CT-Competition 2023

@ 12th International Workshop on Combinatorial Testing

Outline

- Competition aim and participants

- Categories

- Competition results and considerations

- Ideas for the next edition?

Competition organizers

- Andrea Bombarda (University of Bergamo) - *Evaluation*
- Manuel Leithner (SBA Research) - *Execution*
- Michael Wagner (SBA Research) - *Validation*

and thanks for the support and help to

- Angelo Gargantini (University of Bergamo)
- Bernhard Garn(SBA Research)
- Dimitris Simos (SBA Research)

Competition aim and participants (1)

The competition compares state-of-the-art tools for generating combinatorial test suites with respect to the generation time and test suite size.

With this competition, we wanted to motivate implementors to present their work to a broader audience and to compare it with that of others, as well as to improve and test their generators on artificial models.
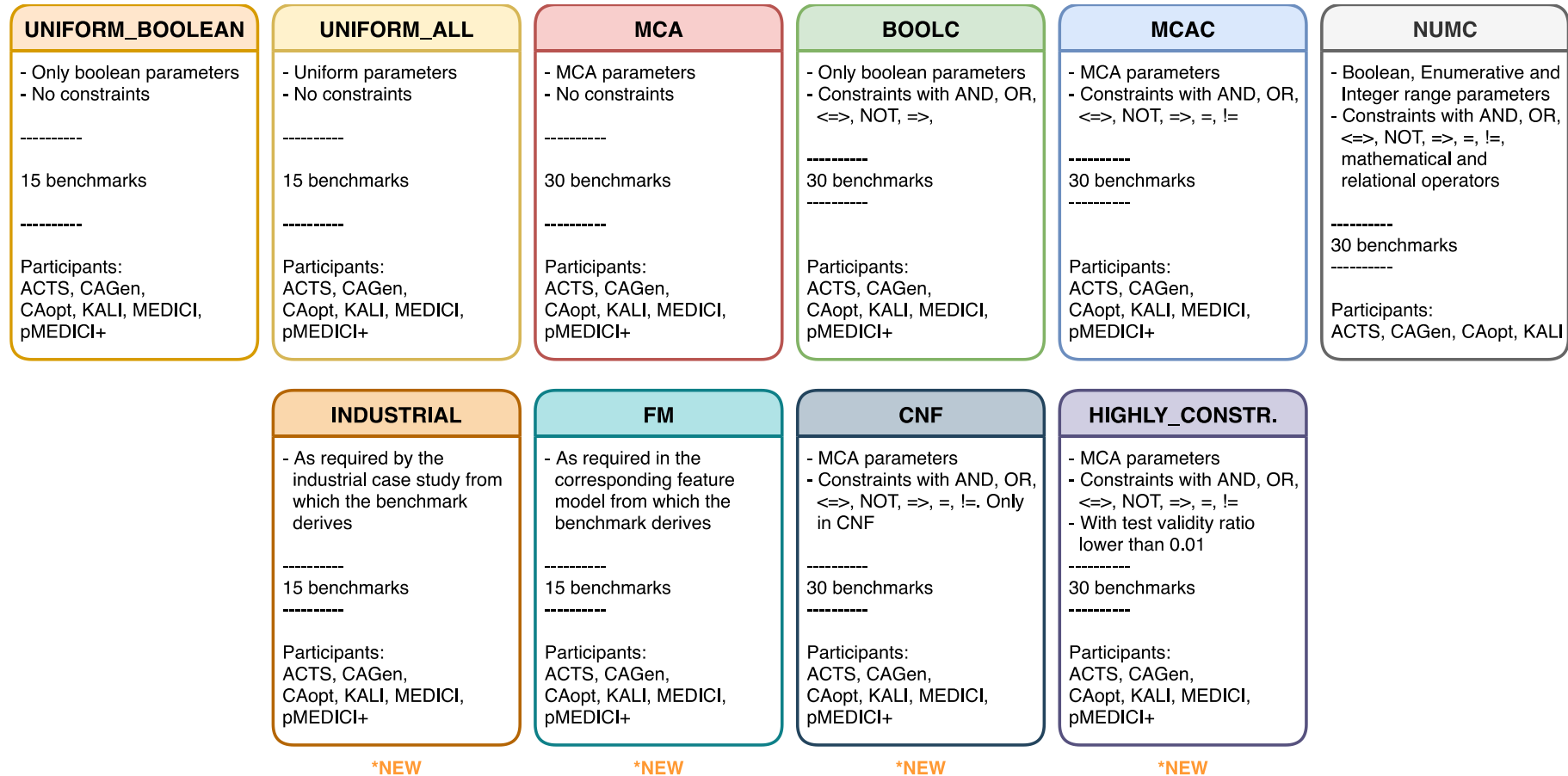
Competition aim and participants (2)

- ACTS: Java Implementation of IPO, one of the most used combinatorial test generation tools. Executed with default settings (IPOG with MFT)

- CAgen: a multithreaded FIPOG implementation written in Rust

- CAopt: a sampling and an optimization phase, based on a SAT solver

- KALI: a java multi-thread tool exploiting SMT solvers

- MEDICI: a C++ tool for combinatorial test generation based on the use of Multi-Valued Decision Diagrams

- pMEDICI: a java multi-thread implementation of the MEDICI tool, based on Multi-valued Decision Diagrams

Categories / Tracks

UNIFORM_BOOLEAN	UNIFORM_ALL	MCA	BOOLC	MCAC	NUMC
- Only boolean parameters - No constraints ---------- 15 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- Uniform parameters - No constraints ---------- 15 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- MCA parameters - No constraints ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- Only boolean parameters - Constraints with AND, OR, <=>, NOT, =>, ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- MCA parameters - Constraints with AND, OR, <=>, NOT, =>, =, != ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- Boolean, Enumerative and Integer range parameters - Constraints with AND, OR, <=>, NOT, =>, =, !=, mathematical and relational operators ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI

INDUSTRIAL	FM	CNF	HIGHLY_CONSTR.
- As required by the industrial case study from which the benchmark derives ---------- 15 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- As required in the corresponding feature model from which the benchmark derives ---------- 15 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- MCA parameters - Constraints with AND, OR, <=>, NOT, =>, =, !=. Only in CNF ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+	- MCA parameters - Constraints with AND, OR, <=>, NOT, =>, =, != - With test validity ratio lower than 0.01 ---------- 30 benchmarks ---------- Participants: ACTS, CAGen, CAopt, KALI, MEDICI, pMEDICI+
*NEW	*NEW	*NEW	*NEW

STRENGTH

2	3	4	5	6

Some statistics

- 240 benchmark models, divided in 10 categories *(last year we had 6 categories)*

- Strength from 2 to 6 *(last year the maximum strength was 4)*

- 3 executions per each model and each strength on each tool

- ~17GB of output files containing test suite results

Evaluation method
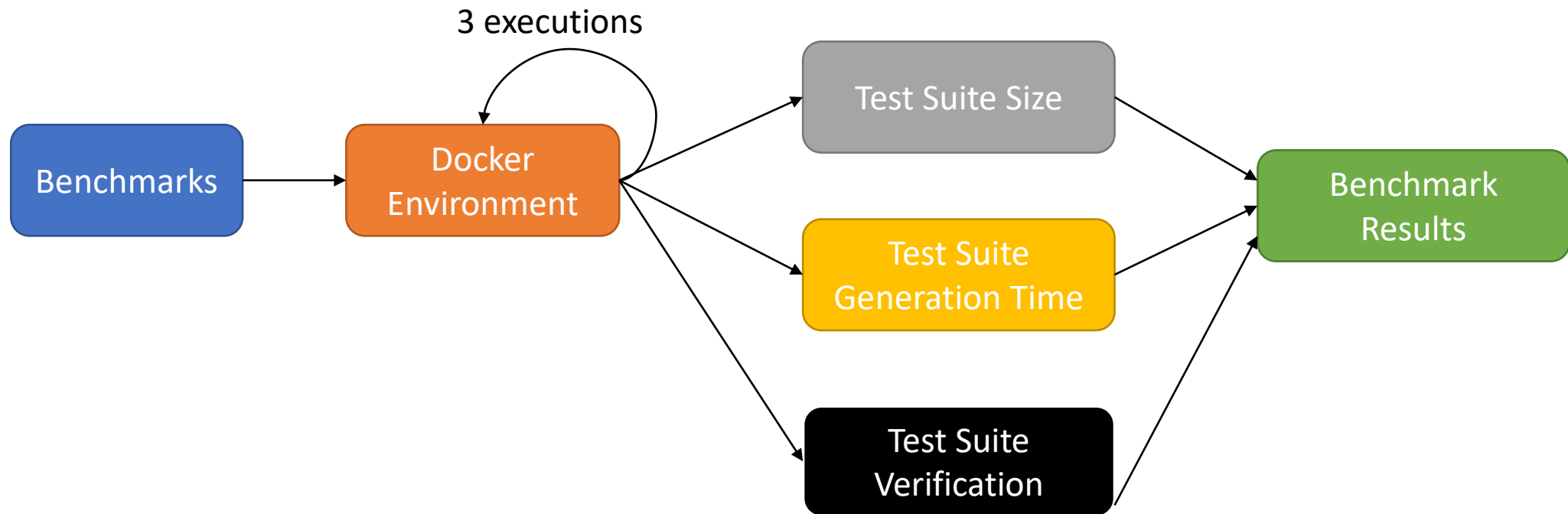
COMPETITION RESULTS

Competition results – Not valid / timed out instances (1)

t =2

Tool	ACTS	CAgen	CAopt	KALI	MEDICI	pMEDICI
UNIFORM_BOOL	0	0	0	0	0	0
UNIFORM_ALL	0	0	0	0	0	0
MCA	0	0	0	10	0	0
BOOLC	0	0	0	14	0	0
MCAC	10	12	0	28	0	14
NUMC	9	9	0	30	--	--
INDUSTRIAL	1	1	4	12	5	2
FM	0	0	0	12	1	0
CNF	0	0	0	30	0	6
HIGHLY_CONSTRAINED	5	6	0	28	0	10

Competition results – Not valid / timed out instances (2)

t =3

Tool	ACTS	CAgen	CAopt	KALI	MEDICI	pMEDICI
UNIFORM_BOOL	0	0	0	0	0	0
UNIFORM_ALL	0	0	4	5	5	5
MCA	0	0	0	10	15	1
BOOLC	0	0	0	23	0	0
MCAC	10	12	3	29	20	20
NUMC	9	9	6	30	--	--
INDUSTRIAL	1	1	4	14	7	4
FM	0	0	0	14	1	0
CNF	0	0	0	30	18	14
HIGHLY_CONSTRAINED	5	6	1	30	17	18

Competition results – Not valid / timed out instances (3)

t =4

Tool	ACTS	CAgen	CAopt	KALI	MEDICI	pMEDICI
UNIFORM_BOOL	0	0	0	0	0	0
UNIFORM_ALL	2	0	8	8	13	8
MCA	0	0	23	24	29	24
BOOLC	0	0	0	29	0	1
MCAC	13	13	23	30	25	27
NUMC	10	10	26	30	--	--
INDUSTRIAL	3	1	6	15	8	6
FM	0	0	0	14	1	0
CNF	0	0	24	30	26	26
HIGHLY_CONSTRAINED	7	7	22	30	23	22

Competition results – Not valid / timed out instances (4)

t =5

Tool	ACTS	CAgen	CAopt	KALI	MEDICI	pMEDICI
UNIFORM_BOOL	0	0	0	0	4	0
UNIFORM_ALL	5	4	13	13	14	13
MCA	7	2	29	26	30	27
BOOLC	0	0	4	29	0	0
MCAC	16	16	26	30	27	27
NUMC	18	16	29	30	--	--
INDUSTRIAL	4	3	6	14	8	6
FM	0	0	0	14	5	4
CNF	15	5	27	30	28	29
HIGHLY_CONSTRAINED	15	9	25	30	24	26

Competition results – Not valid / timed out instances (5)

t =6

Tool	ACTS	CAgen	CAopt	KALI	MEDICI	pMEDICI
UNIFORM_BOOL	0	0	0	4	10	4
UNIFORM_ALL	9	5	13	14	14	14
MCA	16	10	30	30	30	30
BOOLC	0	0	14	30	5	3
MCAC	21	20	30	30	27	28
NUMC	20	18	30	30	--	--
INDUSTRIAL	6	4	8	14	14	7
FM	2	0	0	14	6	5
CNF	23	19	28	30	28	29
HIGHLY_CONSTRAINED	21	17	26	30	26	26

Generation time ranking
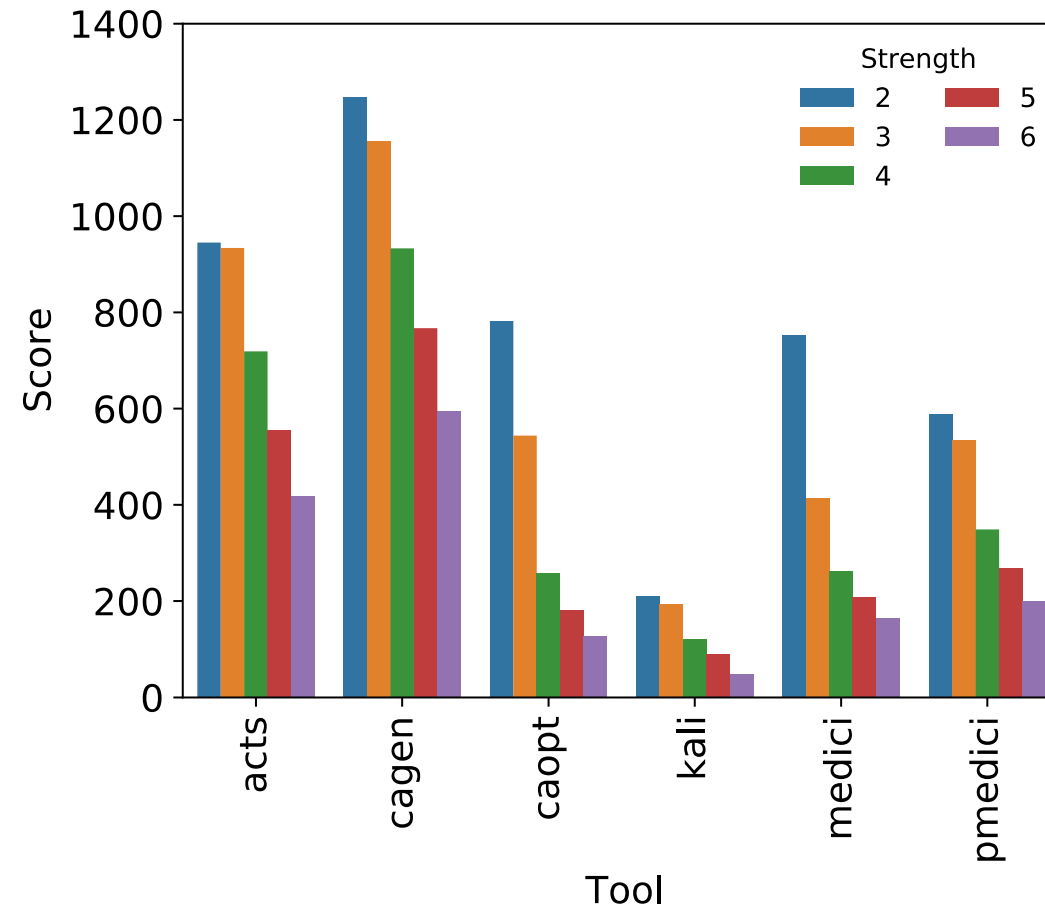
Test suite size ranking
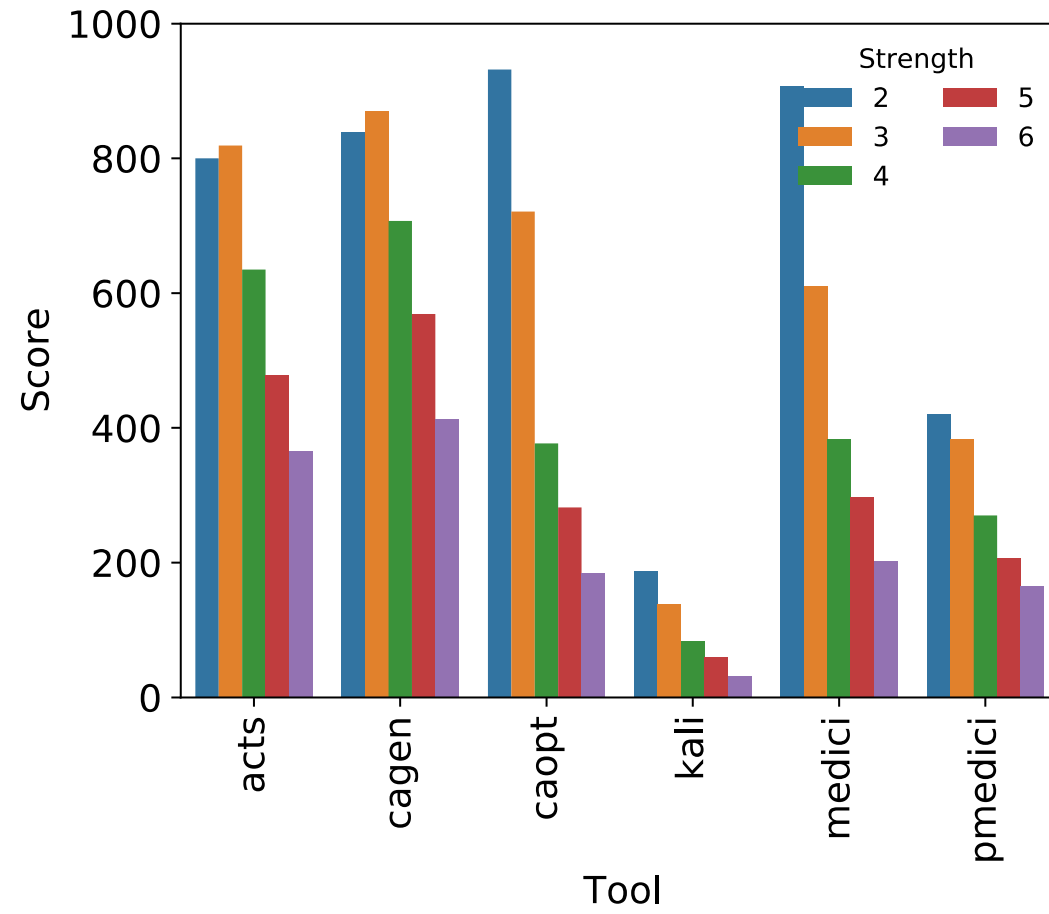
Competition results – Overall
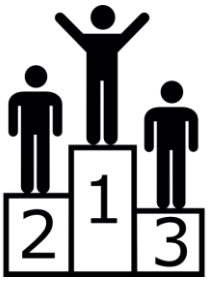
- CAgen (3666.0)
- ACTS (3056.5)
- CAopt (2104.0)
- MEDICI (1997.0)
- pMEDICI (1560.5)
- KALI (537.0)
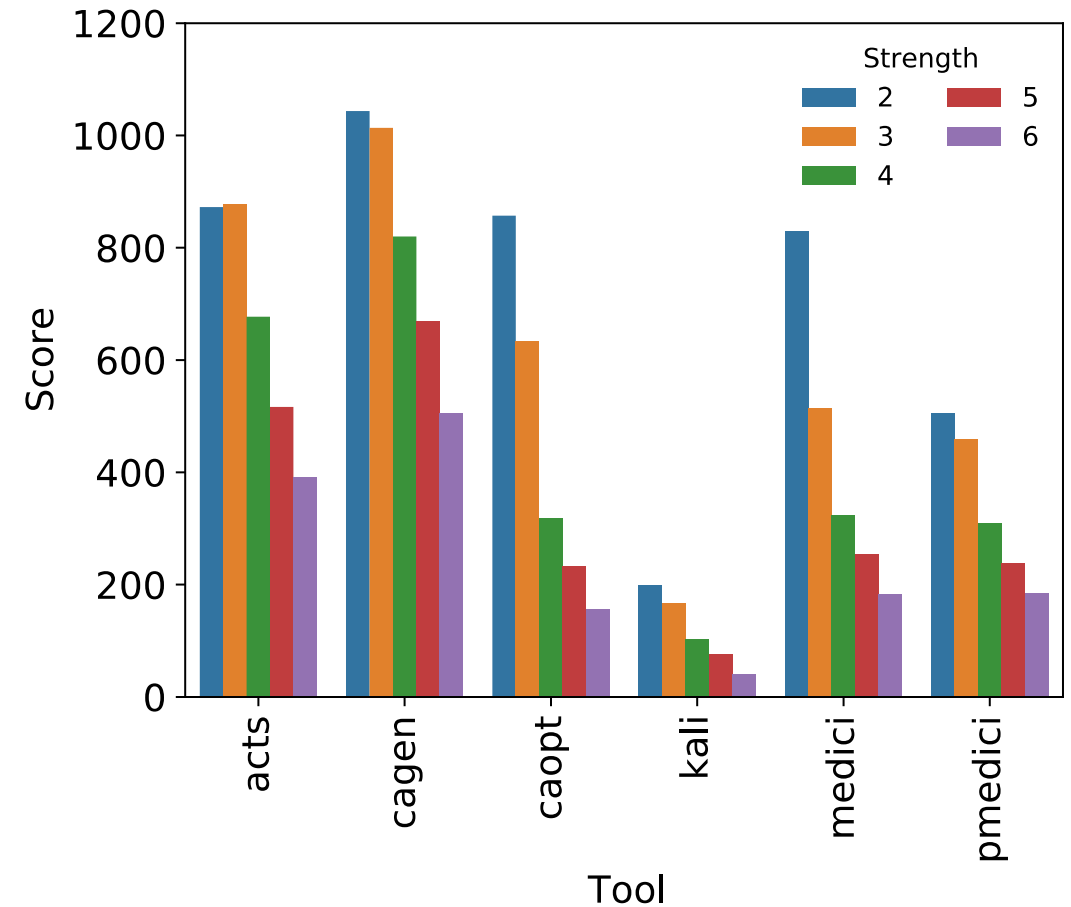
Competition results – UNIFORM_BOOLEAN
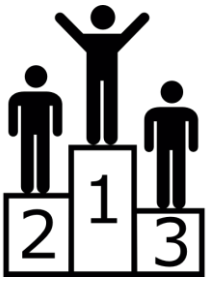
- CAgen (410.5)

- ACTS (363.0)

- CAopt (324.5)

- MEDICI (280.0)

- KALI (223.5)
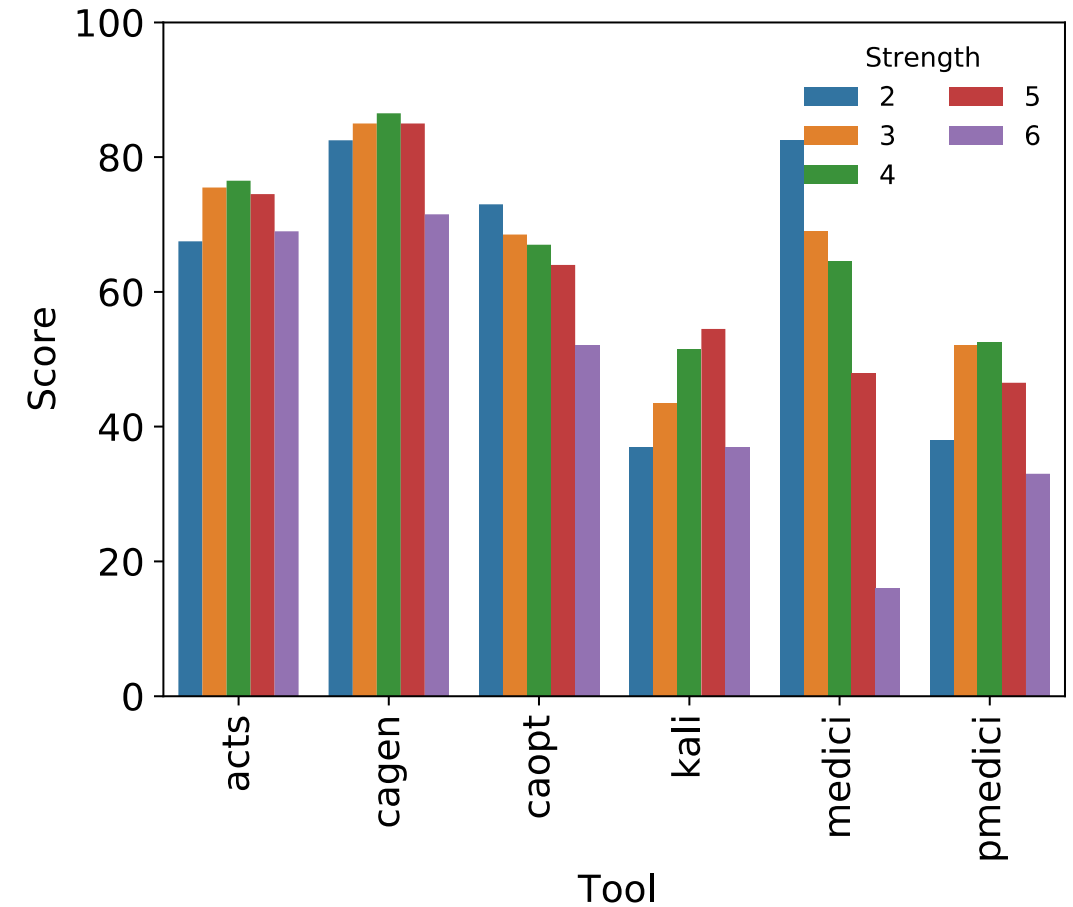
- pMEDICI (220.0)

Competition results – UNIFORM_ALL
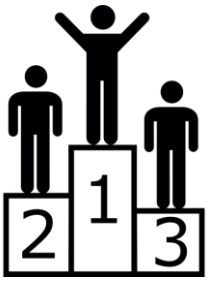
- CAgen (289.5)
- ACTS (235.5)
- CAopt (141.5)
- pMEDICI (131.0)
- MEDICI (126.0)
- KALI (104.5)
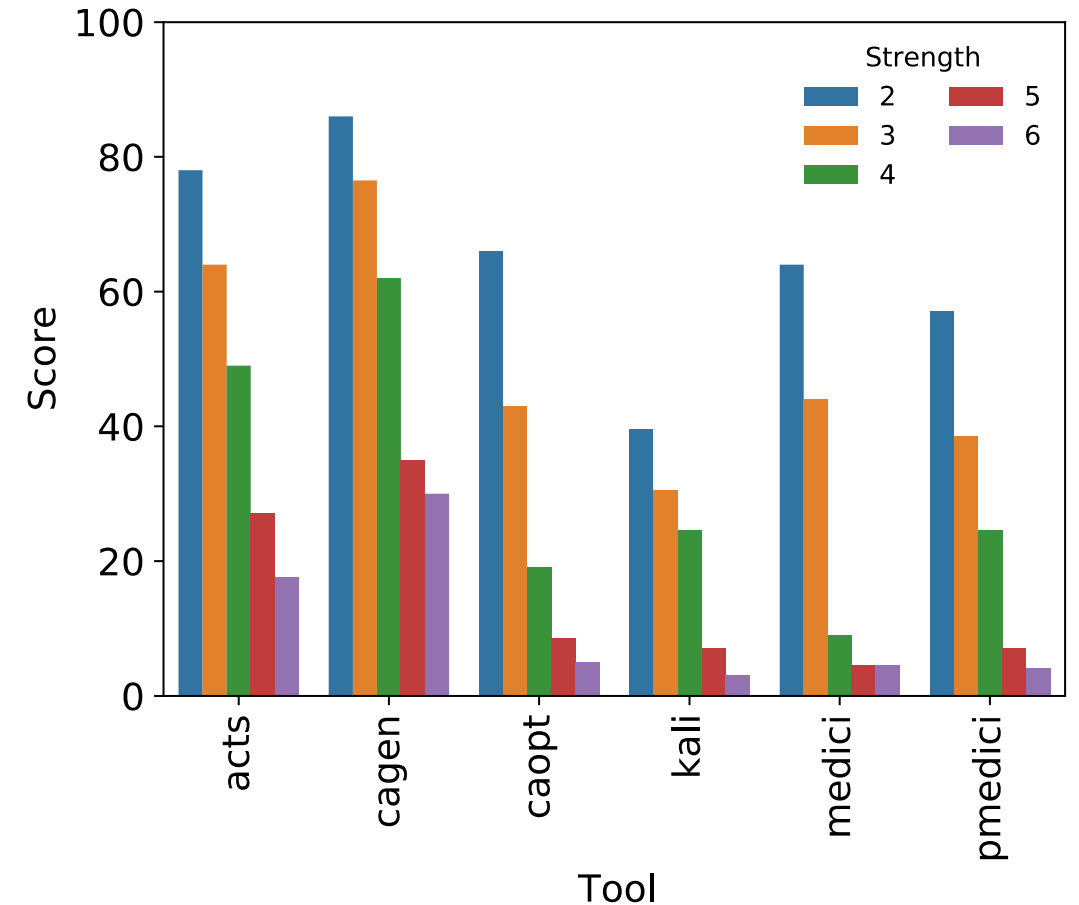
Competition results – MCA

- CAgen (591.5)
- ACTS (468.0)
- pMEDICI (232.0)
- CAopt (211.0)
- MEDICI (172.0)
- KALI (157.5)
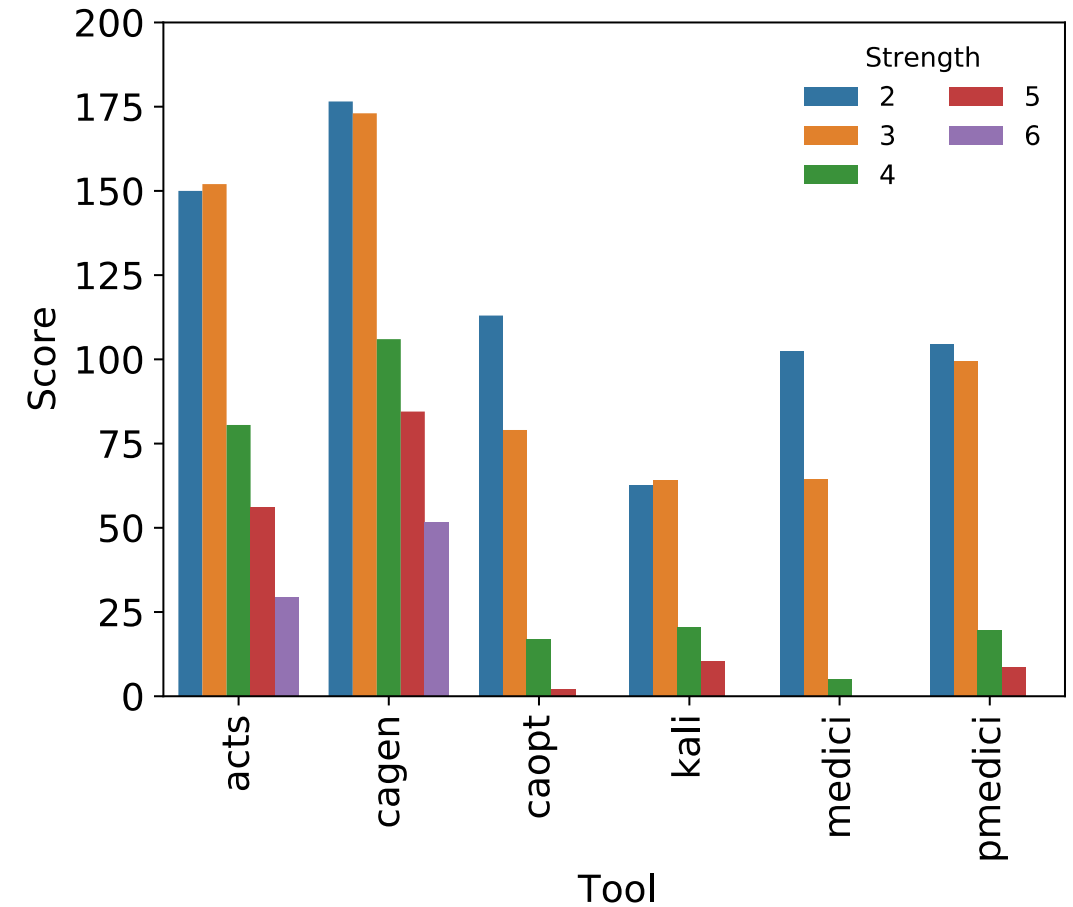
Competition results – BOOLC
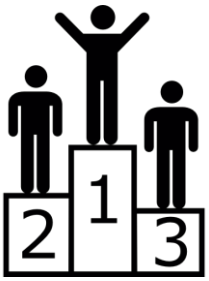
- CAgen (741.5)
- ACTS (627.5)
- MEDICI (600.5)
- pMEDICI (431.5)
- CAopt (380.0)
- KALI (61.5)
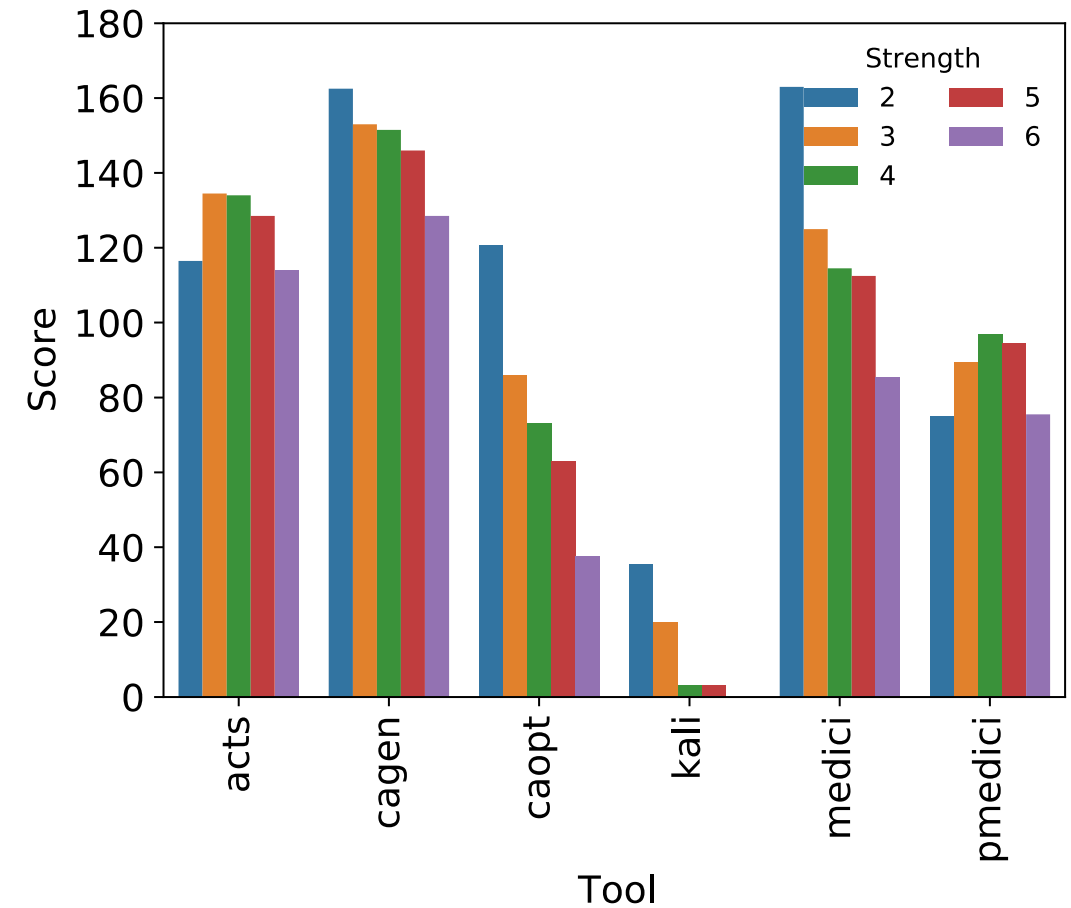
Competition results – MCAC
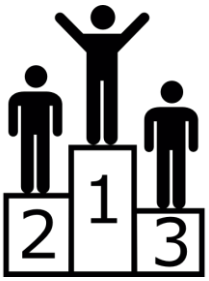
- CAgen (286.5)

- ACTS (263.0)

- MEDICI (178.5)

- CAopt (172.0)

- pMEDICI (100.0)

- KALI (9.5)
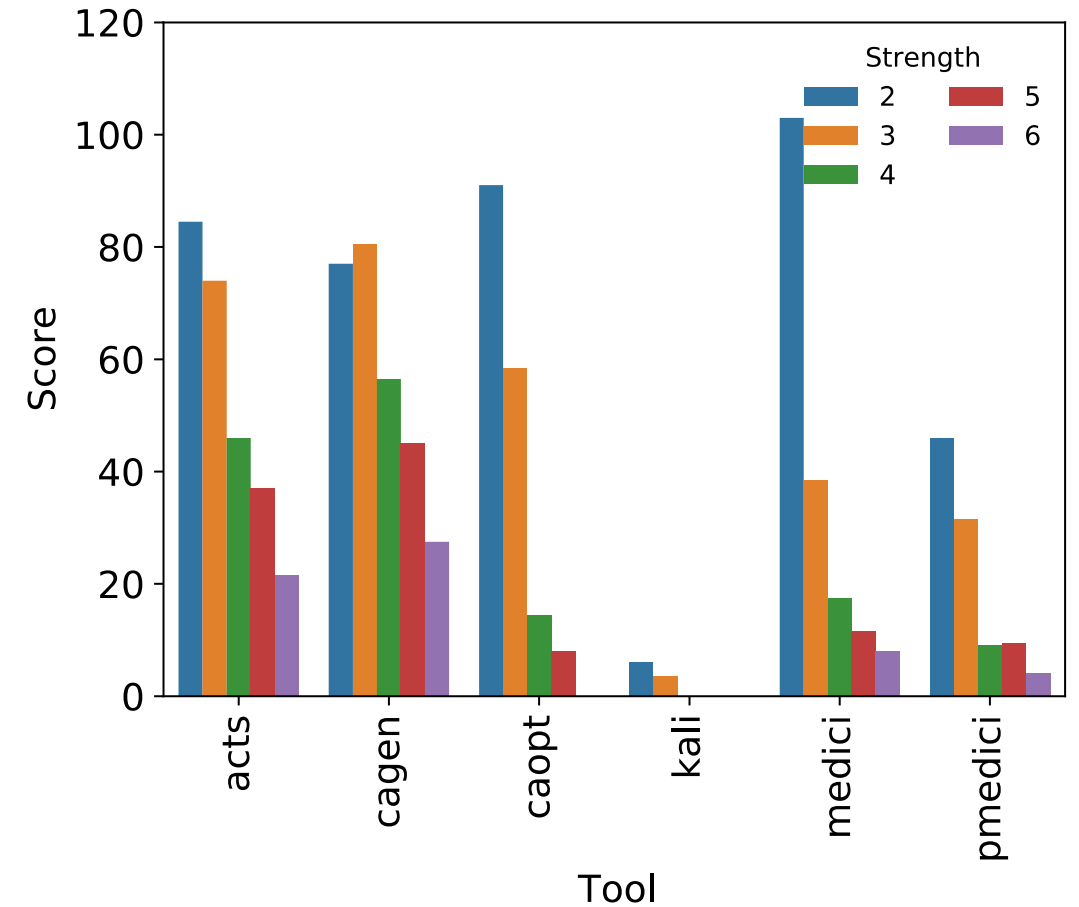
Competition results – NUMC
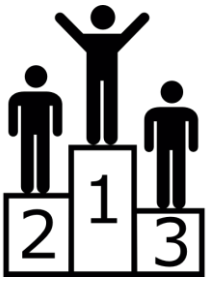
- Cagen (272.0)

- ACTS (216.0)

- CAopt (131.0)

- KALI (0.0)
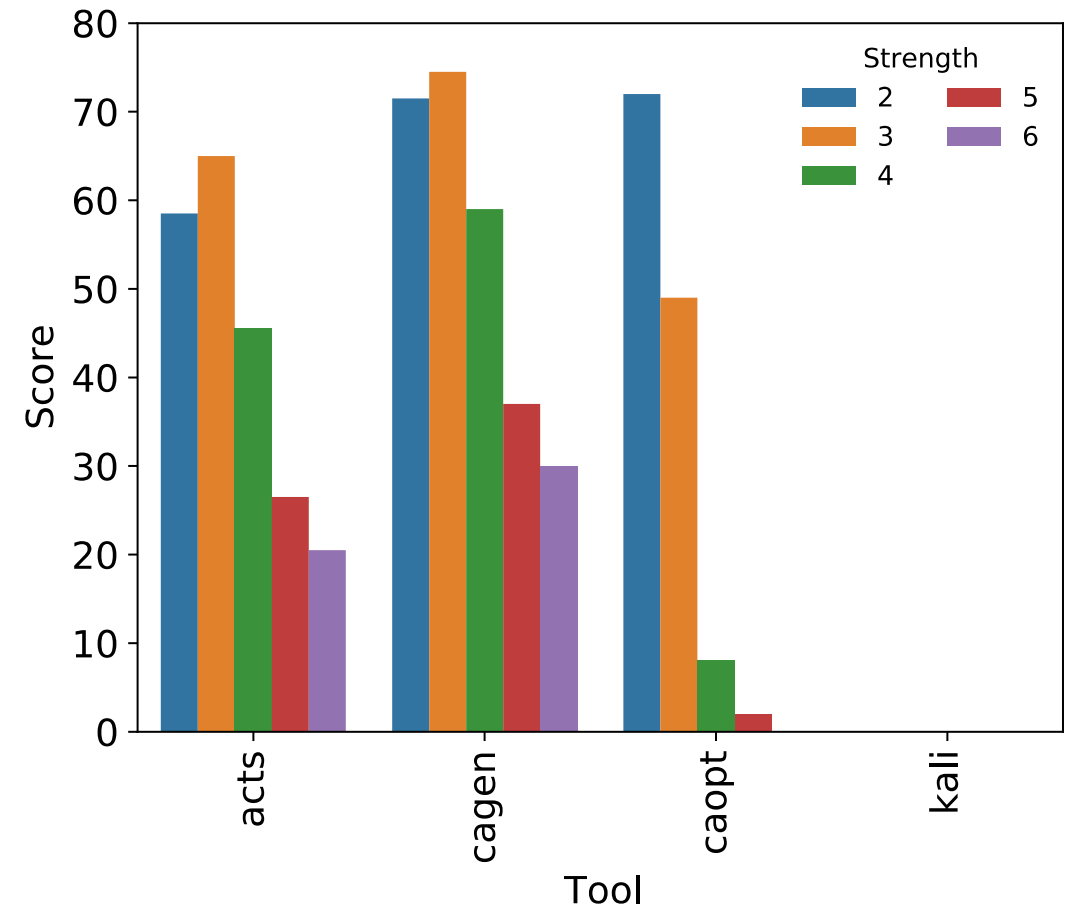
Competition results – INDUSTRIAL
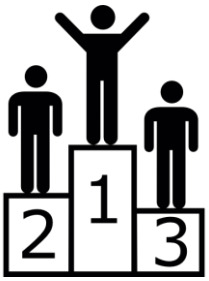
- CAgen (262.0)

- ACTS (217.5)

- CAopt (148.5)

- MEDICI (147.5)

- pMEDICI (140.5)

- KALI (8.5)
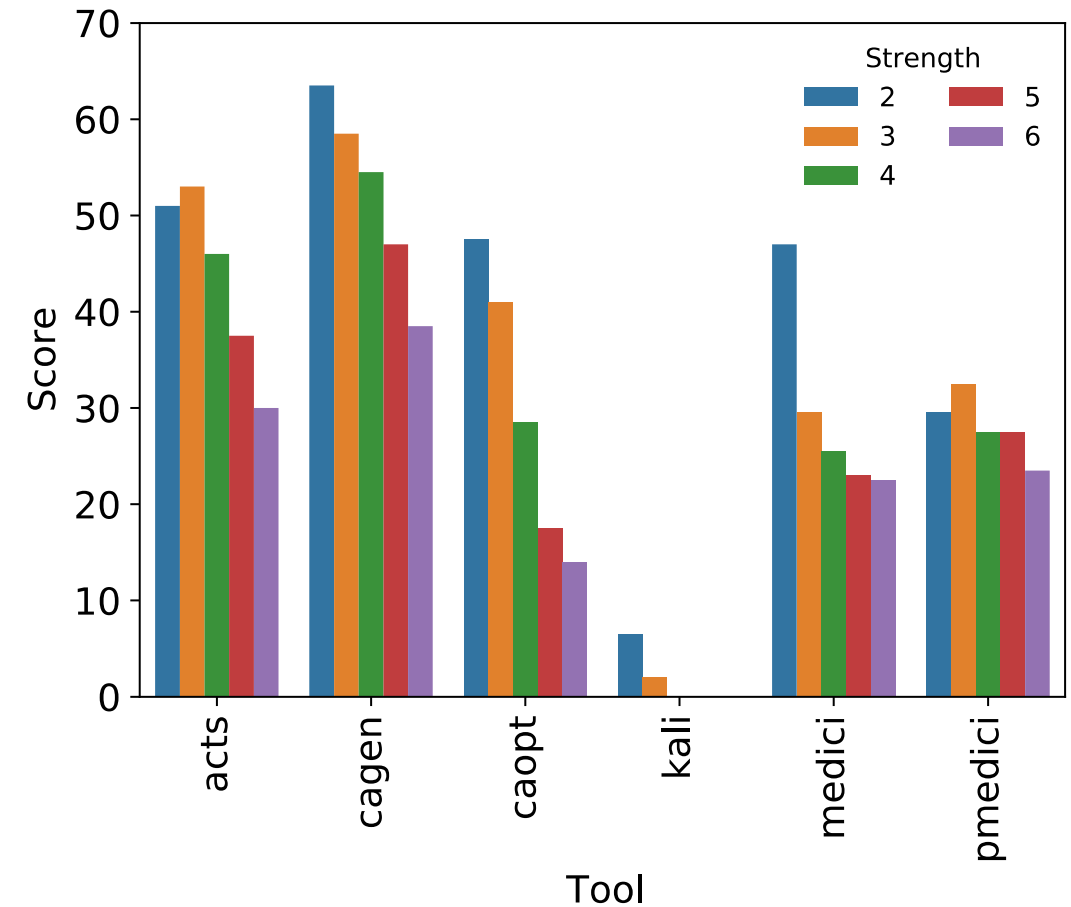
Competition results – FM
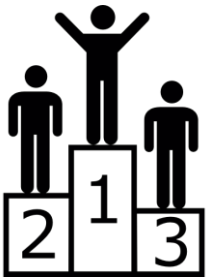
- CAgen (333.5)

- ACTS (271.0)

- CAopt (264.0)

- MEDICI (220.5)

- pMEDICI (169.5)

- KALI (11.5)
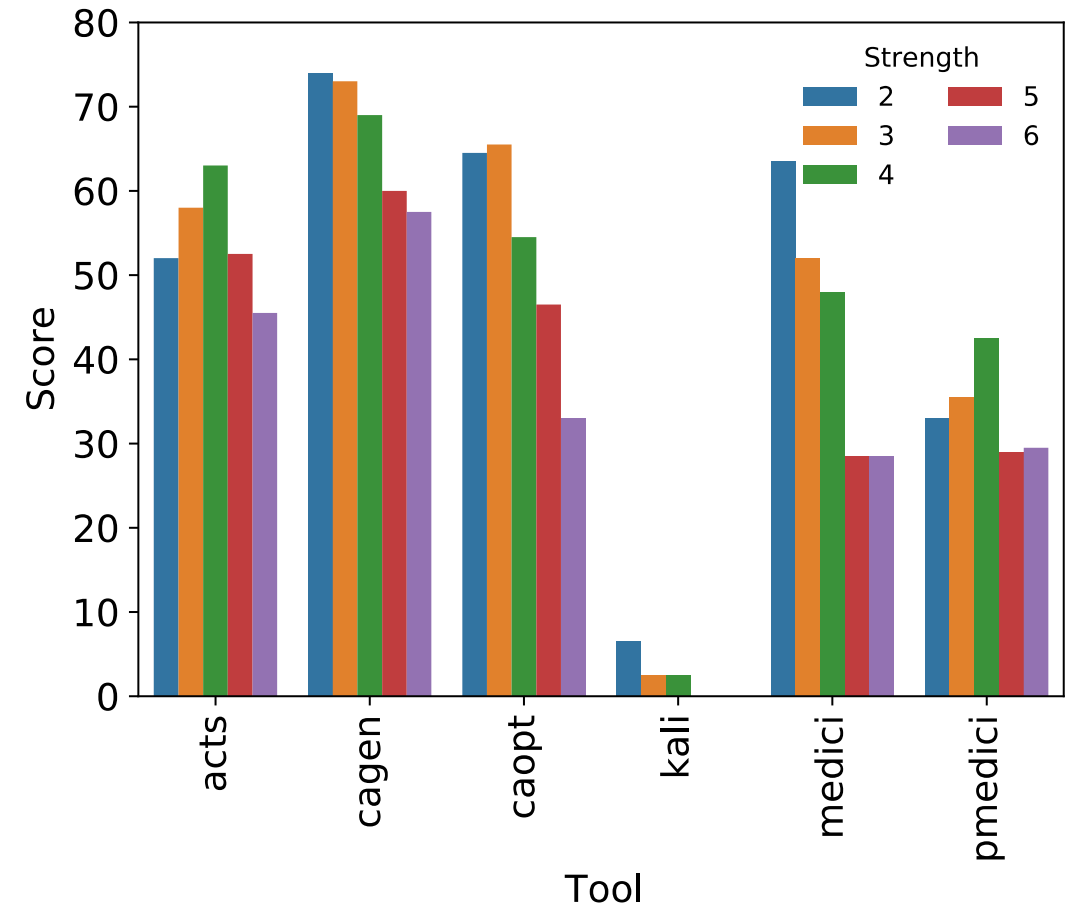
Competition results – CNF
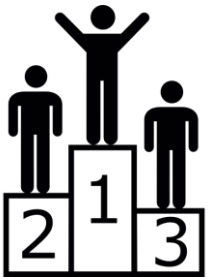
- CAgen (473.5)
- ACTS (364.0)
- CAopt (210.0)
- MEDICI (166.5)
- pMEDICI (133.0)
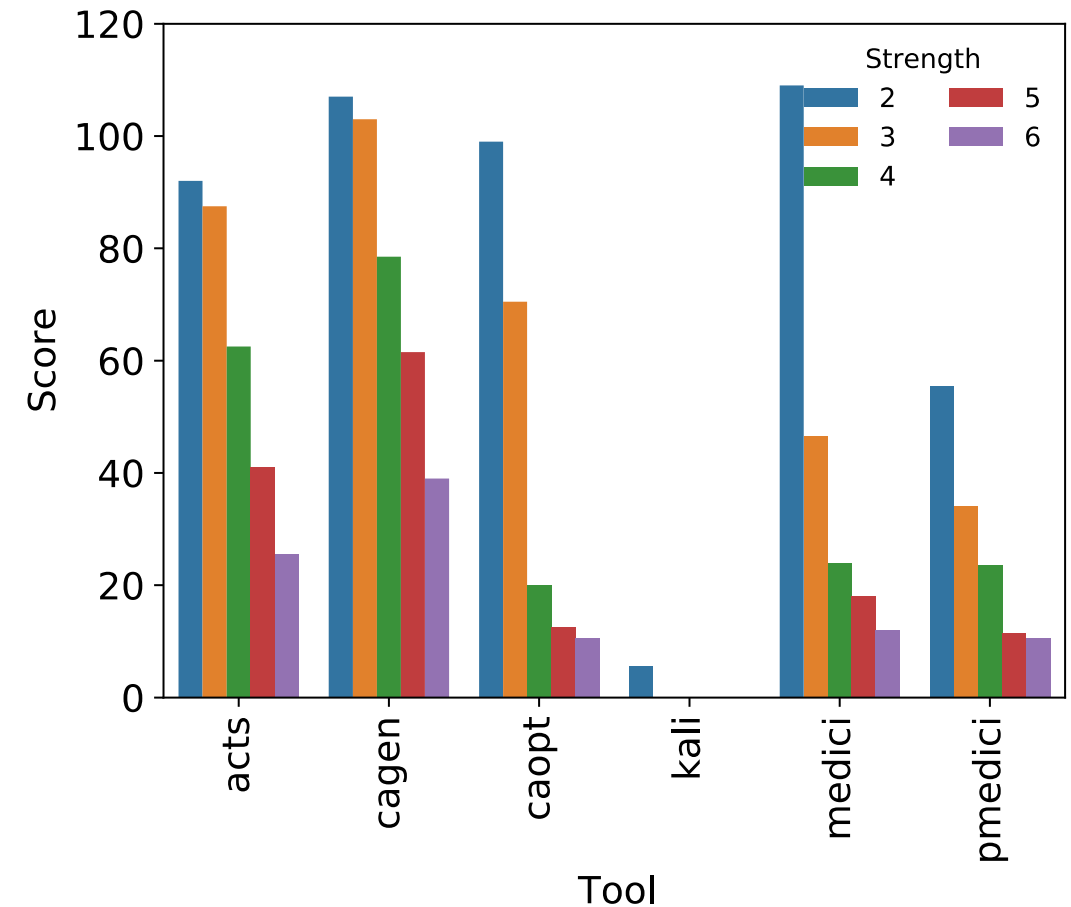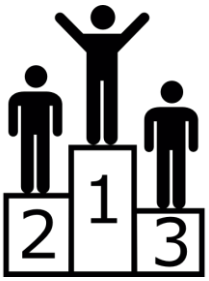- KALI (0.0)
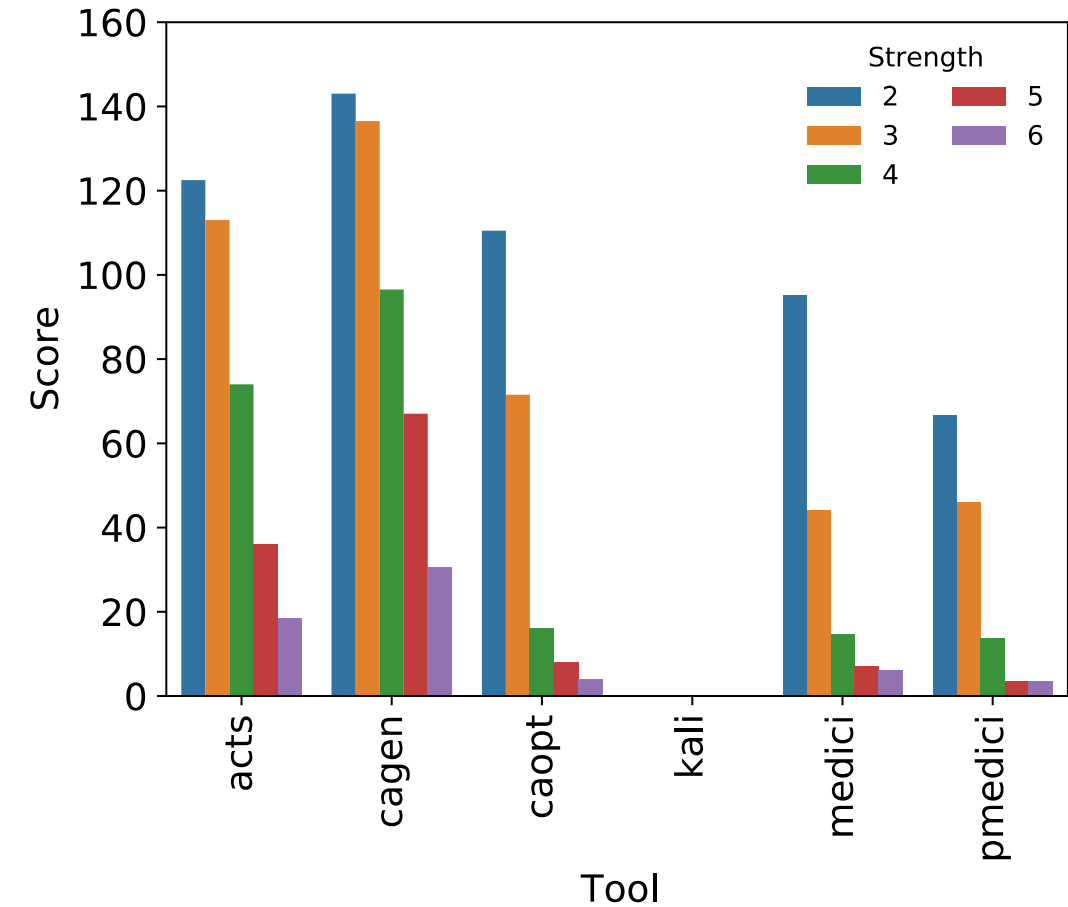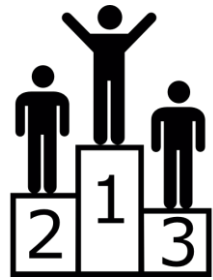
Tools summary

- <u>ACTS</u>: fast generation, MFT derivation causes timeouts for some instances of constraints

- <u>CAgen</u>: fastest tool tested, the same issue as ACTS with MFTs

- <u>CAopt</u>: the smallest number of timeouts for t=2, slower for higher strength

- <u>KALI</u>: a large number of invalid solutions, hard to judge

- <u>MEDICI</u>: the second smallest number of timeouts for t=2, slower for higher strength. In general slower than pMEDICI, but the test suites are smaller

- <u>pMEDICI</u>: a few invalid solutions (fewer than in the previous edition), often somewhat large CAs

FINAL CONSIDERATIONS

Final considerations (1)

- IPO remains the fastest general-purpose CA generation strategy, but a combination with post-optimization is advised

- More tools have been able to compete in all the tracks w.r.t. the first edition of the CT-Competition

- The tools producing invalid instances seem having problem during the randomization phase for filling *don't care* values

Final considerations (2)

- Using MDDs does not require precomputation, but seems to be difficult to use correctly (both by pMEDICI and MEDICI)

- CAopt (based on SAT solvers) seems very effective with constraint handling since it produces less timeouts for strength 2 where pretty much all timeouts should be due to constraint handling

NEXT CT-COMPETITION

Ideas for the next competition?

Thank you!