Coalitions of arguments in bipolar argumentation frameworks

Claudette Cayrol **Marie-Christine Lagasquie-Schiex**
IRIT, UPS, 31062 Toulouse Cédex 09, France
{ccayrol, lagasq}@irit.fr

Abstract

Bipolar argumentation frameworks enable to represent two kinds of interaction between arguments: support and conflict. In this paper, we turn a bipolar argumentation framework into a "meta-argumentation" framework where conflicts occur between sets of arguments, characterized as coalitions of supporting arguments. So, Dung's well-known semantics can be used on this meta-argumentation framework in order to select the acceptable arguments.

1 Introduction

Argumentation has become an influential approach to model defeasible reasoning and dialogues between agents, based on the exchange of interacting arguments (see e.g. [Krause *et al.*, 1995; Prakken and Vreeswijk, 2002; Amgoud *et al.*, 2000; Karacapilidis and Papadias, 2001; Verheij, 2003]).

The following illustrative example presents the arguments exchanged during the meeting of the editorial board of a newspaper.

Ex. 1

Arg. a**:** *If we have the agreement, the important information I on the person X must be published.*

Arg. b**:** *I is a private information and X does not agree for publication.*

Arg. c_1**:** *Yes, I concerns X's family.*

Arg. c_2**:** *Moreover, X is a private person.*

Arg. d**:** *No, X is the new prime minister.*

In most existing systems, the interaction takes the form of a conflict, usually called attack. For example, an argument can be a pair $\langle$set of assumptions, conclusion$\rangle$, where the set of assumptions entails the conclusion according to some logical inference schema. Then, a conflict occurs between two arguments, for instance if the conclusion of one of them contradicts an assumption of the other one. In Ex. 1, b is in conflict with a. The main issue of any argumentation system is the selection of acceptable sets of arguments, based on the way arguments interact. Intuitively, an acceptable set of arguments must be in some sense coherent and strong enough (e.g. able to defend itself against all attacking arguments). The concept of acceptability has been explored through the use of argumentation frameworks, and one which is especially fruitful is Dung's argumentation framework [Dung, 1995], abstracting from the nature of the arguments. In such an abstract framework, from a set of arguments and a binary "attacks" relation, different semantics for acceptability are proposed, each one being characterized by several requirements that a set of arguments must satisfy so that the set could be selected. These selected sets of arguments are called extensions.

Recent work on argumentation [Karacapilidis and Papadias, 2001; Verheij, 2003; Cayrol and Lagasquie-Schiex, 2005a; 2005b] has advocated the representation of another kind of basic interaction between arguments. Indeed, it can be the case in a dialogue that an agent advances an argument which confirms an assumption used by an argument provided by another agent. This kind of interaction is not captured by the notion of defence. It is rather a kind of *support*. In Example 1, we may consider that the argument c_1 given by an agent supports the argument b given by another agent. It is not only a "dialog-like speech act": a new piece of information is really given and it is given *after* the production of the argument b. So taking c_1 into account leads either to modify b, or to find a more intuitive solution for representing the interaction between c_1 and b[1]. In [Cayrol and Lagasquie-Schiex, 2005b], Dung's framework has been extended to cope with both kinds of interaction, into a so-called bipolar abstract argumentation framework. Bipolarity refers to the existence of two independent kinds of information which represent repellent forces[2]. Semantics for acceptability have been defined, based on more complex notions of attack, called the supported and the diverted attacks. In Ex. 1, the fact that c_1 supports an attacker of a may be considered as a supported attack on a by c_1, and the fact that d attacks a supporter of b may be considered as a

[1] We adopt an incremental point of view, considering that pieces of information given by different agents enable them to provide more and more arguments. We do not want to revise already advanced arguments. Contrastedly, we intend to represent as much as possible all the kinds of interaction between these arguments. A comparison between both approaches is a topic for future work, for example from the point of view of computational complexity.

[2] Note that bipolarity already appears in argumentation during the definition of the arguments or the selection of the "best" arguments (even if we only consider one kind of interaction) [Amgoud *et al.*, 2004].

diverted attack on b by d. The new semantics ensure that neither supported attacks, nor diverted attacks can occur within an extension.

However, the definitions are rather complex and choosing the corresponding extensions, then computing them seems hard. So, in this paper, our purpose is to propose more intuitive semantics, using Dung's methodology in a "meta-argumentation" framework, where conflicts occur between coalitions of supporting arguments. Our motivation is the possible reusing of principles, algorithms and properties of Dung's well-known framework, now eponymously known by his name. So, our approach consists in:

- identifying "meta-arguments" (coalitions): conflict-free sets of arguments which are somehow related by the support relation;
- then defining the "meta-attack" relation: attack relation between the coalitions;
- finally using Dung's semantics on this "meta-argumentation" framework in order to identify acceptable sets of arguments (arguments which can be chosen together without conflict) for a bipolar argumentation framework.

Our contribution is twofold: first, we provide a formal method for defining and forming coalitions on the basis of interacting components; then, we propose new semantics for characterizing acceptable sets of arguments in a bipolar framework, taking advantage of Dung's definitions.

The paper is organized as follows: Dung's approach and the bipolar framework are described in Sect. 2; in Sect. 3, coalitions of arguments are introduced and they are used in Sect. 4 as interacting components of a meta-argumentation framework; then, Sect. 5 presents some related works and perspectives.

2 Background

Dung's framework

Let us present some basic definitions at work in Dung's theory of argumentation [Dung, 1995].

Def. 1 *A* finite argumentation framework *is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A}$ is a finite and non-empty set of so-called arguments and $\mathcal{R}$ is a binary relation over $\mathcal{A}$ (a subset of $\mathcal{A} \times \mathcal{A}$), the attacks relation.*

An argumentation framework can be represented by a directed graph in which each argument is a vertex and the edges are defined by the attacks relation: $\forall a, b \in \mathcal{A}$, $a\mathcal{R}b$ is represented by $a \not\to b$[3].

The first important notions are the notion of acceptability and the notion of conflict.

Def. 2 *Let $a \in \mathcal{A}$ and $S \subseteq \mathcal{A}$. a is* acceptable *w.r.t.*[4] *S iff $\forall b \in \mathcal{A}$ s.t. $b\mathcal{R}a$, $\exists c \in S$ s.t. $c\mathcal{R}b$. A set of arguments is acceptable w.r.t. S when each of its elements is acceptable w.r.t. S.*

S is conflict-free *for $\langle \mathcal{A}, \mathcal{R} \rangle$ iff $\nexists a, b \in S$, s.t. $a\mathcal{R}b$.*

Requiring the absence of conflicts and the form of autonomy captured by the notion of self-acceptability (S acceptable w.r.t. S) leads to the notion of admissible set.

Def. 3 *$S \subseteq \mathcal{A}$ is* admissible *for $\langle \mathcal{A}, \mathcal{R} \rangle$ iff S is conflict-free for $\langle \mathcal{A}, \mathcal{R} \rangle$ and acceptable w.r.t. S.*

Every extension of an argumentation framework under the standard semantics introduced by Dung (preferred, stable) is an admissible set, satisfying some form of optimality.

Def. 4 *Let $S \subseteq \mathcal{A}$. S is* a preferred extension *of $\langle \mathcal{A}, \mathcal{R} \rangle$ iff it is maximal w.r.t. $\subseteq$ among the admissible sets for $\langle \mathcal{A}, \mathcal{R} \rangle$. S is* a stable extension *of $\langle \mathcal{A}, \mathcal{R} \rangle$ iff it is conflict-free for $\langle \mathcal{A}, \mathcal{R} \rangle$ and $\forall a \in \mathcal{A} \setminus S$[5], $\exists b \in S$ s.t. $b\mathcal{R}a$.*

For each admissible set S, there exists a preferred extension which contains S. Each stable extension is a preferred extension, the converse is false (for instance when $\mathcal{A} = \{a\}$ and $\mathcal{R}_{\text{att}} = \{(a, a)\}$). The proof of these properties is given in [Dung, 1995].

Bipolar argumentation framework

As already said, arguments may be conflicting. These conflicts are captured by the attacks relation in an argumentation framework, and may be considered as negative interactions. Then, the concept of defence has been introduced in order to reinstate some of the attacked arguments, namely those whose attackers are in turn attacked. So, most logical theories of argumentation assume that if an argument a_3 defends an argument a_1 against an argument a_2 (a_3 attacks a_2 which attacks a_1), then a_3 is a kind of support for a_1, so a positive interaction. It holds in Dung's framework: only negative interaction is explicitly represented by the *attacks* relation, and positive interaction is implicitly represented through the notion of defence. In this case, support and attack are *dependent* notions. It is a parsimonious strategy, but it is not a correct description of the process of argumentation in realistic examples[6]: in Ex. 1, the link between the argument c_1 and the other arguments cannot be expressed with the attacks relation. So, a more complex argumentation framework is needed in order to formalize situations where two *independent* kinds of interaction are available: a positive and a negative one. Following [Karacapilidis and Papadias, 2001; Verheij, 2003], [Cayrol and Lagasquie-Schiex, 2005b; 2005a] propose a bipolar argumentation framework. This new framework is an extension of Dung's basic framework in which a

---

[3] If a does not attack b then, in the directed graph, there is no edge from a to b.

[4] We abbreviate "with respect to" as "w.r.t.", and "such that" as "s.t.".

[5] *i.e.* a belongs to $\mathcal{A}$ and does not belong to S.

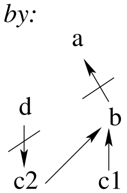
[6] For example, when the arguments are logical explanation arguments, *i.e.* under the form of a pair $\langle \Sigma, \Phi \rangle$ with Σ being a consistent set of formulae and Φ being a formula implied by Σ, the attacks relation can be classically defined by: let $a = \langle \Sigma_a, \Phi_a \rangle$, $b = \langle \Sigma_b, \Phi_b \rangle$ be two arguments, a attacks b iff $\neg\Phi_a \in \Sigma_b$. In the same way, the supports relation can be defined by: a supports b iff $\Phi_a \in \Sigma_b$. In that case, support and defence are different notions: a defends b iff there exists another argument $c = \langle \Sigma_c, \Phi_c \rangle$ such that $\neg\Phi_a \in \Sigma_c$ and $\neg\Phi_c \in \Sigma_b$.

new kind of interaction between arguments is represented by the *supports* relation[7].

Def. 5 *A finite* bipolar argumentation framework *$\langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ consists of a finite and non-empty set $\mathcal{A}$ of arguments, a binary relation $\mathcal{R}_{\text{att}}$ on $\mathcal{A}$ called the* attacks relation *and another binary relation $\mathcal{R}_{\text{sup}}$ on $\mathcal{A}$ called the* supports relation.

Consider a finite bipolar argumentation framework $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$. Note also that BAF can still be represented by a directed graph $\mathcal{G}_b$ called the *bipolar interaction graph* with two kinds of edges[8], one for the attacks relation and another one for the supports relation. Consider $a, b \in \mathcal{A}$, $a\mathcal{R}_{\text{att}}b$ is represented by $a \not\to b$ and $a\mathcal{R}_{\text{sup}}b$ is represented by $a \to b$.

Ex. 1 (cont'd)*The whole discussion during the editorial board meeting can now be formalized by the bipolar framework BAF_1 represented by:*



The fact that c_1 supports an attacker of a may be considered as a kind of negative interaction between c_1 and a, which is however weaker than a direct attack. In the same way, the attack by d of a supporter of b may also be considered as a negative interaction between d and b. From a cautious point of view, such arguments cannot appear together in a same extension. In order to address this problem, [Cayrol and Lagasquie-Schiex, 2005b; 2005a] introduce new kinds of attack.

Def. 6 *Let $a, b \in \mathcal{A}$*
A supported attack *for b by a is a sequence $a_1\mathcal{R}_1 \ldots \mathcal{R}_{n-1}a_n$, $n \geq 3$, with $a_1 = a$, $a_n = b$, s.t. $\forall i = 1 \ldots n-2$, $\mathcal{R}_i = \mathcal{R}_{\text{sup}}$ and $\mathcal{R}_{n-1} = \mathcal{R}_{\text{att}}$.*

A diverted attack *for b by a is a sequence $a_1\mathcal{R}_1 \ldots \mathcal{R}_{n-1}a_n$, $n \geq 3$, with $a_1 = a$, $a_n = b$, s.t. $\mathcal{R}_1 = \mathcal{R}_{\text{att}}$ and $\forall i = 2 \ldots n-1$, $\mathcal{R}_i = \mathcal{R}_{\text{sup}}$.*

In Ex. 1, there are a supported attack for a by c_1 and a diverted attack for b by d.

Combining different notions of conflict (one for each type of attack) with Dung's notion of acceptability, [Cayrol and Lagasquie-Schiex, 2005b] propose various bipolar semantics. However, two problems deserve further investigation: computational issues and the choice of the appropriate semantics depending on the application. For example, with a very simple example, $\mathcal{A} = \{a, b, c\}$, $\mathcal{R}_{\text{att}} = \{(c, b)\}$, $\mathcal{R}_{\text{sup}} = \{(a, b)\}$, the three acceptability semantics proposed in [Cayrol and Lagasquie-Schiex, 2005b] give three different results ($\{a, c\}$ for the d-preferred semantics, $\{a\}$ and $\{c\}$ for the s-preferred semantics, $\{c\}$ for the c-preferred semantics).

We propose a more intuitive (and promising) approach for handling bipolarity: the support relation is first taken into account to form coalitions of arguments, which will interact with a conflict relation in a simple Dung-like framework.

3 Coalitions

Principles which govern the definition of a coalition are the following: each argument belongs to at least one coalition; a coalition satisfies a coherence requirement; if two arguments belong to a same coalition, they are somehow related by the support relation.

Let $\mathsf{BAF} = \langle \mathcal{A}, \mathcal{R}_{\text{att}}, \mathcal{R}_{\text{sup}} \rangle$ be a bipolar argumentation framework represented by the graph $\mathcal{G}_b$. $\mathcal{G}_{\text{sup}}$ will denote the partial graph representing the partial framework $\langle \mathcal{A}, \mathcal{R}_{\text{sup}} \rangle$ (see [Berge, 1973] for a background on graph theory). AF will denote the partial argumentation framework $\langle \mathcal{A}, \mathcal{R}_{\text{att}} \rangle$ associated with BAF.

Def. 7 *$\mathcal{C} \subseteq \mathcal{A}$ is a* coalition *of* BAF *iff*

(i) *The subgraph of $\mathcal{G}_{\text{sup}}$ induced by $\mathcal{C}$ is connected;*

(ii) *$\mathcal{C}$ is conflict-free for* AF*;*

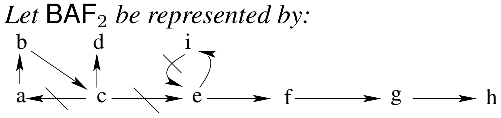
(iii) *$\mathcal{C}$ is maximal (for $\subseteq$) among the sets satisfying* **(i)** *and* **(ii)**.

Note that when $\mathcal{R}_{\text{att}}$ is empty, the coalitions are exactly the connected components of the partial graph $\mathcal{G}_{\text{sup}}$.

Prop. 1 *Each argument which is not self-attacking belongs to at least one coalition.*

> **Proof:** Let a be an argument which is not self-attacking. Then $\{a\}$ is conflict-free and the subgraph of $\mathcal{G}_{\text{sup}}$ induced by $\{a\}$ is connected. So $\{a\}$ satisfies the conditions **(i)** and **(ii)** of Definition 7. Either $\{a\}$ is a coalition, or there exists $\mathcal{C}$ a subset of $\mathcal{A}$ containing a and satisfying the conditions **(i)** and **(ii)**. Since $\mathcal{A}$ is finite, there exists a coalition containing $\mathcal{C}$ and thus containing $\{a\}$. □

Ex. 2 *Let BAF_2 be represented by:*



The coalitions are: $\mathcal{C}_1 = \{b, c, d\}$, $\mathcal{C}_2 = \{i\}$, $\mathcal{C}_3 = \{a, b\}$, $\mathcal{C}_4 = \{e, f, g, h\}$.

Definition 7 respects the desired principles but is not constructive. So an equivalent definition is proposed using the notions of maximal support path and of coalition in a set. The idea is to rely upon the connected components of the partial graph $\mathcal{G}_{\text{sup}}$.

Def. 8 *Let $S \subseteq \mathcal{A}$. $\mathcal{M} = \{a_1, \ldots, a_n\} \subseteq S$ is a* maximal support path *in S iff*

(i) *there exists a permutation $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$ s.t. the sequence of supports $a_{i_1}\mathcal{R}_{\text{sup}}a_{i_2} \ldots \mathcal{R}_{\text{sup}}a_{i_n}$ holds*

(ii) *and $\mathcal{M}$ is maximal (for $\subseteq$) among the subsets of S satisfying* **(i)**.

[7] If the supports relation is removed, Dung's framework is retrieved.

[8] Positive and negative interactions were both envisaged in Wigmore Charts [Wigmore, 1937], which use a complex graphical notation for legal argument structuring. More recently, Yoshimi [Yoshimi, 2004] has published an interesting work on the theory of the structure of debate, where debates and positions are represented by sets of arguments equipped with two inter-argument relations : dispute and support.

In Ex. 1, $\{c_1, b\}$ and $\{c_2, b\}$ are maximal support paths in $\mathcal{A}_1$ but not coalitions (they do not respect the condition **(iii)**). In Ex. 2, $\{a, b, c, d\}$ $\{e, f, g, h\}$ and $\{e, i\}$ are maximal support paths in $\mathcal{A}_2$ but not coalitions (the first one and the third one are not conflict-free for AF_2).

Def. 9 $\mathcal{C} \subseteq S$ *is a* coalition *in* S *iff there exists* $\{\mathcal{M}_1, \ldots, \mathcal{M}_p\}$ *a maximal (for* $\subseteq$*) set of non-disjoint maximal support paths in* S *s.t.* $\mathcal{C} = \mathcal{M}_1 \cup \mathcal{M}_2 \ldots \cup \mathcal{M}_p$.

It can be proved that:

Prop. 2 $\mathcal{C}$ *is a coalition in* S *iff* $\mathcal{C}$ *is a connected component*[9] *of the subgraph of* $\mathcal{G}_{\text{sup}}$ *induced by* S.

Proof:
$\Rightarrow$ Let $\mathcal{C}$ be a coalition in S. We have to prove that $\mathcal{C}$ is a maximal connected subset in the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S. $\mathcal{C}$ is the union of non-disjoint maximal support paths in S, namely $\mathcal{M}_1, \ldots, \mathcal{M}_p$. Let a and b be two distinct elements of $\mathcal{C}$. There exists $\mathcal{M}_i$ (resp. $\mathcal{M}_j$) a maximal support path in S containing a (resp. b).

If $\mathcal{M}_i$ and $\mathcal{M}_j$ are identical, there exists a sequence of supports from a to b (or from b to a).

If $\mathcal{M}_i$ and $\mathcal{M}_j$ are distinct, they are non-disjoint. There exists c which belongs to $\mathcal{M}_i$ and $\mathcal{M}_j$. So, there is a sequence of supports between a and c (in $\mathcal{M}_i$) and also a sequence of supports between b and c (in $\mathcal{M}_j$). So there exists a chain of supports between a and b. So, $\mathcal{C}$ is a connected subset of S in the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S.

If $\mathcal{C}$ is not maximal connected, there exists $\mathcal{C}'$ a connected subset of S which strictly contains $\mathcal{C}$. $\mathcal{C}'$ contains an element d which does not belong to $\mathcal{C}$. Since $\mathcal{C}$ is not empty and $\mathcal{C}'$ is connected, there exists a chain of supports between d and an element e of $\mathcal{C}$. That chain is the concatenation of non-disjoint support paths, each one being part of a maximal support path. Each of these support paths is included in $\mathcal{C}$ since the set $\{\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_p\}$ is maximal. So, d belongs to $\mathcal{C}$. So, $\mathcal{C}$ is maximal connected: $\mathcal{C}$ is a connected component of the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S.

$\Leftarrow$ Consider that $\mathcal{C}$ is a connected component of the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S. Let a be an element of $\mathcal{C}$. $\mathcal{C}$ is also the set of all the elements b in S such that there exists a chain of supports between a and b. Let $\mathcal{M}_1$ be a maximal support path containing a, and $\mathcal{C}'$ be the union of all the maximal support paths which are non-disjoint from $\mathcal{M}_1$. From the first part of the proof, $\mathcal{C}'$ is a maximal connected subset of S in the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S, and $\mathcal{C}'$ is included in $\mathcal{C}$. Since $\mathcal{C}$ is maximal connected, we have $\mathcal{C} = \mathcal{C}'$. So, $\mathcal{C}$ is a coalition in S. $\square$

Prop. 3 $\mathcal{C} \subseteq \mathcal{A}$ *is a coalition of* BAF *iff*

(i) *there exists* $S \subseteq \mathcal{A}$ *maximal (for* $\subseteq$*) conflict-free for* AF *s.t.* $\mathcal{C}$ *is a coalition in* S *and*

(ii) $\mathcal{C}$ *is maximal (for* $\subseteq$*) among the subsets of* $\mathcal{A}$ *satisfying* **(i)**.

[9] $\mathcal{C}$ is a connected component of G iff **(i)** $\forall v_1, v_2 \in \mathcal{C}$, if $v_1 \neq v_2$, $\exists$ a chain from v_1 to v_2 in G and **(ii)** $\mathcal{C}$ is maximal in G w.r.t. $\subseteq$ for **(i)**.

Proof:
$\Rightarrow$ Let $\mathcal{C}$ be a coalition of BAF. $\mathcal{C}$ is conflict-free for AF, so there exists S maximal conflict-free for AF containing $\mathcal{C}$. If the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S is connected, $\mathcal{C} = S$. If it is not the case, it is sufficient to prove that $\mathcal{C}$ is a connected component of the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S (due to Proposition 2). Since $\mathcal{C}$ is a coalition, $\mathcal{C}$ is connected in this subgraph. If $\mathcal{C}$ is not a connected component, there exists $\mathcal{C}'$ connected in S, which strictly contains $\mathcal{C}$.

So, we obtain $\mathcal{C}'$ conflict-free whose induced subgraph is connected. That is in contradiction with the maximality of $\mathcal{C}$. So, $\mathcal{C}$ is a coalition in S, which proves the first item. The second item follows from the definition of a coalition.

$\Leftarrow$ Assume that $\mathcal{C}$ is a coalition in S, maximal conflict-free for AF, and that there does not exist S' maximal conflict-free for AF and $\mathcal{C}'$ coalition in S' such that $\mathcal{C}'$ strictly contains $\mathcal{C}$. Obviously, $\mathcal{C}$ is conflict-free for AF and the subgraph of $\mathcal{G}_{\text{sup}}$ induced by $\mathcal{C}$ is connected. It remains to prove that $\mathcal{C}$ is maximal (condition **(iii)** of Definition 7).

Assume that $\mathcal{C}$ is strictly included in $\mathcal{D}$ connected and conflict-free. There exists S' maximal conflict-free for AF containing $\mathcal{D}$. Since $\mathcal{D}$ is connected, there exists $\mathcal{C}'$ a connected component of S' containing $\mathcal{D}$. So, $\mathcal{C}$ is strictly included in $\mathcal{C}'$, coalition of S'. That is in contradiction with our assumption on $\mathcal{C}$. So, $\mathcal{C}$ is a coalition of BAF. $\square$

Prop. 2 and 3 suggest a procedure for computing the coalitions of BAF:

Step 1: Consider AF and determine the maximal conflict-free sets for AF.

Step 2: For each set of arguments S_i obtained at Step 1, determine the connected components of the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S_i.

Step 3: Keep the maximal (for $\subseteq$) sets obtained at Step 2.

It can be proved that:

Prop. 4 *Let* $a, b \in \mathcal{A}$ *s.t.* $a\mathcal{R}_{\text{sup}}b$ *and* $\{a, b\}$ *is conflict-free in* AF. *Then, there exists a coalition of* BAF *containing both* a *and* b.

Proof: Since $\{a, b\}$ is conflict-free, there exists S maximal conflict-free subset of $\mathcal{A}$ containing a and b. Consider the subgraph of $\mathcal{G}_{\text{sup}}$ induced by S. This subgraph contains the edge (a, b), since a and b belong to S and $a\mathcal{R}_{\text{sup}}b$. So $\{a, b\}$ is included in a connected component of this subgraph, and due to Proposition 2 and Proposition 3, $\{a, b\}$ is included in a coalition. $\square$

4 A meta-argumentation framework

Let $\mathcal{C}(\mathcal{A})$ denote the set of coalitions of BAF. We define a conflict relation on $\mathcal{C}(\mathcal{A})$ as follows.

Def. 10 *Let* $\mathcal{C}_1$ *and* $\mathcal{C}_2$ *be two coalitions of* BAF. $\mathcal{C}_1$ c-attacks $\mathcal{C}_2$ *iff there exists an argument* a_1 *in* $\mathcal{C}_1$ *and an argument* a_2 *in* $\mathcal{C}_2$ *s.t.* $a_1\mathcal{R}_{\text{att}}a_2$.

It can be proved that:

Prop. 5 *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two distinct coalitions of* BAF. *If $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-disjoint then $\mathcal{C}_1$ c-attacks $\mathcal{C}_2$ or $\mathcal{C}_2$ c-attacks $\mathcal{C}_1$.*

> **Proof:** The subgraph of $\mathcal{G}_{\text{sup}}$ induced by $\mathcal{C}_1$ (resp. $\mathcal{C}_2$) is connected. Since $\mathcal{C}_1$ and $\mathcal{C}_2$ are non-disjoint, the subgraph induced by their union is connected. $\mathcal{C}_1$ and $\mathcal{C}_2$ are distinct, so their union strictly contains $\mathcal{C}_1$ (and also $\mathcal{C}_2$). As a coalition is maximal connected conflict-free, $\mathcal{C}_1 \cup \mathcal{C}_2$ cannot be conflict-free. But $\mathcal{C}_1$ and $\mathcal{C}_2$ are conflict-free. So there exists a_1 in $\mathcal{C}_1$ and a_2 in $\mathcal{C}_2$ such that $a_1 \mathcal{R}_{\text{att}} a_2$ ($\mathcal{C}_1$ c-attacks $\mathcal{C}_2$) or $a_2 \mathcal{R}_{\text{att}} a_1$ ($\mathcal{C}_2$ c-attacks $\mathcal{C}_1$). □

So we define a new argumentation framework $\mathsf{CAF} = \langle \mathcal{C}(\mathcal{A}), \text{c-attacks} \rangle$, referred to as the coalition framework associated with BAF. Dung's definitions apply to CAF, and it can be proved that:

Prop. 6 *Let $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ be a finite set of distinct coalitions. $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ is conflict-free for* CAF *iff $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$ is conflict-free for* AF.

> **Proof:** Assume that $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ is not conflict-free for CAF . There exists $1 \leq i \leq p$ and $1 \leq j \leq p$ such that $\mathcal{C}_i$ c-attacks $\mathcal{C}_j$. So there exists a_i in $\mathcal{C}_i$ and a_j in $\mathcal{C}_j$ such that $a_i \mathcal{R}_{\text{att}} a_j$. Then, $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$ is not conflict-free for AF.
>
> Conversely, assume that $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$ is not conflict-free for AF. There exists a and b in $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$ such that $a \mathcal{R}_{\text{att}} b$. Each $\mathcal{C}_i$ is conflict-free (it is a coalition), so a and b do not belong to a same $\mathcal{C}_i$. There exist $i \neq j$ with a belongs to $\mathcal{C}_i$ and b belongs to $\mathcal{C}_j$. Hence, we have $\mathcal{C}_i$ c-attacks $\mathcal{C}_j$ and the set $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ is not conflict-free for CAF. □

However, even if the set of coalitions $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ is conflict-free for CAF, there may exist a supported or a diverted attack in $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$.

Ex. 2 (cont'd) *For* CAF_2, *$\{\mathcal{C}_1, \mathcal{C}_2\}$ is conflict-free. However, there is a diverted attack for i by c in* BAF_2. *$\{\mathcal{C}_3, \mathcal{C}_4\}$ is conflict-free for* CAF_2. *However, there is a supported attack for e by b in* BAF_2.

So, we have a "meta-argumentation" framework (CAF) with a set of "meta-arguments" (the set of coalitions $\mathcal{C}(\mathcal{A})$) and a "meta-attack" relation on these coalitions (the c-attacks relation). A coalition gathers arguments which are close in some sense and can be advanced together. However, as coalitions may conflict, following Dung's methodology, we can compute preferred and stable extensions of CAF. Such an extension contains coalitions which are collectively acceptable. The last step consists in gathering the elements of the different coalitions of an extension. By this way, we are able to select the best groups of arguments (w.r.t. the given interaction relations).

Def. 11 *Let $S \subseteq \mathcal{A}$. S is a* cp-extension[10] *of* BAF *iff there exists $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ a preferred extension of* CAF *s.t. $S = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$. S is a* cs-extension[11] *of* BAF *iff there exists $\{\mathcal{C}_1, \ldots, \mathcal{C}_p\}$ a stable extension of* CAF *s.t. $S = \mathcal{C}_1 \cup \ldots \cup \mathcal{C}_p$.*

[10] cp means coalition-preferred.

[11] cs means coalition-stable.

When the only preferred extension of CAF is the empty set, we define the empty set as the unique cp-extension of BAF.

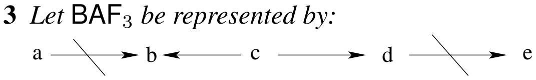
Ex. 2 (cont'd) *There is only one preferred extension of* CAF_2, *which is also stable: $\{\mathcal{C}_1, \mathcal{C}_2\}$. So, $S = \{b, c, d, i\}$ is the cp-extension (and also the cs-extension) of* BAF_2.

Some nice properties of Dung's classical framework are preserved:

- A BAF has always a (at least one) cp-extension. It is a consequence of Def. 11.
- In contrast, there does not always exist a cs-extension of BAF. The reason is that there may be no stable extension of CAF.
- Each cs-extension is also a cp-extension. The converse is false.
- There cannot exist two cp-extensions s.t. one strictly contains the other one. It follows from Def. 7 and 11.

However, other properties are lost. The following example shows that a cp-extension is not always admissible for AF, and a cs-extension is not always a stable extension of AF.

Ex. 3 *Let* BAF_3 *be represented by:*



The coalitions are: $\mathcal{C}_1 = \{a\}$, $\mathcal{C}_2 = \{b, c, d\}$, $\mathcal{C}_3 = \{e\}$. There is only one preferred extension of CAF_3, *which is also stable: $\{\mathcal{C}_1, \mathcal{C}_3\}$. So, $S = \{a, e\}$ is the cp-extension (and also the cs-extension) of* BAF_3. *We have $d \mathcal{R}_{\text{att}} e$, but a does not defend e against d (neither by a direct attack, nor by a diverted or a supported attack). Indeed, a attacks an element of the coalition which attacks e. So, S is not admissible for* AF_3. *S does not contain c, but there is no attack (no supported attack and no diverted attack) of an element of S against c. So, S is not a stable extension of* AF_3.

A coalition is considered as a whole and its members cannot be used separately in an attack process.

Ex. 3 suggests that admissibility is lost due to the size of the coalition $\{b, c, d\}$, and that it would be more fruitful to consider two independent coalitions $\{c, b\}$ and $\{c, d\}$. However, a new formalization of coalitions in terms of conflict-free maximal support paths does not enable to recover Dung's properties as shown below.

Def. 12 *An* elementary coalition *of* BAF *is a subset $\mathcal{EC} = \{a_1, \ldots, a_n\}$ of $\mathcal{A}$ s.t.:*

(i) *there exists a permutation $\{i_1, \ldots, i_n\}$ of $\{1, \ldots, n\}$ s.t. the sequence of supports $a_{i_1} \mathcal{R}_{\text{sup}} a_{i_2} \ldots \mathcal{R}_{\text{sup}} a_{i_n}$ holds;*

(ii) *$\mathcal{EC}$ is conflict-free for* AF*;*

(iii) *$\mathcal{EC}$ is maximal (for $\subseteq$) among the subsets of $\mathcal{A}$ satisfying* **(i)** *and* **(ii)**.

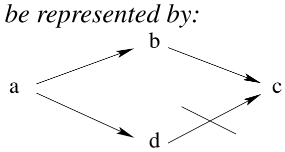
$\mathcal{EC}(\mathcal{A})$ denotes the set of elementary coalitions of BAF. Consider $\mathsf{ECAF} = \langle \mathcal{EC}(\mathcal{A}), \text{c-attacks} \rangle$ the elementary coalition framework associated with BAF.

Def. 13 *Let $S \subseteq \mathcal{A}$. S is a* ecp-extension *of* BAF *iff there exists $\{\mathcal{EC}_1, \ldots, \mathcal{EC}_p\}$ a preferred extension of* ECAF *s.t. $S = \mathcal{EC}_1 \cup \ldots \cup \mathcal{EC}_p$.*

Ex. 3 (cont'd) *In* BAF_3*, there are 4 elementary coalitions* $\{a\}$, $\{c, b\}$, $\{c, d\}$, $\{e\}$. *The unique ecp-extension* $\{a, c, d\}$ *is admissible for* AF_3.

However, some counter-intuitive results are obtained with ecp-extensions:

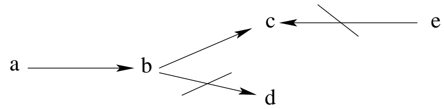
Ex. 4 *Let* BAF_4 *be represented by:*



The coalitions are $\mathcal{C}_1 = \{a, b, d\}$, $\mathcal{C}_2 = \{a, b, c\}$. *The unique cp-extension is* $\{a, b, d\}$. *The elementary coalitions are* $\mathcal{EC}_1 = \{a, d\}$, $\mathcal{EC}_2 = \{a, b, c\}$. *So the unique ecp-extension is* $\{a, d\}$. *It seems difficult to justify the elimination of* b.

And admissibility can still be lost:

Ex. 5 *Let* BAF_5 *be represented by:*



The elementary coalitions are $\mathcal{EC}_1 = \{a, b, c\}$, $\mathcal{EC}_2 = \{d\}$, $\mathcal{EC}_3 = \{e\}$. *The unique ecp-extension is* $S = \{e, d\}$. S *is not admissible for* AF_5*: we have* $b\mathcal{R}_{\text{att}}d$*, but* e *does not defend* d *against* b *(neither by a direct attack, nor by a diverted or a supported attack).*

Note that the lost of admissibility in Dung's sense is neither surprising, nor problematic for us:

- admissibility is lost because it takes into account "individual" attack and defence (d is not defended against b);
- whereas, with meta-argumentation and coalitions, we want to consider "collective" attack and defence (d is "reinstated" because the coalition $\{a, b, c\}$ which attacks d is attacked by another coalition $\{e\}$).

5 Discussion

We have proposed in this paper a "meta-argumentation framework" which takes into account two opposite kinds of interaction:

- the support relation is used in order to identify "coalitions" (sets of arguments which can be used together without conflict and which are related by the support relation);
- then the attack relation is used in order to identify conflicts between coalitions and to define new acceptability semantics as in Dung's framework.

In this meta AF, called "Coalition AF" (CAF), some nice properties of Dung's framework are preserved (link between new stable extensions – cs-extensions – and new preferred extensions – cp-extensions –, existence and maximality for set-inclusion of cp-extensions), but other properties are lost (the cp-extensions and the cs-extensions are not always admissible).

The notion of bipolar argumentation system is a recent notion in the field of argumentation, and, in our opinion, the use of coalitions in this framework is the first. So, we think it is important to compare our approach with other works on that key notion. Indeed, the concept of coalition has already been related to argumentation.

Collective argumentation framework [Bochman, 2003] [Nielsen and Parsons, 2006] A collective argumentation framework is an abstract framework where the initial data are a set of arguments and a binary "attack" relation between sets of arguments. The key idea is the following: a set of arguments can produce an attack against other arguments, which is not reducible to attacks between particular arguments. That is in agreement with our notion of coalition, since in our work, a coalition is considered as a whole and its members cannot be used separately in an attack process. The proposal by Nielsen and Parsons is similar to Bochman's proposal. Both proposals take the attacks between sets of arguments as initial data, and define semantics by properties on subsets of arguments. However, Nielsen and Parsons propose an abstract framework which allows sets of arguments to attack single arguments only, and they stick as close as possible to the semantics provided by Dung. In contrast, Bochman departs from Dung's methodology and give new specific definitions for stable and admissible sets of arguments. Our proposal essentially differs from collective argumentation in two points. First, we keep exactly Dung's construction for defining semantics, but we apply this construction in a meta-argumentation framework (the coalition framework). The second main difference lies in the meaning of a coalition: we intend to gather as many arguments as possible in a coalition, and a coalition cannot be broken in a defence process.

Generation of coalition structures in multi-agent systems [Dang and Jennings, 2004; Amgoud, 2005]

In multi-agent systems, the coalition formation is a process in which independent and autonomous agents come together to act as a collective. A coalition structure (CS) is a partition of the set of agents into coalitions. Each coalition has a value (the utility that the agents in the coalition can jointly get minus the cost which this coalition induces for each agent). So the value of a CS is obtained by aggregating the values of the different coalitions in the structure. One of the main problems is to generate a preferred CS, that is a structure which maximizes the global value. Recently, [Amgoud, 2005] has proposed an abstract framework where the initial data are a set of coalitions equipped with a conflict relation. A preferred CS is a subset of coalitions which is conflict-free and defends itself against attacks. Coalitions may conflict for instance if they are non-disjoint or if they achieve a same task.

However, the generation of the coalitions is not studied in [Amgoud, 2005]. So, one perspective is to apply our work to the formation of coalitions taking into account interactions between the agents. Arguments represent agents in that case. Indeed, it is very important to put together agents which want to cooperate ("supports" relation) and to avoid gathering agents who do not want to cooperate ("attacks" relation). Then, the concept of cp-extension provides a tool for selecting the best groups of agents (w.r.t. the given interaction relations).

More generally, the work reported here is generic and takes place in abstract frameworks, since no assumption is made on the nature of the arguments. Arguments may have a logical structure such as a pair $\langle$explanation, conclusion$\rangle$, may just be positions advanced in a discussion, or may be agents interacting in a multi-agent system. All that we need is the bipolar interaction graph describing how the arguments under consideration are interrelated. We think that this generic work should stimulate discussion across boundaries.

A first perspective is to propose new benchmarks, with new kinds of real-world problems that can be modelled by our approach and not modelled by previous proposals. Another perspective from a computational point of view will be to evaluate the complexity of our approach.

Acknowledgements

We would like to thank the referees for their comments which helped us to improve this paper.

References

[Amgoud *et al.*, 2000] L. Amgoud, N. Maudet, and S. Parsons. Arguments, Dialogue and Negotiation. In *Proc of* 14^{th} *ECAI*, pages 338–342, 2000.

[Amgoud *et al.*, 2004] L. Amgoud, C. Cayrol, and MC. Lagasquie-Schiex. On the bipolarity in argumentation frameworks. In *Proc. of the 10th NMR-UF workshop*, pages 1–9, 2004.

[Amgoud, 2005] L. Amgoud. Towards a formal model for task allocation via coalition formation. In *Proc. of AAMAS*, pages 1185–1186, 2005.

[Berge, 1973] C. Berge. *Graphs and Hypergraphs*. North-Holland Mathematical Library, 1973.

[Bochman, 2003] A. Bochman. Collective argumentation and disjunctive programming. *Journal of Logic and Computation*, 13 (3):405–428, 2003.

[Cayrol and Lagasquie-Schiex, 2005a] C. Cayrol and MC. Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In *Proc of the* 8^{th} *ECSQARU*, pages 366–377, 2005.

[Cayrol and Lagasquie-Schiex, 2005b] C. Cayrol and MC. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proc of the* 8^{th} *ECSQARU*, pages 378–389, 2005.

[Dang and Jennings, 2004] V.D. Dang and N.R. Jennings. Generating coalition structures with finite bound from the optimal guarantees. In *Proc. of AAMAS*, pages 564–571, 2004.

[Dung, 1995] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[Karacapilidis and Papadias, 2001] N. Karacapilidis and D. Papadias. Computer supported argumentation and collaborative decision making: the HERMES system. *Information systems*, 26(4):259–277, 2001.

[Krause *et al.*, 1995] P. Krause, S. Ambler, M. Elvang, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11 (1):113–131, 1995.

[Nielsen and Parsons, 2006] S.H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation. In *Proc. of the* 3^{rd} *WS on Argumentation in multi-agent systems*, 2006.

[Prakken and Vreeswijk, 2002] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In *Handbook of Philosophical Logic*, volume 4, pages 218–319. Kluwer Academic, 2002.

[Verheij, 2003] B. Verheij. Deflog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic in Computation*, 13:319–346, 2003.

[Wigmore, 1937] John Henry Wigmore. *The science of judicial proof as given by logic, psychology, and general experience, and illustrated judicial trials*. Little, Brown and co, 1937.

[Yoshimi, 2004] Jeff Yoshimi. Mapping the structure of debate. *Informal Logic*, 24(1), 2004.