The Sortability of Graphs and Matrices under Context Directed Swaps

Colby Austin Brown — *University of Arizona*

Claudia Sofia Carrillo Vazquez — *University of Rochester*

Rashmika Goswami — *University of Michigan*

Sam Heil — *Washington University in St. Louis*

Marion Scheepers — *Boise State University*

Introduction: Ciliates

Ciliates, unicellular organisms, sort the genome of their micronucleus to assemble a transcriptionally functional macronucleus. The study of sorting algorithms helps in understanding the mechanisms by which this process occurs.

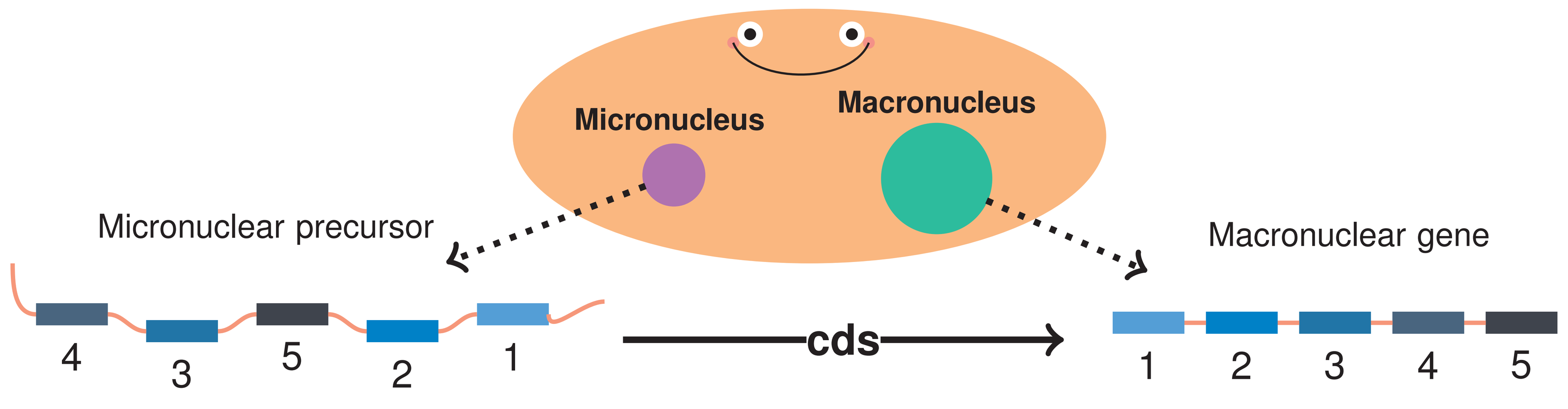

Figure 1: Context Directed Swaps in Ciliate Nuclei

Our research focus was to identify when permutations may be sorted by context directed swaps. We generalize the results of [1] through a novel graph theoretical and linear algebraic framework.

CDS and Breakpoint Graph

In the positive permutation π, assign to each entry a a left pointer $(a-1, a)$ and a right pointer $(a, a+1)$. Let

$$\pi = \Big[\{\alpha_1\ {}_p\}\ \{\boldsymbol{\alpha_2}\ {}_q\}\ \{\alpha_3\}\ \{{}_p\ \boldsymbol{\alpha_4}\}\ \{{}_q\ \alpha_5\}\Big],$$

where each α_i is a block of the permutation and $p = (x,\ x+1)$ and $q = (y,\ y+1)$ are some pointers in the permutation. As in [1], let the **context directed swap** on π with context p and q, $\mathbf{cds}_{\{p,q\}}(\pi)$, be

$$\Big[\{\alpha_1\ {}_p\}\ \{{}_p\ \boldsymbol{\alpha_4}\}\ \{\alpha_3\}\ \{\boldsymbol{\alpha_2}\ {}_q\}\ \{{}_q\ \alpha_5\}\Big].$$

The **breakpoint graph** of π, BG(π), has the pointers of π as vertices and an edge between each pair of identical pointers.

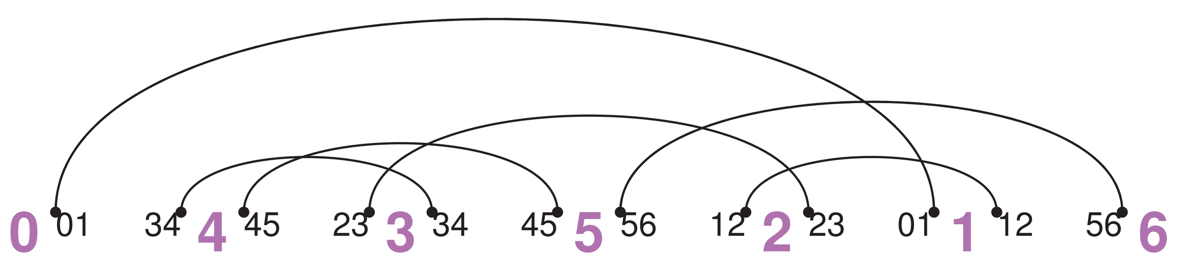

Figure 2: BG(π) of $\pi = [4,3,5,2,1]$

The $(0,1)$ and $(n, n+1)$ pointers are called the **roots** and are not used in valid contexts with which **cds** can be performed.

CDS on the Overlap Graph (GCDS)

The overlap graph of a permutation π, OG(π), has as vertices the edges of BG(π), each labeled by the pointer supporting the corresponding edge in BG(π). Two vertices of OG(π) are adjacent if the respective edges of BG(π) intersect.

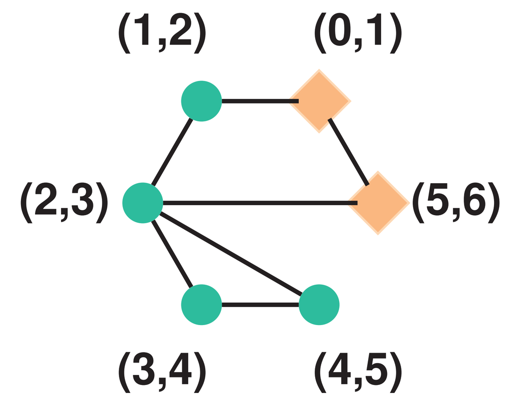

Figure 3: OG(π) of $\pi = [4,3,5,2,1]$

For vertices x and y, let $f_x(y)$ be 1 if x and y are adjacent and 0 otherwise. For any adjacent vertices p and q, let $\mathbf{gcds}_{\{p,q\}}$(OG(π)) be the graph OG$'$(π) with the same vertices as OG(π), where for any vertices u and v, OG$'$(π) includes the edge $\{u, v\}$ if and only if

$$f_p(u)f_q(v) + f_p(v)f_q(u) + f_u(v) \equiv 1 \pmod 2.$$

A graph is **gcds-sortable** if some sequence of **gcds** operations removes all of its edges.

CDS on the Adjacency Matrix (MCDS)

The **adjacency matrix** of OG(π) is the matrix A_π over $\mathbb{F}_2$ such that, for $1 \le i, j \le (n+1)$, $A_\pi(i,j) = 1$ if $\{(i-1, i), (j-1, j)\}$ is an edge in OG(π) and $A_\pi(i,j) = 0$ otherwise.

$$A_\pi = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 4: Adjacency Matrix of $\pi = [4,3,5,2,1]$

For any entry p and q, define the **cds** operation on matrices as

$$\mathbf{mcds}_{p,q}(A_\pi) = A_\pi + A_\pi I_{pq} A_\pi,$$

where $I_{pq}(i,j) = 1$ if $i = p$ and $j = q$ or $i = q$ and $j = p$ and $I_{pq}(i,j) = 0$ otherwise.

A matrix is **mcds-sortable** if applying some sequence of **mcds** operations to it yields the zero matrix.

Generalized Parity Cuts

Extending the definition of parity cut given in [2], let a **generalized parity cut** of a graph be a set S of vertices such that, for each vertex in the graph, the number of edges between that vertex and vertices in S is even.

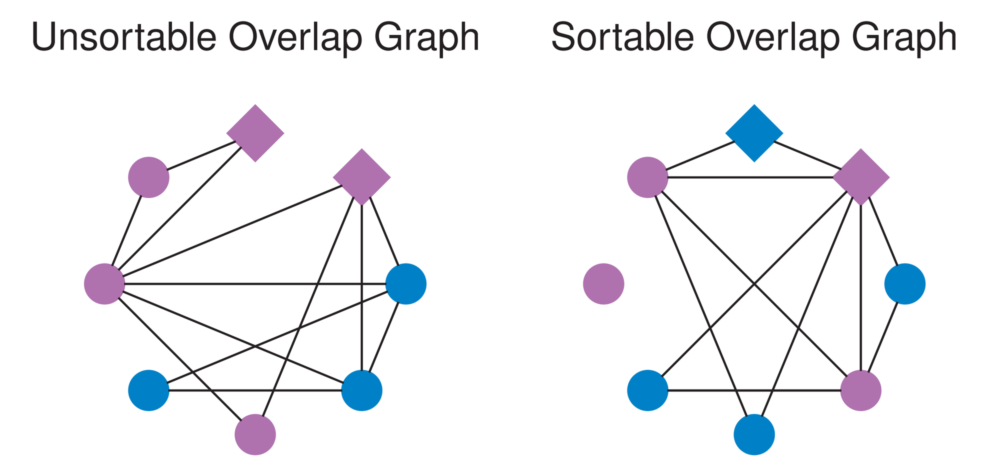

Figure 5: Generalized Parity Cuts (shown in purple)

Results

Theorem 1. *A two-rooted graph with roots x, y is* **gcds***-sortable if and only if for each of x, y there is a generalized parity cut containing that root and not the other.*

The set of generalized parity cuts of a graph corresponds to the kernel of the adjacency matrix of the graph.

Theorem 2. *A matrix is* **mcds***-sortable if and only if its kernel contains an element $\vec{x}$ such that $x_1 = 0$ and $x_n = 1$ and an element $\vec{y}$ such that $y_1 = 1$ and $y_n = 0$.*

Theorem 3. *The number of* **gcds***-sortable two-rooted graphs on n vertices is*

$$\sum_{s=0}^{\lfloor n/2 \rfloor - 1} 2^{s(s+3)} \left(\frac{\prod_{i=0}^{2s-1} (2^{n-2-i} - 1)}{\prod_{i=1}^{s} (2^{2i} - 1)} \right).$$

Let r_n be the proportion of graphs on n vertices that are **gcds**-sortable. Then, the sequences $\{r_{2n}\}$ and $\{r_{2n+1}\}$ converge, and

$$\lim_{n\to\infty} r_{2n} \approx 0.2272 \quad \text{and} \quad \lim_{n\to\infty} r_{2n+1} \approx 0.1061.$$

Let G be a graph on the vertices $(1,2), (2,3), \ldots, (n-1, n)$ with adjacency matrix A taken over the field $\mathbb{F}_2$. Then, the number of ways we can add two roots to G to form a **gcds**-sortable graph is $4^{\text{rank}(A)}$.

Future Work

- Further analyze and classify unsortable permutations, graphs, and matrices.
- Extend the overlap graph to include additional relationships.

References

[1] K.L.M. Adamyk et al., *Sorting Permutations: Games, Genomes and Cycles*, **arXiv:1410.235**.

[2] H.Q. Li et al., *Parity Cuts In 2-Rooted Graphs*, preprint.

[3] J. MacWilliams, *Orthogonal matrices over finite fields*, American Mathematical Monthly **76: 2** (1969), 152–164.

Acknowledgements

This research, conducted at the Complexity Across Disciplines Research Experience for Undergraduates site, was supported by National Science Foundation REU Site Grant DMS-1659872 and by Boise State University.