

# Assessment of benchmarks for abstract argumentation

Jean-Guy Mailly <sup>a,\*</sup> and Marco Maratea <sup>b</sup>

<sup>a</sup> LIPADE, Université de Paris, France

E-mail: [jean-guy.mailly@parisdescartes.fr](mailto:jean-guy.mailly@parisdescartes.fr)

<sup>b</sup> DIBRIS, Università di Genova, Italy

E-mail: [marco@dibris.unige.it](mailto:marco@dibris.unige.it)

**Abstract.** In this paper, we provide an overview of the benchmarks that have been recently employed in Abstract Argumentation. We first describe the benchmark suite from previous editions of the International Competition of Computational Models of Argumentation (ICCMA), and then briefly describe the benchmarks for non-Dung frameworks. This article is a contribution to the new Argument & Computation Community Resources (ACCR) corner.

Keywords: Computational argumentation, community resources, benchmarks

## 1. Introduction

Computational argumentation has been a topic of interest for decades, especially since Dung’s seminal paper on Abstract Argumentation Frameworks (AFs) [18]. It has applications in various domains such as decision making [4], automated negotiation [17], reasoning with inconsistent knowledge [8], legal reasoning [7], and multi-agent systems [26]. However, the main efforts towards the development of efficient computational approaches are quite recent. These efforts have been channeled by the organization of the first International Competition on Computational Models of Argumentation<sup>1</sup> (ICCMA) in 2015 [28]. This event allowed to compare the different approaches to solving the main reasoning tasks for an agent using mainly randomly generated AF benchmarks. The second edition of the competition (ICCMA 2017)<sup>2</sup> [20,21] introduced more semantics, and utilized for the first time a specific *call for benchmarks* for improving the quality and quantity of the benchmarks set, in particular regarding the non-random benchmarks. The third edition of the competition (ICCMA 2019)<sup>3</sup> has added a new kind of reasoning problem: dynamic argumentation, consisting in dealing with a sequence of updated versions of a previously solved instance of a “classical” reasoning problem for AFs.

---

\* Corresponding author. E-mail: [jean-guy.mailly@parisdescartes.fr](mailto:jean-guy.mailly@parisdescartes.fr).

<sup>1</sup><http://argumentationcompetition.org/>

<sup>2</sup><http://argumentationcompetition.org/2017/>

<sup>3</sup><https://www.iccma2019.dmi.unipg.it>

The regular organization of a solver competition is important for theoretical and practical reasons, as proved in a number of neighbor fields, such as SAT [5], ASP [11,22], and SMT.<sup>4</sup> In particular, it allows the comparison of different kinds of algorithms that can solve reasoning tasks. It highlights the recent improvements of solving techniques for argumentation, and assesses the benchmarks that should be used by the community. About benchmarks, we expect good benchmarks to present some features. Especially, to be relevant in the context of a competition, the related instances must not be too easy or too hard to solve. And, more importantly, they must be representative of real world applications as much as possible. Thus, users of argumentation applications are able to pick the best solver for dealing with their concrete problem instances; at least this is the case if the concrete instance is similar to the benchmarks that have been used in the competition. Indeed, the efficiency of methods for solving hard problems is generally highly depending on the structure of the problem instances to be solved [9].

In this paper, we focus our attention on benchmarks, by presenting an overview of such resources that have been recently employed in Abstract Argumentation. In particular, we first describe the benchmark suite from previous editions of ICCMA, followed by a presentation of the benchmarks that have been proposed for variants and extensions of the original AFs proposed by Dung.

## 2. ICCMA benchmarks

In this section we outline the benchmark domains employed in the first two ICCMA competitions.

### 2.1. ICCMA'15 benchmarks

ICCMA'15 introduced three new AF generators, called `GroundedGenerator`, `StableGenerator`, and `ScgGenerator` [15], each of them aiming to produce challenging AFs addressing certain aspects of computational difficulty, which are summarized below:

**GroundedGenerator.** This tool has been developed with the goal of producing AFs with large grounded extensions. It takes the number of arguments  $n$  and probability  $probAttacks$  as parameters, linearly orders the arguments and adds an attack from argument  $a$  to argument  $b$  in case  $a < b$  with probability  $probAttacks$ . It then adds random attacks between the arguments not yet connected and the graph component obtained in the first part.

**ScgGenerator.** This tool has been developed with the goal of producing AFs such that the graph features many Strongly Connected Components. It first partitions the arguments (the number of which is given by parameter  $n$ ) into  $nSCCs$  (also given as parameter) components, which are linearly ordered. Within each component, attacks between pairs of arguments are added with probability given by a parameter. Among arguments of different components, attacks are added with probability given by another parameter, but under the condition that the component of the attacking arguments is ranked lower with respect to the linear order on components than the component of the attacked argument.

**StableGenerator.** This tool has been developed with the goal of producing AFs with a large number of stable extensions. It first identifies a set of arguments to form an acyclic subgraph of the AF and, consequently, to contain the grounded extension. Among the other arguments, subsets are iteratively singled out to form stable extensions by attacking all other arguments. Besides the

---

<sup>4</sup><https://smt-comp.github.io/2019/index.html>

parameter  $n$  for the number of arguments, the tool generates AFs taking into account a number of parameters which determine minimum and maximum values for the number and size of stable extensions and for the size of grounded extensions.

## 2.2. ICCMA'17 benchmarks

Herewith we briefly describe the domains that were evaluated at ICCMA'17, on top of those of ICCMA'15:

**ABA2AF.** These are assumption-based argumentation (ABA) benchmarks translated to AFs. ABA problems are one of the prevalent forms of structured argumentation in which, differently from AFs, the internal structure of arguments is made explicit through derivations from a more basic structure [23,29].

**AdmBuster.** These are crafted benchmark examples for (strong) admissibility featuring a fixed structure composed of 4 sets of arguments and predetermined sets of attacks. Two “starting” and “terminal” sets are composed of only one element, one having only outgoing edges and the other only incoming edges. Given a parameter  $n$  of the generator, the two “intermediate” sets have cardinality  $n$ , and their attack relations are constructed in order to have only one complete labelling [12].

**AFBenchGen2.** This is a set of random AFs generators of three different graph classes, with a configurable number of arguments [14]. The three classes correspond to: (i) Erdős-Rényi [19], which selects attacks randomly; (ii) Watts-Strogatz [30], which aims for a *small-world* topology of networks being not completely random nor regular, and (iii) Barabasi-Albert [6] for large networks.

**Planning2AF.** These are AFs obtained from translating the well-known Blocksworld and Ferry planning domains. Each planning instance is first encoded as a propositional formula, by using the method in [27]; then, each clause is transformed into a material implication; and, finally, to each material implication the transformation in [31] is applied.

**SemBuster.** This is a crafted benchmark example for semi-stable semantics, with a fixed structure composed by 3 sets of arguments of equal cardinality, and predetermined sets of attacks. Attack relations are defined in a way that each instance has exactly  $n + 1$  (with  $n$  the number of arguments of each set) complete labellings that correspond also to preferred labellings, but only one among those corresponds to a semi-stable extension [13].

**Traffic.** These are graphs obtained from real world traffic networks data available at <https://transitfeeds.com/> expressed as AFs. Given a graph, the corresponding AF contains the same set of vertices as the graph, and the attack relation is defined as follows: Given an existing edge, and a probability for the attack of being symmetric, the generator decides whether there are both attacks, or randomly selects the attack.

More detailed descriptions for all such domains can be found in the ICCMA'15 and ICCMA'17 home pages at <http://argumentationcompetition.org>.

## 3. Other benchmarks

In this section, we briefly describe other argumentation benchmarks that have been proposed, and for some of them we point out some directions that could lead to interesting new benchmarks.

*ADF.* Abstract Dialectical Frameworks (ADFs) are one of the main extensions of AFs, where every argument is equipped with an acceptance condition. The benchmarks employed in some of the recent papers on ADFs (e.g., [10,25]) are generated starting from some of the most “practical” benchmarks of ICCMA’17, e.g., Traffic and Planning2AF, and randomly generating acceptance conditions. Some of these benchmarks can be found at the YADF system page (<http://www.dbai.tuwien.ac.at/proj/adf/yadf/>).

*Dynamic argumentation.* Argumentation is highly dynamic by nature. Indeed, debates are situations where agents add new attacks and arguments at each step of the process. However, computational approaches specifically designed for such dynamic situations have only been proposed recently. For different semantics and reasoning problems, the approach to generate benchmarks is the same: AFs from ICCMA’15 or ICCMA’17 are used, and updates are randomly generated (see, e.g., [1]). A similar approach is used to generate dynamic bipolar AFs [2] and dynamic extended AFs [3]. This generation approach allows to easily benefit from the existing ICCMA benchmarks. However, like any random-based benchmarks, they miss some connection with real world instances. A possible direction to obtain more meaningful dynamic AFs is the study of real life debates (e.g., debates coming from a democratic process, from negotiation situations, ...).

*Structured argumentation.* Obtaining an argumentation framework from a knowledge base is not a trivial problem. Indeed, even for some relatively small knowledge base, the number of arguments and attacks can be exponentially high with respect to the size of the knowledge base. We are aware of some efforts to efficiently generate argumentation frameworks from a logical knowledge base. For instance, [34] provides a method for generating AFs from existential rules knowledge bases; a data set and a generator are available online. A similar contribution for AFs based on Datalog $\pm$  has been proposed in [32], while [33] provides such a generator for argumentation hypergraphs (i.e., with collective attacks) based on Datalog $\pm$ .

*Assumption-based argumentation.* Assumption-based Argumentation (ABA) [16] is a well-known approach to structured argumentation, in which arguments are represented in a compact way in terms of graph-based derivations from a given rule-based deductive system over sentences, starting from assumptions, and checking for acceptability queries. Randomly generated ABA frameworks and queries used in a number of papers (e.g., [16,24] can be found at <http://robertcraven.org/proarg/experiments.html>, and correspond to those frameworks utilized for contributing to ICCMA’17.

## 4. Conclusions

In this article we have presented a first contribution to the new born Argument & Computation Community Resources (ACCR) corner, in terms of an assessment of benchmarks for Abstract Argumentation. Subsequent contributions to the corner are expected to be focused on one particular resource (benchmark, software tool, etc), cf. the editorial introducing the corner in this issue, possibly among those we have overviewed here.

## Acknowledgements

We would like to thank the Argument & Computation Editors-in-Chief, Pietro Baroni and Bart Verheij, for giving us the possibility to write this opening contribution to the new ACCR corner. We also would like to thank Martin Caminada for useful comments on a draft of this contribution.

## References

- [1] G. Alfano, S. Greco and F. Parisi, Efficient computation of extensions for dynamic abstract argumentation frameworks: An incremental approach, in: *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017)*, Melbourne, Australia, August 19–25, 2017, pp. 49–55. doi:[10.24963/ijcai.2017/8](https://doi.org/10.24963/ijcai.2017/8).
- [2] G. Alfano, S. Greco and F. Parisi, Computing stable and preferred extensions of dynamic bipolar argumentation frameworks, in: *Proceedings of the 1st Workshop on Advances in Argumentation in Artificial Intelligence Co-Located with XVI International Conference of the Italian Association for Artificial Intelligence, AI<sup>3</sup>@AI\*IA 2017*, Bari, Italy, November 16–17, 2017, pp. 28–42.
- [3] G. Alfano, S. Greco and F. Parisi, Computing extensions of dynamic abstract argumentation frameworks with second-order attacks, in: *Proceedings of the 22nd International Database Engineering & Applications Symposium (IDEAS 2018)*, Villa San Giovanni, Italy, June 18–20, 2018, pp. 183–192. doi:[10.1145/3216122.3216162](https://doi.org/10.1145/3216122.3216162).
- [4] L. Amgoud and H. Prade, Using arguments for making and explaining decisions, *Artif. Intell.* **173**(3–4) (2009), 413–436. doi:[10.1016/j.artint.2008.11.006](https://doi.org/10.1016/j.artint.2008.11.006).
- [5] T. Balyo, M.J.H. Heule and M. Järvisalo, SAT competition 2016: Recent developments, in: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017)*, San Francisco, California, USA, February 4–9, 2017, AAAI Press, 2017, pp. 5061–5063.
- [6] A.L. Barabasi and R. Albert, Emergence of scaling in random networks, *Science* **286** (1999), 509–512. doi:[10.1126/science.286.5439.509](https://doi.org/10.1126/science.286.5439.509).
- [7] T.J.M. Bench-Capon, H. Prakken and G. Sartor, Argumentation in legal reasoning, in: *Argumentation in Artificial Intelligence*, 2009, pp. 363–382. doi:[10.1007/978-0-387-98197-0\\_18](https://doi.org/10.1007/978-0-387-98197-0_18).
- [8] P. Besnard and A. Hunter, *Elements of Argumentation*, MIT Press, 2008, <http://mitpress.mit.edu/books/elements-argumentation>. ISBN 978-0-262-02643-7.
- [9] S. Bistarelli, F. Rossi and F. Santini, Not only size, but also shape counts: Abstract argumentation solvers are benchmark-sensitive, *J. Log. Comput.* **28**(1) (2018), 85–117. doi:[10.1093/logcom/exx031](https://doi.org/10.1093/logcom/exx031).
- [10] G. Brewka, M. Diller, G. Heissenberger, T. Linsbichler and S. Woltran, Solving advanced argumentation problems with answer-set programming, in: *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017)*, AAAI Press, 2017, pp. 1077–1083.
- [11] F. Calimeri, M. Gebser, M. Maratea and F. Ricca, Design and results of the fifth answer set programming competition, *Artif. Intell.* **231** (2016), 151–181. doi:[10.1016/j.artint.2015.09.008](https://doi.org/10.1016/j.artint.2015.09.008).
- [12] M. Caminada, Strong admissibility revisited, in: *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA 2014)*, S. Parsons, N. Oren, C. Reed and F. Cerutti, eds, Frontiers in Artificial Intelligence and Applications, Vol. 266, IOS Press, 2014, pp. 197–208.
- [13] M.W.A. Caminada and B. Verheij, On the existence of semi-stable extensions, in: *Proceedings of the 22nd Benelux Conference on Artificial Intelligence (BNAIC 2010)*, G. Danoy, M. Sredynski, R. Booth, B. Gateau, I. Jars and D. Khadraoui, eds, 2010, <http://bnaic2010.uni.lu/proceedings.html>.
- [14] F. Cerutti, M. Giacomin and M. Vallati, Generating structured argumentation frameworks: AFBenchGen2, in: *Proceedings of the 6th International Conference on Computational Models of Argument (COMMA 2016)*, P. Baroni, T.F. Gordon, T. Scheffler and M. Stede, eds, Frontiers in Artificial Intelligence and Applications, Vol. 287, IOS Press, 2016, pp. 467–468.
- [15] F. Cerutti, N. Oren, H. Strass, M. Thimm and M. Vallati, A benchmark framework for a computational argumentation competition, in: *Proceedings of the 5th International Conference on Computational Models of Argument (COMMA 2014)*, S. Parsons, N. Oren, C. Reed and F. Cerutti, eds, Frontiers in Artificial Intelligence and Applications, Vol. 266, IOS Press, 2014, pp. 459–460.
- [16] R. Craven and F. Toni, Argument graphs and assumption-based argumentation, *Artif. Intell.* **233** (2016), 1–59. doi:[10.1016/j.artint.2015.12.004](https://doi.org/10.1016/j.artint.2015.12.004).
- [17] Y. Dimopoulos, J. Mailly and P. Moraitis, Argumentation-based negotiation with incomplete opponent profiles, in: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2019)*, Montreal, QC, Canada, May 13–17, 2019, pp. 1252–1260.
- [18] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and  $n$ -person games, *Artif. Intell.* **77**(2) (1995), 321–358. doi:[10.1016/0004-3702\(94\)00041-X](https://doi.org/10.1016/0004-3702(94)00041-X).
- [19] P. Erdős and A. Rényi, On random graphs I, *Publicationes Mathematicae Debrecen* **6** (1959), 290–297.
- [20] S.A. Gaggl, T. Linsbichler, M. Maratea and S. Woltran, Introducing the second international competition on computational models of argumentation, in: *Proceedings of the First International Workshop on Systems and Algorithms for Formal Argumentation (SAFA 2016)*, M. Thimm, F. Cerutti, H. Strass and M. Vallati, eds, CEUR Workshop Proceedings, Vol. 1672, CEUR-WS.org, 2016, pp. 4–9.
- [21] S.A. Gaggl, T. Linsbichler, M. Maratea and S. Woltran, Summary report of the second international competition on computational models of argumentation, *AI Magazine* **39**(4) (2019), 77–79. doi:[10.1609/aimag.v39i4.2781](https://doi.org/10.1609/aimag.v39i4.2781).

- [22] M. Gebser, M. Maratea and F. Ricca, The sixth answer set programming competition, *Journal of Artificial Intelligence Research* **60** (2017), 41–95. doi:[10.1613/jair.5373](https://doi.org/10.1613/jair.5373).
- [23] T. Lehtonen, J.P. Wallner and M. Jarvisalo, From structured to abstract argumentation: Assumption-based acceptance via AF reasoning, in: *Proceedings of the 14th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2017)*, A. Antonucci, L. Cholvy and O. Papini, eds, Lecture Notes in Computer Science, Vol. 10369, Springer, 2017, pp. 57–68. doi:[10.1007/978-3-319-61581-3\\_6](https://doi.org/10.1007/978-3-319-61581-3_6).
- [24] T. Lehtonen, J.P. Wallner and M. Jarvisalo, Reasoning over assumption-based argumentation frameworks via direct answer set programming encodings, in: *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019)*, Honolulu, Hawaii, USA, Jan 27–Feb 1, AAAI Press, 2019.
- [25] T. Linsbichler, M. Maratea, A. Niskanen, J.P. Wallner and S. Woltran, Novel algorithms for abstract dialectical frameworks based on complexity analysis of subclasses and SAT solving, in: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI 2018)*, J. Lang, ed., ijcai.org, 2018, pp. 1905–1911. doi:[10.24963/ijcai.2018/263](https://doi.org/10.24963/ijcai.2018/263).
- [26] P. McBurney, S. Parsons and I. Rahwan, *8th International Workshop on Argumentation in Multi-Agent Systems (ArgMAS 2011)*, Taipei, Taiwan, May 3, 2011, Revised Selected Papers. Lecture Notes in Computer Science, Vol. 7543, Springer, 2012. ISBN 978-3-642-33151-0. doi:[10.1007/978-3-642-33152-7](https://doi.org/10.1007/978-3-642-33152-7).
- [27] A. Sideris and Y. Dimopoulos, Constraint propagation in propositional planning, in: *Proceedings of the 20th International Conference on Automated Planning and Scheduling (ICAPS 2010)*, R.I. Brafman, H. Geffner, J. Hoffmann and H.A. Kautz, eds, AAAI, 2010, pp. 153–160.
- [28] M. Thimm and S. Villata, The first international competition on computational models of argumentation: Results and analysis, *Artif. Intell.* **252** (2017), 267–294. doi:[10.1016/j.artint.2017.08.006](https://doi.org/10.1016/j.artint.2017.08.006).
- [29] F. Toni, A tutorial on assumption-based argumentation, *Argument & Computation* **5**(1) (2014), 89–117. doi:[10.1080/19462166.2013.869878](https://doi.org/10.1080/19462166.2013.869878).
- [30] D.J. Watts and S.H. Strogatz, Collective dynamics of “small-world” networks, *Nature* **393** (1998), 440–442. doi:[10.1038/30918](https://doi.org/10.1038/30918).
- [31] A.Z. Wyner, T.J.M. Bench-Capon, P.E. Dunne and F. Cerutti, Senses of ‘argument’ in instantiated argumentation frameworks, *Argument & Computation* **6**(1) (2015), 50–72. doi:[10.1080/19462166.2014.1002535](https://doi.org/10.1080/19462166.2014.1002535).
- [32] B. Yun, M. Croitoru, S. Vesic and P. Bisquert, DAGGER: Datalog+/- argumentation graph generator, in: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*, Stockholm, Sweden, July 10–15, 2018, pp. 1841–1843.
- [33] B. Yun, M. Croitoru, S. Vesic and P. Bisquert, NAKED:  $n$ -ary graphs from knowledge bases expressed in datalog $\pm$ , in: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2019)*, Montreal, QC, Canada, May 13–17, 2019, pp. 2390–2392.
- [34] B. Yun, S. Vesic and M. Croitoru, Toward a more efficient generation of structured argumentation graphs, in: *Proceedings of the 7th International Conference on Computational Models of Argument (COMMA 2018)*, Warsaw, Poland, 12–14 September 2018, Frontiers in Artificial Intelligence and Applications, Vol. 305, IOS Press, 2018, pp. 205–212.