

Constrained Derivation in Assumption-Based Argumentation

Giovanni Buraglio¹, Wolfgang Dvořák¹, Anna Rapberger², Stefan Woltran¹

¹TU Wien, Institute of Logic and Computation, Austria

²Imperial College London, Department of Computing, United Kingdom

{giovanni.buraglio,wolfgang.dvorak,stefan.woltran}@tuwien.ac.at, a.rapberger@imperial.ac.uk

Abstract. Structured argumentation formalisms provide a rich framework to formalise and reason over situations where contradicting information is present. However, in most formalisms the integral step of constructing all possible arguments is performed in an unconstrained way and is thus not under direct control of the user. This can hinder a solid analysis of the behaviour of the system and makes explanations for the results difficult to obtain. In this work, we introduce a general approach that allows constraining the derivation of arguments for assumption-based argumentation.

1 Introduction

Assumption-based argumentation (ABA) [2, 1, 6] is a well-studied formalism in the realm of structured argumentation with applications ranging from medical reasoning and decision-making to eXplainable AI [5, 9, 8]. Argumentative reasoning is hereby performed by instantiating ABA frameworks (ABAF) representing debates through (structured) arguments and an attack relation among them. Arguments are built as forward derivations supported by defeasible sentences called *assumptions*, using (strict) inference rules from the underlying knowledge base. Accordingly, attacks between arguments encode a consistency check among the assumptions that support them. As already noticed by Modgil and Prakken [18], assumption-based argumentation leaves the “set of inference rules unspecified” in the sense that rules are treated equally and no distinction can be made among them. However, in some domains of application, rules might be distinguished on the basis of their function. Such situations can be found, for instance, in the area of normative reasoning. There it may become relevant to tell apart rules that produce obligations and permissions on the one hand from those that produce institutional facts on the other, based on Searle’s famous distinction between regulative and constitutive norms [23]. To prevent instances of deontic paradoxes and fallacious conclusions, the combination of rules is subject to certain restrictions [15, 13]. In the context of multi-agent systems [26, 21], an agent’s frame of reference may differ from that of others, giving rise to individual rule sets for each agent. Another example that requires the separation of rules is the necessity to express a qualitative distinction between them. In order to account for such situations, several argumentation formalisms such as ASPIC+ separate strict and defeasible inference rules [20, 16, 3, 17].

Let us consider the following illustrative example from the domain of normative reasoning.

Example 1 (adapted from [19, Example 3]). *Our protagonist Alice has been accepted to a study program with payment obligations. Every student whose application has been accepted counts as*

eligible student (constitutive norm). Moreover, every eligible student must pay her tuition fee (regulative norm) and every student who pays their tuition fee counts as a self-funding student (constitutive norm). We can derive that Alice must pay her tuition fee (since she is an eligible student), and hence she counts as a self-funding student. However, Alice has furthermore received a study grant which means that she is not a self-funding student after all. Hence we derive a counter-intuitive conflict, deducing Alice to be both self-funding and have received a grant.

In the above example, we end up fallaciously deducing a contradiction from our assumptions. The underlying issue is that the application of constitutive rules after regulative ones may produce fallacious conclusions, called institutional wishful thinking. The undesired situation in Example 1 could be circumvented by preventing the application of the rule “*tuition fee* \rightarrow *self-funding student*” after the rule “*eligible student* \rightarrow *tuition fee*”. In the context of formal argumentation, similar issues have been addressed in recent works, based on an ASPIC-like formalism [24, 19, 14, 13]. Standard ABA is, however, not expressive enough to account for such a qualitative distinction among inference rules. Consequently, it is not possible to constraint the rule combinations on the basis of their kind.

In this work, we propose first steps in order to close this gap. In particular, we (a) extend the ABA formalism with pairwise disjoint sets of rules in order to take into account qualitative differences among them; (b) equip this extension of ABA with formal constraints (called *derivation graphs*) that regulate its deductive machinery, by encoding applicability conditions for inference rules; (c) present two examples where such devices allow to avoid fallacious or paradoxical conclusions; and (d) provide some preliminary results. On the one hand, we examine the definition of constraints as pre-processing operations on the underlying knowledge base, on the other hand, we present a prototype encoding of our formalism in Answer Set Programming (ASP).

2 Constrained ABA with Multiple Rule-Sets

In this section, we introduce our new formalism and discuss preliminary results.

2.1 Background

In order to introduce our formalism, we first need to recall some preliminaries for assumption-based argumentation. In ABA, frameworks representing debates are built up from a rule-based knowledge base and defined in the following way:

Definition 1. *An ABA framework (ABAF) is a tuple $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{})$ where: (i) \mathcal{L} and \mathcal{R} form together a deductive system and are respectively a set of sentences in a language and a set of inference rules; (ii) $\mathcal{A} \subseteq \mathcal{L}$ is a non-empty set of sentences called assumptions; (iii) $\bar{}$ is a total mapping from \mathcal{A} into \mathcal{L} , where \bar{a} is said to be the contrary of a , for each $a \in \mathcal{A}$.*

Following [4], we write rules as $r : \phi \leftarrow \phi_1, \dots, \phi_m$ and we say that ϕ is the head of the rule and $\{\phi_1, \dots, \phi_m\}$ is its body, formally $head(r) = \phi$ and $body(r) = \{\phi_1, \dots, \phi_m\}$. For a set of rules R , we use $head(R)$ to indicate the set of atoms which are head of the rules contained in it. We consider here the finite flat version of ABAF, i.e. frameworks where \mathcal{L} and \mathcal{R} are finite and assumptions do not occur as conclusions of inference rules: there is no $r \in \mathcal{R}$ and $a \in \mathcal{A}$ for which $a = head(r)$. Arguments of an ABAF are based on proof-trees, constructed by forward-derivation from leaf-nodes to the root:

Definition 2 (deduction). *Let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \bar{})$ be an ABAF. A deduction for $p \in \mathcal{L}$ supported by $S \subseteq \mathcal{L}$ and $R \subseteq \mathcal{R}$, denoted $S \vdash^R p$ (or simply $S \vdash p$), is a finite rooted tree t with: i) a*

labelling function that assigns each vertex of t an element from $\mathcal{L} \cup \{\top\}$ such that the root is labelled by p and leaves are labelled by either \top or sentences in S ; ii) a surjective mapping m from the set of internal nodes of t onto rules R satisfying, for each vertex v , that the label of v is the head of the rule $m(v)$ and the children of v are (one-to-one) labelled with the elements of the body of $m(v)$.

In ABA, the attack relation is defined over sets of assumptions.

Definition 3 (attack). *Let $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \neg)$ be an ABAF, let $S, T \subseteq \mathcal{A}$ be two sets of assumptions. S attacks T ($S \rightarrow T$) iff there is a set $S' \subseteq S$ such that $S' \vdash \bar{a}$ for some $a \in T$.*

Semantics can be defined then in the usual way. Here, we recall only admissible semantics.

Definition 4 (semantics). *Given ABAF $D = (\mathcal{L}, \mathcal{R}, \mathcal{A}, \neg)$ and $S, T \in \mathcal{A}$. The set S is conflict-free iff it does not attack itself; and admissible iff it is conflict-free and $T \rightarrow S$ implies $S \rightarrow T$.*

2.2 ABA Frameworks with Multiple Rule-Sets

We can now define ABAFs with multiple rule-sets and derivation graphs. Jointly, these enable to trace rule kinds along with some constraint on their combination. We consider only frameworks where rule-sets are pairwise disjoint. Further, it is often desirable to evaluate scenarios where the same atom cannot be derived by rules of different kinds. For this, we focus on the class of *separated n -ABAFs*, for which heads of rules in different rule-sets are pairwise disjoint.

Definition 5 (*n -rule-sets ABA*). *A n -rule-sets ABAF (n -ABAF) is a tuple $D = (\mathcal{L}, \{\mathcal{R}_i \mid 1 \leq i \leq n\}, \mathcal{A}, \neg)$ such that $(\mathcal{L}, \bigcup_{i=1}^n \mathcal{R}_i, \mathcal{A}, \neg)$ is an ABAF. Moreover, we call D separated whenever $\text{head}(\mathcal{R}_i) \cap \text{head}(\mathcal{R}_j) = \emptyset$ for all i, j with $i \neq j$.*

As mentioned earlier, one might want to represent some constraint on rules combinations on $\bigcup_{i \leq n} \mathcal{R}_i$ depending on the particular application domain. Inspired by input/output combinations presented in [24], we introduce the more expressive concept of derivation graph to formalise combination constraints:

Definition 6 (derivation graph). *Let $D = (\mathcal{L}, \{\mathcal{R}_i \mid 1 \leq i \leq n\}, \mathcal{A}, \neg)$ be an n -ABAF. A derivation graph $G = (V, E)$ for D is a directed graph with $|V| \geq n + 1$ such that V contains: (i) a distinct vertex s (called “starting node”) with no incoming edges; (ii) at least one vertex r_i for each \mathcal{R}_i (called “rule-node” for \mathcal{R}_i) such that there is a surjective mapping from rule-nodes onto the set of rule-sets $\{\mathcal{R}_i \mid 1 \leq i \leq n\}$.*

The outcome of the constraint encoded by some derivation graph is a limitation on the possibility of rules chaining. This affects the derivation process from the underlying deductive system. In particular, the idea consists in allowing only those sequential combinations of rules for which there is a path within the derivation graph. As a result, we extend the usual notion of deduction presented in ABA to accommodate this additional requirement.

Definition 7. *Let $D = (\mathcal{L}, \{\mathcal{R}_i \mid 1 \leq i \leq n\}, \mathcal{A}, \neg)$ be an n -ABAF and let $G = (V, E)$ be a derivation graph for D . A G -deduction for $p \in \mathcal{L}$ supported by $S \subseteq \mathcal{L}$ and $R = \bigcup_{i=1}^n \mathcal{R}_i$ with $R_i \subseteq \mathcal{R}_i$, denoted $S \vdash_G^R a$ (or simply $S \vdash_G a$), is a deduction t with a surjective mapping that maps every v node of t to a rule node w in G such that i) v corresponds to a rule that is in the rule set of w and ii) for each leaf-to-root-path $v_0 \dots v_k$ in t , the corresponding series of nodes w_0, \dots, w_k in G form a path in G with $w_0 = s$.*

Notions of G -attack and G -semantics can be easily adapted from the standard ones accordingly. To show our new adaption at work, let us revisit our introductory example.

Example 2. Consider again Example 1. We construct a 2-ABAF $D = (\mathcal{L}, \mathcal{R}_1, \mathcal{R}_2, \mathcal{A}, \bar{\ })$ where \mathcal{R}_1 and \mathcal{R}_2 contain our constitutive and regulative rules, respectively. We assume the language \mathcal{L} to contain a modality \mathcal{O} where $\mathcal{O}p$ stands for “it is obligatory that p ”. We let $\mathcal{L} = \{a := \text{accepted_application}(x), b := \text{received_grant}(x), p := \text{eligible_student}(x), q := \mathcal{O}(\text{pay_fee}(x)), s := \text{self_funding_student}(x)\}$; $\mathcal{A} = \{a, b\}$; $\mathcal{R}_1 = \{p \leftarrow a, s \leftarrow q\}$ and $\mathcal{R}_2 = \{q \leftarrow p\}$. Moreover, $s = \bar{b}$. One can build the following conflicting derivations:

$$\{b\} \vdash \{\} b \quad \{a\} \vdash \{p \leftarrow a, q \leftarrow p, \bar{b} \leftarrow q\} \bar{b}$$

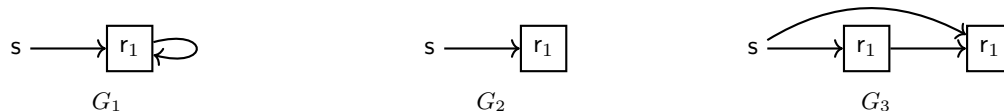
We conclude that $\{a\}$ attacks $\{b\}$. However, there should not be any conflict between a student being accepted and receiving a grant. By prohibiting the application of constitutive rules in the scope of regulative ones, \bar{b} is no longer derived from the assumption a . Indeed, by taking any derivation graph G such that $(r_2, r_1) \notin E(G)$, we get $\{a\} \not\vdash_G^{\{p \leftarrow a, q \leftarrow p, \bar{b} \leftarrow q\}} \bar{b}$. Therefore, $\{a\}$ does not G -attack $\{b\}$ and the assumption set $\{a, b\}$ is an extension under any G -semantics.

Next, let us consider an instance of Forrester’s paradox [10]. In Standard Deontic Logic [25], Forrester’s paradox, also known in the literature as “gentle murderer paradox”, follows from the statements A: “One should not (under the law) commit murder” and B: “if someone commits murder, then they should do it gently”. Moreover, B implies C: “if someone should commit murder gently, then they should commit murder”. Under the assumption that D: “someone commits murder”, this eventually creates a paradoxical situation whereby it is obligatory to commit and not to commit murder at the same time. Therefore, it is contradictory to assume that a law exists and someone violates it. In the following, we show that such an undesired outcome can be avoided by imposing some constraint by means of a derivation graph G .

Example 3. Take a non-flat 1-ABAF $D = (\mathcal{L}, \mathcal{R}_1, \mathcal{A}, \bar{\ })$ such that $\mathcal{L} = \{a, b, p\}$, $\mathcal{A} = \{a, b\}$ where $a := \text{murder}(x)$ and $b := \mathcal{O}(\neg \text{murder}(x))$ and $\mathcal{R}_1 = \{p \leftarrow a, \bar{b} \leftarrow p, b \leftarrow \}$ where $p := \mathcal{O}(\text{murder_gently}(x))$. Then one could build the arguments:

$$\{\} \vdash \{b \leftarrow \} b \quad \{a\} \vdash \{\} a \quad \{a\} \vdash \{p \leftarrow a\} p \quad \{a\} \vdash \{p \leftarrow a, \bar{b} \leftarrow p\} \bar{b}$$

As it can be seen immediately, $\{a\}$ attacks $\{b\}$. Hence the contrary-to-duty paradox: the assumptions that murdering is forbidden and that someone murders are mutually exclusive and their union is not conflict-free. For this example, not every derivation graph will prevent the paradox to arise. Consider the following:



G_1 puts no restriction on deductions, G_2 restricts deductions to using only one rule, and G_3 restricts deductions to using only two rules per branch. In our example both G_1 and G_3 do not prohibit any of the possible deductions, while G_2 does and in fact is the only graph which prevents the paradox. By forbidding the iteration for rules, it blocks the derivation of \bar{b} from the assumption a . Indeed, we get $\{a\} \not\vdash_G^{\{p \leftarrow a, \bar{b} \leftarrow p\}} \bar{b}$. Therefore, $\{a\}$ does not G -attack $\{b\}$ and the assumption set $\{a, b\}$ is an extension under any G -semantics.

2.3 Preliminary Results

We briefly summarise preliminary results. A core feature of ABA is that it comes with guidelines that specify how to instantiate a framework from a given knowledge base. This job is largely

done by the notions of deduction and attack. In turn, derivation graphs work as a device for controlling and manipulate such instantiation process. An interesting research question would be that of asking under which conditions one can obtain an equivalent framework by pre-processing the knowledge base while leaving the derivation process untouched. Initial results show that if a derivation graph contains exactly one rule node for each rule set in the given n -ABAF, it is possible to define a *derivation function* that works in such a way. This is an operation on the knowledge base which automatically identifies and removes rules that would not be allowed under a derivation graph G , allowing unconstrained deductions. For each graph constraint G , there is a derivation function γ that extracts from the rules \mathcal{R} of an n -ABAF the subset of rules whose application is allowed under G . As a result, the set of G -deductions for D is equivalent to the set of standard deductions that can be built using rules in $\gamma(\mathcal{R})$ only. Consequently, the argumentation framework instantiated by means of G -deductions is equivalent to the one instantiated through standard deductions after its rule-set has been restricted by the derivation function. This assures that the same outcome is reached by limiting deductions via some derivation graph or by restricting the knowledge base accordingly.

Moreover, we present an encoding of the n -ABA formalism and a derivation graphs in ASP (available at: <https://www.dbai.tuwien.ac.at/research/argumentation/abasp/>), inspired by the one provided in [12] for regular ABA frameworks and semantics. Given an n -ABAF and derivation graph as input, the encoding provides an answer set M for each σ_G -extension of a given n -ABAF under the graph constraint G . Regarding the n -ABAF in input, we extend the encoding presented in [12] by introducing a new predicate specifying for each rule the (unique) rule-set \mathcal{R}_i it belongs to. The derivation graph G is encoded as a labelled graph using predicates for nodes and edges, specifying which node corresponds to the starting node and each rule set. Eventually, each semantics-related module presented in [12] can be integrated into ours, after being carefully adapted to take into account rule-sets and derivation constraints.

3 Conclusive Remarks

This work introduces an extension of assumption-based argumentation with multiple rule-sets together with some formal constraints on its deductive machinery. These constraints, called *derivation graphs*, regulate the argument construction process from the underlying knowledge base, thereby limiting the procedure for its instantiation into an ABA framework. While this allows to avoid undesired conclusions as shown in Examples 1 and 3, we are currently working on defining constraints that operate directly on the knowledge base. In addition, we presented an encoding of our formalism in ASP, building up on the work presented in [12].

The derivation constraints presented in this work successfully avoid some paradoxes and fallacies in the domain of normative reasoning, but they do so at the expense of the deductive power of the ABA formalism. As a general direction for future research we want to broaden our horizon and investigate different kinds of reasoning constraints that minimise this loss. In doing so, we aim at positioning our formalism with respect to related frameworks: the work by Pigozzi and Van der Torre on constitutive and regulative norms in argumentation [19]; modular ABA [7] as it was proposed in connection with normative reasoning; Deontic ASP [11] encoding input/output logics. Although we restricted our studies on flat ABAFs so far, we anticipate that the full expressiveness of non-flat ABAFs may be needed to capture general instances of normative reasoning (cf. Example 3). Equipping non-flat ABAFs with derivation graphs might pose additional challenges since non-flat ABAFs require certain closure conditions on the set of acceptable assumptions. In addition, we aim at studying how size and complexity of instantiated ABAFs are influenced under our derivation constraints, in line with [22].

References

- [1] Andrei Bondarenko, Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artif. Intell.*, 93:63–101, 1997.
- [2] Andrei Bondarenko, Francesca Toni, and Robert A. Kowalski. An assumption-based framework for non-monotonic reasoning. In Luís Moniz Pereira and Anil Nerode, editors, *Logic Programming and Non-monotonic Reasoning, Proceedings of the Second International Workshop, Lisbon, Portugal, June 1993*, pages 171–189. MIT Press, 1993.
- [3] Martin Caminada, Sanjay Modgil, and Nir Oren. Preferences and unrestricted rebut. In Simon Parsons, Nir Oren, Chris Reed, and Federico Cerutti, editors, *Computational Models of Argument - Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 209–220. IOS Press, 2014.
- [4] Kristijonas Cyras, Xiuyi Fan, Claudia Schulz, and Francesca Toni. Assumption-based argumentation: Disputes, explanations, preferences. *FLAP*, 4(8), 2017.
- [5] Kristijonas Cyras, Tiago Oliveira, Amin Karamlou, and Francesca Toni. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument Comput.*, 12(2):149–189, 2021.
- [6] Phan Minh Dung, Robert A. Kowalski, and Francesca Toni. Assumption-based argumentation. In Guillermo Ricardo Simari and Iyad Rahwan, editors, *Argumentation in Artificial Intelligence*, pages 199–218. Springer, 2009.
- [7] Phan Minh Dung and Phan Minh Thang. Modular argumentation for modelling legal doctrines in common law of contract. *Artif. Intell. Law*, 17(3):167–182, 2009.
- [8] Xiuyi Fan. On generating explainable plans with assumption-based argumentation. In Tim Miller, Nir Oren, Yuko Sakurai, Itsuki Noda, Bastin Tony Roy Savarimuthu, and Tran Cao Son, editors, *PRIMA 2018: Principles and Practice of Multi-Agent Systems - 21st International Conference, Tokyo, Japan, October 29 - November 2, 2018, Proceedings*, volume 11224 of *Lecture Notes in Computer Science*, pages 344–361. Springer, 2018.
- [9] Xiuyi Fan, Francesca Toni, Andrei Mocanu, and Matthew Williams. Dialogical two-agent decision making with assumption-based argumentation. In Ana L. C. Bazzan, Michael N. Huhns, Alessio Lomuscio, and Paul Scerri, editors, *International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014*, pages 533–540. IFAAMAS/ACM, 2014.
- [10] James William Forrester. Gentle murder, or the adverbial samaritan. *The Journal of Philosophy*, 81(4):193–197, 1984.
- [11] Ricardo Gonçalves and José Júlio Alferes. An embedding of input-output logic in deontic logic programs. In Thomas Ågotnes, Jan M. Broersen, and Dag Elgesem, editors, *Deontic Logic in Computer Science - 11th International Conference, DEON 2012, Bergen, Norway, July 16-18, 2012. Proceedings*, volume 7393 of *Lecture Notes in Computer Science*, pages 61–75. Springer, 2012.

- [12] Tuomo Lehtonen, Johannes Peter Wallner, and Matti Järvisalo. Declarative algorithms and complexity results for assumption-based argumentation. *J. Artif. Intell. Res.*, 71:265–318, 2021.
- [13] Beishui Liao, Pere Pardo, Marija Slavkovic, and Leendert van der Torre. The jiminy advisor: Moral agreements among stakeholders based on norms and argumentation. *J. Artif. Intell. Res.*, 77:737–792, 2023.
- [14] Beishui Liao, Marija Slavkovic, and Leendert W. N. van der Torre. Building jiminy cricket: An architecture for moral agreements among stakeholders. *CoRR*, abs/1812.04741, 2018.
- [15] David Makinson and Leendert W. N. van der Torre. Input/output logics. *J. Philos. Log.*, 29(4):383–408, 2000.
- [16] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artif. Intell.*, 195:361–397, 2013.
- [17] Sanjay Modgil and Henry Prakken. The *ASPIC*⁺ framework for structured argumentation: a tutorial. *Argument Comput.*, 5(1):31–62, 2014.
- [18] Sanjay Modgil and Henry Prakken. Abstract rule-based argumentation. *FLAP*, 4(8), 2017.
- [19] Gabriella Pigozzi and Leon van der Torre. Arguing about constitutive and regulative norms. *J. Appl. Non Class. Logics*, 28(2-3):189–217, 2018.
- [20] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument Comput.*, 1(2):93–124, 2010.
- [21] Iyad Rahwan. Guest editorial: Argumentation in multi-agent systems. *Auton. Agents Multi Agent Syst.*, 11(2):115–125, 2005.
- [22] Anna Rapberger, Markus Ulbricht, and Johannes Peter Wallner. Argumentation frameworks induced by assumption-based argumentation: Relating size and complexity. In Ofer Arieli, Giovanni Casini, and Laura Giordano, editors, *Proceedings of the 20th International Workshop on Non-Monotonic Reasoning, NMR 2022, Part of the Federated Logic Conference (FLoC 2022), Haifa, Israel, August 7-9, 2022*, volume 3197 of *CEUR Workshop Proceedings*, pages 92–103. CEUR-WS.org, 2022.
- [23] John Rogers Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, England: Cambridge University Press, 1969.
- [24] Xin Sun and Leendert W. N. van der Torre. Combining constitutive and regulative norms in input/output logic. In Fabrizio Cariani, Davide Grossi, Joke Meheus, and Xavier Parent, editors, *Deontic Logic and Normative Systems - 12th International Conference, DEON 2014, Ghent, Belgium, July 12-15, 2014. Proceedings*, volume 8554 of *Lecture Notes in Computer Science*, pages 241–257. Springer, 2014.
- [25] G. H. von Wright. Deontic logic. *Mind*, 60(237):1–15, 1951.
- [26] Jing Xie and Chen-Ching Liu. Multi-agent systems and their applications. *Journal of International Council on Electrical Engineering*, 7(1):188–197, 2017.