

Acknowledgments

This work is supported in part by Quanta Computer as part of the Qmulus project, MIT's Amar Bose Teaching Fellowship, the Microsoft Research Fellowship, and the NSF Graduate Research Fellowship under Grant No. 1122374.

Feature Engineering for Clustering Student Solutions

Elena Glassman, Rishabh Singh, Rob Miller

MIT CSAIL

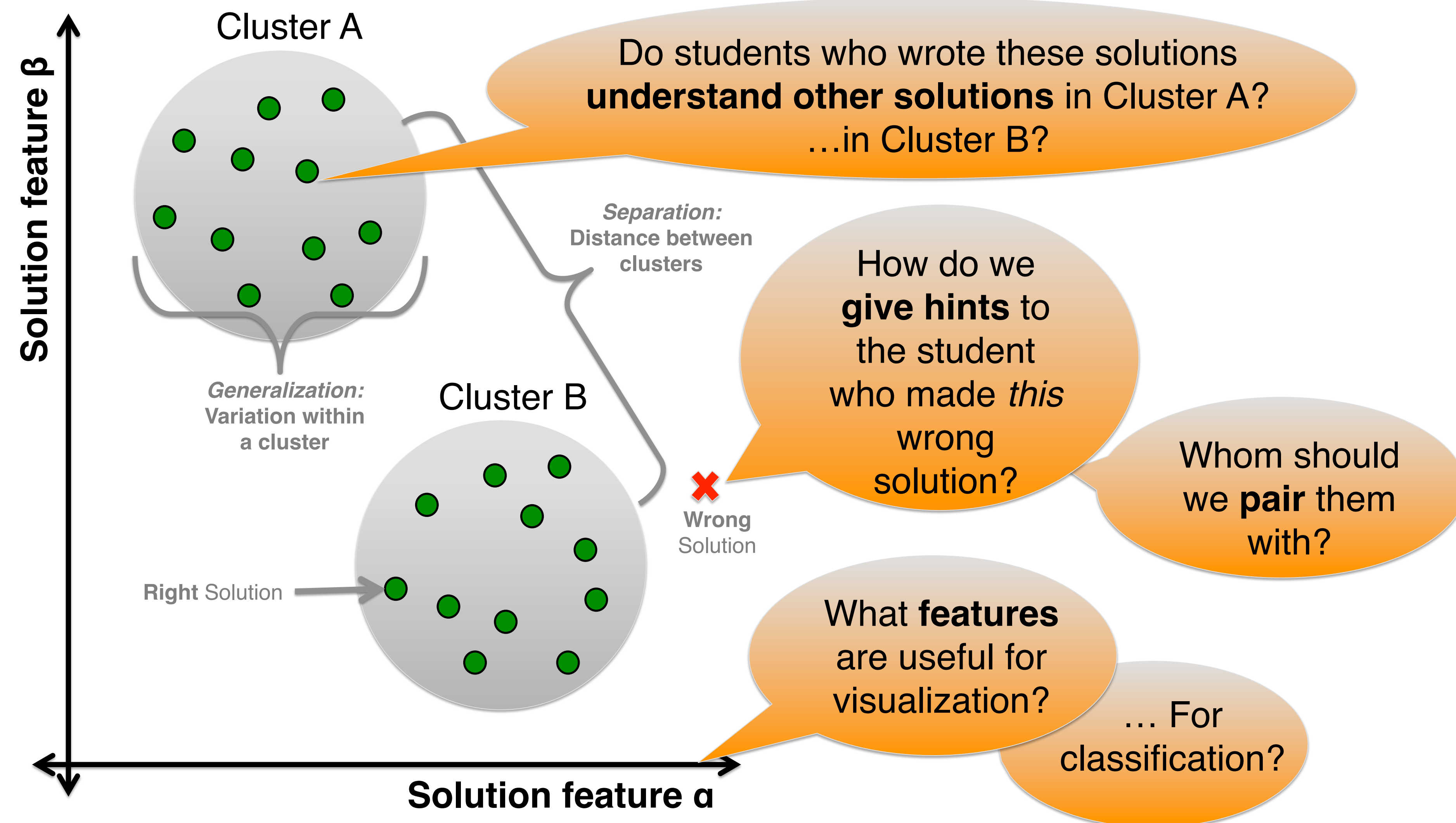
{elg,rishabhs,rcm}@mit.edu

Problem

- There may be **several distinct, correct solutions** to coding assignments.
- Some solutions may be unknown** to the teaching staff or intelligent tutor designer.
- This complicates the task of providing help and hints.

Approach

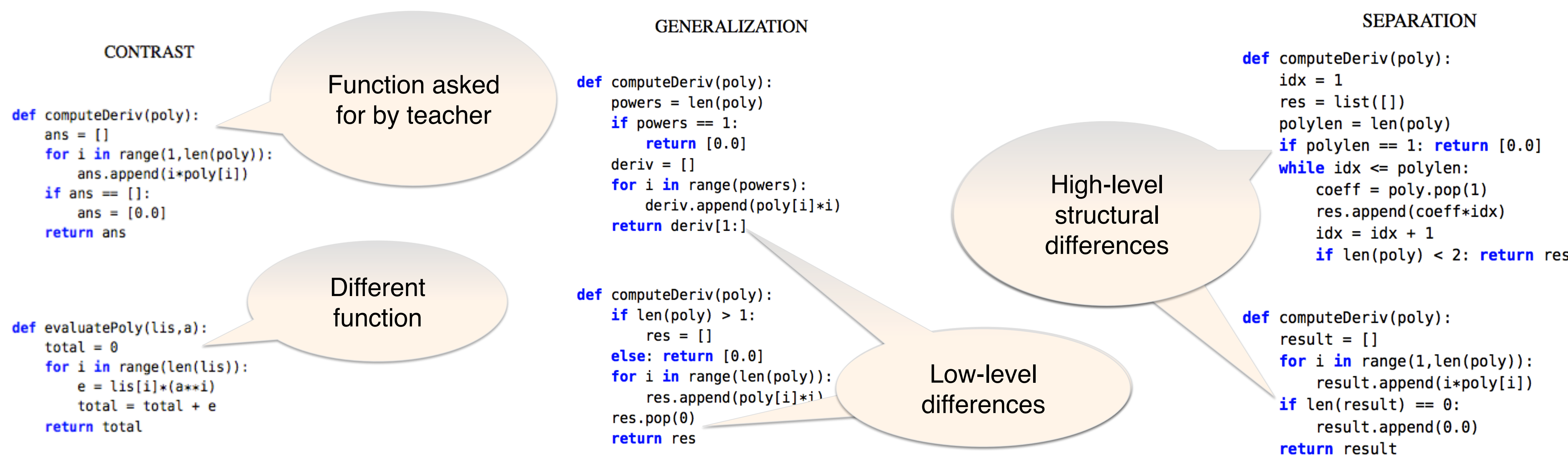
- Visualize** hundreds or thousands of student solutions to find alternatives
- Classify** solutions by their design choices
 - By machine learning and/or human staff
- Use** this knowledge to enhance
 - Activities about design choices and tradeoffs
 - Peer-pairing
 - Automated help



Variation Theory

Variation Theory (Marton et al., '13): Specific dimensions of variation help students learn a concept, e.g., sorting

- Contrast dimension** contains examples that are and are not a sorting function.
- Generalization dimension** includes examples with the same algorithm but different low-level implementations.
- Separation dimension** captures the full variation of implementations that sort.

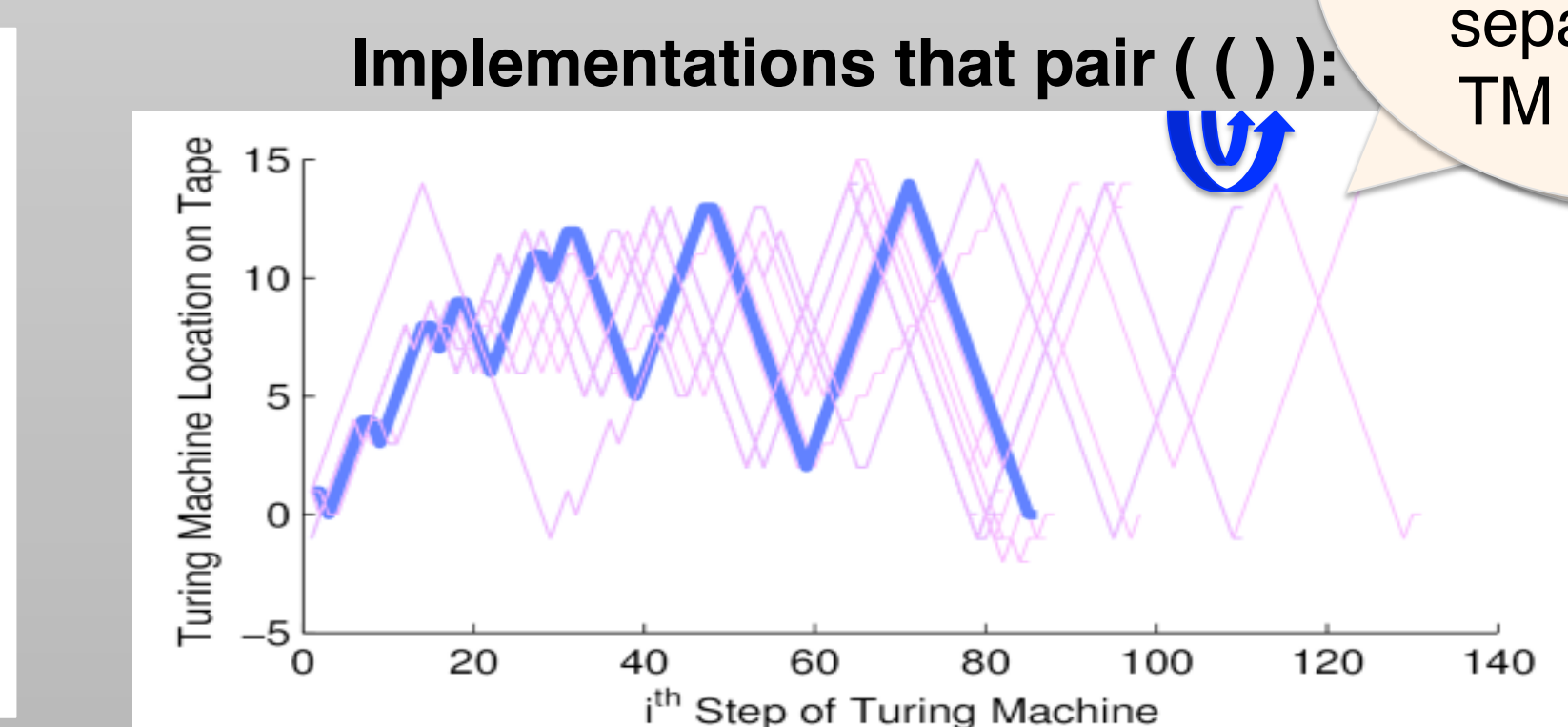
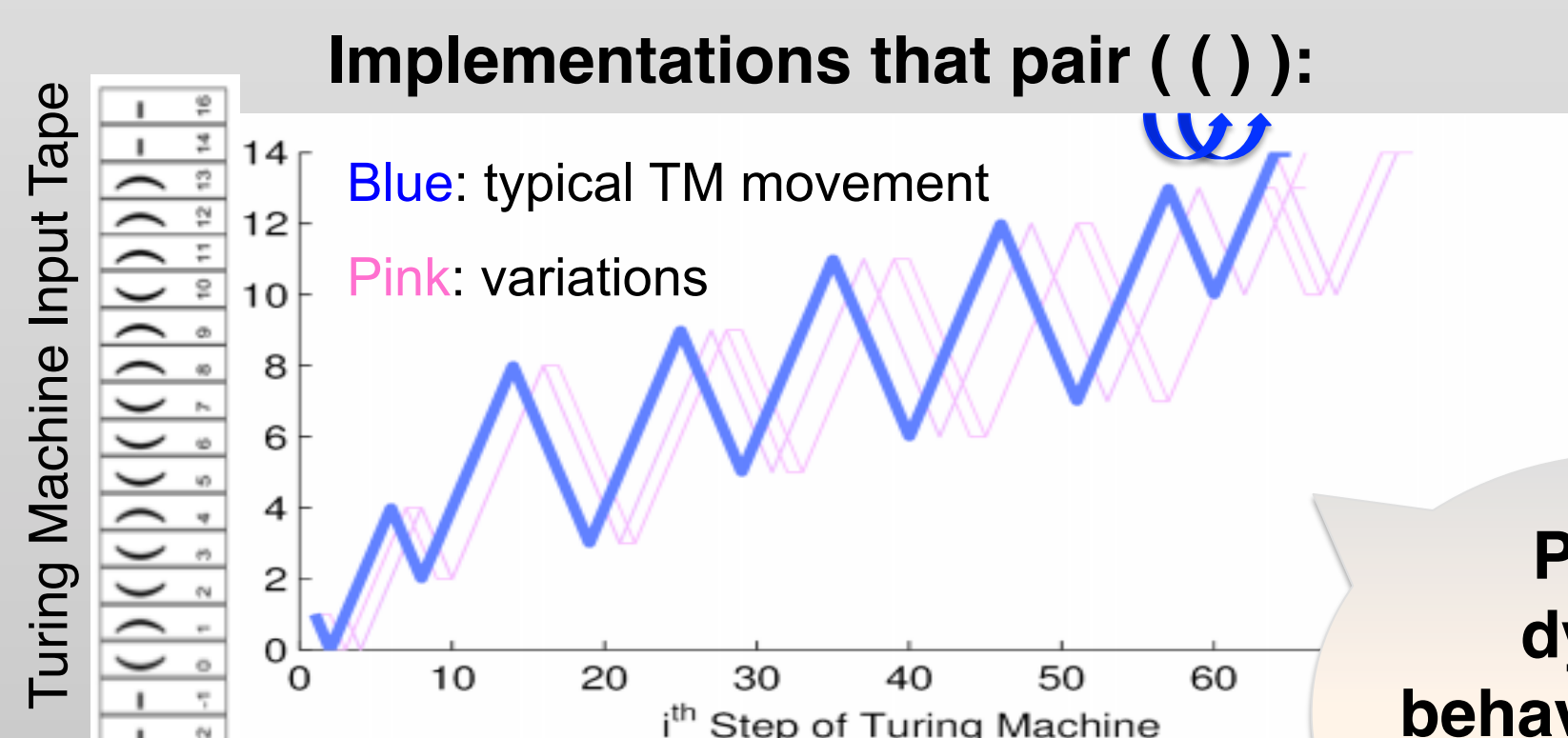


Case Study 1: Turing Machine (TM) Synthesis Lab @ MIT

- Most staff knew of one TM solution.
- Dynamic behavior of TM solutions shows two solution clusters.

Student-written TM code is the original representation.

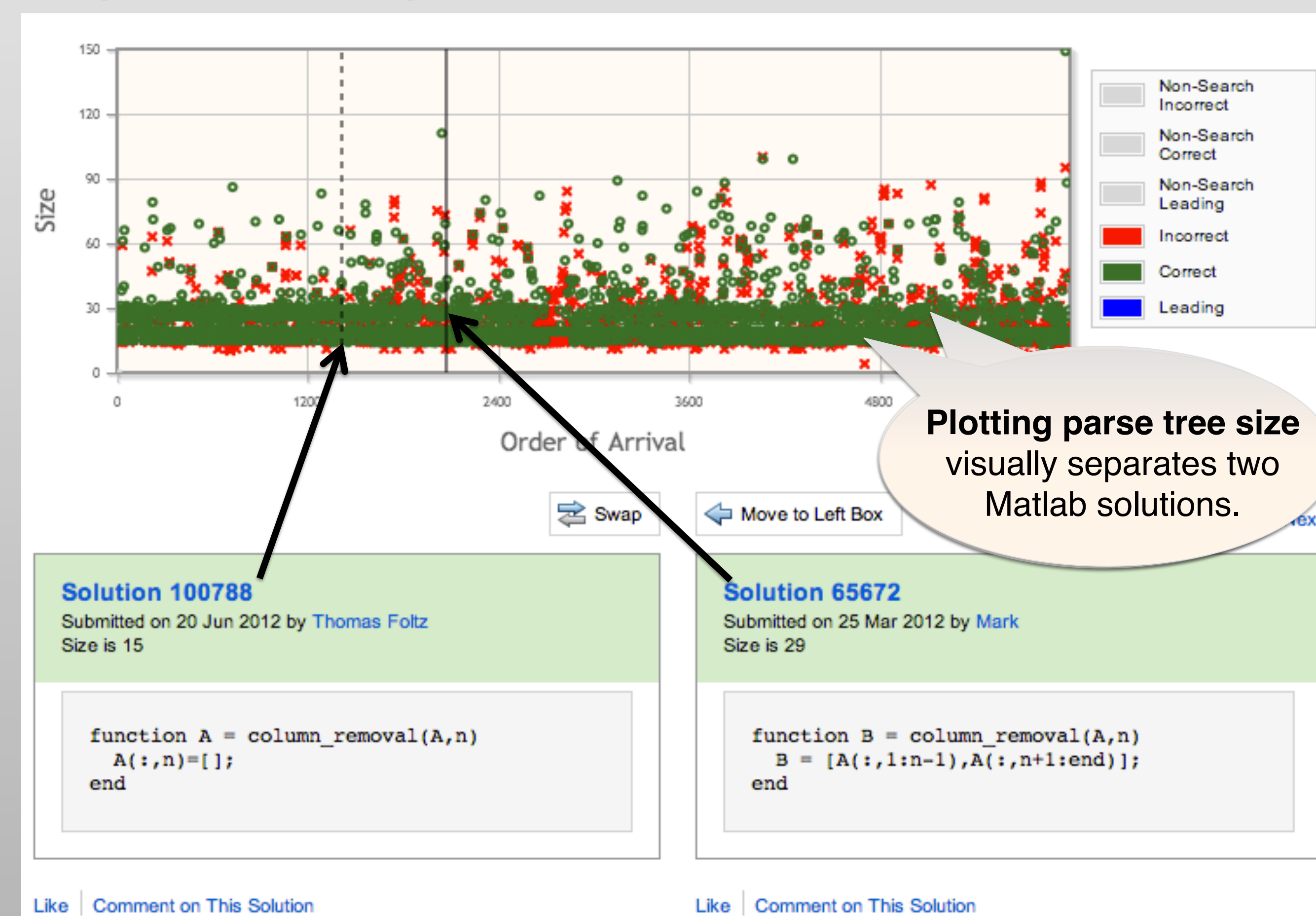
```
// old R new W M
action A - B 1 r
action A 1 C 1 l
action B - A - l
action B 1 D - l
action C - A 1 l
action C 1 *halt* 1 l
action D - B 1 l
action D 1 E 1 r
action E - D - r
action E 1 B - r
```



Plotting dynamic behavior visually separates two TM solutions.

Case Study 2: Matlab Programming Game Visualization

- This **interactive visualization** allows users to **explore and compare** other submitted solutions.



Work In Progress: Visualizing Code Variation in Python for Insight and Feature Discovery

