Some Theoretical Results on the relationship between Argumentation and Coherence Theory

Yannis Dimopoulos[1], Pavlos Moraitis[2], and Carles Sierra[3]

[1] Department of Computer Science, University of Cyprus, Cyprus
[2] LIPADE, Paris Descartes University, France
[3] IIIA-CSIC, Spain

Abstract. This work provides initial results on the relationship between argumentation and Paul Thagard's coherence theory. We study the relationship, via appropriate transformations, between different types of coherent graphs (according to the values in the arcs) and different argumentation frameworks such as Dung's abstract argumentation framework, weighted argument systems or preference-based argumentation. The practical interest of our study is to show that coherence theory and argumentation can be mutually useful.

1 Introduction

This paper studies and provides initial results on the relationship between several models of argumentation and coherence theory.

Coherence theory, as proposed by Paul Thagard [12], assumes that knowledge can be represented as a network where nodes represent claims, and valued edges linking nodes may be labeled with positive or negative values representing respectively the degree of coherence or incoherence between nodes. Every coherence graph is associated with a number called the *coherence of the graph*. Based on Thagard formalism, this can be calculated by partitioning the set of nodes N of the graph in two sets, A and $N \setminus A$, where A contains the accepted elements of N, and $N \setminus A$ contains the rejected ones. The aim is to partition N such that a maximum number of nodes linked by edges with positive values (weights) are in the same set (i.e. A or $N \setminus A$) while a maximum number of nodes linked by edges with negative values are in complementary sets (i.e. A, and $N \setminus A$). The values of edges belong to $[-1, 1] \setminus \{0\}$.

There have been different proposals to represent arguments and their relationships. An Abstract Argumentation Framework (AF) [5] can be considered as a pair of a set arguments and a binary attack relation defined on the set of arguments. Such a theory can be represented as an oriented graph where nodes represent the arguments and edges the attacks between them. In Weighted Argument Systems (WAS) [6] attacks are associated with a weight, indicating the relative strength of the attack. A key concept in this framework is the notion of an *inconsistency budget*, which characterises how much inconsistency we can tolerate when selecting the sets of preferred arguments (extensions). It means that

given an inconsistency budget β, we are prepared to disregard attacks among the arguments up to a total weight of β. In Preference-based Argumentation (PAF), a preference relationship explicitly established between arguments, is used to rank sets of arguments.

Although argumentation and coherence theory strive to understand similar phenomena, such as making sense of contradictory information, their relation has not attracted much attention in the past. The need for a study of the relation between the two formalisms is also evident in the context of specific domains such as legal reasoning. Indeed, there are well established links between argumentation and legal reasoning on the one hand [10, 11], and legal reasoning and coherence on the other hand [1]. Another domain where the combination of coherence theory and argumentation may also prove beneficial, is the domain of argumentative debates. In this context, the coherence of the arguments that are used by the opponents during a debate could be taken into account. For instance, agents may decide to refrain from introducing arguments that decrease the strength of the coherence graph that corresponds to the arguments that has been exchanged in the course of that debate. E-justice or online dispute resolution are specific domains that could benefit from this kind of argumentative debates. Another application domain in which argumentation and coherence can be combined is that of policy analytics. Here the notion of coherence may serve as a measure of the impact of governmental policies on public opinion as it is expressed in social networks, by aggregating arguments supporting or attacking those policies.
As a first step in the direction of resolving these issues, this work provides the first formal results on the relation between coherence and argumentation.

In [8], coherence theory is used to understand the notion of norm adoption and a discussion on the relationship with AF is given although no formal account of this relationship is established. Here we contribute by giving some preliminary results on the relationship between optimal partitions and stable extensions in AF. In [9] argumentation dialogues are used to regain coherence when conflicts arise between agents. Argumentation is considered as a mechanism that permits the interaction between agents endowed with coherence theories. Here, differently from this work, we study the relationships between both approaches as alternative means of representing conflicting views.

In this paper we contribute to the study of the relationships between coherence theory and different argumentation formalisms. In particular we provide three results.

First, we transform classical argumentation theories into particular coherence graphs and show that the optimal partitions of these graphs correspond to stable extensions of the argumentation theory.

Second, we show that some coherent graphs can be understood as a WAS. More precisely, we consider a particular type of coherence graphs, those whose nodes represent atomic arguments, and that contain only maximally negative edges (i.e. -1). We prove that any subset A of arguments of such a coherence graph is an admissible extension with respect to the inconsistency budget β of a particular type of WAS.

Finally, we show that the maximal partitions of coherence graphs that contain edges labeled with $\{-1, 1\}$ can have an interpretation as extensions of PAF systems.

The paper is structured as follows. First, we provide some background knowledge on coherence and argumentation. Then, we study in order the relationship between coherence theory and Dung, WAS, and PAF systems. We conclude with a summary of the results and with the open lines for future work.

2 Background

2.1 Coherence Theory

The theory of coherence is a psychological motivational theory which understands coherence as an intrinsic domain independent motivation to agents. As any other motivational theory it aims at explaining the behaviour of agents at a high-level. We refer to Thagard's interpretation of the theory as he proposed a computational model for an otherwise long disputed concept.

Thagard presents the theory of coherence as a cognitive theory with roots in philosophy that interpret problem solving as the satisfaction of constraints over interconnected entities [12, 13]. The theory of coherence is then the study of associations among different pieces of information and the computation of how do they 'fit' together. Each piece of information puts constraints on other pieces of information; these constraints can be positive or negative. Positive constraints strengthen the connected pieces of information when considered together while negative constraints weaken them. In this theory, the cognitive process to be undertaken by an agent is to put together as many information pieces that have positive constraints while separating from these those that have negative constraints. In other words, coherent-based agents face an optimisation problem.

Several psychological processes can be understood in terms of coherence and constraint optimisation. These processes include stereoscopic vision, word perception, discourse comprehension, analogical mapping, and cognitive dissonance; see [14] for details.

Next we recall the basic definitions of coherence graph, constraint satisfaction and strength.

Definition 1 ([7]). *A* coherence graph *is an edge-weighted undirected graph* $g = \langle N, E, \psi \rangle$*, where*

- N *is a finite set of nodes representing pieces of information*
- $E \subseteq \{\{v, w\} | v, w \in N\}$ *is a finite set of edges representing the coherence or incoherence between pieces of information and that we shall call* constraints
- $\psi : E \to [-1, 1] \setminus \{0\}$ *is an edge-weighted function that assigns a negative or positive value to the coherence between pieces of information, and which we shall call* coherence function

The nodes of coherence graphs can be understood, from a knowledge representation perspective, as representing beliefs, desires, intentions, norms, or other

cognitions an agent may have [7, 9]. How the coherence values are computed depends on what sort of coherence we want to model. Thagard distinguishes among several types of coherence: deductive, explanatory, ..., and suggests different methods of computing these degrees. A coherence-based agent aims at determining which subset of the overall set of information pieces is to be accepted and which is to be rejected, that is, how to partition N into two sets containing accepted and rejected claims.

Definition 2 ([7]). *Given a coherence graph $g = \langle N, E, \psi \rangle$ and a partition of N into (A, R), the set of satisfied constraints $C_A \subseteq E$ is given by:*

$$C_A = \{\{v, w\} \in E | v \in A \text{ iff } w \in A \text{ when } \psi(\{v, w\}) > 0, \\ v \in A \text{ iff } w \in R \text{ when } \psi(\{v, w\}) < 0\}$$

According to Thagard, Coherence-based agents perform a search process to find the *best* partition which is the one that maximises the strength as defined next.

Definition 3 ([7]). *Given a coherence graph $g = \langle N, E, \psi \rangle$ the strength of a partition (A, R) is given by:*

$$Str(g, A) = \frac{\sum_{\{v,w\} \in C_A} |\psi(\{v, w\})|}{|E|}$$

The computation of the best partition does not tell us which one of the two sets is the one to accept, as the computation is symmetric, i.e. $Str(g, A) = Str(g, R)$. To determine which partition to accept an agent should use some ad-hoc criteria (e.g. greater number of nodes, greater average degree, etc.).

Thagard experimented with different computational implementations of coherence. Among them, ECHO [12] uses a neural network approach that, although does not guarantee convergence, has a good behavior on small networks. For very small networks like those in this work, a straightforward algorithm that enumerates all possible partitions is enough and is the algorithm we used.

A major question, left open by Thagard, is how to compute the degrees and links between pieces of information. Some works fill this gap proposing specific domain dependent functions, e.g. deductive relationships in [8]. We are assuming in this paper that these relationships are established and determined before our study can begin.

2.2 Some specific types of coherence graphs

From now onwards when we refer to the partition of a coherence graph we mean the best partition. We finally define the coherence of a graph as its strength assuming we would accept all its elements.

Definition 4. *Given a coherence graph $g = \langle N, E, \psi \rangle$, we define the* coherence *of graph g, noted $Coh(g)$, as the strength of the partition $(N, \emptyset)$, that is the partition with all nodes in N accepted, $Coh(g) = Str(g, N)$.*

Next definition is useful in some of the proofs later on.

Definition 5 (Subgraph). *Given two coherence graphs* $g = \langle N, E, \psi \rangle$ *and* $g' = \langle N', E', \psi' \rangle$ *we say that* g' *is a subgraph of* g*, noted* $g' \sqsubseteq g$ *iff* $N' \subseteq N$*,* $E' = \{\{v, w\} | v, w \in N', \{v, w\} \in E\}$ *and* $\psi' = \psi|_{N'}$*, where* $\psi|_{N'} : E' \to [-1, 1] \setminus \{0\}$*, with* $\psi|_{N'}(\{v, w\}) = \psi(\{v, w\})$*.*

In this paper we will use two particular types of coherence graphs. First, those where the links between nodes are all labeled with -1. This value expresses the fact that the two nodes are maximally incoherent. We call such graphs *negative unipolar* (or *neg-unipolar*). More formally:

Definition 6 (Negative Unipolar Coherence Graphs). *We say that a coherence graph* $g = \langle N, E, \psi \rangle$ *is negative unipolar (or neg-unipolar) if and only if for all* $e \in E$*,* $\psi(e) = -1$*.*

Second, those where the links between nodes are all labeled with -1 or 1. We call such graphs *Bipolar*. More formally:

Definition 7 (Bipolar Coherence Graphs). *Given a coherence graph* $g = \langle N, E, \psi \rangle$*, we say it is a* Bipolar Coherence Graph *iff (1) it is connected and (2)* $\psi(e) \in \{1, -1\}$ *for all* $e \in E$*.*

2.3 Argumentation Systems

An *argumentation system*, as introduced by Dung in [5], is a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, where $\mathcal{A}$ is a set of *arguments*, and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an *attack relation*. The relation a *attacks* b, or b *is attacked by* a, is denoted by $a \mathcal{R} b$ or $(a, b) \in \mathcal{R}$.

In [5], different acceptability semantics were introduced. They are based on two basic concepts: *defence* and *conflict-freeness*, defined as follows:

Definition 8 (Defence/Conflict-freeness). *Let* $T = \langle \mathcal{A}, \mathcal{R} \rangle$ *be an argumentation system. Let* $A' \subseteq \mathcal{A}$*.*

- A' *is conflict free iff* $\nexists\, a, b \in A'$ *s.t* $(a, b) \in \mathcal{R}$*.*
- A' *defends* $a \in \mathcal{A}$ *iff* $\forall b \in \mathcal{A}$*, if* $(b, a) \in \mathcal{R}$*, then* $\exists c \in A'$ *s.t* $(c, b) \in \mathcal{R}$*.*

The basic idea behind these concepts is the following: for a rational agent, an argument a is acceptable if he can defend a against all attacks. All the arguments acceptable for a rational agent will be gathered in a so-called *extension*. An extension must satisfy a consistency requirement and must defend all its elements.

Definition 9 (Acceptability Semantics). *Let* $T = \langle \mathcal{A}, \mathcal{R} \rangle$ *be an argumentation system and* A' *a conflict free set of arguments.*

- A' *is an admissible extension iff* A' *defends every element in* A'*.*
- A' *is a preferred extension iff* A' *is a maximal (w.r.t set* $\subseteq$*) admissible set.*

– *A' is a stable extension iff it is a preferred extension that attacks any argument in $A \setminus A'$.*

In [6] the authors have proposed an extension of classical Dung's argument systems in which attacks are associated with a *weight* which indicates the relative strength of each attack. A key idea in weighted argument systems is that of an *inconsistency budget*, characterizing how much inconsistency we are prepared to tolerate. More formally:

Definition 10 (Weighted Argument Systems (WAS) [6]). *A* weighed argument system *is a triple $W = \langle \mathcal{A}, \mathcal{R}, w \rangle$ where $\langle \mathcal{A}, \mathcal{R} \rangle$ is a Dung-style abstract system and $w : \mathcal{R} \to \Re_{>}$ is a function assigning real-valued weights to attacks.*

An *inconsistency budget* β characterizes how much inconsistency we are prepared to tolerate. Thus, accepting an inconsistency budget β means that we are prepared to disregard attacks up to a total weight of β. Dung systems implicitly assume an inconsistency budget of $\beta = 0$. An increasing number of extensions can be found for increasing values of β. We note a WAS system with budget β as $W^{\beta} = (\langle \mathcal{A}, \mathcal{R}, w \rangle, \beta)$.

Definition 11 ([6]). *Let $W = \langle \mathcal{A}, \mathcal{R}, w \rangle$ be a weighted argument system. Given $R \subseteq \mathcal{R}$, we define the* budget *of R as:*

$$wt(R, w) = \sum_{(a_1, a_2) \in R} w(a_1, a_2)$$

And the sets of links under budget β as:

$$sub(\mathcal{R}, w, \beta) = \{R : R \subseteq \mathcal{R} \text{ and } wt(R, w) \leq \beta\}$$

3 Coherence theory and Classic Argumentation (AF)

In this section we establish results on the relation of Dung classic argumentation [5] and coherence theory. Given a symmetric Dung system, i.e. $T = \langle \mathcal{A}, \mathcal{R} \rangle$, such that $(a, b) \in \mathcal{R}$ *iff* $(b, a) \in \mathcal{R}$, we define its associated coherence graph as $g_T = \langle \mathcal{A}, \mathcal{R}, \psi \rangle$, where $\psi(e) = -1$ for all $e \in \mathcal{R}$. Obviously, g_T is neg-unipolar.

In the particular case we are considering in this work, namely arguments correspond to the nodes of a coherence graph and attacks to its arcs, it is reasonable to consider that the non-oriented negative arcs in a neg-unipolar graph correspond to symmetric attacks in the associated argumentation system.

The coherence graph g_T associated with a symmetric argumentation theory T is a classic undirected graph. A *bipartite* graph is a graph whose nodes can be divided into two disjoint sets A and B such that every edge connects a node in A to one in B. Clearly, if a coherence graph g is bipartite it admits an optimal partition (A, B) with $Str(g, A) = 1$. On the other hand, it is well known that a graph is bipartite iff it contains no odd cycles. The above leads to the following

observation: A neg-unipolar graph g has a partition (A, R) with $Str(g, A) = 1$ iff it contains no odd cycles.

Clearly, the coherence graph g_T of a symmetric Dung argumentation theory T contains an odd cycle iff T contains an odd cycle. The next proposition states that an optimal partition of the coherence graph associated to a symmetric Dung theory without odd cycles induces two stable extensions for the theory.

Proposition 1. *Let $T = (\mathcal{A}, \mathcal{R})$ be a symmetric Dung theory, (A, R) an optimal partition of its corresponding neg-unipolar graph g_T, and $i(\mathcal{A}) \subseteq \mathcal{A}$ the set of nodes with degree 0. Then $A \cup i(\mathcal{A})$ and $R \cup i(\mathcal{A})$ are stable extensions of T iff T does not contain odd cycles.*

Proof. If T does not contain odd cycles, then g_T is a bipartite graph, i.e. there is an optimal partition (A, R) with $Str(g, A) = 1$. It suffices to show that $A \cup i(\mathcal{A})$ is a stable extension. First, $A \cup i(\mathcal{A})$ is conflict-free because otherwise $Str(g, A) \neq 1$. Now assume that A is not a stable extension because there is $b \in R$ such that there is no $a \in A$ with $\{a, b\} \in \mathcal{R}$. Clearly, b cannot have degree 0 because then $b \in i(\mathcal{A})$. Therefore, there must be $b' \in R$ such that $\{b', b\} \in \mathcal{R}$, which means that $Str(g, A) \neq 1$, and thus we get a contradiction. On the other hand, if T contains an odd cycle, g_T is not bipartite, and therefore the arguments of T cannot be partitioned in two sets that are conflict-free. Similar arguments hold for R.

We now study a relation between non-symmetric Dung frameworks and coherence theories based on a different coherence theory construction that is described in the next definition and used in the rest of this section.

Definition 12. *Given an argumentation framework $T = (\mathcal{A}, \mathcal{R})$, we define its corresponding coherence theory $g_T = \langle N, E, \psi \rangle$ as follows*

- $N = \mathcal{A} \cup \{x_{ij} | (a_i, a_j) \in \mathcal{R}\}$
- $E = \{\{a_i, x_{ij}\}, \{x_{ij}, a_j\} | (a_i, a_j) \in \mathcal{R}\}$
- $\psi(\{a_i, x_{ij}\}) = 1$, $\psi(\{x_{ij}, a_j\}) = -1$, $\forall (a_i, a_j) \in \mathcal{R}$

We say that a Dung argumentation theory $T = (\mathcal{A}, \mathcal{R})$ is *connected* if there is a directed path from any node in $\mathcal{A}$ to any other node in $\mathcal{A}$.

Proposition 2. *Let T be a connected Dung argumentation theory, and g_T its corresponding coherence theory. Any partition (A, R) on g_T such that $Str(g_T, A) = 1$ induces two stable extensions on T.*

Proof. Let $T = (\mathcal{A}, \mathcal{R})$ and let (A, R) be a partition of g_T with $Str(g_T, A) = 1$. We consider A, as similar arguments hold for R. Clearly, A contains a set of nodes $S \subseteq \mathcal{A}$ that correspond to arguments of $\mathcal{A}$. We show that this set $S = A \cap \mathcal{A}$ is a stable extension of T.
First observe that for each node $a_i \in S$ all nodes x_{ij} for arguments a_j s.t. $(a_i, a_j) \in \mathcal{R}$ must also belong to S, since $\psi(\{a_i, x_{ij}\}) = 1$ and $Str(g_T, A) = 1$. The same holds for the nodes of $R \cap \mathcal{A}$. We first show that S is conflict-free. By

way of contradiction, suppose that $a_i, a_j \in S$ and $(a_i, a_j) \in \mathcal{R}$. Then, A must contain the nodes a_i, a_j, x_{ij} with $\psi(\{x_{ij}, a_j\}) = -1$, therefore $Str(g_T, A) \neq 1$, contradiction.
We now prove that for all $a_j \in R \cap \mathcal{A}$ there is a node $a_i \in A \cap \mathcal{A}$ s.t. $(a_i, a_j) \in \mathcal{R}$. Since T is connected, there must be an argument $a_k \in \mathcal{A}$ s.t. $(a_k, a_j) \in \mathcal{R}$. If $a_k \in A$, the result holds. Assume that $a_k \in R$. Then there is a node $x_{kj} \in R$ s.t. $\psi(\{x_{kj}, a_j\}) = -1$ therefore $Str(g_T, A) \neq 1$, contradiction.

The above property leads to the following correspondence between the optimal partitions of the coherence graph of a Dung theory without odd cycles and its stable extensions.

Proposition 3. *Let T be a connected argumentation theory without odd cycles, and g_T its corresponding coherence theory. An optimal partition of g_T induces two stable extensions of T.*

Proof. Given g_T we construct an undirected graph g' as follows. For node a_i and all nodes x_{ij} connected to a_i with a positive link, we introduce a node a'_i in g'. A node a'_i is connected to node a'_j in g' if there is a node x_{ij} in g_T such that $\{a_i, x_{ij}\}, \{x_{ij}, a_j\} \in E$ for the nodes a_i, a_j that correspond to a'_i, a'_j. Clearly, g' is isomorphic to (the graph that corresponds to) T, therefore does not contain odd cycles. Moreover, a bipartition of g' induces an optimal partition (A, R) of g_T with $Str(g_T, A) = 1$. Then the claim follows by proposition 2.

4 Coherence theory and Weighted Argument Systems (WAS)

In this section we study a relationship between coherence theory and weighted argument systems (WAS). We consider the particular case of *neg-unipolar* graphs. We consider that negative arcs linking nodes in a neg-unipolar graph represent symmetric weighted attacks of equal value (e.g. $w = 1$) between arguments in an associated weighted argument system. More formally:

Definition 13. *Given a neg-unipolar graph $g = \langle N, E, \psi \rangle$ we define the* weighted argument system *associated to g with inconsistency budget β as $W^\beta(g) = (\langle N, E, w \rangle, \beta)$ where $w(a, b) = w(b, a) = 1$ for all $(a, b) \in E$.*

When $\beta = 0$ the weighted argument system associated to a neg-unipolar graph corresponds to a symmetric Dung abstract argumentation system.

Based on the above we can define formally a $WAS(g)$ as follows:

Definition 14. *Given a neg-unipolar graph $g = (N, E, \psi)$ we define the weighted argument system of g as $WAS(g) = W^{2*|\Sigma\psi(e)|}(g)$.*

We need now to define a notion of *internal inconsistency* of a coherence graph which is simply the sum of the weights of its negative links. More formally:

Definition 15 (Internal Inconsistency (INC)). *Given a graph* $g = \langle N, E, \psi \rangle$ *the* internal inconsistency *of graph* g *is defined as* $INC(g) = |\Sigma_{\psi(e)<0}\psi(e)|$

Based on the above notions we can now formulate a relation between coherence and weighted argument systems.

Proposition 4. *Let* (A, R) *be a partition of a neg-unipolar graph* $g = \langle N, E, \psi \rangle$. *Then* A *is an admissible extension of* $W^k(g)$, *where* $k = 2 * INC(\langle A, E|_A, \psi \rangle)$.

Proof. Clearly, A, as any subset of N, is an admissible extension. On the other hand the budget of A is the number of negative edges that link its nodes, i.e. $INC(\langle A, E|_A, \psi \rangle)$, multiplied by 2, since every undirected edge of g corresponds to a pair of directed edges in $W^k(g)$. □

It is then obvious that all admissible extensions of $WAS(g)$ are also parts of the possible bipartitions of the associated neg-unipolar coherence graph g.

We will now show that the strength of coherence graphs induces a ranking on the bipartitions of the nodes of neg-unipolar graphs that has an interesting meaning from an argumentation perspective. The following result shows that the order of the bipartitions induced by $Str(\cdot)$ of a neg-unipolar graph induces a ranking over Dung's stable extensions (i.e. for inconsistency budget $\beta = 0$ of the associated WAS.)

Theorem 1 (Ranking of stable extensions). *Given a neg-unipolar graph* $g = \langle N, E, \psi \rangle$, *let* $P = \langle P_1, .., P_n \rangle$ *be the partially ordered set (or poset), according to* $Str(\cdot)$, *of all possible partitions of* g *where* $P_i = (A_i, R_i)$. *Then, for any pair* $\mathcal{E}_i$ *and* $\mathcal{E}_j$ *of stable extensions of* $W^0(g)$, $INC(\langle N \setminus \mathcal{E}_i, E|_{N\setminus\mathcal{E}_i}, \psi \rangle) < INC(\langle N \setminus \mathcal{E}_j, E|_{N\setminus\mathcal{E}_j}, \psi \rangle)$ *if there are* k, l *such that* $P_k = (\mathcal{E}_i, N \setminus \mathcal{E}_i)$ *and* $P_l = (\mathcal{E}_j, N \setminus \mathcal{E}_j)$ *and* $k < l$.

Proof. : Let $P = \langle P_1, .., P_n \rangle$ be the partially ordered set, according to $Str(\cdot)$, of all possible partitions of the neg-unipolar graph $g = \langle N, E, \psi \rangle$. Let's consider two partitions $P_i = (\mathcal{E}_i, N \setminus \mathcal{E}_i)$ and $P_j = (\mathcal{E}_j, N \setminus \mathcal{E}_j)$ s.t. $\mathcal{E}_i$, $\mathcal{E}_j$ are stable extensions of $W^0(g)$. Following definition 3 the strength of the partition P_i is $Str(g, \mathcal{E}_i)$=$Str(g, N \setminus \mathcal{E}_i)$ and the strength of P_j is $Str(g, \mathcal{E}_j)$=$Str(g, N \setminus \mathcal{E}_j)$. We must prove that $INC(\langle N \setminus \mathcal{E}_i, E|_{N\setminus\mathcal{E}_i}, \psi \rangle) < INC(\langle N \setminus \mathcal{E}_j, E|_{N\setminus\mathcal{E}_j}, \psi \rangle)$ if $Str(g, N \setminus \mathcal{E}_i) > Str(g, N \setminus \mathcal{E}_j)$ (i.e $i < j$). Following definitions 2 and 3 the strength of a partition P depends on the number of satisfied constraints namely a) how many negative arcs are cut, splitting the linked arguments in the two subparts of a partition and b) how many positive arcs are protected i.e keeping the linked arguments in the same subpart of the partition. In our case the graph g is a neg-unipolar graph and therefore only negative arcs (i.e. for all e ∈ E, $\psi(e) = -1$) exist between the arguments. That means that the number of not satisfied constraints only relies on the number of negative arcs that link arguments in any subpart of the partition. As $\mathcal{E}_i$, $\mathcal{E}_j$ are stable extensions we know that $INC(\mathcal{E}_i)$=$INC(\mathcal{E}_j)$=0. Thus there is no violated constraints (i.e. arguments linked by negative arcs). So the value of the strength of P_i (resp. P_j) depends exclusively on the number of not satisfied constraints (i.e. number of negative

arcs) in $N \backslash \mathcal{E}_i$ (resp. $N \backslash \mathcal{E}_j$). As the total number of negative arcs is $|E|$, the lower the number of negative arcs appearing in e.g. $N \setminus \mathcal{E}_i$, the greater the number of satisfied constraints (i.e. negative arcs cut) and thus, according to definition 3, the higher the value of $Str(g, N \setminus \mathcal{E}_i)$. Thus if $Str(g, N \setminus \mathcal{E}_i) > Str(g, N \setminus \mathcal{E}_j)$ that means that $INC(\langle N \setminus \mathcal{E}_i, E|_{N \setminus \mathcal{E}_i}, \psi \rangle) < INC(\langle N \setminus \mathcal{E}_j, E|_{N \setminus \mathcal{E}_j}, \psi \rangle)$. □

The above result implies an ranking on Dung's extensions according to the internal inconsistency of the arguments that are left out of the extensions. The following example illustrates this ranking.

Example 1. Consider the neg-unipolar graph g of Figure 1 and its associated weighted argument system $W^{18}(g)$ (i.e. $18 = 2 * INC(g)$ with $INC(g) = 9$) in Table 1. On the left hand column of the table we see the partitions of the graph ranked according to their strength and on the right hand column the Dung's stable extensions (i.e. $\beta = 0$) of the associated weighted argument system $W^{18}(g)$.

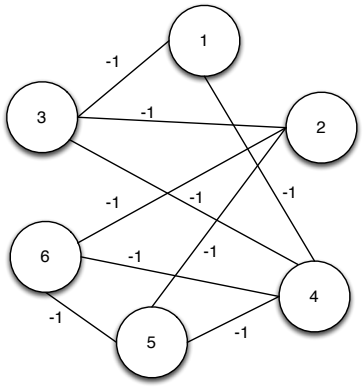

Fig. 1. A neg-unipolar graph.

The set of Dung stable extensions of $W^{18}(g)$ is $SE = \{\{2,4\}, \{3,6\}, \{3,5\}, \{1,6\}, \{1,5\}, \{1,2\}\}$. These extensions are ranked wrt the internal inconsistency of their complementary parts. So we can observe that (by abusing slightly the notation) for $\{2,4\}$ we have $INC[1,3,5,6] = 2 * 2 = 4$, for $\{3,6\}$ we have $INC[1,3,4,5] = 2*3 = 6$, for $\{3,5\}$ we have $INC[1,2,4,6] = 2*3 = 6$, for $\{1,6\}$ we have $INC[2,3,4,5] = 2 * 4 = 8$, for $\{1,5\}$ we have $INC[2,3,4,6] = 2 * 4 = 8$ and finally for $\{1,2\}$ we have $INC[3,4,5,6] = 2 * 4 = 8$.

Partitions	*Strength*	*Ranking*
$[3,5,6],[1,2,4]$	0.77	
$[1,3,5,6],[2,4]$	0.77	rank 1 for $[2,4]$
$[1,2,6],[3,4,5]$	0.66	
$[3,6],[1,2,4,5]$	0.66	rank 2 for $[3,6]$
$[1,3,6],[2,4,5]$	0.66	
$[2,4,6],[1,3,5]$	0.66	
$[1,2,4,6],[3,5]$	0.66	rank 2 for $[3,5]$
$[3,4,6],[1,2,5]$	0.66	
$[1,5,6],[2,3,4]$	0.66	
$[1,6],[2,3,4,5]$	0.55	rank 3 for $[1,6]$
$[2,3,6],[1,4,5]$	0.55	
$[1,2,3,6],[4,5]$	0.55	
$[4,6],[1,2,3,5]$	0.55	
$[1,4,6],[2,3,5]$	0.55	
$[2,3,4,6],[1,5]$	0.55	rank 3 for $[1,5]$
$[1,2,5,6],[3,4]$	0.55	
$[3,4,5,6],[1,2]$	0.55	rank 3 for [1,2]
$[2,6],[1,3,4,5]$	0.44	
$[1,3,4,6],[2,5]$	0.44	
$[5,6],[1,2,3,4]$	0.44	
$[2,3,5,6],[1,4]$	0.44	
$[1,2,3,5,6],[4]$	0.44	
$[4,5,6],[1,2,3]$	0.44	
$[1,4,5,6],[2,3]$	0.44	
$[6],[1,2,3,4,5]$	0.33	
$[1,2,3,4,6],[5]$	0.33	
$[2,5,6],[1,3,4]$	0.33	
$[2,4,5,6],[1,3]$	0.33	
$[1,2,4,5,6],[3]$	0.33	
$[1,3,4,5,6],[2]$	0.33	
$[2,3,4,5,6],[1]$	0.22	
$[1,2,3,4,5,6]$	0	

Table 1. Partitions of graph in Figure 1

5 Coherence theory and Preference based Argumentation (PAF)

In this section we present a relationship between coherence theory and *preference-based argumentation* (PAF) (see e.g. [2],[3]).

Before recalling the definition of a PAF, we provide a quick reminder on notions related to *preference* relations. We use the symbol $\succeq \subseteq \mathcal{A} \times \mathcal{A}$ to denote a preference relation on the set of arguments $\mathcal{A}$. $\succeq$ is a *partial preorder* i.e. a *reflexive* and *transitive* binary relation. So $a \succeq b$ means that a is *preferred* over b (or a is at *least as good as* b). We also use $\succ$ for representing a *strict preference* relation. More precisely, a is *strictly preferred* over b and it is represented as $a \succ b$ iff $a \succeq b$ and $b \not\succeq a$. Finally, we use the symbol $\sim$ for expressing the *indifference* relation between a and b. We say that $a \sim b$ iff $a \succeq b$ and $b \succeq a$. We are now ready to define a *PAF* as follows.

Definition 16 (PAF). *A preference-based argumentation framework is a tuple* $PAF = \langle \mathcal{A}, Att, \succeq, \rhd \rangle$ *where* $\mathcal{A}$ *is a set of arguments,* $Att \subseteq \mathcal{A} \times \mathcal{A}$ *is an irreflexive and symmetric attack (or conflict) relation,* $\succeq \subseteq \mathcal{A} \times \mathcal{A}$ *is a preference relation on the set of arguments* $\mathcal{A}$ *and* $\rhd$ *is a defeat relation composed by Att and* $\succeq$*. Here we define a defeat relation* $\rhd$ *s.t.* $\forall a, b \in \mathcal{A}, a \rhd b$ *iff* $(a, b) \in Att$ *and* $a \succ b$*.*

It follows directly from the definition that if $(a, b) \in Att$ and $a \sim b$, then $(a, b) \notin \rhd$. We note that different ways of defining the *defeat* relation may lead to different *PAFs*.
Based on the definition of *PAF* given above we can now establish a relationship between a coherence graph g and a $PAF(g)$ theory associated to it and defined as follows:

Definition 17 (Neg-unipolar graph-PAF relation). *Let* $g = \langle N, E, \psi \rangle$ *be a neg-unipolar graph, and* (A, R) *a partition of* g*. The PAF theory associated to* g *and* A *is* $PAF_g^A = \langle N, Att, \succeq, \rhd \rangle$*, where*

- $(a, b) \in Att$ *iff* $\{a, b\} \in E$
- $\forall a, b \in A$ $(a, b \in R)$ *it holds that* $a \sim b$
- $\forall a, b, a \in A$ *and* $b \in R$ *it holds that* $a \succ b$*.*

We can now interpret partitions of neg-unipolar coherence graphs in terms of extensions in PAF.

Proposition 5. *Let* $g = \langle N, E, \psi \rangle$ *be a neg-unipolar graph,* (A, R) *an optimal partition of* g *and* $i(N)$ *the nodes of* g *with degree 0. Then* $A \cup i(N)$ *is the unique grounded, preferred and stable extension of* PAF_g^A*.*

Proof. For any pair of nodes $a, b \in A$, it holds by construction that $a \not\rhd b$ and $b \not\rhd a$. Similarly for R. Therefore, A (and R) is conflict-free, and therefore $A \cup i(N)$ is conflict-free as well. On the other hand, the only attacks are from

nodes in A to nodes in R, therefore PAF_g^A is acyclic. Therefore, its unique stable extension coincides with its grounded extension, so we need to show that $A \cup i(N)$ is a stable extension.
Since it has already been proved that $A \cup i(N)$ is conflict-free, it suffices to show that for any $a_i \in N \setminus A \cup i(N) = R \setminus i(N)$, there is some $a_j \in A$ s.t. $a_j \rhd a_i$. Clearly, there must be a node $a_k \in N$ such that $\{a_i, a_k\} \in E$, because otherwise $a_i \in i(N)$. If $a_k \in A$ the result holds. Otherwise, it must be the case that for all nodes $a_k \in N$ such that $\{a_i, a_k\} \in E$, it holds that $a_k \in R$. But then $Str(g, A \cup \{a_i\}) > Str(g, A)$ which contradicts the assumption that (A, R) is optimal. □

Next, we introduce a relation between *bipolar coherence graphs* and *preference-based argumentation* (PAF).

Based on definitions 7 and 16 we propose a PAF construction for bipolar graphs. To do this, we consider that a negative arc represents an attack (or conflict) between the linked arguments (similar to the case of neg-unipolar graphs) while a positive link represents a mutual support between the linked arguments.

Definition 18 (Bipolar graph-PAF relation). *Let $g = \langle N, E, \psi \rangle$ be a bipolar graph and (A, R) a maximally coherent partition such that $|A| \geq |R|$. Then we define the associated preference-based argumentation framework $PAF_g^A = \langle N, Att, \succeq, \rhd \rangle$ as follows:*

- $\forall \{a, b\} \in E$ *s.t.* $\psi(\{a, b\}) = -1$, $(a, b), (b, a) \in Att$
- $\forall a, b \in A$ $(a, b \in R)$ *if* $(a, b) \in Att$ *it holds that* $a \sim b$
- $\forall a, b$, $a \in A$ *and* $b \in R$ *if* $(a, b) \in Att$ *it holds that* $a \succ b$

We can now interpret partitions of bipolar coherence graphs in terms of extensions in PAF.

Proposition 6. *Let $g = \langle N, E, \psi \rangle$ be a bipolar graph and (A, R) a maximally coherent partition such that $|A| \geq |R|$. Then A is the unique grounded, preferred and stable extension of PAF_g^A.*

Proof. Let (A, R) a maximally coherent partition and A be the accepted part s.t. $|A| \geq |R|$. Let also $Str(g, A)$ be the strength of this partition and C_A the set of satisfied constraints (see definition 2). We have to prove that A is the unique grounded, preferred and stable extension of the associated PAF_g^A. We know by construction that $\forall a, b \in A$, $(a, b) \notin \rhd$. So A is conflict-free. The same holds for R. We also know by construction that $\forall a, b$, if $a \in A$ and $b \in R$, then $(a, b) \in \rhd$. We know that g is a connected graph so it holds that $\forall b \in R$ there exists at least a negative link coming from an argument $a \in A$ and therefore it holds that $\forall b \in R, \exists a \in A$ s.t. $(a, b) \in \rhd$. Otherwise, we could have an argument $x \in R$ that could be added to A so that we would have $A' = A \cup \{x\}$. By definition 3 we know that $Str(g, A)$ is maximal which means that in that case we would have $Str(g, A) = Str(g, A')$ with $|A'| > |A|$. However this cannot be true because we know that the partition (A, R) is a maximally coherent partition. Contradiction. Thus A is also a maximal (wrt $\subseteq$) admissible extension and therefore it is stable extension. From the above we can also conclude that PAF_g^A is acyclic. Therefore A is also grounded and unique. □

6 Conclusion

In this work we have presented a theoretical analysis of the relation between argumentation and Paul Thagard's coherence theory. We studied several connections between the two theories by defining transformations between coherence graphs and some well known argumentation frameworks (classical systems (AF), weighted argument systems (WAS), and preference based argumentation frameworks (PAF)). We showed that coherence theory can be interpreted as a weighed argument system (WAS) and that partition maximization generates a ranking of extensions. We also saw that some coherence graphs can be translated into PAF systems and its partitions interpreted as PAF extensions.

We would like to complete the study of links between the two fields, as we believe there are many interesting relations that are left unexplored. For instance, we plan to study the relationship between coherence theory and bipolar argumentation [4]. Furthermore, we would like to extend the notion of argument to sets of nodes of a coherence graph, i.e. sets of claims that are internally coherent. Moreover, a study of the computational aspects of both fields may reveal potential gains that can be obtained by applying algorithms from one field to the other. Finally, we reiterate that the ultimate goal of this line of research is to integrate argumentation and coherence in applications domains such as legal reasoning and policy analytics.

References

1. Amaya, A.: The tapestry of reason: An inquiry into the nature of coherence and its role in legal argument. Bloomsbury Publishing (2015)
2. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Ann. Math. Artif. Intell. 34(1-3), 197–215 (2002)
3. Amgoud, L., Dimopoulos, Y., Moraitis, P.: Making decisions through preference-based argumentation. In: Brewka, G., Lang, J. (eds.) KR. pp. 113–123. AAAI Press (2008)
4. Cayrol, C., Lagasquie-Schiex, M.: On the acceptability of arguments in bipolar argumentation frameworks. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 8th European Conference, ECSQARU 2005, Barcelona, Spain, July 6-8, 2005, Proceedings. pp. 378–389 (2005)
5. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77(2), 321–358 (1995)
6. Dunne, P.E., Hunter, A., McBurney, P., Parsons, S., Wooldridge, M.: Weighted argument systems: Basic definitions, algorithms, and complexity results. Artif. Intell. 175(2), 457–486 (2011)
7. Joseph, S.: Coherence-Based Computational Agency, Monografies de l'Institut d'Investigació en Intelligència Artificial, vol. 45. CSIC (2011)
8. Joseph, S., Prakken, H.: Coherence-driven argumentation to norm consensus. In: ICAIL. pp. 58–67. ACM (2009)
9. Pasquier, P., Rahwan, I., Dignum, F., Sonenberg, L.: Argumentation and persuasion in the cognitive coherence theory. In: Dunne, P.E., Bench-Capon, T.J.M.

(eds.) COMMA. Frontiers in Artificial Intelligence and Applications, vol. 144, pp. 223–234. IOS Press (2006)
10. Prakken, H., Sartor, G.: Logical models of legal argumentation. Springer (1997)
11. Prakken, H., Sartor, G.: The role of logic in computational models of legal argument: a critical survey. In: Computational logic: Logic programming and beyond, pp. 342–381. Springer (2002)
12. Thagard, P.: Coherence in Thought and Action. MIT Press (2002)
13. Thagard, P.: Hot Thought. MIT Press (2006)
14. Thagard, P., Verbeurgt, K.: Coherence as constraint satisfaction. Cognitive Science 22(1), 1–24 (1998)