# Can we count on Deep Learning: Characterizing Combinatorial Structures with Interpretability

Helen Jenne, Davis Brown, Herman Chau, Jackson Warley, Tim Doster, Henry Kvinge

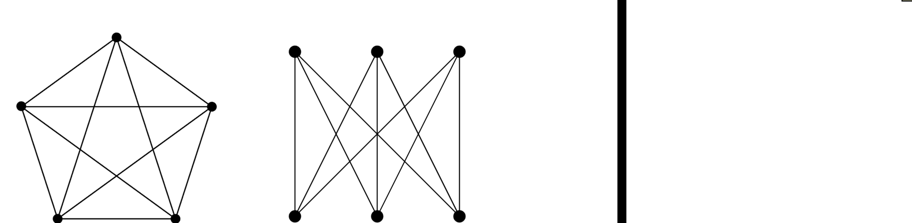
Pacific Northwest National Laboratory, University of Washington

**Feature Attribution Clustering for Exploration (FACE) is a method to help mathematicians characterize sets of mathematical objects via prototypical feature attribution maps.**

**Examples of characterization problems:**

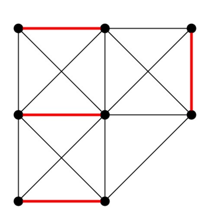
Q: *Which graphs are planar?*

A: Planar ⟷ no $K_5$, $K_{3,3}$ minors

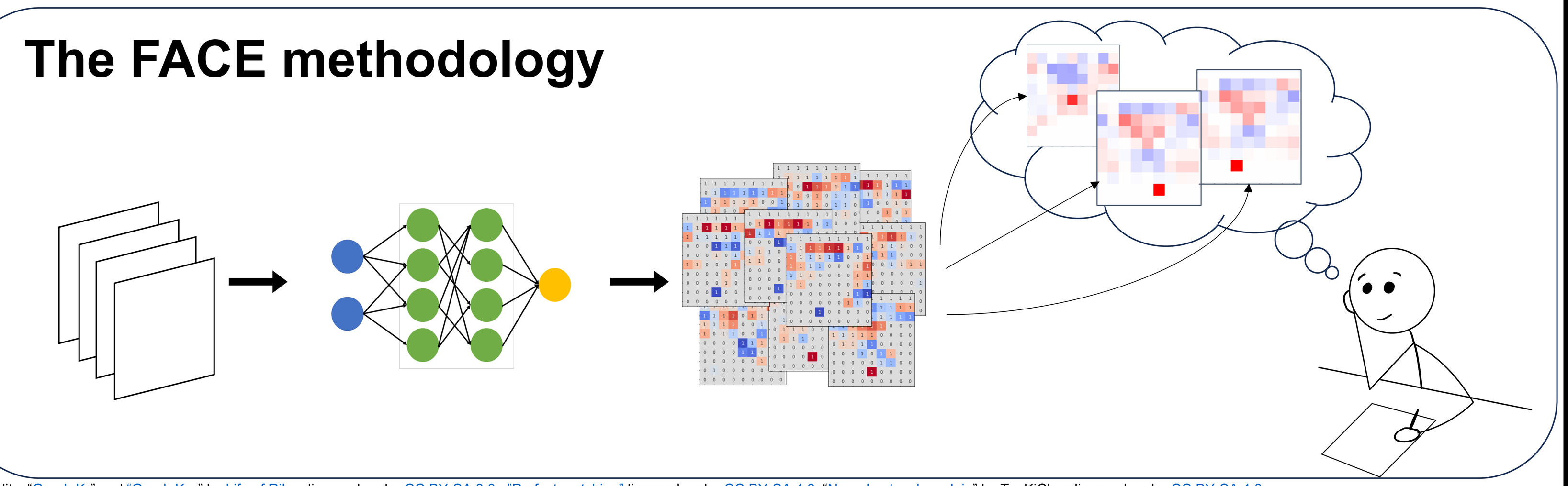


Q: *Which graphs have perfect matchings?*

A: $G$ has a perfect matching ⟷ for every subset of vertices $S$, $G \setminus S$ has at most $|S|$ odd components

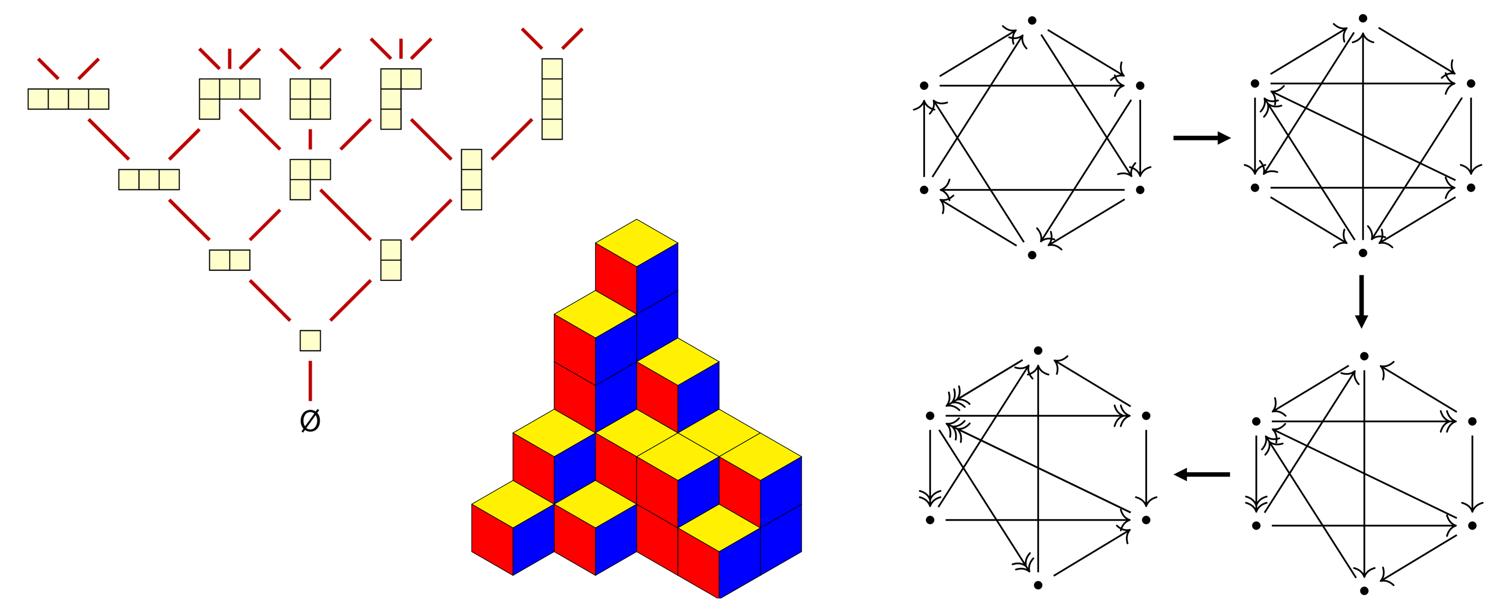


**The FACE methodology**

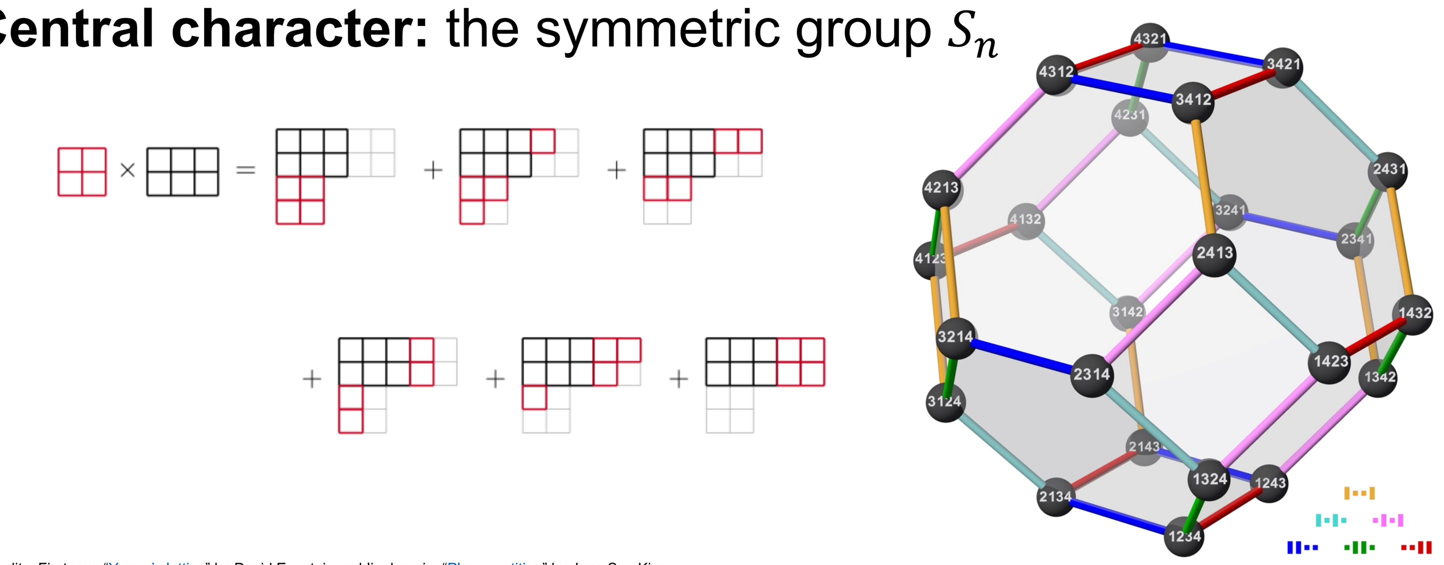




## Algebraic combinatorics



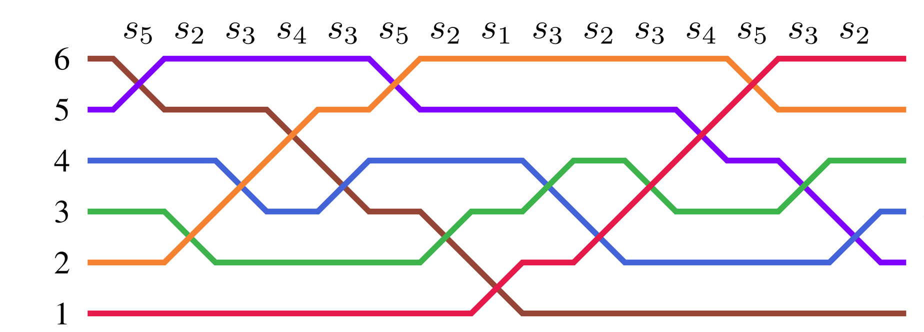
**Central character:** the symmetric group $S_n$





## Weaving patterns

Any permutation can be written as a product of adjacent transpositions, called a *reduced word.*



| | | | | | |
|---|---|---|---|---|---|
| 6 | 1 | 1 | 1 | 1 | 1 |
| 5 | 0 | 1 | 1 | 1 | 1 |
| 4 | 1 | 0 | 1 | 1 | 0 |
| 3 | 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 |

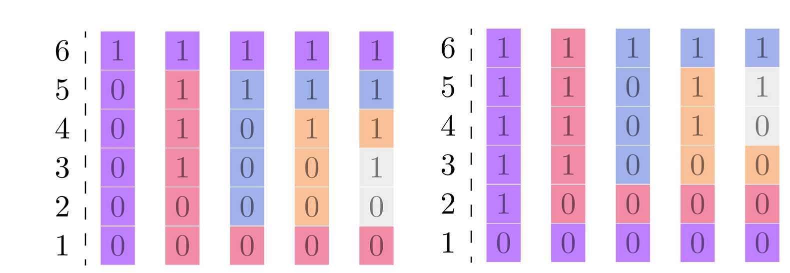
Every reduced word that differs by a *commutation relation* has the same weaving pattern, e.g. $s_1s_2s_3s_1s_2s_1 \sim s_1s_2s_1s_3s_2s_1$

Q: Which {0,1} matrices are weaving patterns?

Some families of reduced words have weaving patterns with nice characterizations.

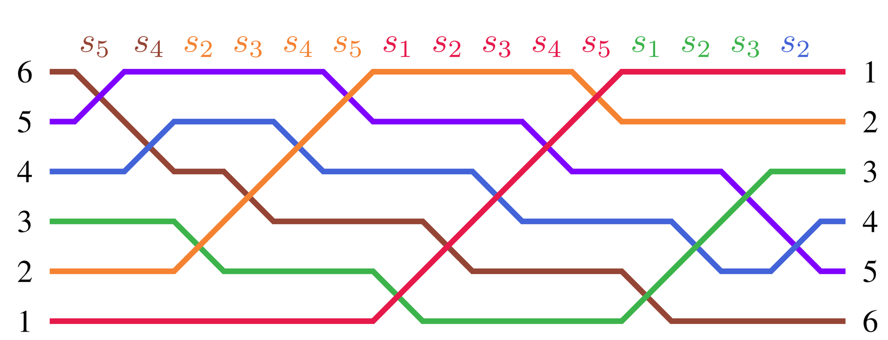
| | | | | | |
|---|---|---|---|---|---|
| 6 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 0 | 1 | 1 | 1 |
| 4 | 0 | 1 | 1 | 1 | 0 |
| 3 | 1 | 0 | 0 | 1 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 0 | 0 | 0 |

{0, 1} matrix that is not a weaving pattern



Examples of weaving patterns of *ordered words*

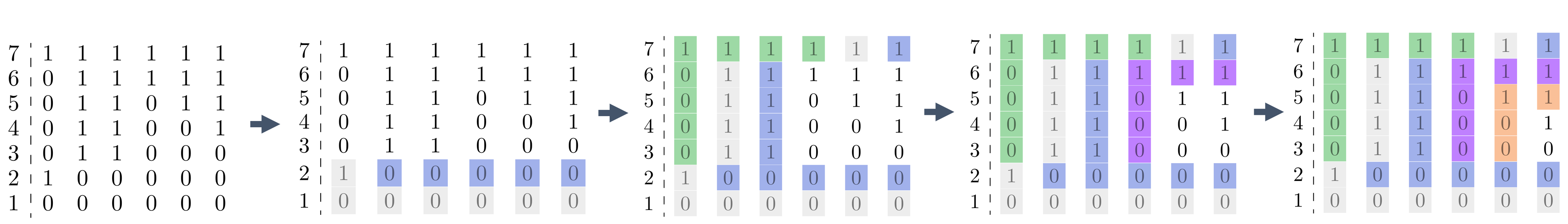
## Case Study: Two-Sided Ordered Words



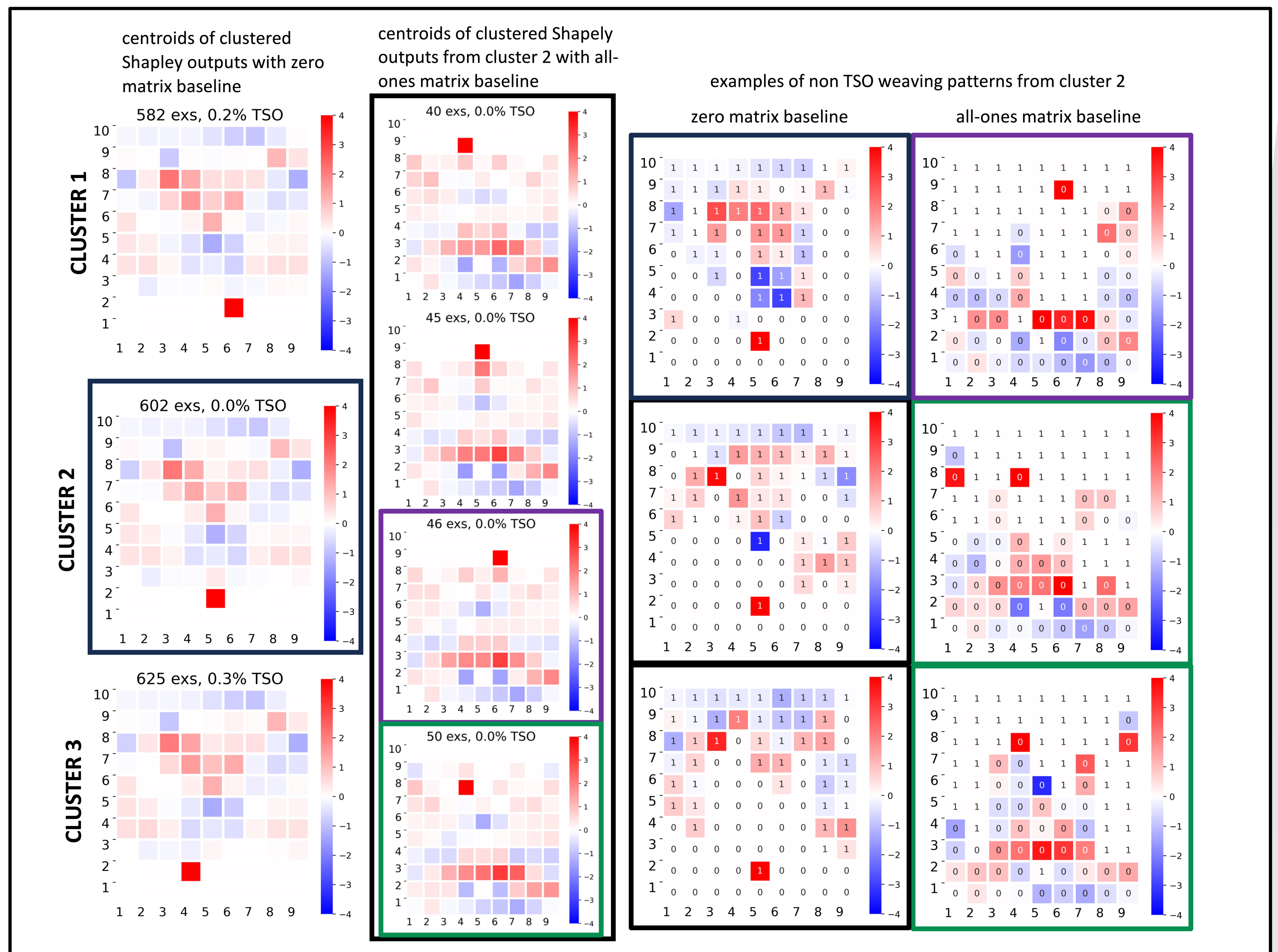
| | | | | | |
|---|---|---|---|---|---|
| 6 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 1 | 1 | 0 |
| 4 | 0 | 0 | 1 | 1 | 1 |
| 3 | 1 | 0 | 0 | 0 | 1 |
| 2 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 |

### Application of FACE

- **Create dataset:** Generated all TSO weaving patterns for $n = 9$ (48,896 $10 \times 9$ {0,1} arrays) and the same number of non-TSO weaving patterns.
- **Train model:** Trained a CNN with two convolutional layers; 99% test accuracy
- **Calculate feature attribution representations:** Applied Shapley to 16,000 weaving patterns from the test set.
- **Find prototype feature attributions through clustering:** Clustered Shapley outputs using $k$-means and calculated centroids for each cluster.
- **Mathematician analysis of prototypes**



Example of application of algorithm to determine whether weaving pattern is TSO




U.S. DEPARTMENT OF ENERGY
BATTELLE
PNNL is operated by Battelle for the U.S. Department of Energy

For additional information, contact:
Helen Jenne | helen.jenne@pnnl.gov

Pacific Northwest NATIONAL LABORATORY
www.pnnl.gov
PNNL-SA-192589