihedral action



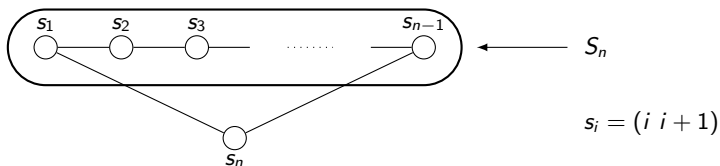
# A recurring problem

Even if the groups are well-studied, moving to a non-standard biological generating set makes the problem a lot harder.

Brute force calculations have their limit at 9 points. The number of regions in a biological study is around 80.

## Affine symmetric group for circular permutations

$\tilde{S}_n$  has generators  $s_1, s_2, \dots, s_n$ , and relations as implied by the Dynkin diagram:



E-N, Volker Gebhardt, Mark M. Tanaka, Andrew Francis:

**Group-theoretic models of the inversion process in bacterial genomes**, *Journal of Mathematical Biology*, paper online June 2013.

## New direction

Is the distance a good enough measure?

Can we use the number of shortest evolutionary paths?

Maybe the 'shape' how these paths are put together...



## Definitions

Let  $G$  be a group with generators  $S = \{s_1, \dots, s_n\}$ . We say  $S$  is a set of *semigroup generators* of the group  $G$  if  $\langle S \rangle = G$ , and it is a set of *group generators* if  $\langle S \cup S^{-1} \rangle = G$ .

$S^*$  is the free monoid generated by  $S$ , i.e. the set of all finite sequences, *words* of the elements of  $S$ . The group element realized by the word  $w$  is denoted by  $\bar{w}$ , thus  $w \in S^*$  and  $\bar{w} \in G$ .

The *geodesic distance* defined by  $\ell_S(g_1, g_2) = |u|$ , where  $u$  is a minimal length word in  $S^*$  with the property that  $g_1 \bar{u} = g_2$  also denoted by  $g_1 \xrightarrow{u} g_2$ , and  $u$  is called a *geodesic word*.  $\text{Geo}_S(g_1, g_2)$  is the set of all geodesic words from  $g_1$  to  $g_2$ . If no confusion arises, then we will use  $\ell(g_1, g_2)$  and  $\text{Geo}(g_1, g_2)$

Let  $w$  be a geodesic word in  $\text{Geo}(1, g_1^{-1})$ . Then applying  $w$  on the right maps the whole path  $g_1 \xrightarrow{u} g_2$  bijectively to  $1 \xrightarrow{uw} g_2 g_1^{-1}$ . Thus, for the distances of translated geodesics we can simply write  $\ell(g)$  instead of  $\ell(1, g)$ . Similarly, we use  $\text{Geo}(g)$  instead of  $\text{Geo}(1, g)$ .

The *Cayley graph*  $\Gamma(G, S)$  of  $G$  with respect to the generating set  $S$  is the directed graph with group elements as nodes and the labeled edges encoding the action of  $G$  on itself. Thus  $g \xrightarrow{s} gs$  is an edge. The *diameter* of  $G$  is defined by  $\text{diam}(G) = \max_{g \in G} \ell(g)$ .

## A partial order defined by the geodesics

For group elements  $g_1, g_2 \in G$  we write  $g_1 \leq g_2$  if

$$\ell(g_2) = \ell(g_1) + \ell(g_1^{-1}g_2).$$

Alternatively, we can give a more explicit definition using geodesic words.

For group elements  $g_1, g_2 \in G = \langle S \rangle$  we write  $g_1 \leq g_2$  if  $\exists w = uv \in S^*$  such that  $\bar{w} = g_2$ ,  $\bar{u} = g_1$ ,  $w \in \text{Geo}(g_2)$ , i.e. there is a geodesic from the identity to  $g_2$  and  $g_1$  is on it.

Also called the *prefix* order, or *weak* order for Coxeter groups.

Combinatorics of Coxeter Groups.



Andreas Björner and Francesco Brenti.

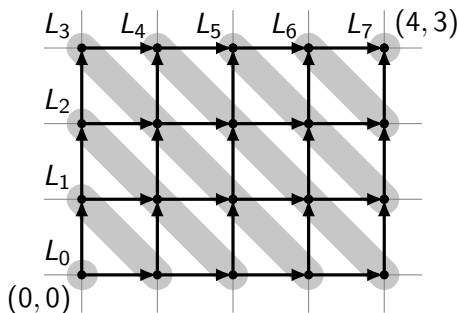
*Combinatorics of Coxeter Groups.*

Graduate Texts in Mathematics. Springer, 2005.

## Graded poset

With the partial order closed intervals are defined in the obvious way

$$[g_1, g_2] := \{g \in G \mid g_1 \leq g \leq g_2\}.$$



The graded structure of the interval  $[(0,0), (3,4)]$  in  $\mathbb{Z} \times \mathbb{Z}$ .

---

**Algorithm 1:** Constructing the graded interval  $[g, h]$ .

---

**input** :  $g, h \in G$ ,  $S$  generator set,  $d$  distance function

**output:**  $[g, h]$  interval,  $L_i$  layers,  $I_{g'} = \{(s, g's) \mid s \in S, g's \in [g, h]\}$

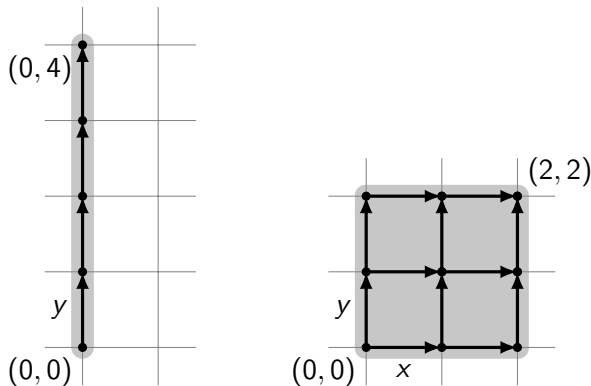
GradedInterval ( $g, h, S, d$ ):

```
1  $n \leftarrow d(g, h)$ ;  
2  $L_0 \leftarrow \{g\}$ ;  
3 foreach  $i \in \{1, \dots, n\}$  do  
4      $L_i \leftarrow \emptyset$ ;  
5     foreach  $g' \in L_{i-1}$  do  
6          $I_{g'} \leftarrow \emptyset$ ;  
7         foreach  $s \in S$  do  
8             if  $d(g's, h) = n - i$  then  
9                  $L_i \leftarrow L_i \cup g's$ ;  
0                  $I_{g'} \leftarrow I_{g'} \cup (s, g's)$ ;
```

---

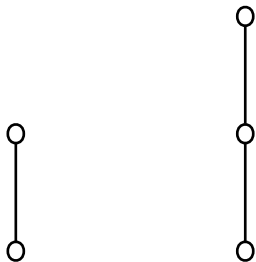
## Length and size

In general there is no connection.



In  $\mathbb{Z}^2$  two group elements with same length can have intervals of different size.  $|[(0,0), (0,4)]| = 5$ ,  $|[(0,0), (2,2)]| = 9$ .

$$S_3 = \langle (1, 2, 3), (1, 2) \rangle$$

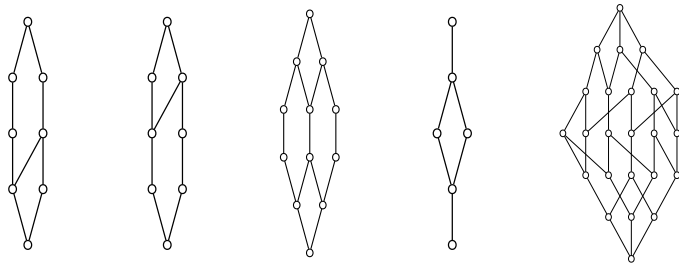
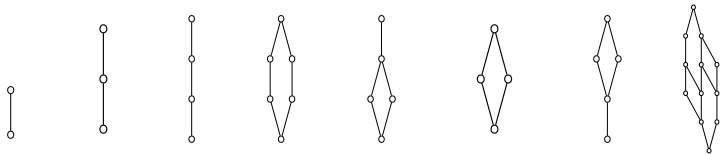


$$S_3 = \langle (1, 2), (2, 3) \rangle$$

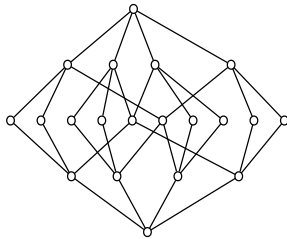
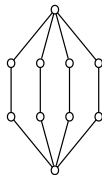
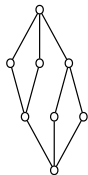
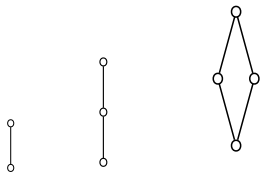




$$S_4 = \langle (1, 2), (2, 3), (3, 4) \rangle$$

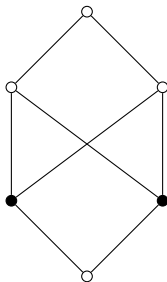


$$S_4 = \langle (1, 2), (2, 3), (3, 4), (1, 4) \rangle$$

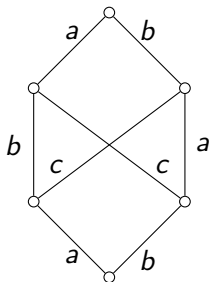


## Is it a lattice?

An obvious mathematical but biologically not so relevant question.  
A minimal counterexample would be:



## Trying with involutions



$$ab = bc = ca,$$

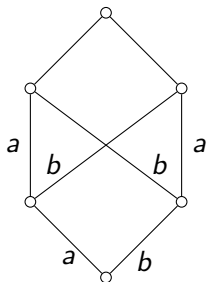
$$ac = ba = cb.$$

But since they are involutions,


$$ba = cb \implies c = bab$$

# Trying it with 2 generators

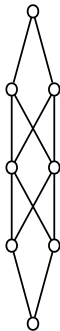
Minimal counterexamples



$$a^2 = b^2, ab = ba$$

For instance,  $a = (3, 4, 5)$ ,  $b = (1, 2)(3, 4, 5)$ . 

$$C_4 \times C_2 = \langle (3, 4, 5, 6), (1, 2)(3, 4, 5, 6) \rangle$$



$$[( ), (1, 2)]$$

# Changing generators is disruptive

The ultimate goal is to find equivalence classes of group elements.

$g \sim h$  if  $[1, g] \cong [1, h]$  or  $[1, g] \cong [1, h]^{\partial}$

Is it the case that adding generators yields a coarser partition of group elements?

NO



# Conclusion

We have to study individual cases as a general theory do not seem to be available/possible.



# Thank You!