# EXTENDING THE COMPUTATIONAL HORIZON

Attila Egri-Nagy

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

- Bacterial Genomics (with Andrew Francis and Volker Gebhardt)
  - How to reconstruct phylogeny trees?
- Foundations of Computing (with James East, James D. Mitchell, Chrystopher L. Nehaniv).
  - What is computable with $n$ states?
  - What is the structure of finite computations?

Precise answers can be obtained in *abstract algebra*, in computational group and semigroup theory.

# WHY COMPUTATIONAL?

- More examples, more raw data for the mathematical reasoning.

# BACTERIAL CHANGES

Single celled organisms with circular chromosome.

- *Local* changes such as *single nucleotide polymorphisms* (SNPs):

$$ACGGCCCTTAGG \quad \longrightarrow \quad ACGGCCATTAGG$$

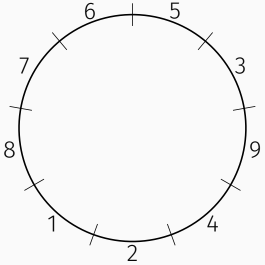
- *Regional* changes such as *inversion* that affect whole regions along the chromosome.

$$\overset{1}{\longrightarrow}\overset{2}{\longrightarrow}\overset{3}{\longrightarrow}\overset{4}{\longrightarrow}\overset{5}{\longrightarrow}\overset{6}{\longrightarrow} \quad \xrightarrow{\qquad} \quad \overset{1}{\longrightarrow}\overset{2}{\longrightarrow}\overset{5}{\longleftarrow}\overset{4}{\longleftarrow}\overset{3}{\longleftarrow}\overset{6}{\longrightarrow}$$

(Regional changes include inversion, translocation, deletion and others.)

- *Topological* changes that produce knots and links in the DNA.

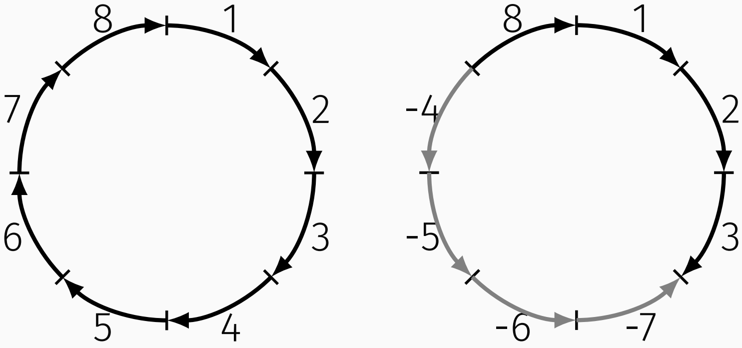
Genome→ permutations




Sequences of evolutionary → Sequences of generators

Genomic distance→ Length of geodesic words

Genomic space → Cayley-graph




Reference genome and the signed permutation
$[1, 2, 3, -7, -6, -5, -4, 8]$.

# WHERE DOES THE DIFFICULTY FROM?

- The groups are finite and well-studied (symmetric, hyperoctahedral), but big, e.g. $\mathcal{S}_{80}$
- The generating sets are unusual, "biological".

For example,

- 2-inversions of the circular genome (vs. linear)
- looking at the "width" as well

Strategy: Calculate and look at small, but non-trivial examples to get insights.

# SYMMETRIES OF IRREDUCIBLE GENERATING SETS

| | 1 | $\mathbb{Z}_2$ | $\mathbb{Z}_2 \times \mathbb{Z}_2$ | $\mathbb{Z}_3$ | $D_8$ | $\mathcal{S}_3$ | $\mathcal{S}_4$ | $\mathcal{S}_5$ |
|---|---|---|---|---|---|---|---|---|
| $\mathcal{S}_2$ | | 1 | | | | | | |
| $\mathcal{S}_3$ | 1 | 1 | | | | | | |
| $\mathcal{S}_4$ | 8 | 5 | | | | 1 | | |
| $\mathcal{S}_5$ | 150 | 25 | | 1 | | 1 | 1 | |
| $\mathcal{S}_6$ | 7931 | 645 | 11 | 6 | 4 | 20 | 2 | 2 |

**Definition**

A *semigroup* is a set $S$ with an associative binary operation $S \times S \to S$.

**Example (Flip-flop monoid)**

| | 1 | a | b |
|---|---|---|---|
| 1 | 1 | a | b |
| a | a | a | b |
| b | b | a | b |

**Definition**

A *semigroup* is a set $S$ with an associative binary operation $S \times S \to S$.

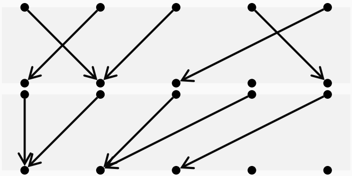
**Example (Flip-flop monoid)**

| | 1 | a | b |
|---|---|---|---|
| 1 | 1 | a | b |
| a | a | a | b |
| b | b | a | b |

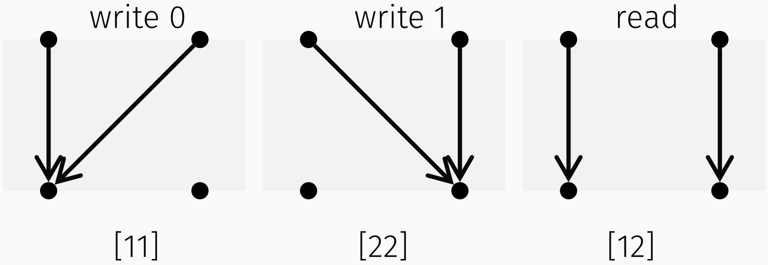
**Definition**

A *transformation semigroup* $(X, S)$ is a set of states $X$ and a set $S$ of transformations $s : X \to X$ closed under function composition.

**Example (Transformations)**



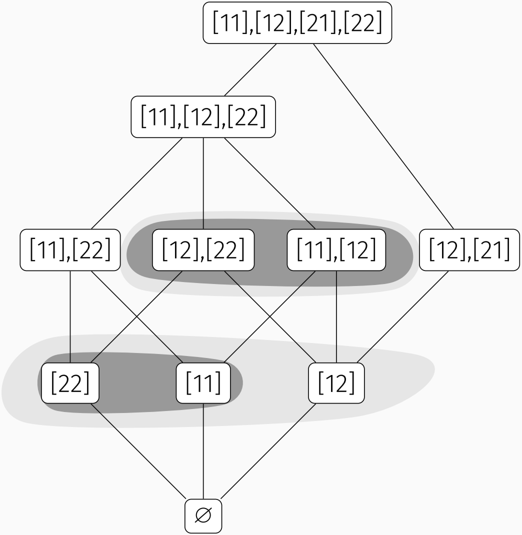
# FLIP-FLOP, THE 1-BIT MEMORY SEMIGROUP




So these are computational devices… $\approx$ automata

With transformation semigroups, we get all semigroups.
(Cayley's theorem)

# DEGREE 2 TRANSFORMATION SEMIGROUPS

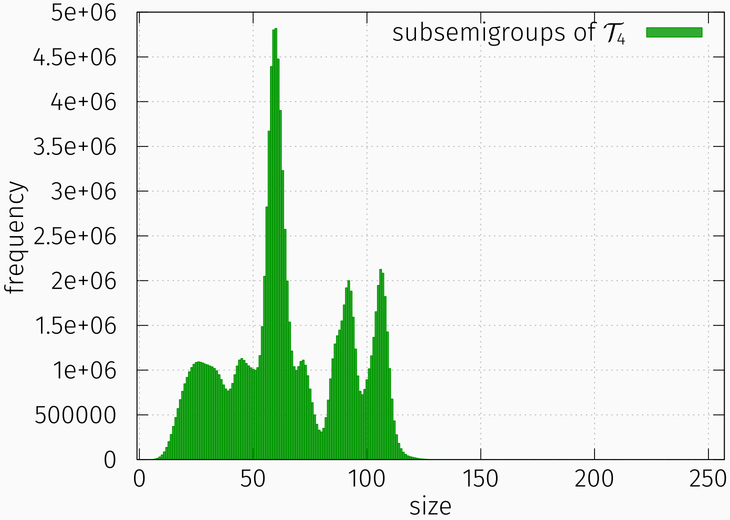


Number of subsemigroups of full transformation semigroups.

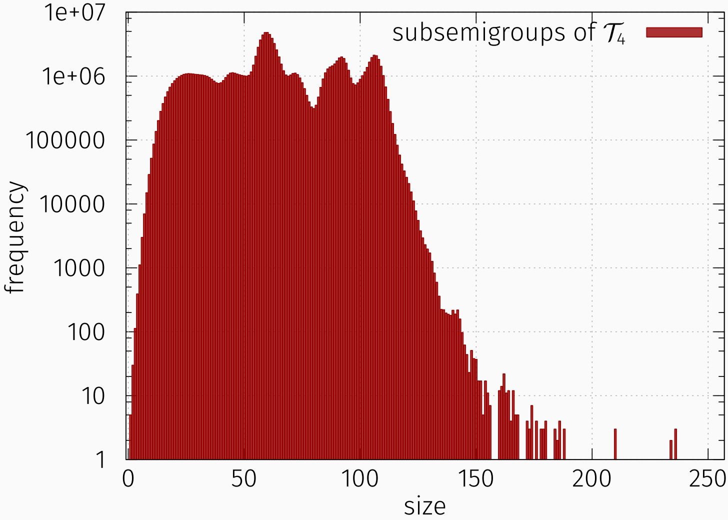
| | #subsemigroups | #conjugacy classes | #isomorphism classes |
|---|---|---|---|
| $\mathcal{T}_0$ | 1 | 1 | 1 |
| $\mathcal{T}_1$ | 2 | 2 | 2 |
| $\mathcal{T}_2$ | 10 | 8 | 7 |
| $\mathcal{T}_3$ | 1 299 | 283 | 267 |
| $\mathcal{T}_4$ | 3 161 965 550 | 132 069 776 | 131 852 491 |

After discounting the state-relabelling symmetries the database of degree 4 transformation semigroups is still around 9GB.

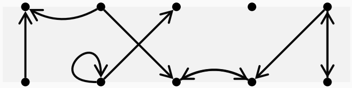
# SIZE DISTRIBUTION

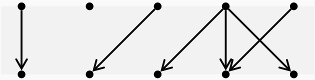


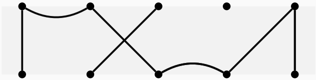
# SIZE DISTRIBUTION – LOGARITHMIC SCALE

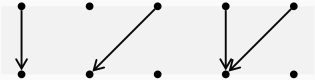


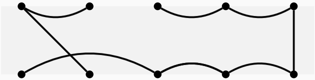
# DIAGRAM SEMIGROUPS – TYPICAL ELEMENTS

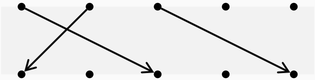

$\in \mathcal{P}\mathsf{B}_n$,


$\in \mathsf{B}_n$,

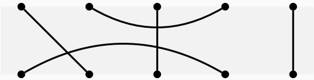

$\in \mathcal{P}_n$


$\in \mathsf{P}\mathcal{T}_n$,

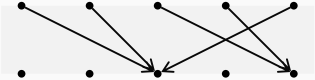

$\in \mathcal{I}_n^*$



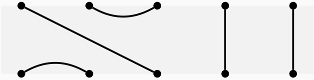
$\in \mathcal{I}_n$,



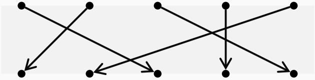
$\in \mathfrak{B}_n$



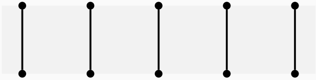
$\in \mathcal{T}_n$,



$\in \mathsf{TL}_n$



$\in \mathcal{S}_n$,



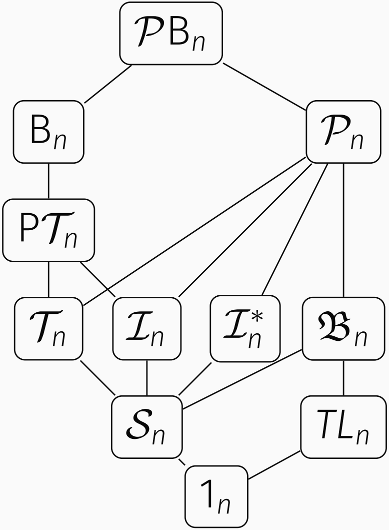
$1_n$

$\mathcal{P}_1 \hookrightarrow \mathcal{T}_2$
$\mathcal{P}_2 \hookrightarrow \mathcal{T}_5$

$\mathfrak{B}_1 \cong \mathcal{T}_1$
$\mathfrak{B}_2 \hookrightarrow \mathcal{T}_3$

$\mathrm{TL}_1 \cong \mathcal{T}_1$
$\mathrm{TL}_2 \hookrightarrow \mathcal{T}_2$
$\mathrm{TL}_3 \hookrightarrow \mathcal{T}_4$

$\mathcal{P}_1 \hookrightarrow \mathfrak{B}_2$



| | Order | $n=1$ | $n=2$ | $n=3$ | $n=4$ | $n=5$ | $n=6$ |
|---|---|---|---|---|---|---|---|
| $\mathcal{P}\mathrm{B}_n$ | $2^{(2n)^2}$ | 16 | 65536 | $2^{36}$ | $2^{64}$ | $2^{100}$ | $2^{144}$ |
| $\mathrm{B}_n$ | $2^{n^2}$ | 2 | 16 | 512 | 65536 | $2^{25}$ | $2^{36}$ |
| $\mathcal{P}_n$ | $B_{2n} = \sum_1^{2n} S(2n,k)$ | 2 | 15 | 203 | 4140 | 115975 | 4213597 |
| $\mathrm{P}\mathcal{T}_n$ | $(n+1)^n$ | 2 | 9 | 64 | 625 | 7776 | 117649 |
| $\mathcal{I}_n^*$ | $\sum_1^n k!\big(S(n,k)\big)^2$ | 1 | 3 | 25 | 339 | 6721 | 179643 |
| $\mathcal{T}_n$ | $n^n$ | 1 | 4 | 27 | 256 | 3125 | 46656 |
| $\mathcal{I}_n$ | $\sum_0^n k!\binom{n}{k}^2$ | 2 | 7 | 34 | 209 | 1546 | 13327 |
| $\mathfrak{B}_n$ | $(2n-1)!!$ | 1 | 3 | 15 | 105 | 945 | 10395 |
| $\mathcal{S}_n$ | $n!$ | 1 | 2 | 6 | 24 | 120 | 720 |
| $TL_n, J_n$ | $C_n = \frac{1}{n+1}\binom{2n}{n}$ | 1 | 2 | 5 | 14 | 42 | 132 |

# COMPUTATIONAL HORIZON

| | $n=1$ | $n=2$ | $n=3$ | $n=4$ | $n=5$ | $n=6$ |
|---|---|---|---|---|---|---|
| $\mathcal{P}\mathsf{B}_n$ | 1262 | | | | | |
| $\mathsf{B}_n$ | 4 | 385 | | | | |
| $\mathcal{P}_n$ | 4 | 272 | | | | |
| $\mathsf{P}\mathcal{T}_n$ | 4 | 50 | 94232 | | | |
| $\mathcal{I}_n$ | 4 | 23 | 2963 | | | |
| $\mathcal{I}_n^*$ | 2 | 6 | 795 | | | |
| $\mathcal{T}_n$ | 2 | 8 | 283 | 132069776 | | |
| $\mathfrak{B}_n$ | 2 | 6 | 42 | 10411 | | |
| $\mathsf{TL}_n$ | 2 | 4 | 12 | 232 | 12592 | 324835618 |
| $\mathcal{S}_n$ | 1 | 2 | 4 | 11 | 19 | 56 |

Given a semigroup $S$, the equivalence relation $\mathcal{J}$ is defined by

$$t \; \mathcal{J} \; s \iff S^1 t S^1 = S^1 s S^1,$$

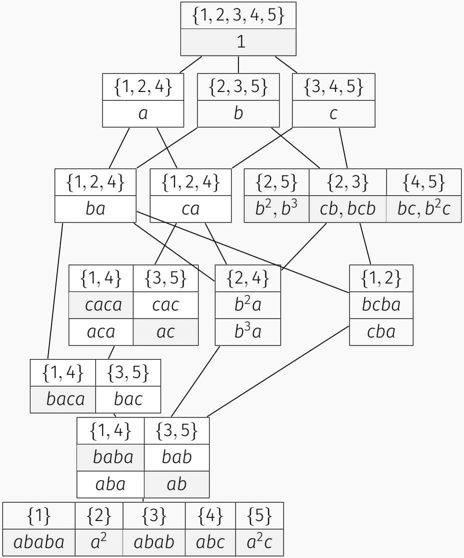
where $S^1$ is $S$ with an identity adjoined in case $S$ is not a monoid.

In other words,

$$t \; \mathcal{J} \; s \iff \exists p, q, u, v \in S^1 \text{ such that } t = psq \text{ and } s = utv$$

The equivalence classes of $\mathcal{J}$ are *"local pools of reversibility"*.

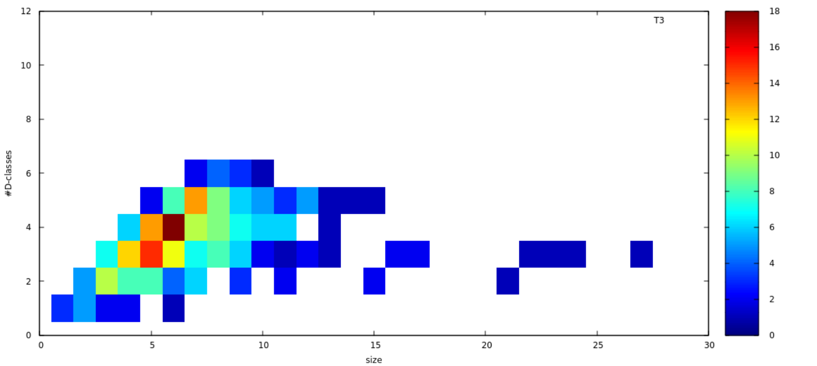
$a = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 2 & 1 & 2 & 4 \end{pmatrix}$, $b = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 3 & 2 \end{pmatrix}$, $b = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 4 & 5 & 4 \end{pmatrix}$ and $M = \langle a, b, c \rangle$. $|M| = 31$



# TRANSFORMATION SEMIGROUPS OF DEGREE 3

x axis : size of the semigroups

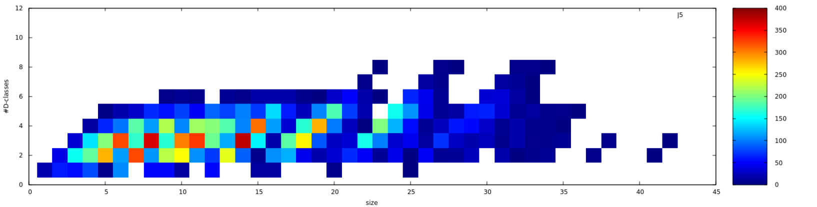
y axis : the number of $\mathcal{D}$-classes



# SUBSEMIGROUPS OF THE DEGREE 5 JONES MONOID

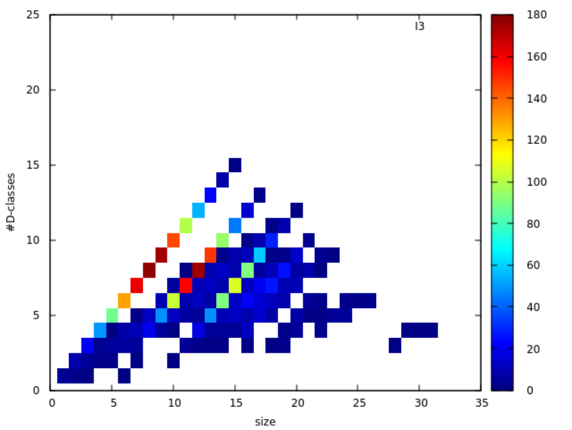
x axis : size of the semigroups

y axis : the number of $\mathcal{D}$-classes



x axis : size of the semigroups

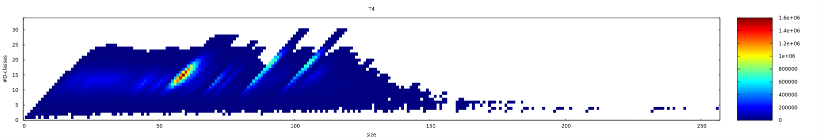
y axis : the number of $\mathcal{D}$-classes



# TRANSFORMATION SEMIGROUPS OF DEGREE 4

x axis : size of the semigroups

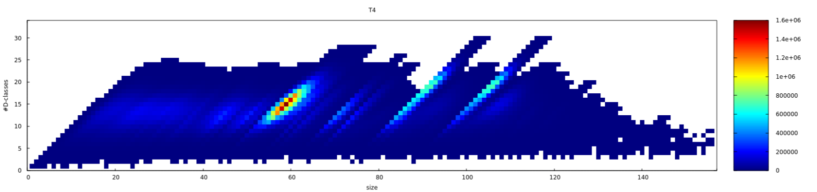
y axis : the number of $\mathcal{D}$-classes



## TRANSFORMATION SEMIGROUPS OF DEGREE 4

x axis : size of the semigroups

y axis : the number of $\mathcal{D}$-classes



# EVEN MORE STRUCTURE, ALL GREEN'S RELATIONS

$\mathcal{L}$, $\mathcal{R}$ equivalence relations

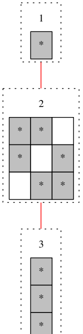
$$t \,\mathcal{R}\, s \iff tS^1 = sS^1,$$
$$t \,\mathcal{L}\, s \iff S^1t = S^1s,$$
$$t \,\mathcal{J}\, s \iff S^1tS^1 = S^1sS^1$$
$$\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L} = \mathcal{D}$$

$\mathcal{J} = \mathcal{D}$ in the finite case

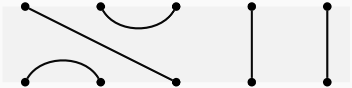
$$t \,\mathcal{R}\, s \iff \exists p, q \in S^1 \text{ such that } t = sp \text{ and } s = tq$$
$$t \,\mathcal{L}\, s \iff \exists p, q \in S^1 \text{ such that } t = ps \text{ and } s = qt$$

# “EGGBOX” PICTURE

Tables are $\mathcal{D}$-classes. Columns are $\mathcal{L}$-classes, rows are $\mathcal{R}$-classes. Shaded cells are $\mathcal{H}$-classes that contain idempotents – used for locating subgroups of the semigroup. $\mathcal{T}_3$:

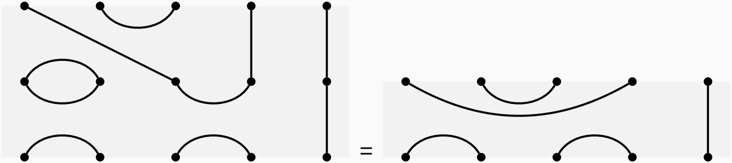


Catalan numbers, sequences of well-formed parentheses.



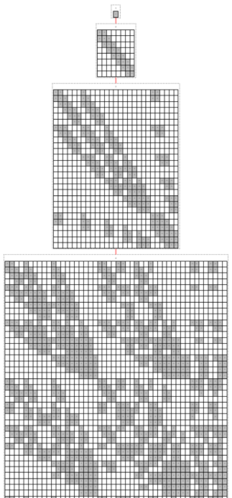
corresponds to (()(()))()

Applications in Physics: statistical mechanics, percolation problem.



# FIRST GLIMPSE THROUGH THE USUAL EGGBOX DIAGRAMS

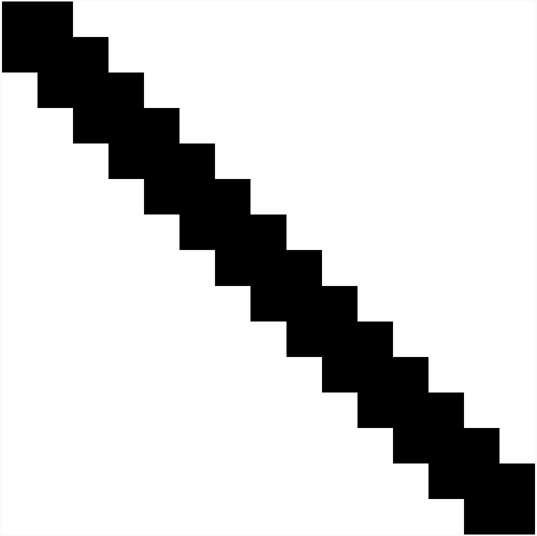
$J_9$

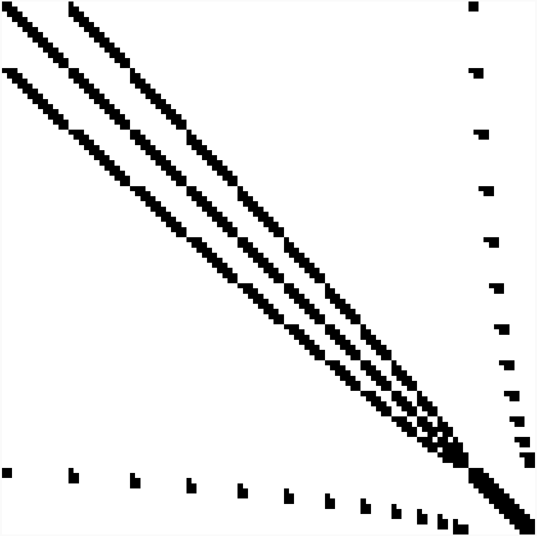


## TOP LEVEL

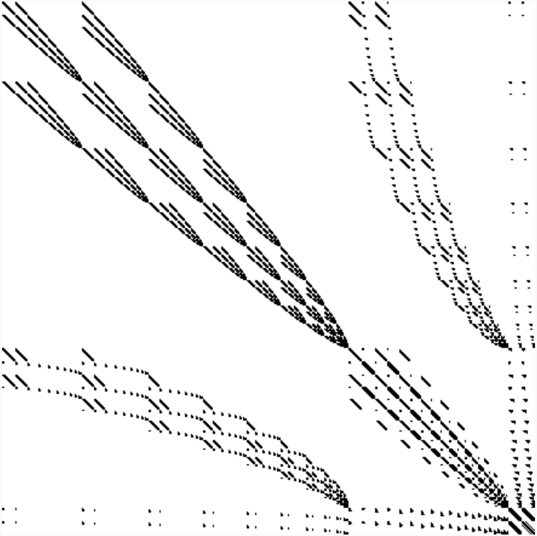
$J_{16}$

# LEVEL 2

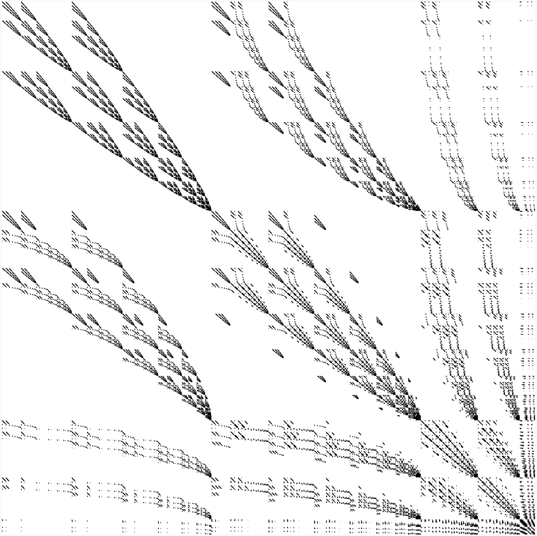




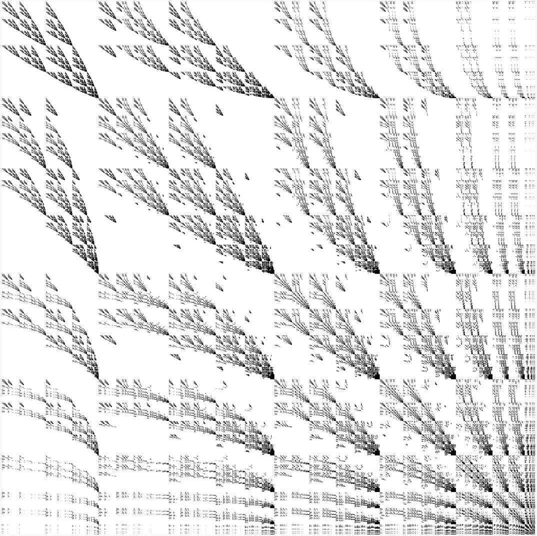
# LEVEL 4



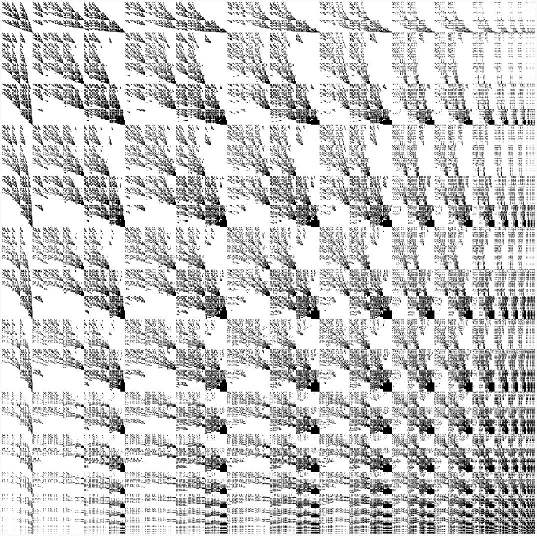
# LEVEL 5



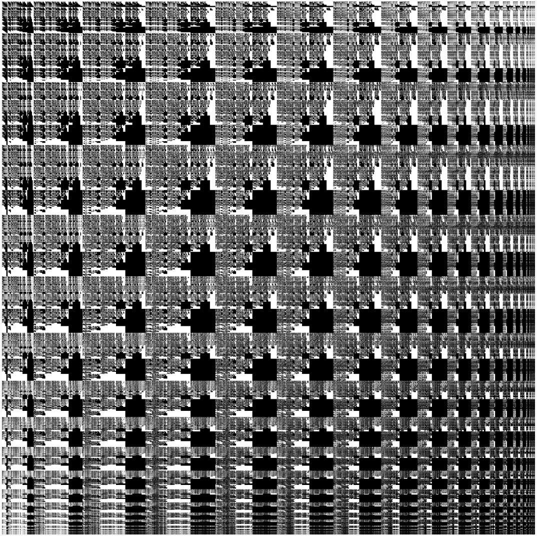
# LEVEL 6



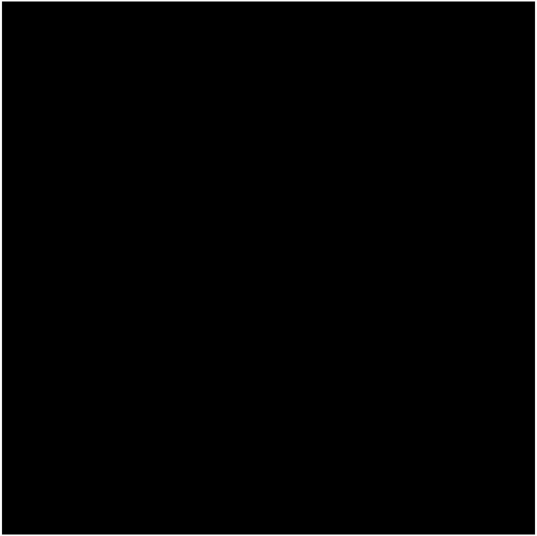
# LEVEL 7



# LEVEL 8



# LEVEL 9



**The Good** We can discover/construct more and more new, interesting and useful mathematics by using computers.

**The Bad** There is a gap between mathematical rigour and the correctness of software implementations and the physicality of computation.

**and The Ugly** Developing software is still detrimental to academic career.

Blog on computational semigroup theory:

`compsemi.wordpress.com`

Blog on computational semigroup theory:

`compsemi.wordpress.com`

# Thank You!