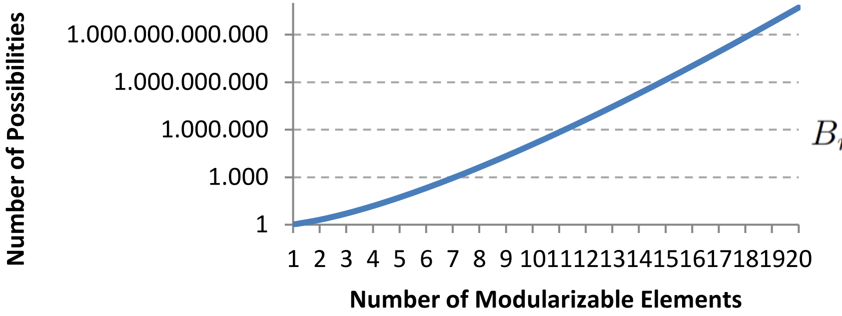
Towards a Multi-Objective Modularization Approach for Entity-Relationship Models

Dominik Bork, Antonio Garmendia, Manuel Wimmer

November, 4th 2020

Research Context

- Problem
  - Data models often evolve into large monolithic artefacts
  - Modularization enables handling complexity, but how?



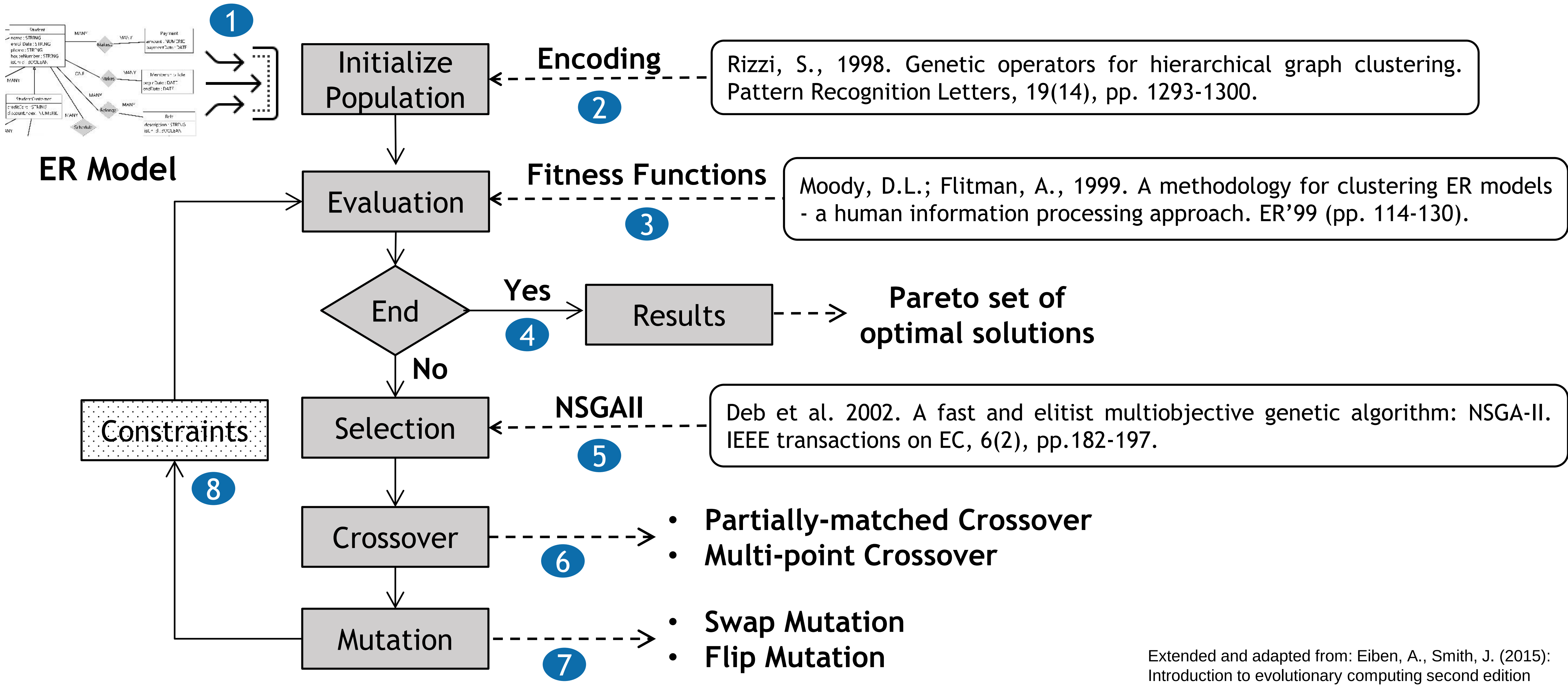
$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k$$

$$B_0 = 1$$

- Objective
  - Intelligent approach to automatically modularize large ER models
  - Formulation as Many-objective problem to be solved by Search-based Software Engineering using Genetic Algorithms
- Solution Overview
  - We use Conceptual Modeling and Model-driven Development
  - **Focus** We realize a Generic Encoding
  - We implement the *ModulER* Tool

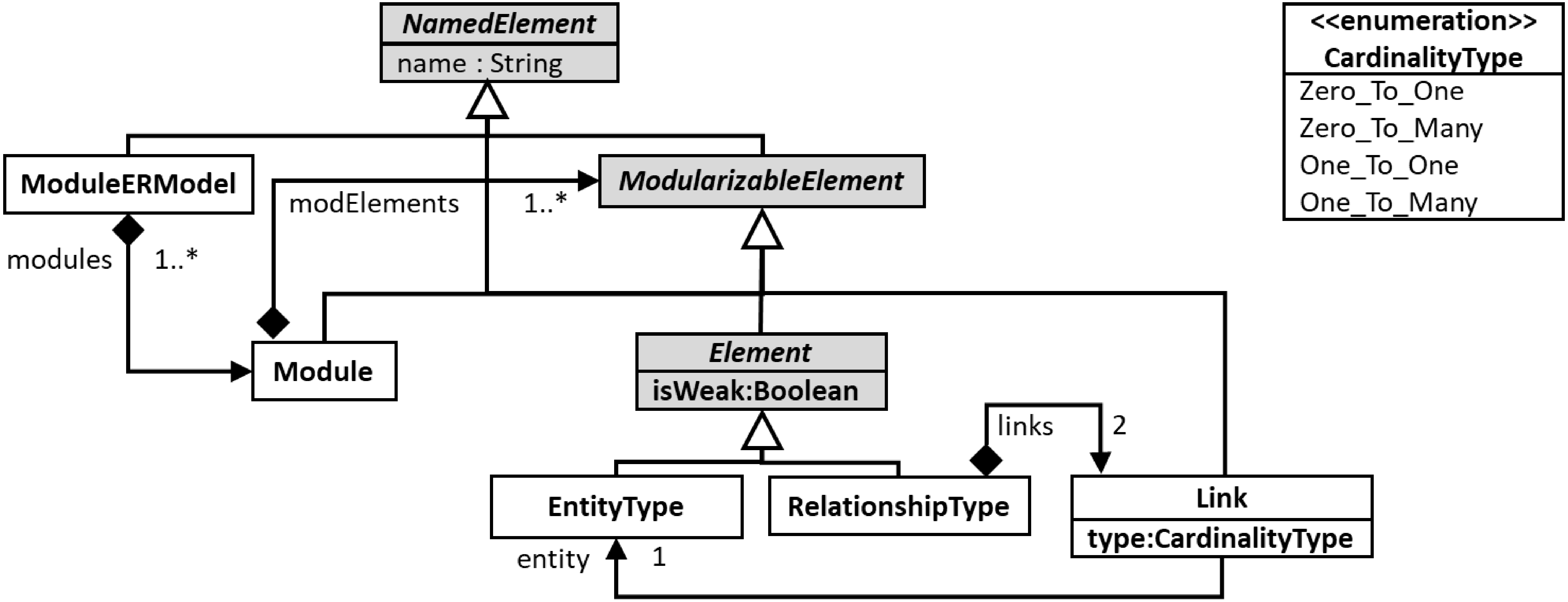
Jenetics eclipse

The *ModulER* Approach: Overview

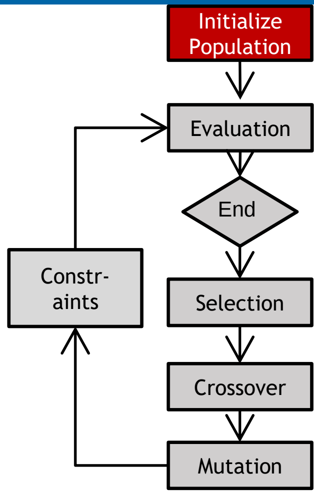


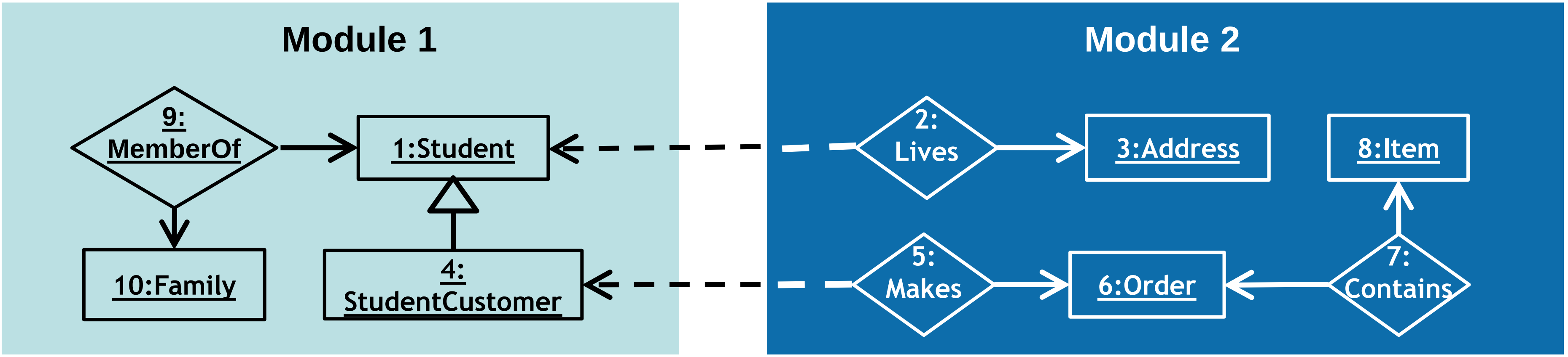
Extended and adapted from: Eiben, A., Smith, J. (2015): Introduction to evolutionary computing second edition

The *ModulER* Metamodel



ModulER Encoding





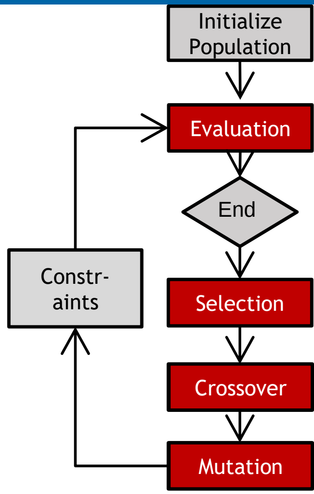
Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chromosome	1	0	4	0	9	0	10	1	2	0	3	0	5	0	6	0	7	0	8

Transform into ↓

Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chromosome	1	4	9	10	2	3	5	6	7	8	0	0	0	1	0	0	0	0	0

Enumerable Chromosome (positions 1–10) **Bit Chromosome** (positions 11–19)

ModulER Fitness Functions & Alterers

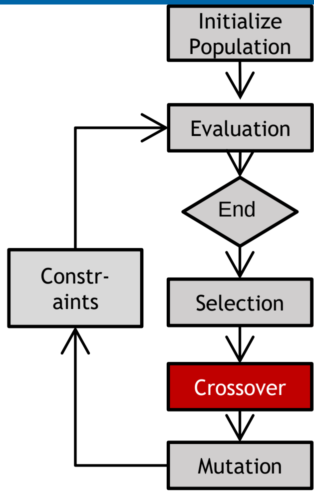


ID	Description	Goal
COH	Cohesion: the sum of links within modules.	MAXIMIZE
COP	Coupling: the sum of links between modules.	MINIMIZE
NMOD	Number of modules.	MINIMIZE
AVGMODEL	Average number of modularizable elements per module.	MINIMIZE
BAL	The standard deviation of module size of all modules.	MINIMIZE

Alterers

- Crossover and Mutation used to alter
  - the assignment of modularizable elements to modules
    - Partially-matched Crossover and Swap Mutator
  - the number of modules
    - Multi-point Crossover and Flip Mutator

ModulER Alterers: Crossover



Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chromosome	1	4	9	10	2	3	5	6	7	8	0	0	0	1	0	0	0	0	0

	Enumerable Chromosome	Bit Chromosome
Crossover	Partially-matched Crossover (PMX)	Multi-point Crossover (MX)
Mutators	SwapMutator	FlipMutator

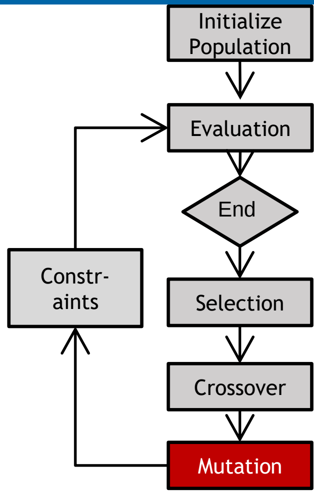
Individual \ Position	1	2	3	4	5	6	7	8	9
P1	1	2	3	4	5	6	7	8	9
P2	11	10	9	8	7	6	5	4	3
P1’	1	2	9	8	7	6	5	4	3
P2’	11	10	3	4	5	6	7	8	9

Partially-matched Crossover (PMX)

Individual \ Position	1	2	3	4	5	6	7	8	9
P1	1	2	3	4	5	6	7	8	9
P2	11	10	9	8	7	6	5	4	3
P1’	1	2	9	8	7	6	7	8	9
P2’	11	10	3	4	5	6	5	4	3

Multi-point Crossover (MX)

ModulER Alterers: Mutators



Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Chromosome	1	4	9	10	2	3	5	6	7	8	0	0	0	1	0	0	0	0	0

Enumerable Chromosome (positions 1–10) | **Bit Chromosome** (positions 11–19)

	Enumerable Chromosome	Bit Chromosome
Crossover →	**Partially-matched Crossover (PMX)**	**Multi-point Crossover (MX)**
Mutators →	**SwapMutator**	**FlipMutator**

Position	1	2	3	4	5	6	7	8	9
Chromosome	11	10	3	4	5	6	7	8	9

↓

Position	1	2	3	4	5	6	7	8	9
Chromosome	11	10	7	4	5	6	3	8	9

SwapMutator

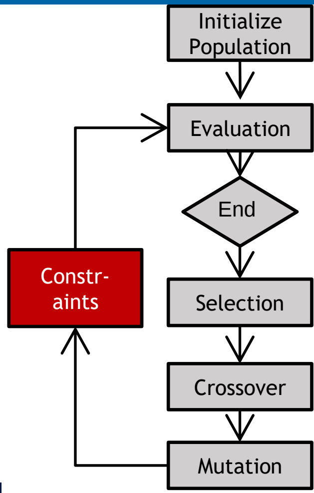
Position	11	12	13	14	15	16	17	18	19
Chromosome	0	0	0	1	0	0	0	0	0

↓

Position	11	12	13	14	15	16	17	18	19
Chromosome	0	1	0	1	0	0	0	0	0

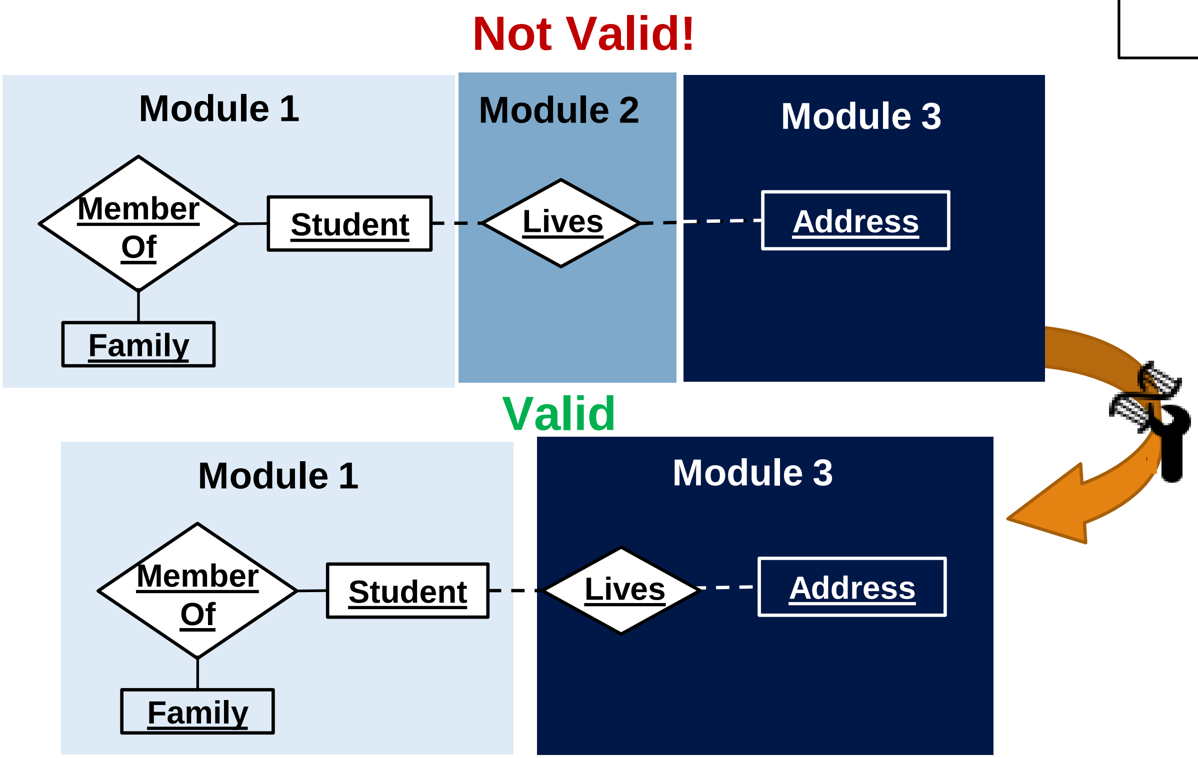
Flip Mutator

ModulER Constraints



Two constraints to steer the GA toward valid solutions

- **RelationshipToEntity**
  A RelationshipType should always be in one of the modules the two entities it relates are in.
- **NumberOfModules**
  The number of modules and the module size should be comprehensible for humans.



Selected Evaluation Results I

- **RQ1 Search Validation** ✓
  Is the approach capable of efficiently modularizing ER models?

Table 2: Characteristics of the two example cases and modularization performance.

Case	Entity Types	Relationship Types	Inheritance	Attributes	Avg. Computation time [Seconds]
Karate	12	13	1	29	1.8
Finance	116	17	97	67	6.8

✓ Efficient computation of the Pareto set
✓ Only valid solutions w.r.t. the ModulER metamodel and the ModulER constraints

Selected Evaluation Results II

- **RQ2 Solution Correctness** ✓
  How good is the quality of the produced solutions and how good are the solutions compared to ones conceptual modelers create?

Pareto set (n = 70)

- Heterogeneous solutions
- Fulfilling the different objectives to different extends

Avg. Edits: 5 ←

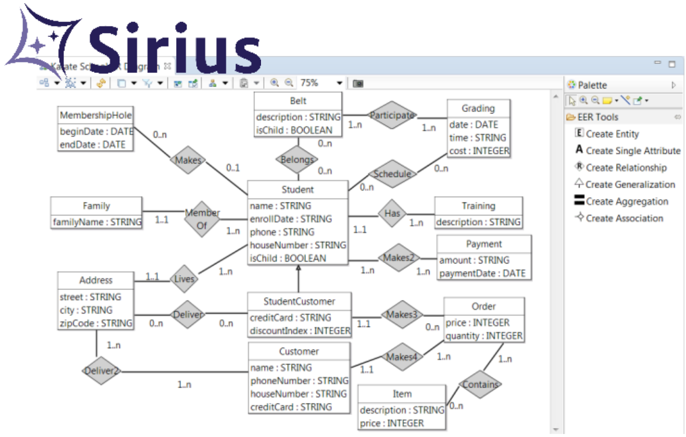
→ Outperform in all objectives

Modeler solutions (n = 7)

- Homogeneous solutions
  - All with 3 modules
- Focus on cohesion

Conclusions and Future Research

- The algorithm efficiently produces good modularizations
- More empirical research required
  - Larger sample size
  - More complex examples
  - Thematic clustering
- Next steps: Integrating the Modeler in-the-loop[1]
- Extending the approach to
  - Consider Extended Entity Relationship model concepts
  - Consider hierarchical modularizations



Open Source!

[1] Kessentini, Wimmer, Sahraoui (2018): Integrating the designer in-the-loop for metamodel/model co-evolution via interactive computational search. In MODELS’18 (pp. 101-111)

Thank you!

TU WIEN

JKU
JOHANNES KEPLER
UNIVERSITY LINZ

JKU
JOHANNES KEPLER
UNIVERSITY LINZ