EXTENDING THE COMPUTATIONAL HORIZON

Attila Egri-Nagy

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

- **Bacterial Genomics** (with Andrew Francis and Volker Gebhardt)
  - How to reconstruct phylogeny trees?
  - 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.
· More examples, more raw data for the mathematical reasoning.
Single celled organisms with circular chromosome.

- **Local** changes such as *single nucleotide polymorphisms* (SNPs):
  
  \[ \text{ACGGCCCTTAGG} \rightarrow \text{ACGGCCATTAGG} \]

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

  \[
  \begin{array}{cccccc}
  1 & 2 & 3 & 4 & 5 & 6 \\
  \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \\
  \end{array}
  \quad \rightarrow \quad
  \begin{array}{cccccc}
  1 & 2 & 5 & 4 & 3 & 6 \\
  \rightarrow \leftrightarrow \leftrightarrow \leftrightarrow \leftrightarrow \\
  \end{array}
  \]

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

- **Topological** changes that produce knots and links in the DNA.
Genome $\rightarrow$ permutations

Sequences of evolutionary $\rightarrow$ Sequences of generators
Genomic distance $\rightarrow$ Length of geodesic words
Genomic space $\rightarrow$ 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. $S_{80}$
- 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

<table>
<thead>
<tr>
<th>Set</th>
<th>1</th>
<th>(\mathbb{Z}_2)</th>
<th>(\mathbb{Z}_2 \times \mathbb{Z}_2)</th>
<th>(\mathbb{Z}_3)</th>
<th>(D_8)</th>
<th>(S_3)</th>
<th>(S_4)</th>
<th>(S_5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(S_2)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(S_3)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(S_4)</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(S_5)</td>
<td>150</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(S_6)</td>
<td>7931</td>
<td>645</td>
<td>11</td>
<td>6</td>
<td>4</td>
<td>20</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Definition
A semigroup is a set $S$ with an associative binary operation $S \times S \to S$.

Example (Flip-flop monoid)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>
Definition

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

Example (Flip-flop monoid)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>b</td>
<td>b</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

Definition

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

Example (Transformations)
So these are computational devices... ≈ automata

With transformation semigroups, we get all semigroups. (Cayley’s theorem)
Degree 2 Transformation Semigroups
Number of subsemigroups of full transformation semigroups.

<table>
<thead>
<tr>
<th>$\mathcal{T}_0$</th>
<th>#subsemigroups</th>
<th>#conjugacy classes</th>
<th>#isomorphism classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{T}_1$</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>$\mathcal{T}_2$</td>
<td>10</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>$\mathcal{T}_3$</td>
<td>1 299</td>
<td>283</td>
<td>267</td>
</tr>
<tr>
<td>$\mathcal{T}_4$</td>
<td>3 161 965 550</td>
<td>132 069 776</td>
<td>131 852 491</td>
</tr>
</tbody>
</table>

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

subsemigroups of $\mathcal{T}_4$
SIZE DISTRIBUTION – LOGARITHMIC SCALE

The graph shows the frequency distribution of subsemigroups of $T_4$ on a logarithmic scale. The x-axis represents the size of the subsemigroups, ranging from 0 to 250, and the y-axis represents the frequency, ranging from $1 \times 10^0$ to $1 \times 10^7$. The distribution peaks in the mid-size range and decreases rapidly as the size increases.
\[ \in \mathcal{PB}_n, \]
\[ \in \mathcal{B}_n, \]
\[ \in \mathcal{PT}_n, \]
\[ \in \mathcal{P}_n \]
\[ \in \mathcal{I}_n^* \]
DIAGRAM SEMIGROUPS – TYPICAL ELEMENTS

$\in \mathcal{I}_n$, $\in \mathcal{B}_n$

$\in \mathcal{T}_n$, $\in \mathcal{TL}_n$

$\in \mathcal{S}_n$, $1_n$
$\mathcal{P}_1 \hookrightarrow \mathcal{T}_2$
$\mathcal{P}_2 \hookrightarrow \mathcal{T}_5$
$\mathfrak{B}_1 \cong \mathcal{T}_1$
$\mathfrak{B}_2 \hookrightarrow \mathcal{T}_3$
$\mathcal{T}_L_1 \cong \mathcal{T}_1$
$\mathcal{T}_L_2 \hookrightarrow \mathcal{T}_2$
$\mathcal{T}_L_3 \hookrightarrow \mathcal{T}_4$
$\mathcal{P}_1 \hookrightarrow \mathfrak{B}_2$
<table>
<thead>
<tr>
<th></th>
<th>Order</th>
<th>( n = 1 )</th>
<th>( n = 2 )</th>
<th>( n = 3 )</th>
<th>( n = 4 )</th>
<th>( n = 5 )</th>
<th>( n = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mathcal{P} B_n )</td>
<td>( 2^{(2n)^2} )</td>
<td>16</td>
<td>65536</td>
<td>2(^{36} )</td>
<td>2(^{64} )</td>
<td>2(^{100} )</td>
<td>2(^{144} )</td>
</tr>
<tr>
<td>( B_n )</td>
<td>( 2^n )</td>
<td>2</td>
<td>16</td>
<td>512</td>
<td>65536</td>
<td>2(^{25} )</td>
<td>2(^{36} )</td>
</tr>
<tr>
<td>( \mathcal{P}_n )</td>
<td>( B_{2n} = \sum_{1}^{2n} S(2n, k) )</td>
<td>2</td>
<td>15</td>
<td>203</td>
<td>4140</td>
<td>115975</td>
<td>4213597</td>
</tr>
<tr>
<td>( \mathcal{P} T_n )</td>
<td>( (n + 1)^n )</td>
<td>2</td>
<td>9</td>
<td>64</td>
<td>625</td>
<td>7776</td>
<td>117649</td>
</tr>
<tr>
<td>( \mathcal{I}_n^* )</td>
<td>( \sum_{1}^{n} k! (S(n, k))^2 )</td>
<td>1</td>
<td>3</td>
<td>25</td>
<td>339</td>
<td>6721</td>
<td>179643</td>
</tr>
<tr>
<td>( \mathcal{I}_n )</td>
<td>( n^n )</td>
<td>1</td>
<td>4</td>
<td>27</td>
<td>256</td>
<td>3125</td>
<td>46656</td>
</tr>
<tr>
<td>( \mathcal{I}_n )</td>
<td>( \sum_{0}^{n} k! \binom{n}{k}^2 )</td>
<td>2</td>
<td>7</td>
<td>34</td>
<td>209</td>
<td>1546</td>
<td>13327</td>
</tr>
<tr>
<td>( \mathcal{B}_n )</td>
<td>( (2n - 1)!! )</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>105</td>
<td>945</td>
<td>10395</td>
</tr>
<tr>
<td>( S_n )</td>
<td>( n! )</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>24</td>
<td>120</td>
<td>720</td>
</tr>
<tr>
<td>( TL_n, J_n )</td>
<td>( C_n = \sum_{1}^{n+1} \binom{2n}{n} )</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>14</td>
<td>42</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td>$n = 1$</td>
<td>$n = 2$</td>
<td>$n = 3$</td>
<td>$n = 4$</td>
<td>$n = 5$</td>
<td>$n = 6$</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>$\mathcal{P}B_n$</td>
<td>1262</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{B}_n$</td>
<td>4</td>
<td>385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{P}_n$</td>
<td>4</td>
<td>272</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{P}\mathcal{T}_n$</td>
<td>4</td>
<td>50</td>
<td>94232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{I}_n$</td>
<td>4</td>
<td>23</td>
<td>2963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{I}_n^*$</td>
<td>2</td>
<td>6</td>
<td>795</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{T}_n$</td>
<td>2</td>
<td>8</td>
<td>283</td>
<td>132069776</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{B}_n$</td>
<td>2</td>
<td>6</td>
<td>42</td>
<td>10411</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathcal{T}\mathcal{L}_n$</td>
<td>2</td>
<td>4</td>
<td>12</td>
<td>232</td>
<td>12592</td>
<td>324835618</td>
<td></td>
</tr>
<tr>
<td>$\mathcal{S}_n$</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>11</td>
<td>19</td>
<td>56</td>
<td></td>
</tr>
</tbody>
</table>
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”.
\[ a = (\frac{1}{2} 2 \frac{3}{4} 4 \frac{5}{5}), \quad b = (\frac{1}{3} 2 3 \frac{4}{4} 5), \quad b = (\frac{1}{3} 2 3 \frac{4}{4} 5) \] and
\[ M = \langle a, b, c \rangle. \quad |M| = 31 \]
TRANSFORMATION SEMIGROUPS OF DEGREE 3

x axis: size of the semigroups

y axis: the number of $\mathcal{D}$-classes
SUBSEMGROUPS OF THE DEGREE 5 JONES MONOID

x axis: size of the semigroups
y axis: the number of $\mathcal{D}$-classes
INVERSE SEMIGROUPS (OF PARTIAL PERMUTATIONS)

x axis : size of the semigroups
y axis : the number of $D$-classes
TRANSFORMATION SEMIGROUPS OF DEGREE 4

x axis: size of the semigroups

y axis: the number of $D$-classes
TRANSFORMATION SEMIGROUPS OF DEGREE 4

x axis : size of the semigroups
y axis : the number of $\mathcal{D}$-classes
\[ \mathcal{L}, \mathcal{R} \text{ equivalence relations} \]

\[ t \mathcal{R} s \iff tS^1 = sS^1, \]
\[ t \mathcal{L} s \iff S^1t = S^1s, \]
\[ t \mathcal{J} s \iff S^1tS^1 = S^1sS^1 \]

\[ \mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L} = \mathcal{D} \]

\[ \mathcal{J} = \mathcal{D} \text{ in the finite case} \]

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

\[
\text{corresponds to } (())((()))()
\]

Applications in Physics: statistical mechanics, percolation problem.
$I_9$
$J_{16}$
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!