HAL
open science

Classes of Explanations for the Verification Problem in Abstract Argumentation

Sylvie Doutre, Théo Duchatelle, Marie-Christine Lagasquie-Schiex

To cite this version:

Sylvie Doutre, Théo Duchatelle, Marie-Christine Lagasquie-Schiex. Classes of Explanations for the Verification Problem in Abstract Argumentation. IRIT/RR–2022–09–FR, IRIT - Institut de Recherche en Informatique de Toulouse. 2022. hal-04016550

HAL Id: hal-04016550

https://ut3-toulouseinp.hal.science/hal-04016550v1

Submitted on 6 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Classes of Explanations for the Verification Problem in Abstract Argumentation

Sylvie Doutre
Théo Duchatelle
Marie-Christine Lagasquie-Schiex

IRIT, Université de Toulouse
118 route de Narbonne, 31062 Toulouse, France
{doutre, theo.duchatelle, lagasq}@irit.fr

Tech. Report
IRIT/RR- -2022- -09- -FR

October 2022

Abstract

The Verification Problem in abstract argumentation consists in checking whether a set is acceptable under a given semantics in a given argumentation graph. Explaining why the answer is so is the challenge tackled by this paper. Visual explanations in the form of subgraphs of the initial argumentation framework are defined. These explanations are grouped into classes, allowing an agent to select the explanation that suits them best among the several offered possibilities. Results are provided on how to use the visual aspects of our explanations to support the acceptability of a set of arguments under a semantics. Computational aspects of specific explanations are also investigated.

Contents

1 Introduction **4**

2 Background notions **6**
2.1 Argumentation and Graph Theory 6
2.2 Explanations in Argumentation 9

3 Classes of explanations **11**
3.1 Explanation about Conflict-freeness 12
3.2 Explanation about Defence 13
3.3 Explanation about Reinstatement 15
3.4 Explanation about Complement Attack 18

4 Properties of Explanations **19**
4.1 Some specific explanations 20
4.2 Properties of specific explanations 21

5 Computation of Explanations **36**
5.1 Maximal Explanations 36
5.2 From Maximal to Minimal Explanations 38

6 Conclusion and Future Work **45**

1 Introduction

Abstract Argumentation is increasingly studied as a formal tool to provide explanations of decisions made using an Artificial Intelligence system in the context of eXplainable Artificial Intelligence (XAI). The recent survey by [ČRA+21a] indicates that Argumentation can be used to generate explanations in various domains, notably in multi-agent systems as in [GTWX16], and that explanations for the argumentative process itself are also necessary.

The basic argumentation process is based on an abstract structure which takes the form of a directed graph, whose nodes are arguments and edges represent attacks between arguments [Dun95]. Characterising the acceptability of arguments can take the form of extension-based semantics: they define sets (extensions) of arguments which are collectively acceptable according to the semantics. The main questions which have been addressed so far in this context concern the global acceptability status of an argument or of a set of arguments, that is, why, under a given semantics, they belong to at least one extension (credulous acceptance) or to every extension (skeptical acceptance). The most common explanation approach consists in identifying set(s) of arguments which act as explanation(s), as in [FT15, BB21a, BB21b, UW21, LvdT20, BU21]. However, since the argumentative process of Abstract Argumentation already provides ways for selecting arguments, explaining this process by more selection of arguments (although different ones) may not be fully helpful. Moreover, this set approach does not highlight the attacks which are involved in the explanations.

Another question regarding the argumentation process concerns the *Verification Problem* Ver, defined as follows: given an Argumentation Framework $\mathcal{A}$, a set of arguments S and an extension-based semantics σ, "Is S an extension under σ in $\mathcal{A}$?". The answer to this problem is "yes" or "no". In order to explain why the answer is so, the *eXplanation Verification Problem* XVer can be defined using the question Q_σ: "Why is S (not) an extension under σ in $\mathcal{A}$?".

As far as we know, [BDDLS22] is one of the only approaches which has addressed this problem and which has provided answers for some acceptability semantics of [Dun95] in the form of relevant subgraphs, as in [SWW20, NJ20, RT21]. Such a visual approach is particularly of interest for human agents, graphs having been shown to be helpful for humans to comply with argumentation reasoning principles [VYT22]. This graph-based approach not only highlights arguments, but also attacks. In [BDDLS22], properties

that these answers satisfy have been established, depending on whether the answer to the corresponding verification problem is "yes" or "no". This methodology follows the line of [CČRT18] in that an explanation for a set S satisfying a semantic σ is a (set of) subgraph(s) G of $\mathcal{A}$ such that G satisfies a given graph property C. Another interesting point in [BDDLS22] is that the considered semantics are based on a modular definition, which allows the explanations to be decomposed.

A limitation of [BDDLS22] is however that, for each semantic principle, a *single* explanation subgraph is defined. It could be more realistic to consider classes (sets) of explanations. Indeed such classes would be particularly meaningful and useful when several agents, human or artificial, are involved around the explanation of a same problem, in that they offer a variety of answers, which all follow a same schema, but which may differ on their exact content. Any agent can choose or can be presented an explanation that suits them best, and any agent can understand an explanation given by another agent, different from theirs. Classes of explanations adapt to a wide set of agents. As an example, Fig. 1 shows three agents Ag1, Ag2, Ag3, which face an explanation question regarding a result output by a system, an argumentation solver for instance. The considered set is an extension, and they wonder why it is so. The three agents agree on the first part of the explanation (explaining why there is no conflict between the arguments), but they differ on the second (concerning the defence). However, their explanations all follow a same pattern, which consists in showing that every argument of the set is defended. A class of explanations is thus defined. It can be noticed that in this example, the agents specifically ask for an explanation on the result computed by the system. As in [BDDLS22], the approach that will be presented in this paper goes further, by considering the possibility that the answer to the Verification Problem is not known before an explanation be asked and given. In this case, the explanation graph and its interpretation offer at a same time the answer to the problem and a justification to this answer.

Only few related works can be found concerning this notion of classes of explanation. Such classes have already been proposed in [BU21] for the problem of credulous acceptance of an argument, where the authors consider explanation schemes made of several elements, one of them being fixed, the other ones varying from one explanation to another. Another related work is [BB21a] in which the authors define a parametric computation of explanations. As such, it is more the computation processes that are grouped in

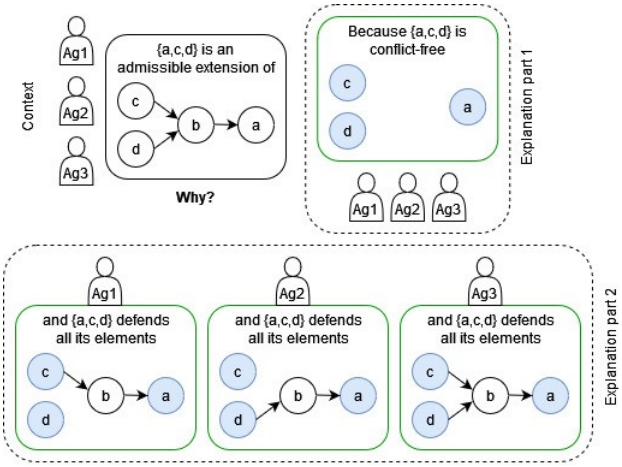

Figure 1: Explanations in a multi-agent context

classes, rather than the explanations (i.e. results of the processes) themselves.

Thus, our aim in the current paper is to build up on the approach of [BDDLS22] and to go further by defining classes of explanations following a generic methodology, applied to the semantics that [BDDLS22] addressed (conflict-free, admissible, stable, complete). Additional properties (emptyness, uniqueness, maximality, minimality, computation) of explanations on these newly defined classes will be investigated.

Sec. 2 recalls background notions relative to abstract argumentation, graph theory, and presents the explanation approach defined in [BDDLS22]. The definition of our classes of explanations is given in Sec. 3, Sec. 4 studies their properties, and Sec. 5 shows how to compute their maximal and minimal explanations. Sec. 6 concludes and presents some future works.

2 Background notions

2.1 Argumentation and Graph Theory

We begin by recalling basic notions on Abstract Argumentation.

Definition 1 ([Dun95])**.** A *Dung's argumentation framework* (AF) is an ordered pair (A, R) such that $R \subseteq A \times A$.

Each element $a \in A$ is called an *argument* and aRb means that a *attacks*

b. For $S \subseteq A$, we say that S attacks $a \in A$ iff bRa for some $b \in S$. Any argumentation framework can be represented as a directed graph (the nodes are the arguments and the edges correspond to the attack relation).

The main asset of Dung's approach is the definition of semantics using some basic properties in order to define sets of acceptable arguments, as follows.

Definition 2 ([Dun95])**.** Let $\mathcal{A} = (A, R)$. An argument $a \in A$ is *acceptable* wrt $S \subseteq A$ iff for all $b \in A$, if bRa then $\exists c \in S$ st cRb.

Definition 3 ([Dun95])**.** Given $\mathcal{A} = (A, R)$, a subset S of A is:

- a *conflict-free set* iff there are no a and b in S such that a attacks b,
- an *admissible set* iff S is conflict-free and for any $a \in S$, a is acceptable wrt S,
- a *complete extension* iff S is admissible and for any $a \in A$, if a is acceptable wrt S then $a \in S$,
- a *stable extension* iff S is conflict-free and S attacks any $a \in A \setminus S$.

The Verification Problem for the four semantics given in Def. 3 can be solved in polynomial time, as indicated by [DD18].

Since an AF can be represented using directed graphs, we also need to recall some basic notions of Graph Theory.

Definition 4. Let $G = (V, E)$ and $G' = (V', E')$ be two graphs.

- G' is a *subgraph* of G iff $V' \subseteq V$ and $E' \subseteq E$.[1]
- G' is a *strict subgraph* of G iff it is a subgraph of G and either $V' \subset V$ or $E' \subset E$.[2]
- G' is an *induced subgraph of G by V'* if G' is a subgraph of G and for all $a, b \in V', (a, b) \in E'$ iff $(a, b) \in E$. G' is denoted as $G[V']_V$.
- G' is a *spanning subgraph of G by E'* if G' is a subgraph of G and $V' = V$. G' is denoted as $G[E']_E$.

[1] G is then a *supergraph* of G'

[2] G is then a *strict supergraph* of G'

A subgraph G' of G is included in G. In an induced subgraph G' of G by a set of *vertices* S, some vertices of G can be missing but all the edges concerning the kept vertices are present. In a spanning subgraph G' of G by a set of *edges* S, all the vertices of G are present but some edges of G can be missing.

Induced and spanning subgraphs are examples of ways to compute a graph from another single graph. Another operation producing a new graph from other ones is the union that represents the aggregation of the information contained in the two graphs:

Definition 5 (Graph union)**.** Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two graphs. The *union* of G_1 and G_2 is defined by $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$.

Let us consider also a particular kind of graphs, bipartite graphs, whose set of vertices can be split in two disjoint sets and in which every arc connects a vertex of one part to a vertex of the other part:

Definition 6 (Bipartite Graph)**.** Let $G = (V, E)$ be a graph. G is *bipartite* (with *parts* T and U) iff there exist $T, U \subseteq V$ such that $T \cup U = V$ and $T \cap U = \varnothing$ (T and U are a partition of V) and for every $(a, b) \in E$, either $a \in T$ and $b \in U$, or $a \in U$ and $b \in T$. G will be denoted with (T, U, E) and U is the *complement part of* T (and vice-versa).

Some important functions can be defined over graphs.

Definition 7 (Successor and Predecessor functions)**.** Let $G = (V, E)$ be a graph. The *successor* function of G is the function $E^+ : V \mapsto 2^V$ such that $E^+(v) = \{u \mid (v, u) \in E\}$ and the *predecessor* function of G is the function $E^- : V \mapsto 2^V$ such that $E^-(v) = \{u \mid (u, v) \in E\}$. Let S be a set of vertices, $E^+(S) = \bigcup_{v \in S} E^+(v)$ and $E^-(S) = \bigcup_{v \in S} E^-(v)$.

Let $n \geq 0$. The *n-step successor* (resp. *predecessor*) *function* of G is $E^{+n}(v) = \overbrace{E^+ \circ \cdots \circ E^+}^{\text{n times}}(v)$ (resp. $E^{-n}(v) = \overbrace{E^- \circ \cdots \circ E^-}^{\text{n times}}(v)$). By convention, we have $E^{+0}(v) = E^{-0}(v) = \{v\}$.[3]

Considering an argumentation framework, the successor (resp. predecessor) function represents the arguments that are attacked by (resp. are the attackers of) some argument(s). An AF being usually denoted by (A, R),

[3]Note that $E^{+1}(v) = E^+(v)$ and $E^{-1}(v) = E^-(v)$

the successor and predecessor functions are thus denoted R^+ and R^- in this context.

We then recall some notions on vertices having a particular status in a graph.

Definition 8 (Source, Sink, Isolated vertex)**.** Let $G = (V, E)$ be a graph and v be a vertex of G. v is said to be a *source* iff $E^-(v) = \varnothing$ and it is said to be a *sink* iff $E^+(v) = \varnothing$. v is said to be *isolated* iff it is both a source and a sink.

Thus, *sources* (resp. *sinks*) are vertices that may only be origins (resp. endpoints) of arcs. *Isolated* vertices are those that are connected to no other vertices.

2.2 Explanations in Argumentation

We recall the main definitions of what are explanations in [BDDLS22] but only for those answering the questions about semantics results in abstract argumentation. These questions are defined as follows: let σ represent a semantics among conflict-freeness, admissibility, completeness and stability, and given an argumentation framework $\mathcal{A} = (A, R)$ and some set $S \subseteq A$,

Q_σ: `Why is S (not) an extension under` σ `in` $\mathcal{A}$`?`

In order to answer these questions, and hence to provide explanations, the authors in [BDDLS22] use the decomposition of semantics into principles. The idea is to identify some properties that can be used to provide a modular characterization of semantics. We refer the reader to [DM16] for further details. Given a set S, the following principles are considered:

Conflict-freeness (CF):	No internal conflicts in S
Defence (Def):	$\forall x \in S$, x is acceptable wrt S
Reinstatement ($Reins$):	$\forall x$ acceptable wrt S, $x \in S$
Complement Attack (CA):	S attacks all arguments not in S

Note that, in [BDDLS22], the reinstatement principle has been splitted into two sub-principles. Indeed, to decide whether a set S of arguments contains all the arguments acceptable wrt S, one must consider on the one hand the arguments that are unattacked and thus acceptable by lack of attackers (sub-principle denoted by $Reins_1$), and on the other hand the arguments for which S defeats all the attackers (sub-principle denoted by $Reins_2$).

The following has been proven in [DM16].

Proposition 1. *Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. S is:*

Conflict-free	*iff S respects*	$\{CF\}$
Admissible	*iff S respects*	$\{CF,\ Def\}$
Complete	*iff S respects*	$\{CF,\ Def,\ Reins_1,\ Reins_2\}$
Stable	*iff S respects*	$\{CF,\ CA\}$

With this result, a straightforward answer arises for Q_σ: a set S is an extension under semantics σ because it respects all the principles listed for σ in Prop. 1. This moves the burden of explanation from semantics to principles. So, in order to answer Q_σ, we are going to answer intermediate questions on principles. Let $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$ represent a principle. Given an argumentation framework $\mathcal{A} = (A, R)$ and some set $S \subseteq A$, the questions we will define answers for are:

Q_π: `Why does (not) S respect principle` π`?`

[BDDLS22] provides definitions for visual answers to these questions. These answers take the form of a graph. This allows for the answers to be drawn, as well as to study their visual (i.e. structural) properties. More precisely, as argumentation frameworks are graphs themselves, the answers given in [BDDLS22] are subgraphs of an argumentation framework.

Definition 9 ([BDDLS22])**.** Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$. $G_\pi(S)$ is defined as:

$$
\begin{aligned}
G_{CF}(S) &= \mathcal{A}[S]_V \\
G_{Def}(S) &= (\mathcal{A}[S \cup R^{-1}(S)]_V) \\
&\quad [\{(a,b) \in R \mid (a \in R^{-1}(S) \text{ and } b \in S) \\
&\qquad \text{or } (a \in S \text{ and } b \in R^{-1}(S))\}]_E \\
G_{Reins_1}(S) &= \mathcal{A}[\{a \in A | R^{-}(a) = \varnothing\}]_V \\
G_{Reins_2}(S) &= (\mathcal{A}[S \cup R^2(S) \cup R^{-1}(R^2(S))]_V) \\
&\quad [\{(a,b) \in R \mid (a \in R^{-1}(R^2(S)) \text{ and } b \in R^2(S)) \\
&\qquad \text{or } (a \in S \text{ and } b \in R^{-1}(R^2(S)))\}]_E \\
G_{CA}(S) &= \mathcal{A}[\{(a,b) \in R \mid a \in S \text{ and } b \notin S\}]_E
\end{aligned}
$$

Moreover the interpretation of these subgraphs can be done using a "checking procedure" in order to explicitly identify if the given subset satisfies or not the concerned principle:

Definition 10 ([BDDLS22])**.** Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$. Let G be a subgraph of $\mathcal{A}$. The checking procedure $C_\pi(G)$ is defined as:

$$\begin{array}{rcl}
C_{CF}(G) & = & \text{there are no attacks in } G \\
C_{Def}(G) & = & \text{there are no source vertices in } R^{-1}(S) \text{ in } G \\
C_{Reins_1}(G) & = & \text{all vertices in } G \text{ are in } S \\
C_{Reins_2}(G) & = & \text{all vertices in } R^2(S) \setminus S \text{ are the endpoint of an arc whose} \\
 & & \text{origin is a source vertex in } G \\
C'_{Reins_2}(G) & = & \text{all vertices in } R^2(S) \setminus S \text{ are the endpoint of an arc whose} \\
 & & \text{origin is a source vertex or is in } R^2(S), \text{ in } G \\
C_{CA}(G) & = & \text{there are no isolated vertices in the complement part of} \\
 & & S \text{ in } G
\end{array}$$

In [BDDLS22], it has been proven that, for each principle π, the subgraph G_π associated with the corresponding checking procedure C_π provides an explanation that answers question Q_π.[4] More precisely, if a set S respects a principle π, then G_π verifies C_π, otherwise it does not. When the principles are combined into a semantics σ, the answer to Q_σ is the corresponding set of subgraphs along with their corresponding checking procedures.

This allows this form of explanation to be used for two purposes as indicated in the introduction: when the answer to the corresponding verification problem is known, that is, when we know that a set is (resp. is not) acceptable under a given semantics or principle, G_π on which C_π is (resp. is not) verified, offers a visual explanation of the situation, answering XVer. When the answer to the verification problem is not known, G_π and the verification of whether C_π holds or not offers at the same time an answer to Ver and an explanation of this answer.

3 Classes of explanations

In this section, we give the essential elements that compose our approach: the classes of explanations. Knowing that we want to extend the notions and the results presented in [BDDLS22], the definition of these classes allows to recover the explanations described in [BDDLS22] but also it results in the *possibility of producing several explanations for the same question.* Thus, it

[4] This result is slightly more complex in the case of reinstatement. See [BDDLS22] and Sec. 3.3.

takes into account the different points of view that may emerge and focus on different aspects.

Hence, for each principle π, we first define our explanations so that they contain at least enough information to be able to decide whether or not S respects π. Then, we also prove that our explanations can be used in conjunction with the checking procedures defined in [BDDLS22] (and recalled in Def. 10) in order to obtain useful explanations as in [BDDLS22].

3.1 Explanation about Conflict-freeness

To decide whether a set S of arguments is conflict-free, one must know whether or not there are attacks among its arguments. Thus, we firstly require our explanation to contain only arguments of S, and secondly to contain only attacks between these arguments. However, with only these two constraints, it may happen that no attacks are displayed on the explanation when there are some in the original framework, leading at best to an impossibility to decide or at worst, an incorrect decision. Hence, we add a third constraint, which is that if conflicts exist between arguments of S, then at least one must be present in the explanation.

Definition 11. Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{(a, b) \in R \mid a, b \in S\}$. The subgraph (A', R') of $\mathcal{A}$ is an explanation to Q_{CF} iff

- $A' = S$
- $R' \subseteq X$
- If $X \neq \varnothing$, then $R' \neq \varnothing$

Note that the subgraph G_{CF} defined in [BDDLS22] (and recalled in Def. 9) obviously belongs to our class of explanations for conflict-freeness. Moreover, in [BDDLS22], a result concerning the structural property of explanations for conflict-freeness has been given: a set of arguments is conflict-free iff there is no attack in the subgraph corresponding to its explanation (checking procedure C_{CF} recalled in Def. 10). This result can be extended to all the subgraphs captured by our class of explanations.

Theorem 1. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an explanation to Q_{CF}. S is conflict-free iff $C_{CF}(A', R')$ is satisfied by S.*

Proof. Let $X = \{(a, b) \in R \mid a, b \in S\}$.

Suppose that S is conflict-free and that there is an attack (a, b) in (A', R'). By Def. 11, we have that $R' \subseteq X$, so $a, b \in S$ and that $(a, b) \in R$. This contradicts Def. 3 on conflict-freeness.

Suppose now that there are no attacks in (A', R') and that S is not conflict-free. By Def. 3, there exists $a, b \in S$ such that $(a, b) \in R$. Thus, $X \neq \varnothing$ and by Condition 3 of Def. 11, $R' \neq \varnothing$. This contradicts the absence of attacks in (A', R'). □

This provides a way of deciding whether a set is conflict-free based on an explanation to Q_{CF}. Note that this also provides a way of deciding whether a set is *not* conflict-free, hence the possibility of handling the negative version of Q_{CF}. The same goes for all the other equivalence results concerning the other principles.

3.2 Explanation about Defence

To decide whether a set S of arguments contains only arguments that are acceptable wrt S, one must know whether or not this set defeats all its attackers. Thus, we firstly require our explanation to contain only arguments of S and its attackers, and secondly to contain only attacks from S to its attackers and vice versa. To make sure the attackers are spotted as such, we further require that all the attacks of the second type are contained in the explanation. However, with only these two constraints, it may happen that no attacks targeting a specific attacker are displayed on the explanation when there are some in the original framework. As we wish the explanation to show how S defends itself, this situation is certainly undesirable. Hence, we add a third constraint, which is that if an attacker is attacked by S, then at least one attack from S to this attacker must be present in the explanation.

Definition 12. Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. Consider $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. The subgraph (A', R') of $\mathcal{A}$ is an explanation to Q_{Def} iff

- $A' = S \cup R^{-1}(S)$
- $X \subseteq R' \subseteq X \cup Y$
- $\forall b \in R^{-1}(S)$, if $b \in R^{+1}(S)$, then $\exists (a, b) \in R'$ with $a \in S$

Note that the subgraph G_{Def} defined in [BDDLS22] (and recalled in Def. 9) obviously belongs to our class of explanations for defence. Moreover it was shown in [BDDLS22] that a conflict-free set of arguments defends all its arguments iff there is no source vertex among its attackers in $G_{Def}(S)$ (checking procedure C_{Def} recalled in Def. 10). This result can be extended to all the subgraphs captured by our class of explanations.

Theorem 2. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ be a conflict-free set of arguments and (A', R') be an explanation to Q_{Def}. $S \subseteq F_{\mathcal{A}}(S)$ iff $C_{Def}(A', R')$ is satisfied by S.*

Proof. Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$.

Assume that $S \subseteq F_{\mathcal{A}}(S)$. Suppose now that there is a source vertex b in $R^{-1}(S)$ in (A', R'). By Def. 8, we have that $R'^{-1}(b) = \varnothing$, which means there exists no $a \in A'$ such that $(a, b) \in R'$. Because S is conflict-free, $S \cap R^{-1}(S) = \varnothing$. As $b \in R^{-1}(S)$ and $R' \subseteq X \cup Y$ (Def. 12), it must be the case that $a \in S$. Hence, there exists no $a \in S$ such that $(a, b) \in R'$. So following Condition 3 of Def. 12, there exists no $a \in S$ such that $(a, b) \in R$. As $b \in R^{-1}(S)$, there exists $c \in S$ such that $(b, c) \in R$. Hence, we know that there exists $b \in A$ with $(b, c) \in R$ for some $c \in S$ and such that there exists no $a \in S$ with $(a, b) \in R$. This contradicts the assumption that $S \subseteq F_{\mathcal{A}}(S)$.

Assume now that there are no source vertices in $R^{-1}(S)$ in (A', R'). Suppose that there is some $c \in S$ such that c is not acceptable wrt S. By Def. 2, this means that there exists $a \in A$ such that $(a, c) \in R$ and there is no $b \in S$ with $(b, a) \in R$. Firstly, notice that by Def. 7, $a \in R^{-1}(c)$ and so $a \in R^{-1}(S)$. Secondly, since $c \in S$, $a \in R^{-1}(S)$ and $(a, c) \in R$, $(a, c) \in X$ and so, by Def. 12, it holds that $c, a \in A'$ and $(a, c) \in R'$. Thus, by assumption, a is not a source vertex in (A', R'). Subsequently, there exists $b \in A'$ such that $(b, a) \in R'$. Since $a \in R^{-1}(S)$ and S is conflict-free (i.e. $S \cap R^{-1}(S) = \varnothing$), it holds that $b \in S$. In addition, as $R' \subseteq R$, we deduce that $(b, a) \in R$. Thus, we have $c \in S$ such that c is not acceptable w.r.t. S and for any $a \in A$ with $(a, c) \in R$, there is $b \in S$ with $(b, a) \in R$, a contradiction. □

Additionally, the next result extends a similar result given in [BDDLS22] providing more insight on the behavior of an explanation for defence: when computed using a conflict-free set, the explanation for defence takes the form of a bipartite graph.

Proposition 2. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an explanation to Q_{Def}. If S is conflict-free, (A', R') is a bipartite graph and S can always be one of its parts.*

Proof. Suppose S is conflict-free. So, $S \cap R^{-1}(S) = \varnothing$. Since by Def. 12 $A' = S \cup R^{-1}(S)$, S and $R^{-1}(S)$ then form a partition of A'. According to Def. 6, we must then show that for every $(a, b) \in R'$, either $a \in S$ and $b \in R^{-1}(S)$ or $a \in R^{-1}(S)$ and $b \in S$. This is given by Def. 12. □

The two previous results can thus be used to decide whether a set of arguments effectively defends all its arguments or if it is not conflict-free.

3.3 Explanation about Reinstatement

The first part of the reinstatement principle concerns unattacked arguments. All these arguments are acceptable wrt S and should thus belong to S. Thus, we firstly require our explanation to contain only unattacked arguments, and secondly to contain no attacks (which results from the only arguments displayed being unattacked). However, with only these two constraints, it may happen that an unattacked argument not belonging to S is not displayed on the explanation. Hence, we add a third constraint, which is that if there exists unattacked arguments that are not in S, then at least one must be present in the explanation.

Definition 13. Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. The subgraph (A', R') of $\mathcal{A}$ is an explanation to Q_{Reins_1} iff

- $S \cap X \subseteq A' \subseteq X$
- $R' = \varnothing$
- If $(A \setminus S) \cap X \neq \varnothing$, then $\exists a \in (A \setminus S) \cap X$ with $a \in A'$

The second part concerns arguments for which S defeats the attackers. These arguments must belong to S if S defeats all of their attackers. Thus, we firstly require our explanation to contain the arguments of S, the arguments that S defends (two steps of the attack relation from S), and the attackers of these arguments. Secondly, we require it to contain only the attacks from S to the attackers and from the attackers to the arguments S defends. In addition, we require that all the attacks of the second type are displayed

on the explanation, so that none is missed. However, with only these two constraints, it may happen that no attacks targeting a specific attacker are displayed on the explanation when there are some in the original framework. Hence, we add a third constraint, which is that if an attacker is attacked by S, then at least one attack from S to this attacker must be present in the explanation.

Definition 14. Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. Consider $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. The subgraph (A', R') of $\mathcal{A}$ is an explanation to Q_{Reins_2} iff

- $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$
- $X \subseteq R' \subseteq X \cup Y$
- For every $b \in R^{-1}(R^{+2}(S))$, if $b \in R^{+1}(S)$, then $\exists (a, b) \in R'$ with $a \in S$

Note that the subgraph G_{Reins_1} (resp. G_{Reins_2}) defined in [BDDLS22] (and recalled in Def. 9) obviously belongs to our class of explanations for the first (resp. second) part of the principle of reinstatement. Moreover in the case of reinstatement, two results have been proven in [BDDLS22] and can be extended to all the subgraphs captured by our class of explanations.

The first one shows how to conclude that a set contains all the arguments that it effectively defends from both parts of the explanation on reinstatement.

Theorem 3. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$, (A', R') be an explanation to Q_{Reins_1} and (A'', R'') be an explanation to Q_{Reins_2}. If $C_{Reins_1}(A', R')$ and $C_{Reins_2}(A'', R'')$ are satisfied by S then $F_{\mathcal{A}}(S) \subseteq S$.*

Proof. Let $X_1 = \{a \in A \mid R^{-1}(a) = \varnothing\}$, $X_2 = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$.

Assume that $A' \subseteq S$ and that all vertices in $R^{+2}(S) \setminus S$ are the endpoint of an arc whose origin is a source vertex in (A'', R''). In other words, $A' \subseteq S$ and for every $x \in R^{+2}(S) \setminus S$, there exists $y \in A''$ such that $(y, x) \in R''$ and $R''^{-1}(y) = \varnothing$. Consider $a \in F_{\mathcal{A}}(S)$. This means that for every $b \in A$ such that $(b, a) \in R$, there exists $c \in S$ with $(c, b) \in R$. We must show that $a \in S$. Suppose first that a is not attacked in $\mathcal{A}$. That is to say, $R^{-1}(a) = \varnothing$, and so, $a \in X_1$. By assumption, $A' \subseteq S$. This means that $(A' \setminus S) = \varnothing$. In particular, $(A' \setminus S) \cap X_1 = \varnothing$, so by Condition 3 of Def. 13, we have

$(A \setminus S) \cap X_1 = \varnothing$. Since $a \in X_1$, $X_1 \neq \varnothing$, so we deduce that either $A \setminus S = \varnothing$ or $X_1 \subseteq S$. In the first case, we conclude $A \subseteq S$ and thus $a \in S$. In the second case, we have that $X_1 \cap S = X_1$, and so $X_1 \subseteq A'$. As $A' \subseteq S$ by assumption, $a \in S$. Suppose now that $R^{-1}(a) \neq \varnothing$. By Def. 7, we have $a \in R^{+2}(S)$ and for every $b \in A$ such that $(b, a) \in R$, $b \in R^{-1}(R^{+2}(S))$. As such, $(b, a) \in X_2$ and $(c, b) \in Y$. So, by Def. 14, we have that $a, b, c \in A''$ and $(b, a) \in R''$. In addition, by Condition 3 of Def. 14, as $b \in R^{+1}(S)$, there exists $(c'', b) \in R''$ with $c'' \in S$. Thus, for every $b \in A''$ such that $(b, a) \in R''$, $R''^{-1}(b) \neq \varnothing$. Hence, all $b \in A''$ such that $(b, a) \in R''$ are not source vertices. Consequently, by assumption, $a \notin R^{+2}(S) \setminus S$, but we know that $a \in R^{+2}(S)$. It follows that $a \in R^{+2}(S) \cap S$, and thus that $a \in S$. □

The second results shows the behavior of both parts of the explanation on reinstatement if computed on a set that contains all the arguments it effectively defends.

Theorem 4. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$, (A', R') be an explanation to Q_{Reins_1} and (A'', R'') be an explanation to Q_{Reins_2}. If $F_{\mathcal{A}}(S) \subseteq S$ then $C_{Reins_1}(A', R')$ and $C'_{Reins_2}(A'', R'')$ are satisfied by S.*

Proof. Let $X_1 = \{a \in A \mid R^{-1}(a) = \varnothing\}$, $X_2 = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$.

Assume that $F_{\mathcal{A}}(S) \subseteq S$. Suppose now that either $A' \not\subseteq S$ or there is a vertex a in $R^{+2}(S) \setminus S$ that is not the endpoint of an arc whose origin is a source vertex or in $R^{+2}(S)$ in (A'', R''). In the first case, by Def. 13 $A' \subseteq X_1$. So, we have that there exists $x \in A$ such that $R^{-1}(x) = \varnothing$ and $x \notin S$. However, by Def. 2, this means that $x \in F_{\mathcal{A}}(S)$ and $x \notin S$, a contradiction. In the second case, we have $a \notin S$ and for every $b \in A''$ such that $(b, a) \in R''$, $R''^{-1}(b) \neq \varnothing$ and $b \notin R^{+2}(S)$. In other words, $b \notin R^{+2}(S)$ and there exists $c \in A''$ with $(c, b) \in R''$. By Def. 14, $X_2 \subseteq R'' \subseteq X_2 \cup Y$. So, since $a \in R^{+2}(S) \setminus S$ and $b \notin R^{+2}(S)$, we thus know that $b \in R^{-1}(R^{+2}(S))$. In addition, also because $b \notin R^{+2}(S)$, it must be the case that $c \in S$. So, for every $b \in A''$ such that $(b, a) \in R''$, there exists $c \in S$ with $(c, b) \in R''$. By Def. 14 again, we deduce that for every $b \in A$ such that $(b, a) \in R$, there exists $c \in S$ with $(c, b) \in R$. By Def. 2, this means that a is acceptable wrt S and so that $a \in F_{\mathcal{A}}(S)$. Hence, by assumption, $a \in S$, a contradiction. □

From Th. 3 and 4 follows the next corollary, which shows an equivalence result:

Corollary 1. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ such that $R^2(S)$ is conflict-free, (A', R') be an explanation to Q_{Reins_1} and (A'', R'') be an explanation to Q_{Reins_2}. $F_{\mathcal{A}}(S) \subseteq S$ iff $C_{Reins_1}(A', R')$ and $C_{Reins_2}(A'', R'')$ are satisfied by S.*

3.4 Explanation about Complement Attack

To decide whether a set S of arguments attacks its complement, one must know whether or not all the arguments not in S are attacked by S. Thus, we firstly require our explanation to contain all the arguments of the original framework (S and its complement), and secondly to contain only attacks from S to arguments not in S. However, with only these two constraints, it may happen that no attacks targeting a specific argument outside of S are displayed on the explanation when there are some in the original framework. Hence, we add a third constraint, which is that if an argument not in S is attacked by S, then at least one attack from S to this argument must be present in the explanation.

Definition 15. Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. The subgraph (A', R') of $\mathcal{A}$ is an explanation to Q_{CA} iff

- $A' = A$
- $R' \subseteq X$
- $\forall b \in A \setminus S$, if $b \in R^{+1}(S)$, then $\exists (a, b) \in R'$ with $a \in S$

Note that the subgraph G_{CA} defined in [BDDLS22] (and recalled in Def. 9) obviously belongs to our class of explanations for the principle of complement attack. Moreover concerning this principle, it was proven in [BDDLS22] that a set of arguments attacks its complement iff there are no isolated vertices in $G_{CA}(S)$. We extend this result to our class of explanations for complement attack.

Theorem 5. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an explanation to Q_{CA}. $A \setminus S \subseteq R^{+1}(S)$ iff $C_{CA}(A', R')$ is satisfied by S.*

Proof. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$.

Assume that $A \setminus S \subseteq R^{+1}(S)$. Suppose now that there is an isolated vertex a in $A' \setminus S$. By Def. 15, we know that $A' = A$. As such, there exists $a \in A \setminus S$ such that a is isolated in (A', R'). By Def. 8, this means in particular

that $R'^{-1}(a) = \varnothing$ and thus that there is no $b \in A'$ with $(b, a) \in R'$. Again, in particular, we have that there is no $b \in S$ with $(b, a) \in R'$. However, by Condition 3 of Def. 15, we deduce that there is no $b \in S$ with $(b, a) \in R$. Since $a \in A \setminus S$, this contradicts the assumption that $A \setminus S \subseteq R^{+1}(S)$.

Suppose now that there are no isolated vertices in $A' \setminus S$ in (A', R') and that $A \setminus S \not\subseteq R^{+1}(S)$. From the first assumption, by Def. 15, we have that there are no isolated vertices in $A \setminus S$ in (A', R'). In particular, by Def. 8, we know that there is no $a \in A \setminus S$ such that $R'^{-1}(a) = \varnothing$, or equivalently, for every $a \in A \setminus S$, there exists $b \in A$ such that $(b, a) \in R'$. By Def. 15, we have that $R' \subseteq X$, thus we deduce that for every $a \in A \setminus S$, there exists $b \in S$ such that $(b, a) \in R'$. From the second assumption, we have that there exists some $c \in A \setminus S$ such that there is no $b \in S$ with $(b, c) \in R$. By Def. 15 (Conditions 1 and 2), we deduce that there exists some $c \in A \setminus S$ such that there is no $b \in S$ with $(b, c) \in R'$, a contradiction of the first assumption. □

In addition, like for defence, an additional result gives more precision on the general behavior of an explanation for complement attack (extending a similar result given in [BDDLS22]): the explanation subgraph is always a bipartite graph with the arguments of S being the only possible origins for attacks.

Proposition 3. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an explanation to Q_{CA}. (A', R') is a bipartite graph, S can always be one of its parts and all vertices in S are sources in it.*

Proof. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$.

By Def. 15, we have that $A' = A$ and $R' \subseteq X$. An obvious partition of A based on S is of course S and $A \setminus S$. As $R' \subseteq X$, by Def. 6, (A', R') is a bipartite graph. In addition, since there is no $(b, a) \in R'$ such that $b \in A \setminus S$ and $a \in S$, it holds that for every $a \in S$, $R'^{-1}(a) = \varnothing$. Thus, by Def. 8, every vertex of S is a source vertex in (A', R'). □

4 Properties of Explanations

We now turn to a formal study of our explanations according to several general properties. This will allow to highlight some particular kinds of explanations, as well as better understand their behavior. The notions we are interested in are the following: minimality, maximality, emptyness and uniqueness.

4.1 Some specific explanations

In this section, we identify some specificities that could be respected by our explanations.

Minimality A minimal explanation is an explanation which contains the least possible amount of information. In a sense, a minimal explanation only provides what is required to explain. Since our explanations are graphs, this will rely on the notion of subgraphs: a minimal explanation is an explanation such that none of its strict subgraphs is also an explanation.

Definition 16. Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. The subgraph (A', R') of $\mathcal{A}$ is a minimal explanation that answers Q_π iff there is no subgraph (A'', R'') of $\mathcal{A}$ which is also an explanation that answers Q_π such that (A'', R'') is a strict subgraph of (A', R').

Maximality A maximal explanation is an explanation which contains all the possible amount of information. One could argue that a maximal explanation in fact provide everything that might be relevant to explain, even if it might be redundant. Since our explanations are graphs, this will rely on the notion of supergraphs: a maximal explanation is an explanation such that none of its strict supergraphs is also an explanation.

Definition 17. Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. The subgraph (A', R') of $\mathcal{A}$ is a maximal explanation that answers Q_π iff there is no subgraph (A'', R'') of $\mathcal{A}$ which is also an explanation that answers Q_π such that (A', R') is a strict subgraph of (A'', R'').

Example. Going back to Fig. 1, there are 3 different explanations for defence. The explanations of Agents 1 and 2 are minimal, while the explanation of Agent 3 is maximal.

Emptyness The notion of empty explanation is one that should be avoided when providing explanations, in the sense that it somewhat represents the incapacity of the system to answer the question that has been asked. Hence, we will study whether or not such empty explanations can occur, and if so, in which circumstances. Since our explanations are graphs, an empty explanation is quite straightforwardly the empty graph.

Definition 18. Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. The subgraph (A', R') is an empty explanation that answers Q_π iff $(A', R') = (\varnothing, \varnothing)$.

Uniqueness We consider an explanation to be unique when there is only one of its kind. Although we defined classes of explanations in an attempt to represent all the different points of view that could emerge as to how to answer a question, in some situations, there can only be one way to answer that question. Since our explanations are graphs, a unique explanation is such that no other graph satisfies the same conditions.

Definition 19. Let $\mathcal{A} = (A, R)$ be a graph. The subgraph (A', R') is a unique explanation that answers Q_π iff there is no subgraph (A'', R'') with $(A'', R'') \neq (A', R')$ which is also an explanation that answers Q_π.

4.2 Properties of specific explanations

Here, we provide the results of our formal study on our explanations using the properties we mentioned above. We begin with empty explanations. The results show that, although empty explanations can occur, they only do so in very specific situations.

The following theorem establishes a characterisation of empty explanations. Note that it extends a similar result given in [BDDLS22], confirming that our approach generalises [BDDLS22].

Theorem 6. *Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. $(\varnothing, \varnothing)$ is an explanation that answers*

1. *Q_π with $\pi \in \{CF, Def, Reins_2\}$ iff $S = \varnothing$.*
2. *Q_{Reins_1} iff $\{a \in A \mid R^{-1}(a) = \varnothing\} = \varnothing$.*
3. *Q_{CA} iff $\mathcal{A} = (\varnothing, \varnothing)$.*

Proof.

1. Firstly, consider $\pi = CF$. Suppose that $(\varnothing, \varnothing)$ is an answer to Q_{CF}. According to Def. 11, we have that $A' = \varnothing = S$. So the "only if" part is satisfied. Suppose now that $S = \varnothing$ and consider the empty graph $(A' = \varnothing, R' = \varnothing)$. We must prove that it is an answer to Q_{CF}. The first condition, $A' = \varnothing = S$, is respected by supposition. Since $R' = \varnothing$, the second condition is respected as well. Moreover, since $S = \varnothing$, by definition $X = \varnothing$ and the third condition is trivially satisfied. Thus, the "if" part is satisfied.

Secondly, consider $\pi = \mathit{Def}$. Suppose that $(\varnothing, \varnothing)$ is an answer to Q_{Def}. By Def. 12, we have that $S \cup R^{-1}(S) = \varnothing$ and thus $S = \varnothing$. So the "only if" part is satisfied. Suppose now that $S = \varnothing$ and consider the empty graph $(A' = \varnothing, R' = \varnothing)$. We must prove that it is an answer to Q_{Def}. By supposition, $S = \varnothing$, so $R^{-1}(S) = \varnothing$ and the first condition is respected. Since $R^{-1}(S) = \varnothing$, the third condition is also respected. Since $S = \varnothing$, by definition $X = \varnothing$ and the second condition is trivially respected. Thus, the "if" part is satisfied.

Lastly, consider $\pi = \mathit{Reins}_2$. Suppose that $(\varnothing, \varnothing)$ is an answer to Q_{Reins_2}. By Def. 14, we have that $S \cup R^{+1}(S) \cup R^{-1}(R^{+2}(S)) = \varnothing$ and thus $S = \varnothing$. So the "only if" part is satisfied. Suppose now that $S = \varnothing$ and consider the empty graph $(A' = \varnothing, R' = \varnothing)$. We must prove that it is an answer to Q_{Reins_2}. By supposition, $S = \varnothing$, so $R^{-1}(S) = \varnothing$ and $R^{-1}(R^{+2}(S)) = \varnothing$ as well, which means that the first condition is respected. As $R^{-1}(R^{+2}(S)) = \varnothing$, the third condition is also respected. Since $S = \varnothing$, by definition $X = \varnothing$ and the second condition is trivially respected. Thus, the "if" part is satisfied.

2. Let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$ and assume that $(\varnothing, \varnothing)$ is an answer to Q_{Reins_1}. Suppose now that $X \neq \varnothing$. By Def. 13, since $A' = \varnothing$, it must be the case that $S \cap X = \varnothing$. Since $X \subseteq A$ and $S \subseteq A$, this means that $(A \setminus S) \cap X = X$, and so by supposition that $(A \setminus S) \cap X \neq \varnothing$. So by Condition 3 of Def. 13, $\exists a \in (A \setminus S) \cap X$ with $a \in A'$, which contradicts the hypothesis that $(\varnothing, \varnothing)$ is an answer to Q_{Reins_1}. As such, we deduce $X = \varnothing$. Thus, the "only if" is satisfied.

   Assume now that $X = \varnothing$. Consider the empty graph $(\varnothing, \varnothing)$. We must prove that it is an answer to Q_{Reins_1}. Condition 2 of Def. 13 is obviously respected. Moreover, since $X = \varnothing$, Condition 1 of Def. 13 is also obviously respected. Again, from $X = \varnothing$ we deduce $(A \setminus S) \cap X = \varnothing$, and thus Condition 2 of Def. 13 is respected as well. Hence, it follows that $(\varnothing, \varnothing)$ is an answer to Q_{Reins_1}. Thus, the "if" is satisfied.

3. Assume that $(\varnothing, \varnothing)$ is an answer to Q_{CA}. By Def. 15 we thus know that $A' = \varnothing = A$. Hence, since $A = \varnothing$ and $R \subseteq A \times A$, it follows that $R = \varnothing$ and so $\mathcal{A} = (\varnothing, \varnothing)$. Thus, the "only if" is satisfied.

   Assume now that $\mathcal{A} = (\varnothing, \varnothing)$. Consider the empty graph $(\varnothing, \varnothing)$. We must prove that it is an answer to Q_{CA}. First and second condition

of Def. 15 follow immediately from $A' = R' = \varnothing$. The third condition follows from $A = \varnothing$. Thus, the "if" is satisfied.

□

To stress that empty explanations only occur in specific situations, we provide the following result. It states for each principle that when the empty explanation occurs, it is the only possible explanation.

Theorem 7. *Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. If $(\varnothing, \varnothing)$ is an explanation to Q_π with $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$, then it is unique.*

Proof. Assume $(\varnothing, \varnothing)$ is an answer to Q_π and let (A', R') be a subgraph of $\mathcal{A}$. Suppose (A', R') is also an answer to Q_π. We must prove that $(A', R') = (\varnothing, \varnothing)$. Since $R' \subseteq A' \times A'$, this can be reduced to proving that $A' = \varnothing$.

Consider $\pi = CF$ or Def or $Reins_2$. Following Th. 6 and since $(\varnothing, \varnothing)$ is an answer to Q_π, $S = \varnothing$ and so $R^{-1}(S) = R^{+2}(S) = R^{-1}(R^{+2}(S)) = \varnothing$. So following Condition 1 in Def. 11 (resp. Def. 12, Def. 14), $A' = \varnothing$.

Consider $\pi = Reins_1$ and let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. By assumption $(\varnothing, \varnothing)$ is an answer to Q_{Reins_1}, thus by Th. 6, we know that $X = \varnothing$, so by Condition 1 in Def. 13, $A' = \varnothing$.

Consider $\pi = CA$. By assumption $(\varnothing, \varnothing)$ is an answer to Q_{CA}, thus by Th. 6, we know that $A = \varnothing$. Since (A', R') is a subgraph of $\mathcal{A}$, we deduce that $A' = \varnothing$. □

Now, we turn to our study of maximal explanations. The next theorem we provide states for each principle that there is only one possible maximal explanation.

To get this result, we need a number of intermediate results, that are provided in the following lemmas. There is one lemma by principle, and each gives a characterization of a maximal explanation in terms of conditions either on the attack relation or the arguments.

The first lemma states that a maximal explanation for conflict-freeness is one that includes all the attacks between the arguments of S.

Lemma 1. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{(a, b) \in R \mid a, b \in S\}$. If (A', R') is a maximal answer to Q_{CF}, then $X \subseteq R'$.*

Proof. Suppose that (A', R') is a maximal answer to Q_{CF}. Assume firstly that $X = \varnothing$. Then we trivially have $X \subseteq R'$. Assume secondly that $X \neq \varnothing$ and

that $X \not\subseteq R'$. Then, there exists $(a,b) \in R$ such that $a, b \in S$ and $(a,b) \notin R'$. Consider the graph (A'', R'') with $A'' = A'$ and $R'' = R' \cup \{(a,b)\}$. Obviously, (A', R') is a strict subgraph of (A'', R''). Since (A', R') is an answer to Q_{CF}, by Def. 11, $A' = S$, $R' \subseteq X$ and because $X \neq \varnothing$, $R' \neq \varnothing$ (Condition 3 in Def. 11). Thus, we deduce that $A'' = S$, $R'' \neq \varnothing$ and since $(a,b) \in X$, $R'' \subseteq X$. Hence, by Def. 11, (A'', R'') is an answer to Q_{CF}, a contradiction with the supposition that (A', R') is a maximal answer to Q_{CF}. □

The next lemma states that a maximal explanation for defence is one that includes all the attacks from S to the attackers of S.

Lemma 2. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $Y = \{(a,b) \in R \mid a \in S, b \in R^{-1}(S)\}$. If (A', R') is a maximal answer to Q_{Def}, then $Y \subseteq R'$.*

Proof. Suppose that (A', R') is a maximal answer to Q_{Def} and let $X = \{(b,a) \mid b \in R^{-1}(S), a \in S\}$. Two cases must be considered.

Assume firstly that for all $b \in R^{-1}(S)$, $b \notin R^{+1}(S)$. In other words, $\nexists (a,b) \in R$ such that $a \in S, b \in R^{-1}(S)$. So, we trivially have $Y = \varnothing \subseteq R'$.

Assume secondly that for some $b \in R^{-1}(S)$, $b \in R^{+1}(S)$. In this case, $Y \neq \varnothing$. Assume additionally that $Y \not\subseteq R'$. Then, there exists $(x,y) \in R$ such that $x \in S$, $y \in R^{-1}(S)$ and $(x,y) \notin R'$. Consider the graph (A'', R'') with $A'' = A'$ and $R'' = R' \cup \{(x,y)\}$. Obviously, (A', R') is a strict subgraph of (A'', R''). Since (A', R') is an answer to Q_{Def}, by Def. 12, $A' = S \cup R^{-1}(S)$, $X \subseteq R' \subseteq X \cup Y$ and for all $b \in R^{-1}(S)$ such that $b \in R^{+1}(S)$, $\exists (a,b) \in R'$ with $a \in S$. Thus, we deduce that $A'' = S \cup R^{-1}(S)$ and since $(x,y) \in Y$, $X \subseteq R'' \subseteq X \cup Y$. Moreover, it is obvious that for all $b \in R^{-1}(S)$ such that $b \in R^{+1}(S)$, $\exists (a,b) \in R''$ with $a \in S$ (since $R' \subseteq R''$). Hence, by Def. 12, (A'', R'') is an answer to Q_{Def}, a contradiction with the supposition that (A', R') is a maximal answer to Q_{Def}. □

The following lemma states that a maximal explanation for the first part of reinstatement is one that includes all the unattacked arguments.

Lemma 3. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. If (A', R') is a maximal answer to Q_{Reins_1}, then $X \subseteq A'$.*

Proof. Suppose that (A', R') is a maximal answer to Q_{Reins_1}. Two cases must be considered.

Assume firstly that $(A \setminus S) \cap X = \varnothing$. In other words, $S \cap X = X$. In this case, as (A', R') is an answer to Q_{Reins_1}, by Def. 13, we have $S \cap X \subseteq A'$ and so $X \subseteq A'$.

Assume secondly that $(A \setminus S) \cap X \neq \varnothing$ and that $X \not\subseteq A'$. Then, there exists $x \in A$ such that $R^{-1}(x) = \varnothing$ (so $x \in X$), and $x \notin A'$. Consider the graph (A'', R'') with $A'' = A' \cup \{x\}$ and $R'' = R'$. Obviously, (A', R') is a strict subgraph of (A'', R''). Since (A', R') is an answer to Q_{Reins_1}, by Def. 13, $S \cap X \subseteq A' \subseteq X$, and $R' = \varnothing$. Thus, since $x \in X$, conditions 1 and 2 of Def. 13 are also satisfied for (A'', R''): $S \cap X \subseteq A'' \subseteq X$ and $R'' = \varnothing$. In addition, as by assumption $(A \setminus S) \cap X \neq \varnothing$, by Condition 3 of Def. 13 we know that $\exists a \in (A \setminus S) \cap X$ such that $a \in A'$, and so $a \in A''$; thus Condition 3 of Def. 13 is also satisfied for (A'', R''). Hence, by Def. 13, (A'', R'') is an answer to Q_{Reins_1}, a contradiction with the supposition that (A', R') is a maximal answer to Q_{Reins_1}. □

This lemma states that a maximal explanation for the second part of reinstatement is one that includes all the attacks from S to the attackers of the arguments S defends.

Lemma 4. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. If (A', R') is a maximal answer to Q_{Reins_2}, then $Y \subseteq R'$.*

Proof. Suppose that (A', R') is a maximal answer to Q_{Reins_2} and let $X = \{(b, c) \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$. Two cases must be considered.

Assume firstly that for all $b \in R^{-1}(R^{+2}(S))$, $b \notin R^{+1}(S)$. In other words, $\nexists (a, b) \in R$ such that $a \in S, b \in R^{-1}(R^{+2}(S))$. So, we trivially have $Y = \varnothing \subseteq R'$.

Assume secondly that for some $b \in R^{-1}(R^{+2}(S))$, $b \in R^{+1}(S)$. In this case, $Y \neq \varnothing$. Assume additionally that $Y \not\subseteq R'$. Then, there exists $(x, y) \in R$ such that $x \in S$, $y \in R^{-1}(R^{+2}(S))$ and $(x, y) \notin R'$. Consider the graph (A'', R'') with $A'' = A'$ and $R'' = R' \cup \{(x, y)\}$. Obviously, (A', R') is a strict subgraph of (A'', R''). Since (A', R') is an answer to Q_{Reins_2}, by Def. 14, $A' = S \cup R^{-1}(S)$, $X \subseteq R' \subseteq X \cup Y$ and for all $b \in R^{-1}(S)$ such that $b \in R^{+1}(S)$, $\exists (a, b) \in R'$ with $a \in S$. Thus, we deduce that $A'' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and since $(x, y) \in Y$, $X \subseteq R'' \subseteq X \cup Y$. Moreover, it is obvious that for all $b \in R^{-1}(R^{+2}(S))$ such that $b \in R^{+1}(S)$, $\exists (a, b) \in R''$ with $a \in S$ (since $R' \subseteq R''$). Hence, by Def. 14, (A'', R'') is an answer to Q_{Reins_2}, a contradiction with the supposition that (A', R') is a maximal answer to Q_{Reins_2}. □

Finally, the last lemma states that a maximal explanation for complement attack is one that includes all the attacks from S to the arguments that are not in S.

Lemma 5. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. If (A', R') is a maximal answer to Q_{CA}, then $X \subseteq R'$.*

Proof. Suppose that (A', R') is a maximal answer to Q_{CA}. Two cases must be considered.

Assume firstly that for all $b \notin S$, $b \notin R^{+1}(S)$. In other words, $\nexists (a, b) \in R$ such that $a \in S, b \notin S$. So, we trivially have $X = \varnothing \subseteq R'$.

Assume secondly that for some $b \notin S$, $b \in R^{+1}(S)$. In this case, $X \neq \varnothing$. Assume additionally that $X \not\subseteq R'$. Then, there exists $(x, y) \in R$ such that $x \in S$, $y \notin S$ and $(x, y) \notin R'$. Consider the graph (A'', R'') with $A'' = A'$ and $R'' = R' \cup \{(x, y)\}$. Obviously, (A', R') is a strict subgraph of (A'', R''). Since (A', R') is an answer to Q_{CA}, by Def. 15, $A' = A$, $R' \subseteq X$ and for all $b \notin S$ such that $b \in R^{+1}(S)$, $\exists (a, b) \in R'$ with $a \in S$. Thus, we deduce that $A'' = A$ and since $(x, y) \in X$, $R'' \subseteq X$. Moreover, it is obvious that for all $b \notin S$ such that $b \in R^{+1}(S)$, $\exists (a, b) \in R''$ with $a \in S$ (since $R' \subseteq R''$). Hence, by Def. 15, (A'', R'') is an answer to Q_{CA}, a contradiction with the supposition that (A', R') is a maximal answer to Q_{CA}. □

Using the previous Lemmas, we can now give the result on the uniqueness of maximal explanations.

Theorem 8. *Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. If (A', R') is a maximal explanation that answers Q_π with $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$, then it is the unique maximal explanation that answers Q_π.*

Proof. Assume (A', R') is a maximal answer to Q_π and let (A'', R'') be a subgraph of $\mathcal{A}$. Suppose (A'', R'') is also a maximal answer to Q_π. We must prove that $(A'', R'') = (A', R')$.

Consider $\pi = CF$ and let $X = \{(a, b) \in R \mid a, b \in S\}$. Since (A'', R'') is an answer to Q_{CF}, we know by Def. 11 that $A'' = S$ and $R'' \subseteq X$. But (A'', R'') is a maximal answer to Q_{CF}. Thus, by Lem. 1, $X \subseteq R''$ and so $R'' = X$. However, (A', R') is also an answer to Q_{CF}, so $A' = S$ and $R' \subseteq X$. Thus, $A' = A''$ and $R' \subseteq R''$, so we conclude that (A', R') is a subgraph of (A'', R''). If $R' \subset R''$, we conclude further that (A', R') is a strict subgraph of (A'', R''), a contradiction with the assumption that (A', R') is a maximal answer to Q_{CF}. So it must be the case that $R' = R''$.

Consider $\pi = Def$ and let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. Since (A'', R'') is an answer to Q_{Def}, we know by Def. 12 that $A'' = S \cup R^{-1}(S)$, $X \subseteq R''$ and $R'' \subseteq X \cup Y$.

But (A'', R'') is a maximal answer to Q_{CF}. Thus, by Lem. 2, $Y \subseteq R''$, so $X \cup Y \subseteq R''$, and so $R'' = X \cup Y$. However, (A', R') is also an answer to Q_{Def}, so $A' = S \cup R^{-1}(S)$ and $R' \subseteq X \cup Y$. Thus, $A' = A''$ and $R' \subseteq R''$, so we conclude that (A', R') is a subgraph of (A'', R''). If $R' \subset R''$, we conclude further that (A', R') is a strict subgraph of (A'', R''), a contradiction with the assumption that (A', R') is a maximal answer to Q_{Def}. So it must be the case that $R' = R''$.

Consider $\pi = Reins_1$ and let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. Since (A'', R'') is an answer to Q_{Reins_1}, we know by Def. 13 that $A'' \subseteq X$ and $R'' = \varnothing$. But (A'', R'') is a maximal answer to Q_{Reins_1}. Thus, by Lem. 3, $X \subseteq A''$, and so $A'' = X$. However, (A', R') is also an answer to Q_{Reins_1}, so $A' \subseteq X$ and $R' = \varnothing$. Thus, $A' \subseteq A''$ and $R' = R''$, so we conclude that (A', R') is a subgraph of (A'', R''). If $A' \subset A''$, we conclude further that (A', R') is a strict subgraph of (A'', R''), a contradiction with the assumption that (A', R') is a maximal answer to Q_{Reins_1}. So it must be the case that $A' = A''$.

Consider $\pi = Reins_2$ and let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. Since (A'', R'') is an answer to Q_{Reins_2}, we know by Def. 14 that $A'' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$, $X \subseteq R''$, and $R'' \subseteq X \cup Y$. But (A'', R'') is a maximal answer to Q_{Reins_2}. Thus, by Lem. 4, $Y \subseteq R''$, so $X \cup Y \subseteq R''$, and so $R'' = X \cup Y$. However, (A', R') is also an answer to Q_{Reins_2}, so $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and $R' \subseteq X \cup Y$. Thus, $A' = A''$ and $R' \subseteq R''$, so we conclude that (A', R') is a subgraph of (A'', R''). If $R' \subset R''$, we conclude further that (A', R') is a strict subgraph of (A'', R''), a contradiction with the assumption that (A', R') is a maximal answer to Q_{Reins_2}. So it must be the case that $R' = R''$.

Consider $\pi = CA$ and let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. Since (A'', R'') is an answer to Q_{CA}, we know by Def. 15 that $A'' = A$ and $R'' \subseteq X$. But (A'', R'') is a maximal answer to Q_{CA}. Thus, by Lem. 5, $X \subseteq R''$, and so $R'' = X$. However, (A', R') is also an answer to Q_{CA}, so $A' = A$ and $R' \subseteq X$. Thus, $A' = A''$ and $R' \subseteq R''$, so we conclude that (A', R') is a subgraph of (A'', R''). If $R' \subset R''$, we conclude further that (A', R') is a strict subgraph of (A'', R''), a contradiction with the assumption that (A', R') is a maximal answer to Q_{CA}. So it must be the case that $R' = R''$. □

As it turns out, there can be multiple minimal explanations in general for each principle (cf. Fig. 1: the explanations for defence of Agents 1 and 2 are both minimal). The next theorem studies the relation between minimal and maximal explanations and shows that the maximal explanation is exactly

the union of all the minimal explanations.

To obtain it, we again need a number of intermediate results, this time on minimal explanations. These results are provided in the following Lemmas. There is one Lemma by principle, and each gives a characterization of a minimal explanation in terms of conditions either on the attack relation or the arguments.
The first lemma states that a minimal explanation for conflict-freeness is one that contains at most one attack.

Lemma 6. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an answer to Q_{CF}. (A', R') is a minimal answer to Q_{CF} iff $|R'| \leq 1$.*

Proof. Consider a minimal answer to Q_{CF} (A', R') and suppose that $|R'| > 1$. Let $(x, y) \in R'$ and consider (A', R'') such that $R'' = R' \setminus \{(x, y)\}$. Obviously, (A', R'') is a strict subgraph of (A', R'). Let $X = \{(a, b) \in R \mid a, b \in S\}$. Since (A', R') is an answer to Q_{CF}, by Def. 11, we know that $A' = S$, $R' \subseteq X$, and because $|R'| > 1$, $X \neq \varnothing$. However, $R'' \subset R'$, so $R'' \subset X$ and $|R''| \geq 1$, so $R'' \neq \varnothing$. Thus, by Def. 11, (A', R'') is an answer to Q_{CF}, which contradicts the minimality of (A', R').

Consider now an answer to Q_{CF} (A', R') such that $|R'| \leq 1$. Suppose there exists a strict subgraph (A'', R'') of (A', R') such that (A'', R'') is also an answer to Q_{CF}. Let $X = \{(a, b) \in R \mid a, b \in S\}$. Assume firstly that $A'' \subset A'$. By Def. 11, we know that $A'' = S$, so we have $S \subset A'$, which contradicts the fact that (A', R') is an answer to Q_{CF}. Assume secondly that $R'' \subset R'$. By supposition $|R'| \leq 1$, so in this case, $|R'| = 1$ and $R'' = \varnothing$. As such, $R'' \subseteq X$. Since (A'', R'') is an answer to Q_{CF} and $R'' = \varnothing$, by Def. 11, $X = \varnothing$. However, as $|R'| = 1$, this would mean that $R' \not\subseteq X$, which contradicts the fact that (A', R') is an answer to Q_{CF}. Consequently, it must be the case that $R'' = R'$, and so $(A'', R'') = (A', R')$ and (A', R') is a minimal answer to Q_{CF}. □

The next lemma states that a minimal explanation for defence is one such that all attackers of S that are not in S have at most one attack targeting them.

Lemma 7. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an answer to Q_{Def}. (A', R') is a minimal answer to Q_{Def} iff for all $x \in R^{-1}(S) \setminus S$, $|R'^{-1}(x)| \leq 1$.*

Proof. Consider a minimal answer to Q_{Def} (A', R') and suppose that there exists $x \in R^{-1}(S) \setminus S$ such that $|R'^{-1}(x)| > 1$. Let $(w, x) \in R'$ and consider

(A', R'') such that $R'' = R' \setminus \{(w, x)\}$. Obviously, (A', R'') is a strict subgraph of (A', R'). Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. Since (A', R') is an answer to Q_{Def}, by Def. 12, we know that $A' = S \cup R^{-1}(S)$, $X \subseteq R' \subseteq X \cup Y$. In addition, because $x \in R^{-1}(S)$ and $x \notin S$, it cannot be that $(w, x) \in X$, so $(w, x) \in (Y \setminus X)$ and $x \in R^{+1}(S)$. However, $R'' \subset R'$, so $R'' \subset X \cup Y$ and since $(w, x) \in (Y \setminus X)$, $X \subseteq R''$. Moreover, as $|R'^{-1}(x)| > 1$, we have that $|R''^{-1}(x)| \geq 1$, so $\exists (w', x) \in R''$; consequently, knowing that $x \in R^{-1}(S)$ and $x \in R^{+1}(S)$, the third condition of Def. 12 is satisfied for each $b \in R^{-1}(S)$ including x. Thus, by Def. 12, (A', R'') is an answer to Q_{Def}, which contradicts the minimality of (A', R').

Consider now an answer to Q_{Def} (A', R') such that for all $x \in R^{-1}(S) \setminus S$, $|R'^{-1}(x)| \leq 1$. Suppose there exists a strict subgraph (A'', R'') of (A', R') such that (A'', R'') is also an answer to Q_{Def}. Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. Assume firstly that $A'' \subset A'$. By Def. 12, we know that $A'' = S \cup R^{-1}(S)$, so we have $S \cup R^{-1}(S) \subset A'$, which contradicts the fact that (A', R') is an answer to Q_{Def}. Assume secondly that $R'' \subset R'$. By Def. 12, we have $X \subseteq R' \subseteq X \cup Y$ and $X \subseteq R'' \subseteq X \cup Y$; so since $R'' \subset R'$ there exists at least $(a, b) \in R' \setminus R''$ and by definition $(a, b) \in Y \setminus X$, so $b \in R^{-1}(S) \setminus S$. Moreover, by supposition, for all $x \in R^{-1}(S) \setminus S$, $|R'^{-1}(x)| \leq 1$, so in this case, we have $|R'^{-1}(b)| = 1$ and $R''^{-1}(b) = \varnothing$. In addition, as (A'', R'') is an answer to Q_{Def}, $b \in R^{-1}(S)$ and $R''^{-1}(b) = \varnothing$, by the third condition of Def. 12, $b \notin R^{+1}(S)$ and so $(a, b) \notin Y$. This would mean that $R' \not\subseteq X \cup Y$, which contradicts the fact that (A', R') is an answer to Q_{Def}. Consequently, it must be the case that $R'' = R'$, and so $(A'', R'') = (A', R')$ and (A', R') is a minimal answer to Q_{Def}. □

The following lemma states that a minimal explanation for the first part of reinstatement is one that includes at most one argument that is not in S.

Lemma 8. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an answer to Q_{Reins_1}. (A', R') is a minimal answer to Q_{Reins_1} iff $|A' \setminus S| \leq 1$.*

Proof. Consider a minimal answer to Q_{Reins_1} (A', R') and suppose that $|A' \setminus S| > 1$. Let $x \in A' \setminus S$ and consider (A'', R') such that $A'' = A' \setminus \{x\}$. Obviously, (A'', R') is a strict subgraph of (A', R'). Let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. Since (A', R') is an answer to Q_{Reins_1}, by Def. 13, we know that $R' = \varnothing$, $S \cap X \subseteq A' \subseteq X$, and because $x \notin S$, $x \in X \setminus S$. However, $A'' \subset A'$, so

$A'' \subset X$ and since $x \in A' \setminus S$, $S \cap X \subseteq A''$. Moreover, as $|A' \setminus S| > 1$ and $x \in X$, we have that $|(A'' \setminus S) \cap X| \geq 1$. Thus, by Def. 13, (A'', R') is an answer to Q_{Reins_1}, which contradicts the minimality of (A', R').

Consider now an answer to Q_{Reins_1} (A', R') such that $|A' \setminus S| \leq 1$. Suppose there exists a strict subgraph (A'', R'') of (A', R') such that (A'', R'') is also an answer to Q_{Reins_1}. Let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. Assume firstly that $R'' \subset R'$. By Def. 13, we know that $R'' = \varnothing$, so we have $|R'| > 0$, which contradicts the fact that (A', R') is an answer to Q_{Reins_1}. Assume secondly that $A'' \subset A'$. As (A', R') and (A'', R'') are answers to Q_{Reins_1}, by Def. 13, $S \cap X \subseteq A' \subseteq X$ and $S \cap X \subseteq A'' \subseteq X$. Since $A'' \subset A'$, we have that $A'' \setminus S \subset A' \setminus S$. By supposition $|A' \setminus S| \leq 1$, so in this case, $|A' \setminus S| = 1$ and $A'' \setminus S = \varnothing$. As $A'' \setminus S = \varnothing$, $\nexists a \in (A \setminus S) \cap X$ with $a \in A''$. Since (A'', R'') is an answer to Q_{Reins_1}, by Def. 13, $(A \setminus S) \cap X = \varnothing$. However, as $|A' \setminus S| = 1$, this would mean that $A' \not\subseteq X$, which contradicts the fact that (A', R') is an answer to Q_{Reins_1}. Consequently, it must be the case that $A'' = A'$, and so $(A'', R'') = (A', R')$ and (A', R') is a minimal answer to Q_{Reins_1}. □

This lemma states that a minimal explanation for the second part of reinstatement is one such that all attackers, not defended by S, of arguments that S defend have at most one attack targeting them.

Lemma 9. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an answer to Q_{Reins_2}. (A', R') is a minimal answer to Q_{Reins_2} iff for all $x \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $|R'^{-1}(x)| \leq 1$.*

Proof. Consider a minimal answer to Q_{Reins_2} (A', R') and suppose that there exists $x \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$ such that $|R'^{-1}(x)| > 1$. Let $(w, x) \in R'$ and consider (A', R'') such that $R'' = R' \setminus \{(w, x)\}$. Obviously, (A', R'') is a strict subgraph of (A', R'). Let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. Since (A', R') is an answer to Q_{Reins_2}, by Def. 14, we know that $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and $X \subseteq R' \subseteq X \cup Y$. In addition, because $x \in R^{-1}(R^{+2}(S))$ and $x \notin R^{+2}(S)$, it cannot be that $(w, x) \in X$, so $(w, x) \in (Y \setminus X)$ and $x \in R^{+1}(S)$. However, $R'' \subset R'$, so $R'' \subset X \cup Y$ and since $(w, x) \in (Y \setminus X)$, $X \subseteq R''$. Moreover, as $|R'^{-1}(x)| > 1$, we have that $|R''^{-1}(x)| \geq 1$, so $\exists (w', x) \in R''$; consequently, knowing that $x \in R^{-1}(R^{+2}(S))$ and $x \in R^{+1}(S)$, the third condition of Def. 14 is satisfied for each $b \in R^{-1}(R^{+2}(S))$ including x. Thus, by Def. 14, (A', R'') is an answer to Q_{Reins_2}, which contradicts the minimality of (A', R').

Consider now an answer to Q_{Reins_2} (A', R') st for all $x \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $|R'^{-1}(x)| \leq 1$. Suppose there exists a strict subgraph (A'', R'') of (A', R') such that (A'', R'') is also an answer to Q_{Reins_2}. Let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. Assume firstly that $A'' \subset A'$. By Def. 14, we know that $A'' = S \cup R^{-1}(S)$, so we have $S \cup R^{-1}(S) \subset A'$, which contradicts the fact that (A', R') is an answer to Q_{Reins_2}. Assume secondly that $R'' \subset R'$. By Def. 14, we have $X \subseteq R' \subseteq X \cup Y$ and $X \subseteq R'' \subseteq X \cup Y$; so since $R'' \subset R'$ there exists at least $(a, b) \in R' \setminus R''$ and by definition $(a, b) \in Y \setminus X$, so $b \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$. Moreover, by supposition, for all $x \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $|R'^{-1}(x)| \leq 1$, so in this case, we have $|R'^{-1}(b)| = 1$ and $R''^{-1}(b) = \varnothing$. In addition, as (A'', R'') is an answer to Q_{Reins_2}, $b \in R^{-1}(R^{+2}(S))$ and $R''^{-1}(b) = \varnothing$, by the third condition of Def. 14, $b \notin R^{+1}(S)$ and so $(a, b) \notin Y$. This would mean that $R' \not\subseteq X \cup Y$, which contradicts the fact that (A', R') is an answer to Q_{Reins_2}. Consequently, it must be the case that $R'' = R'$, and so $(A'', R'') = (A', R')$ and (A', R') is a minimal answer to Q_{Reins_2}. □

Finally, the last lemma states that a minimal explanation for complement attack is one such that all arguments that are not in S have at most one attack targeting them.

Lemma 10. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and (A', R') be an answer to Q_{CA}. (A', R') is a minimal answer to Q_{CA} iff for all $x \notin S$, $|R'^{-1}(x)| \leq 1$.*

Proof. Consider a minimal answer to Q_{CA} (A', R') and suppose that there exists $x \notin S$ such that $|R'^{-1}(x)| > 1$. Let $(w, x) \in R'$ and consider (A', R'') such that $R'' = R' \setminus \{(w, x)\}$. Obviously, (A', R'') is a strict subgraph of (A', R'). Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. Since (A', R') is an answer to Q_{CA}, by Def. 15, we know that $A' = A$, $R' \subseteq X$, and because $x \notin S$, $(w, x) \in X$ so $x \in R^{+1}(S)$. However, $R'' \subset R'$, so $R'' \subset X$ and as $|R'^{-1}(x)| > 1$, we have that $|R''^{-1}(x)| \geq 1$, so $\exists (w', x) \in R''$ and because $R'' \subset X$ and $x \notin S$, $w' \in S$. Thus, by Def. 15, (A', R'') is an answer to Q_{CA}, which contradicts the minimality of (A', R').

Consider now an answer to Q_{CA} (A', R') such that for all $x \notin S$, $|R'^{-1}(x)| \leq 1$. Suppose there exists a strict subgraph (A'', R'') of (A', R') such that (A'', R'') is also an answer to Q_{CA}. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. Assume firstly that $A'' \subset A'$. By Def. 15, we know that $A'' = A$, so we have $A \subset A'$, which contradicts the fact that (A', R') is an answer to Q_{CA}. Assume

secondly that $R'' \subset R'$. By Def. 15, we have $R' \subseteq X$ and by supposition, for all $x \notin S$, $|R'^{-1}(x)| \leq 1$, so in this case, there exists $b \notin S$ such that $|R'^{-1}(b)| = 1$ and $R''^{-1}(b) = \varnothing$. As such, $R'' \subseteq X$. Since (A'', R'') is an answer to Q_{CA}, $b \notin S$ and $R''^{-1}(b) = \varnothing$, by Def. 15, $b \notin R^{+1}(S)$. However, as $|R'^{-1}(b)| = 1$, this would mean that $R' \not\subseteq X$, which contradicts the fact that (A', R') is an answer to Q_{CA}. Consequently, it must be the case that $R'' = R'$, and so $(A'', R'') = (A', R')$ and (A', R') is a minimal answer to Q_{CA}. □

Using the previous Lemmas, we can now prove that, for each principle, the maximal explanation is exactly the union of all the minimal explanations.

Theorem 9. *Let $\mathcal{A} = (A, R)$ and $S \subseteq A$. Consider $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$ and let (A', R') be the maximal explanation that answers Q_π and M be the set of all minimal explanations that answers Q_π. Then, $(A', R') = \bigcup_{G \in M} G$.*

Proof.
• $(A', R') \subseteq \cup_{G \in M} G$. We denote $M = \{G_1, \ldots, G_n\}$ with $G_1 = (A_1, R_1)$, $\ldots$, $G_n = (A_n, R_n)$. Suppose that $(A', R') \not\subseteq \bigcup_{G \in M} G$. This means that $A' \not\subseteq A_1 \cup \cdots \cup A_n$ or $R' \not\subseteq R_1 \cup \cdots \cup R_n$.

Consider $\pi = CF$ and let $X = \{(a, b) \in R \mid a, b \in S\}$. By supposition, $(A', R'), G_1, \ldots, G_n$ are all answers to Q_{CF}. So, by Def. 11, we have $A' = A_1 = \cdots = A_n = S$. Thus, it must be the case that $R' \not\subseteq R_1 \cup \cdots \cup R_n$. In addition, by Lem. 1, we know that $R' = X$ and so that $R_1 \subseteq R'$, $\ldots$, $R_n \subseteq R'$. Assume firstly that $X = \varnothing$. Then, by Def. 11, we have $R' = R_1 = \cdots = R_n = \varnothing$, a contradiction. Assume secondly that $X \neq \varnothing$. In this case we have $R' \neq \varnothing$, and by Def. 11, $R_1 \neq \varnothing$, $\ldots$, $R_n \neq \varnothing$. This means that there exists $R'' \subseteq X$ with $R'' \neq \varnothing$, such that $R'' \cap R_1 = \cdots = R'' \cap R_n = \varnothing$. Let $R''' \subseteq R''$ with $|R'''| = 1$. Consider (A', R'''). We already know that $A' = S$, $R''' \subseteq X$ and we have both $X \neq \varnothing$ and $R''' \neq \varnothing$. So, by Def. 11, (A', R''') is an answer to Q_{CF}. In addition, since $R'' \cap R_1 = \cdots = R'' \cap R_n = \varnothing$, we have $(A', R''') \neq (A_1, R_1)$, $\ldots$, $(A', R''') \neq (A_n, R_n)$. However, $|R'''| = 1$, so by Lem. 6, (A', R''') is a minimal answer to Q_{CF}. A contradiction with the hypothesis that M is the set of all minimal answers to Q_{CF}.

Consider $\pi = Def$ and let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. By supposition, $(A', R'), G_1, \ldots, G_n$ are all answers to Q_{Def}. So, by Def. 12, we have $A' = A_1 = \cdots = A_n = S \cup R^{-1}(S)$. Thus, it must be the case that $R' \not\subseteq R_1 \cup \cdots \cup R_n$. In addition, by Lem. 2,

we know that $R' = X \cup Y$ and so that $R_1 \subseteq R'$, ..., $R_n \subseteq R'$. Assume firstly that for all $y \in R^{-1}(S)$, $R'^{-1}(y) = \varnothing$. Then, by Def. 12, we have for all $y \in R^{-1}(S)$, $y \notin R^{+1}(S)$. Since $X \subseteq R_1 \subseteq X \cup Y$, ..., $X \subseteq R_n \subseteq X \cup Y$, we deduce that for all $y \in R^{-1}(S)$, $R_1^{-1}(y) = \cdots = R_n^{-1}(y) = R'^{-1}(y) = \varnothing$, and so that $R' = R_1 = \cdots = R_n = X$, a contradiction. Assume secondly that for some $y \in R^{-1}(S)$, $R'^{-1}(y) \neq \varnothing$. In this case we have $R' \neq \varnothing$ and $y \in R^{+1}(S)$, so by Def. 12, $R_1^{-1}(y) \neq \varnothing$, ..., $R_n^{-1}(y) \neq \varnothing$ and thus, $R_1 \neq \varnothing$, ..., $R_n \neq \varnothing$. Since, $R' \not\subseteq R_1 \cup \ldots \cup R_n$ and $\forall i, R_i \subseteq R'$, this means that there exists $R'' \subseteq Y$ with $R'' \neq \varnothing$, such that $R'' \cap R_1 = \cdots = R'' \cap R_n = \varnothing$. In particular, this means that there exists $y_0 \in R^{-1}(S)$ such that $R''^{-1}(y_0) \neq \varnothing$ and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$. Moreover, because $X \subseteq R_1$, $\ldots X \subseteq R_n$, we know that $y_0 \notin S$. Let R''' such that: (1) $X \subseteq R''' \subset R'$, (2) for all $y \in R^{-1}(S) \setminus S, y \neq y_0$, $|R'''^{-1}(y)| = 1$ if $y \in R^{+1}(S)$ and $|R'''^{-1}(y)| = 0$ otherwise, (3) $|R'''^{-1}(y_0)| = 1$ with $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$. Consider (A', R'''). We already know that $A' = S \cup R^{-1}(S)$ and $X \subseteq R''' \subseteq X \cup Y$. In addition, we have that for all $y \in R^{-1}(S)$ such that $y \in R^{+1}(S)$, $R'''^{-1}(y) \neq \varnothing$ (definition of R'''). So, by Def. 12, (A', R''') is an answer to Q_{Def}. In addition, since $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$ and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$, we have $(A', R''') \neq (A_1, R_1)$, ..., $(A', R''') \neq (A_n, R_n)$. However, for all $y \in R^{-1}(S) \setminus S$, $|R'''-1(y)| \leq 1$, so by Lem. 7, (A', R''') is a minimal answer to Q_{Def}. A contradiction with the hypothesis that M is the set of all minimal answers to Q_{Def}.

Consider $\pi = Reins_1$ and let $X = \{a \in A \,|R^{-1}(a) = \varnothing\}$. By supposition, $(A', R'), G_1, \ldots, G_n$ are all answers to Q_{Reins_1}. So, by Def. 13, we have $R' = R_1 = \cdots = R_n = \varnothing$. Thus, it must be the case that $A' \not\subseteq A_1 \cup \cdots \cup A_n$. In addition, by Lem. 3, we know that $A' = X$ and so that $A_1 \subseteq A'$, ..., $A_n \subseteq A'$. Assume firstly that $X \setminus S = \varnothing$ (or, written differently, $S \cap X = X$ and $(A \setminus S) \cap X = \varnothing$). Then, by Def. 13, we have $A' = A_1 = \cdots = A_n = X$, a contradiction. Assume secondly that $X \setminus S \neq \varnothing$, so $(A \setminus S) \cap X \neq \varnothing$. By assumption $A' \not\subseteq A_1 \cup \cdots \cup A_n$, so there exists $A'' \subseteq X \setminus S$ with $A'' \neq \varnothing$, such that $A'' \cap A_1 = \cdots = A'' \cap A_n = \varnothing$. Let $x \in A''$ and $A''' = (S \cap X) \cup \{x\}$. Consider (A''', R'). We already know that $R' = \varnothing$, $S \cap X \subseteq A''' \subseteq X$ and we have both $(A \setminus S) \cap X \neq \varnothing$ and $x \in ((A \setminus S) \cap X) \cap A'''$. So, by Def. 13, (A''', R') is an answer to Q_{Reins_1}. In addition, since $A'' \cap A_1 = \cdots = A'' \cap A_n = \varnothing$, we have $(A''', R') \neq (A_1, R_1)$, ..., $(A''', R') \neq (A_n, R_n)$. However, $|A''' \setminus S| = 1$, so by Lem. 8, (A''', R') is a minimal answer to Q_{Reins_1}. A contradiction with the hypothesis that M is the set of all minimal answers to Q_{Reins_1}.

Consider $\pi = Reins_2$ and let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in$

$R^{+2}(S)\}$ and $Y = \{(a,b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. By supposition, $(A', R'), G_1, \ldots, G_n$ are all answers to Q_{Reins_2}. So, by Def. 14, we have $A' = A_1 = \cdots = A_n = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$. Thus, it must be the case that $R' \not\subseteq R_1 \cup \cdots \cup R_n$. In addition, by Lem. 4, we know that $R' = X \cup Y$ and so that $R_1 \subseteq R'$, ..., $R_n \subseteq R'$. Assume firstly that for all $y \in R^{-1}(R^{+2}(S))$, $R'^{-1}(y) = \varnothing$. Then, by Def. 14, we have for all $y \in R^{-1}(R^{+2}(S))$, $y \notin R^{+1}(S)$ (so $Y = \varnothing$). Since $X \subseteq R_1 \subseteq X \cup Y$, ..., $X \subseteq R_n \subseteq X \cup Y$, we deduce that $R' = R_1 = \cdots = R_n = X$, a contradiction. Assume secondly that for some $y \in R^{-1}(R^{+2}(S))$, $R'^{-1}(y) \neq \varnothing$. In this case we have $R' \neq \varnothing$ and $y \in R^{+1}(S)$, so by Def. 14, $R_1^{-1}(y) \neq \varnothing$, ..., $R_n^{-1}(y) \neq \varnothing$ and thus, $R_1 \neq \varnothing$, ..., $R_n \neq \varnothing$. Since, $R' \not\subseteq R_1 \cup \ldots \cup R_n$ and $\forall i, R_i \subseteq R'$, this means that there exists $R'' \subseteq Y$ with $R'' \neq \varnothing$, such that $R'' \cap R_1 = \cdots = R'' \cap R_n = \varnothing$. In particular, this means that there exists $y_0 \in R^{-1}(R^{+2}(S))$ such that $R''^{-1}(y_0) \neq \varnothing$ and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$. Moreover, because $X \subseteq R_1, \ldots X \subseteq R_n$, we know that $y_0 \notin R^{+2}(S)$. Let R''' such that: (1) $X \subseteq R''' \subset R'$, (2) for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S), y \neq y_0$, $|R'''^{-1}(y)| = 1$ if $y \in R^{+1}(S)$ and $|R'''^{-1}(y)| = 0$ otherwise, (3) $|R'''^{-1}(y_0)| = 1$ and $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$. Consider (A', R'''). We already know that $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and $X \subseteq R''' \subseteq X \cup Y$. In addition, we have that for all $y \in R^{-1}(R^{+2}(S))$ such that $y \in R^{+1}(S)$, $R'''^{-1}(y) \neq \varnothing$ (definition of R'''). So, by Def. 14, (A', R''') is an answer to Q_{Reins_2}. In addition, since $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$ and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$, we have $(A', R''') \neq (A_1, R_1)$, ..., $(A', R''') \neq (A_n, R_n)$. However, for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $|R'''^{-1}(y)| \leq 1$, so by Lem. 9, (A', R''') is a minimal answer to Q_{Reins_2}. A contradiction with the hypothesis that M is the set of all minimal answers to Q_{Reins_2}.

Consider $\pi = CA$ and let $X = \{(a,b) \in R \mid a \in S, b \notin S\}$. By supposition, $(A', R'), G_1, \ldots, G_n$ are all answers to Q_{CA}. So, by Def. 15, we have $A' = A_1 = \cdots = A_n = A$. Thus, it must be the case that $R' \not\subseteq R_1 \cup \cdots \cup R_n$. In addition, by Lem. 5, we know that $R' = X$ and so that $R_1 \subseteq R'$, ..., $R_n \subseteq R'$. Assume firstly that for all $y \notin S$, $R'^{-1}(y) = \varnothing$. Then, by Def. 15, we have for all $y \notin S$, $y \notin R^{+1}(S)$ (so $X = \varnothing$). Since $R_1 \subseteq X$, ..., $R_n \subseteq X$, we deduce that $R' = R_1 = \cdots = R_n = \varnothing$, a contradiction. Assume secondly that for some $y \notin S$, $R'^{-1}(y) \neq \varnothing$. In this case we have $R' \neq \varnothing$ and $y \in R^{+1}(S)$, so by Def. 15, $R_1^{-1}(y) \neq \varnothing$, ..., $R_n^{-1}(y) \neq \varnothing$ and thus, $R_1 \neq \varnothing$, ..., $R_n \neq \varnothing$. Since, by assumption, $R' \not\subseteq R_1 \cup \ldots \cup R_n$, there exists $R'' \subseteq X$ with $R'' \neq \varnothing$, such that $R'' \cap R_1 = \cdots = R'' \cap R_n = \varnothing$. In particular, this means that there exists $y_0 \notin S$ such that $R''^{-1}(y_0) \neq \varnothing$

and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$. Let $R''' \subset R'$ such that: (1) for all $y \notin S, y \neq y_0$, $|R'''^{-1}(y)| \leq 1$ if $y \in R^{+1}(S)$ and $|R'''^{-1}(y)| = 1$ otherwise, (2) $|R'''^{-1}(y_0)| = 1$ with $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$. Consider (A', R'''). We already know that $A' = A$, $R''' \subseteq X$ and for all $y \notin S$ such that $y \in R^{+1}(S)$, we have $R'''-1(y) \neq \varnothing$. So, by Def. 15, (A', R''') is an answer to Q_{CA}. In addition, since $R'''^{-1}(y_0) \subseteq R''^{-1}(y_0)$ and $R''^{-1}(y_0) \cap R_1^{-1}(y_0) = \cdots = R''^{-1}(y_0) \cap R_n^{-1}(y_0) = \varnothing$, we have $(A', R''') \neq (A_1, R_1)$, ..., $(A', R''') \neq (A_n, R_n)$. However, for all $y \notin S$, $|R'''-1(y)| \leq 1$, so by Lem. 10, (A', R''') is a minimal answer to Q_{CA}. A contradiction with the hypothesis that M is the set of all minimal answers to Q_{CA}.

• $(A', R') \supseteq \cup_{G \in M} G$. We denote $M = \{G_1, \ldots, G_n\}$ with $G_1 = (A_1, R_1)$, ..., $G_n = (A_n, R_n)$.

Consider $\pi = CF$ and let $X = \{(a, b) \in R \mid a, b \in S\}$. By Def. 11, we have $A' = A_1 = \cdots = A_n = S$. Moreover, by Lem. 1, we know that $R' = X$. Finally, by Def. 11, we know that $R_1 \subseteq X$, ..., $R_n \subseteq X$. Thus, $R_1 \subseteq R'$, ..., $R_n \subseteq R'$ and so $\bigcup_{G \in M} G \subseteq (A', R')$.

Consider $\pi = Def$ and let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. By Def. 12, we have $A' = A_1 = \cdots = A_n = S \cup R^{-1}(S)$. Moreover, by Lem. 2, we know that $R' = X \cup Y$. Finally, by Def. 12, we know that $R_1 \subseteq X \cup Y$, ..., $R_n \subseteq X \cup Y$. Thus, $R_1 \subseteq R'$, ..., $R_n \subseteq R'$ and so $\bigcup_{G \in M} G \subseteq (A', R')$.

Consider $\pi = Reins_1$ and let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. By Def. 13, we have $R' = R_1 = \cdots = R_n = \varnothing$. Moreover, by Lem. 3, we know that $A' = X$. Finally, by Def. 13, we know that $A_1 \subseteq X$, ..., $A_n \subseteq X$. Thus, $A_1 \subseteq A'$, ..., $A_n \subseteq A'$ and so $\bigcup_{G \in M} G \subseteq (A', R')$.

Consider $\pi = Reins_2$ and let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. By Def. 14, we have $A' = A_1 = \cdots = A_n = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$. Moreover, by Lem. 4, we know that $R' = X \cup Y$. Finally, by Def. 14, we know that $R_1 \subseteq X \cup Y$, ..., $R_n \subseteq X \cup Y$. Thus, $R_1 \subseteq R'$, ..., $R_n \subseteq R'$ and so $\bigcup_{G \in M} G \subseteq (A', R')$.

Consider $\pi = CA$ and let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. By Def. 15, we have $A' = A_1 = \cdots = A_n = A$. Moreover, by Lem. 5, we know that $R' = X$. Finally, by Def. 15, we know that $R_1 \subseteq X$, ..., $R_n \subseteq X$. Thus, $R_1 \subseteq R'$, ..., $R_n \subseteq R'$ and so $\bigcup_{G \in M} G \subseteq (A', R')$. □

5 Computation of Explanations

This section investigates how to compute the maximal and minimal explanations of a class.

5.1 Maximal Explanations

It turns out that maximal explanations exactly correspond to the explanations defined in [BDDLS22] (recalled in Def. 9).

Proposition 4. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$. $G_\pi(S)$ is the maximal explanation that answers Q_π.*

Proof. Consider $\pi = CF$. We denote $G_{CF}(S)$ with (A', R'). Suppose $G_{CF}(S)$ is not a maximal answer to Q_{CF}. Then, there exists a subgraph (A'', R'') of $\mathcal{A}$ that is an answer to Q_{CF} and such that $G_{CF}(S)$ is a subgraph of (A'', R''). The case when $(A'', R'') = (A', R')$ is trivial, so we will assume that $(A'', R'') \neq (A', R')$. Let $X = \{(a, b) \in R \mid a, b \in S\}$. By Def. 11, we know that $A'' = S$ and $R'' \subseteq X$. By definition of $G_{CF}(S)$, we have that $A' = S = A''$. Thus, since $G_{CF}(S)$ is a subgraph of (A'', R''), it must be the case that $\exists (a, b) \in R''$ with $a, b \in S$ s.t. $(a, b) \notin R'$. However (A'', R'') is a subgraph of $\mathcal{A}$, so $(a, b) \in R$. Hence, there exists $(a, b) \in R$ with $a, b \in S$ such that $(a, b) \notin R'$, a contradiction of Def. 4.

Consider $\pi = Def$. We denote $G_{Def}(S)$ with (A', R'). Suppose $G_{Def}(S)$ is not a maximal answer to Q_{Def}. Then, there exists a subgraph (A'', R'') of $\mathcal{A}$ that is an answer to Q_{Def} and such that $G_{Def}(S)$ is a subgraph of (A'', R''). The case when $(A'', R'') = (A', R')$ is trivial, so we will assume that $(A'', R'') \neq (A', R')$. Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. By Def. 12, we know that $A'' = S \cup R^{-1}(S)$ and $X \subseteq R'' \subseteq X \cup Y$. By definition of $G_{Def}(S)$, we have that $A' = S \cup R^{-1}(S) = A''$. Thus, since $G_{Def}(S)$ is a subgraph of (A'', R''), it must be the case that either $\exists (a, b) \in R''$ with $a \in S$ and $b \in R^{-1}(S)$ s.t. $(a, b) \notin R'$, or $\exists (b, a) \in R''$ with $b \in R^{-1}(S)$ and $a \in S$ s.t. $(b, a) \notin R'$. However (A'', R'') is a subgraph of $\mathcal{A}$, so $(a, b), (b, a) \in R$. Hence, either there exists $(a, b) \in R$ with $a \in S$ and $b \in R^{-1}(S)$ such that $(a, b) \notin R'$, or there exists $(b, a) \in R$ with $b \in R^{-1}(S)$ and $a \in S$ such that $(b, a) \notin R'$. This is a contradiction of Def. 4.

Consider $\pi = Reins_1$. We denote $G_{Reins_1}(S)$ with (A', R'). Suppose $G_{Reins_1}(S)$ is not a maximal answer to Q_{Reins_1}. Then, there exists a subgraph

(A'', R'') of $\mathcal{A}$ that is an answer to Q_{Reins_1} and such that $G_{Reins_1}(S)$ is a subgraph of (A'', R''). The case when $(A'', R'') = (A', R')$ is trivial, so we will assume that $(A'', R'') \neq (A', R')$. Let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$. By Def. 13, we know that $S \cap X \subseteq A'' \subseteq X$ and $R'' = \varnothing$. Since $G_{Reins_1}(S)$ is a subgraph of (A'', R''), it must then be the case that $\exists a \in X$ s.t. $a \notin A'$. This is a contradiction of Def. 4.

Consider $\pi = Reins_2$. We denote $G_{Reins_2}(S)$ with (A', R'). Suppose $G_{Reins_2}(S)$ is not a maximal answer to Q_{Reins_2}. Then, there exists a subgraph (A'', R'') of $\mathcal{A}$ that is an answer to Q_{Reins_2} and such that $G_{Reins_2}(S)$ is a subgraph of (A'', R''). The case when $(A'', R'') = (A', R')$ is trivial, so we will assume that $(A'', R'') \neq (A', R')$. Let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. By Def. 14, we know that $A'' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and $X \subseteq R'' \subseteq X \cup Y$. By definition of $G_{Reins_2}(S)$, we have that $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S)) = A''$. Thus, since $G_{Reins_2}(S)$ is a subgraph of (A'', R''), it must be the case that either $\exists (a, b) \in R''$ with $a \in S$ and $b \in R^{-1}(R^{+2}(S))$ s.t. $(a, b) \notin R'$, or $\exists (b, c) \in R''$ with $b \in R^{-1}(R^{+2}(S))$ and $c \in R^{+2}(S)$ s.t. $(b, c) \notin R'$. However (A'', R'') is a subgraph of $\mathcal{A}$, so $(a, b), (b, c) \in R$. Hence, either there exists $(a, b) \in R$ with $a \in S$ and $b \in R^{-1}(R^{+2}(S))$ such that $(a, b) \notin R'$, or there exists $(b, c) \in R$ with $b \in R^{-1}(R^{+2}(S))$ and $c \in R^{+2}(S)$ such that $(b, c) \notin R'$. This is a contradiction of Def. 4.

Consider $\pi = CA$. We denote $G_{CA}(S)$ with (A', R'). Suppose $G_{CA}(S)$ is not a maximal answer to Q_{CA}. Then, there exists a subgraph (A'', R'') of $\mathcal{A}$ that is an answer to Q_{CA} and such that $G_{CA}(S)$ is a subgraph of (A'', R''). The case when $(A'', R'') = (A', R')$ is trivial, so we will assume that $(A'', R'') \neq (A', R')$. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. By Def. 15, we know that $A'' = A$ and $R'' \subseteq X$. By definition of $G_{CA}(S)$, we have that $A' = A = A''$. Thus, since $G_{CA}(S)$ is a subgraph of (A'', R''), it must be the case that $\exists (a, b) \in R''$ with $a \in S$ and $b \notin S$ s.t. $(a, b) \notin R'$. This contradicts Def. 4. $\square$

This result entails that maximal explanations can be computed using only the graph operators of induced and spanning subgraphs, thus ensuring an efficient computation.

Note that Prop. 4 aggregated with Th. 8 allows to recover a unicity result given in [BDDLS22] confirming once again the relationship between our approach and that given in [BDDLS22].

5.2 From Maximal to Minimal Explanations

In order to compute the minimal explanations, we will start from the maximal explanation, and gradually remove elements until obtaining a minimal explanation. This leads to five algorithms (one for each principle) that are built following the same schema. They also use the same condition for stopping the removal: "it remains at most one element to remove". The only differences between these algorithms concern the "nature" of the removed elements:[5]

For *CF***:** removal of attacks between elements of S

For *Def***:** for each attacker of S that is not in S, removal of attacks that target it

For $Reins_1$**:** removal of unattacked arguments that are not in S

For $Reins_2$**:** for each argument that is an attacker of the arguments S defends and that is not defended by S, removal of attacks that target it

For *CA***:** for each argument that is not in S, removal of attacks that target it

Alg$_{CF}$ Computation of a minimal answer to Q_{CF}

Require: $\mathcal{A} = (A, R), S \subseteq A$

1: $(A', R') \leftarrow G_{CF}(S)$
2: **while** $|R'| > 1$ **do**
3: $\quad (x, y) \leftarrow choose(R')$
4: $\quad R' \leftarrow R' \setminus \{(x, y)\}$
5: **end while**
6: **return** (A', R')

Our algorithms are sound and complete for the computation of minimal explanations as shown by the following proposition.

[5]Note that these elements are generally attacks except in the case of the principle $Reins_1$.

Alg$_{Def}$ Computation of a minimal answer to Q_{Def}

Require: $\mathcal{A} = (A, R), S \subseteq A$

```
(A', R') ← G_Def(S)
for y ∈ R^{-1}(S) \ S do
    while |R'^{-1}(y)| > 1 do
        x ← choose(R'^{-1}(y))
        R' ← R' \ {(x, y)}
    end while
end for
return (A', R')
```

Alg$_{Reins_1}$ Computation of a minimal answer to Q_{Reins_1}

Require: $\mathcal{A} = (A, R), S \subseteq A$

```
(A', R') ← G_Reins1(S)
while |A' \ S| > 1 do
    x ← choose(A' \ S)
    A' ← A' \ {x}
end while
return (A', R')
```

Proposition 5. *Let $\mathcal{A} = (A, R)$, $S \subseteq A$ and $\pi \in \{CF, Def, Reins_1, Reins_2, CA\}$. Algorithm Alg_π using $\mathcal{A}$ and S as inputs is sound and complete for the computation of a minimal explanation that answers Q_π.*

Proof.

• *Alg_{CF}.* It begins by computing $(A', R') = G_{CF}(S)$ which, by Prop. 4 is a maximal answer to Q_{CF}. So, in particular, it is an answer to Q_{CF}. Obviously $(x, y) \leftarrow choose(R')$ implies that $(x, y) \in R'$. This would mean that $R' \setminus \{(x, y)\} \subset R'$, and thus that $|R' \setminus \{(x, y)\}| < |R'|$. As such, lines 2-5 compute (A'', R'') such that $A'' = A'$, $R'' \subseteq R'$ (in cases $R' = \varnothing$ and $|R'| = 1$, we have $R'' = R'$) and it holds that $|R''| \leq 1$. Let $X = \{(a, b) \in R \mid a, b \in S\}$. As (A', R') is an answer to Q_{CF}, by Def. 11, we know that $A' = S$ and $R' \subseteq X$. But $A'' = A'$, so $A'' = S$ and $R'' \subseteq R'$, so $R'' \subseteq X$. In addition, if $R' = \varnothing$, then $X = \varnothing$ by Def. 11, but $R'' \subseteq X$, so we have $R'' = \varnothing$ as well. Thus, by Def. 11, (A'', R'') is an answer to Q_{CF}. Moreover, we know that $|R''| \leq 1$, so by Lem. 6, (A'', R'') is a minimal answer to Q_{CF}. So Alg_{CF} is sound.

By Lem. 6, we know that $|R''| \leq 1$. Alg_{CF} begins by computing $(A', R') =$

Alg$_{Reins_2}$ Computation of a minimal answer to Q_{Reins_2}

Require: $\mathcal{A} = (A, R), S \subseteq A$

1: $(A', R') \leftarrow G_{Reins_2}(S)$
2: **for** $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$ **do**
3: **while** $|R'^{-1}(y)| > 1$ **do**
4: $x \leftarrow choose(R'^{-1}(y))$
5: $R' \leftarrow R' \setminus \{(x, y)\}$
6: **end while**
7: **end for**
8: **return** (A', R')

Alg$_{CA}$ Computation of a minimal answer to Q_{CA}

Require: $\mathcal{A} = (A, R), S \subseteq A$

1: $(A', R') \leftarrow G_{CA}(S)$
2: **for** $y \in A \setminus S$ **do**
3: **while** $|R'^{-1}(y)| > 1$ **do**
4: $x \leftarrow choose(R'^{-1}(y))$
5: $R' \leftarrow R' \setminus \{(x, y)\}$
6: **end while**
7: **end for**
8: **return** (A', R')

$G_{CF}(S)$ which, by Prop. 4, is a maximal answer to Q_{CF}. Since (A'', R'') and (A', R') are both answers to Q_{CF}, by Def. 11, we know that $A'' = S = A'$. In addition, by Th. 9, we know that $(A'', R'') \subseteq (A', R')$. Let $X = \{(a, b) \in R \mid a, b \in S\}$ and (A''', R''') be the result computed by Alg$_{CF}$. In the case where $|R''| = 0$, by Def. 11 we have $X = \varnothing$, and thus, still by Def. 11, $R' = \varnothing$. So, in this case, $R' = R''$. Lines 2-5 are ignored and Alg$_{CF}$ computes (A''', R''') with $(A''', R''') = (A', R') = (A'', R'')$. In the case where $|R''| = 1$, we denote $R'' = \{(x_0, y_0)\}$. Since $(A'', R'') \subseteq (A', R')$, $R'' \subseteq R'$ and so $(x_0, y_0) \in R'$. We denote $R' = \{(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)\}$. Obviously $(x, y) \leftarrow choose(R')$ implies that $(x, y) \in R'$. This would mean that $R' \setminus \{(x, y)\} \subset R'$, and thus that $|R' \setminus \{(x, y)\}| < |R'|$. As such, lines 2-5 compute (A''', R''') such that $A''' = A' = A''$ and $R''' = R' \setminus \Delta$ with $|R'''| = 1$. As we already know that $A''' = A''$, we only need to find a set Δ such that $R' \setminus \Delta = R''$. $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ is such a set. So Alg$_{CF}$ is complete.

• Alg_{Def}. It begins by computing $(A', R') = G_{Def}(S)$ which, by Prop. 4 is a maximal answer to Q_{Def}. So, in particular, it is an answer to Q_{Def}. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute (A'', R'') such that, $A'' = A'$, for some $y \in R^{-1}(S) \setminus S$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ (in cases $R'^{-1}(y) = \varnothing$ and $|R'^{-1}(y)| = 1$, we have $R''^{-1}(y) = R'^{-1}(y)$) and it holds that $|R''^{-1}(y)| \leq 1$. Thus, lines 2-7 compute (A'', R'') such that, $A'' = A'$ and for all $y \in R^{-1}(S) \setminus S$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ and $|R''^{-1}(y)| \leq 1$. Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$. As (A', R') is an answer to Q_{Def}, by Def. 12, we know that $A' = S \cup R^{-1}(S)$ and $X \subseteq R' \subseteq X \cup Y$. But $A'' = A'$, so $A'' = S \cup R^{-1}(S)$, $R'' \subseteq R'$, so $R'' \subseteq X \cup Y$, and as $y \in R^{-1}(S) \setminus S$ and $(x, y) \in R'$, we deduce that $(x, y) \in Y \setminus X$ and so $X \subseteq R''$. In addition, if $R'^{-1}(y) \cap S = \varnothing$, then $y \notin R^{+1}(S)$ by Def. 12, but $R'' \subseteq X \cup Y$, so we have $R''^{-1}(y) \cap S = \varnothing$ as well. Thus, by Def. 12, (A'', R'') is an answer to Q_{Def}. Moreover, we know that for all $y \in R^{-1}(S) \setminus S$, $|R''^{-1}(y)| \leq 1$, so by Lem. 7, (A'', R'') is a minimal answer to Q_{Def}. So Alg_{Def} is sound.

By Lem. 7, we know that for all $y \in R^{-1}(S) \setminus S$, $|R''^{-1}(y)| \leq 1$. Alg_{Def} begins by computing $(A', R') = G_{Def}(S)$ which, by Prop. 4, is a maximal answer to Q_{Def}. Since (A'', R'') and (A', R') are both answers to Q_{Def}, by Def. 12, we know that $A'' = S \cup R^{-1}(S) = A'$. In addition, by Th. 9, we know that $(A'', R'') \subseteq (A', R')$. Let $X = \{(b, a) \in R \mid b \in R^{-1}(S), a \in S\}$, $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(S)\}$, (A''', R''') be the result computed by Alg_{Def} and consider $y \in R^{-1}(S) \setminus S$. In the case where $|R''^{-1}(y)| = 0$, in particular, $\nexists (x, y) \in R''$ with $x \in S$. So, by Def. 12 we have $y \notin R^{+1}(S)$, and thus, still by Def. 12, $R'^{-1}(y) = \varnothing$. So, in this case, $R'^{-1}(y) = R''^{-1}(y)$ and lines 3-6 are ignored. Thus, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$, $R''' = R'$ and so, $R'''^{-1}(y) = R''^{-1}(y)$. In the case where $|R''^{-1}(y)| = 1$, we denote $R''^{-1}(y) = \{x_0\}$. Since $(A'', R'') \subseteq (A', R')$, $R'' \subseteq R'$, so $(x_0, y) \in R'$ and in particular, $x_0 \in R'^{-1}(y)$. We denote $R'^{-1}(y) = \{x_0, x_1, \ldots, x_n\}$. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$ and $R''' = R' \setminus \Delta$ with $|R'''^{-1}(y)| = 1$. So, we only need to find a set Δ such that $(R' \setminus \Delta)^{-1}(y) = R''^{-1}(y)$. $\{(x_1, y), \ldots, (x_n, y)\}$ is such a set. So, using $\Delta = \{(x_1, y), \ldots, (x_n, y)\}$ in the second case, lines 3-6 compute (A''', R''') such that $A''' = A''$ and for some $y \in R^{-1}(S) \setminus S$,

$R'''^{-1}(y) = R''^{-1}(y)$. Thus, lines 2-7 compute (A''', R''') such that $A''' = A''$ and for all $y \in R^{-1}(S) \setminus S$, $R'''^{-1}(y) = R''^{-1}(y)$. Since $X \subseteq R''' \subseteq X \cup Y$ and for all $y \in R^{-1}(S) \setminus S$, $R'''^{-1}(y) = R''^{-1}(y)$, we deduce that $R''' = R''$ and so $(A''', R''') = (A'', R'')$. So Alg_{Def} is complete.

• Alg_{Reins_1}. It begins by computing $(A', R') = G_{Reins_1}(S)$ which, by Prop. 4 is a maximal answer to Q_{Reins_1}. So, in particular, it is an answer to Q_{Reins_1}. Obviously $x \leftarrow choose(A' \setminus S)$ implies that $x \in A' \setminus S$. This would mean that $A' \setminus \{x\} \subset A'$, and thus that $|A' \setminus \{x\}| < |A'|$. In addition, since $x \in A' \setminus S$, $|(A' \setminus S) \setminus \{x\}| < |A' \setminus S|$. As such, lines 2-5 compute (A'', R'') such that $R'' = R'$, $A'' \subseteq A'$ (in cases $A' \setminus S = \varnothing$ and $|A' \setminus S| = 1$, we have $A'' = A'$) and it holds that $|A'' \setminus S| \leq 1$. Let $X = \{a \in A \,|R^{-1}(a) = \varnothing\}$. As (A', R') is an answer to Q_{Reins_1}, by Def. 13, we know that $R' = \varnothing$ and $S \cap X \subseteq A' \subseteq X$. But $R'' = R'$, so $R'' = \varnothing$, $A'' \subseteq A'$, so $A'' \subseteq X$ and since $x \in A' \setminus S$, $S \cap X \subseteq A''$. In addition, if $(A' \setminus S) \cap X = \varnothing$, then $(A \setminus S) \cap X = \varnothing$ by Def. 13, but $A'' \subseteq X$, so we have $(A'' \setminus S) \cap X = \varnothing$ as well. Thus, by Def. 13, (A'', R'') is an answer to Q_{Reins_1}. Moreover, we know that $|A'' \setminus S| \leq 1$, so by Lem. 8, (A'', R'') is a minimal answer to Q_{Reins_1}. So Alg_{Reins_1} is sound.

By Lem. 8, we know that $|A'' \setminus S| \leq 1$. Alg_{Reins_1} begins by computing $(A', R') = G_{Reins_1}(S)$ which, by Prop. 4, is a maximal answer to Q_{Reins_1}. Since (A'', R'') and (A', R') are both answers to Q_{Reins_1}, by Def. 13, we know that $R'' = \varnothing = R'$. In addition, by Th. 9, we know that $(A'', R'') \subseteq (A', R')$. Let $X = \{a \in A \mid R^{-1}(a) = \varnothing\}$ and (A''', R''') be the result computed by Alg_{Reins_1}. In the case where $|A'' \setminus S| = 0$, by Def. 13 we have $(A \setminus S) \cap X = \varnothing$, and thus, $A' \setminus S = \varnothing$. So, in this case, $A' = S \cap X = A''$. In addition, lines 2-5 are ignored and Alg_{Reins_1} computes (A''', R''') with $(A''', R''') = (A', R') = (A'', R'')$. In the case where $|A'' \setminus S| = 1$, we denote $A'' \setminus S = \{x_0\}$. Since $(A'', R'') \subseteq (A', R')$, $A'' \subseteq A'$ and so $x_0 \in A'$. We denote $A' \setminus S = \{x_0, x_1, \ldots, x_n\}$. Obviously $x \leftarrow choose(A' \setminus S)$ implies that $x \in A' \setminus S$. This would mean that $A' \setminus \{x\} \subset A'$, and thus that $|A' \setminus \{x\}| < |A'|$. In addition, since $x \in A' \setminus S$, $|(A' \setminus S) \setminus \{x\}| < |A' \setminus S|$. Thus, lines 2-5 compute (A''', R''') such that $R''' = R' = R''$ and $A''' \setminus S = (A' \setminus S) \setminus \Delta$ with $|A''' \setminus S| = 1$. As we already know that $R''' = R''$, we only need to find a set Δ such that $(A' \setminus S) \setminus \Delta = A'' \setminus S$. $\{x_1, \ldots, x_n\}$ is such a set. So Alg_{Reins_1} is complete.

• Alg_{Reins_2}. It begins by computing $(A', R') = G_{Reins_2}(S)$ which, by Prop. 4 is a maximal answer to Q_{Reins_2}. So, in particular, it is an answer to Q_{Reins_2}. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute

(A'', R'') such that, $A'' = A'$, for some $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ (in cases $R'^{-1}(y) = \varnothing$ and $|R'^{-1}(y)| = 1$, we have $R''^{-1}(y) = R'^{-1}(y)$) and it holds that $|R''^{-1}(y)| \leq 1$. Thus, lines 2-7 compute (A'', R'') such that, $A'' = A'$ and for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ and $|R''^{-1}(y)| \leq 1$. Let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$ and $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$. As (A', R') is an answer to Q_{Reins_2}, by Def. 14, we know that $A' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$ and $X \subseteq R' \subseteq X \cup Y$. But $A'' = A'$, so $A'' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S))$, $R'' \subseteq R'$, so $R'' \subseteq X \cup Y$, and as $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$ and $x \in S$, we deduce that $(x, y) \in Y \setminus X$ and so $X \subseteq R''$. In addition, if $R'^{-1}(y) = \varnothing$, then $y \notin R^{+1}(S)$ by Def. 14, but $R'' \subseteq X \cup Y$, so we have $R''^{-1}(y) = \varnothing$ as well. Thus, by Def. 14, (A'', R'') is an answer to Q_{Reins_2}. Moreover, we know that for all $y \in R^{-1}(S) \setminus R^{+2}(S)$, $|R''^{-1}(y)| \leq 1$, so by Lem. 9, (A'', R'') is a minimal answer to Q_{Reins_2}. So Alg_{Reins_2} is sound.

By Lem. 9, we know that for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $|R''^{-1}(y)| \leq 1$. Alg_{Reins_2}, begins by computing $(A', R') = G_{Reins_2}(S)$ which, by Prop. 4 is a maximal answer to Q_{Reins_2}. Since (A'', R'') and (A', R') are both answers to Q_{Reins_2}, by Def. 14, we know that $A'' = S \cup R^{+2}(S) \cup R^{-1}(R^{+2}(S)) = A'$. In addition, by Th. 9, we know that $(A'', R'') \subseteq (A', R')$. Let $X = \{(b, c) \in R \mid b \in R^{-1}(R^{+2}(S)), c \in R^{+2}(S)\}$, $Y = \{(a, b) \in R \mid a \in S, b \in R^{-1}(R^{+2}(S))\}$, (A''', R''') be the result computed by Alg_{Reins_2} and consider $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$. In the case where $|R''^{-1}(y)| = 0$, $\nexists (x, y) \in R''$ with $x \in S$. So, by Def. 14 we have $y \notin R^{+1}(S)$, and thus, still by Def. 14, $R'^{-1}(y) = \varnothing$. So, in this case, $R'^{-1}(y) = R''^{-1}(y)$ and lines 3-6 are ignored. Thus, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$, $R''' = R'$ and so, $R'''^{-1}(y) = R''^{-1}(y)$. In the case where $|R''^{-1}(y)| = 1$, we denote $R''^{-1}(y) = \{x_0\}$. Since $(A'', R'') \subseteq (A', R')$, $R'' \subseteq R'$, so $x_0 \in R'$ and in particular, $x_0 \in R'^{-1}(y)$. We denote $R'^{-1}(y) = \{x_0, x_1, \ldots, x_n\}$. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$ and $R''' = R' \setminus \Delta$ with $|R'''^{-1}(y)| = 1$. So, we only need to find a set Δ such that $(R' \setminus \Delta)^{-1}(y) = R''^{-1}(y)$. $\{(x_1, y), \ldots, (x_n, y)\}$ is such a set. So, using $\Delta = \{(x_1, y), \ldots, (x_n, y)\}$ in the second case, lines 3-6 compute (A''', R''') such that $A''' = A''$ and for some $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $R'''^{-1}(y) = R''^{-1}(y)$. As such, lines 2-7 thus compute (A''', R''') such that $A''' = A''$ and for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $R'''^{-1}(y) = R''^{-1}(y)$. Since $X \subseteq R''' \subseteq X \cup Y$ and for all $y \in R^{-1}(R^{+2}(S)) \setminus R^{+2}(S)$, $R'''^{-1}(y) = R''^{-1}(y)$,

we deduce that $R''' = R''$ and so $(A''', R''') = (A'', R'')$. So Alg_{Reins_2} is complete.

• Alg_{CA}. It begins by computing $(A', R') = G_{CA}(S)$ which, by Prop. 4 is a maximal answer to Q_{CA}. So, in particular, it is an answer to Q_{CA}. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute (A'', R'') such that, $A'' = A'$, for some $y \in A \setminus S$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ (in cases $R'^{-1}(y) = \varnothing$ and $|R'^{-1}(y)| = 1$, we have $R''^{-1}(y) = R'^{-1}(y)$) and it holds that $|R''^{-1}(y)| \leq 1$. Thus, lines 2-7 compute (A'', R'') such that, $A'' = A'$ and for all $y \notin S$, $R''^{-1}(y) \subseteq R'^{-1}(y)$ and $|R''^{-1}(y)| \leq 1$. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$. As (A', R') is an answer to Q_{CA}, by Def. 15, we know that $A' = A$ and $R' \subseteq X$. But $A'' = A'$, so $A'' = A$ and $R'' \subseteq R'$, so $R'' \subseteq X$. In addition, if $R'^{-1}(y) \cap S = \varnothing$, then $y \notin R^{+1}(S)$ by Def. 15, but $R'' \subseteq X$, so we have $R''^{-1}(y) \cap S = \varnothing$ as well. Thus, by Def. 15, (A'', R'') is an answer to Q_{CA}. Moreover, we know that for all $y \notin S$, $|R''^{-1}(y)| \leq 1$, so by Lem. 10, (A'', R'') is a minimal answer to Q_{CA}. So Alg_{CA} is sound.

By Lem. 10, we know that for all $y \notin S$, $|R''^{-1}(y)| \leq 1$. Alg_{CA} begins by computing $(A', R') = G_{CA}(S)$ which, by Prop. 4, is a maximal answer to Q_{CA}. Since (A'', R'') and (A', R') are both answers to Q_{CA}, by Def. 15, we know that $A'' = A = A'$. In addition, by Th. 9, we know that $(A'', R'') \subseteq (A', R')$. Let $X = \{(a, b) \in R \mid a \in S, b \notin S\}$, (A''', R''') be the result computed by Alg_{CA} and consider $y \notin S$. In the case where $|R''^{-1}(y)| = 0$, in particular, $\nexists (x, y) \in R''$ with $x \in S$. So, by Def. 15 we have $y \notin R^{+1}(S)$, and thus, still by Def. 15, $R'^{-1}(y) = \varnothing$. So, in this case, $R'^{-1}(y) = R''^{-1}(y)$ and lines 3-6 are ignored. Thus, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$, $R''' = R'$ and so, $R'''^{-1}(y) = R''^{-1}(y)$. In the case where $|R''^{-1}(y)| = 1$, we denote $R''^{-1}(y) = \{x_0\}$. Since $(A'', R'') \subseteq (A', R')$, $R'' \subseteq R'$, so $x_0 \in R'$ and in particular, $x_0 \in R'^{-1}(y)$. We denote $R'^{-1}(y) = \{x_0, x_1, \ldots, x_n\}$. Obviously $x \leftarrow choose(R'^{-1}(y))$ implies that $x \in R'^{-1}(y)$. In particular, it implies that $(x, y) \in R'$. This would mean that $R'^{-1}(y) \setminus \{(x, y)\} \subset R'^{-1}(y)$, and thus that $|R'^{-1}(y) \setminus \{(x, y)\}| < |R'^{-1}(y)|$. As such, lines 3-6 compute (A''', R''') such that $A''' = A' = A''$ and $R''' = R' \setminus \Delta$ with $|R'''^{-1}(y)| = 1$. So, we only need to find a set Δ such that $(R' \setminus \Delta)^{-1}(y) = R''^{-1}(y)$. $\{(x_1, y), \ldots, (x_n, y)\}$ is such a set. So, using $\Delta = \{(x_1, y), \ldots, (x_n, y)\}$ in the second case, lines 3-6 compute (A''', R''') such that $A''' = A''$ and for some $y \notin S$, $R'''^{-1}(y) = R''^{-1}(y)$. As such, lines 2-7 thus compute (A''', R''') such that $A''' = A''$ and for all $y \notin S$, $R'''^{-1}(y) = R''^{-1}(y)$. Since $R''' \subseteq X$ and for all $y \notin S$, $R'''^{-1}(y) = R''^{-1}(y)$, we

deduce that $R''' = R''$ and so $(A''', R''') = (A'', R'')$. So Alg_{CA} is complete. $\square$

The computation of minimal explanations thus relies on the computation of maximal explanations, and the removal of some arcs (or arguments) in them. The computation of maximal explanations is already known to be polynomial (see [BDDLS22]). Moreover the complexity of the removal operation in the worst case is linear in the number of removed elements and this number is either quadratic in the number of vertices in the graph when these elements are attacks (so for any principle except the one for the first part of reinstatement), or linear in the number of vertices in the graph when these elements are vertices (for the first part of reinstatement). So globally, our algorithms can be considered as computationally efficient.

6 Conclusion and Future Work

This paper has presented an extension of the work done in [BDDLS22] regarding the Explanation Verification Problem (XVer). In particular, we defined classes of answers to questions regarding the principles of abstract argumentation semantics. We extended the results of [BDDLS22] to our classes, providing ways of using them as explanations of argumentative results. In addition, we studied these classes of explanations according to general properties such as maximality, minimality, emptyness and uniqueness. We showed that the explanations studied in [BDDLS22] correspond to the maximal explanations of our classes, thus providing a way to compute them using graph operators. We also provided a way to compute minimal explanations from the maximal ones, and proved this procedure to be sound and complete for each class of explanation. A complete summary of the results is given in Tab. 1. These results make an implementation of the proposed approach ready to be done. From this implementation, like in any XAI approach, as underlined by [ČRA+21b], an empirical evaluation should be conducted to assess whether these visual explanations actually are helpful for human agents to understand the answer to the Verification Problem. Such an evaluation should be done in future work.

By defining classes of explanations, the system adapts to a wider set of agents, be they human or artificial. This adaptation can go even further if the question of the "realizability" of an explanation would be considered: an agent may have in mind parts of an explanation (some arguments, some

<table>
<tr><th rowspan="2">Principle π</th><th rowspan="2">defined in</th><th colspan="6">involved in:</th></tr>
<tr><th colspan="2">link with [BDDLS22]</th><th colspan="2">specific explanations</th><th colspan="2">computation</th></tr>
<tr><td>CF</td><td>Def. 11</td><td>Th. 1</td><td rowspan="5">Prop. 4</td><td>Lem. 1, Lem. 6</td><td rowspan="5">Th. 6, 7, 8, 9</td><td>Alg_{CF}</td><td rowspan="5">Prop. 4, 5</td></tr>
<tr><td>Def</td><td>Def. 12</td><td>Th. 2, Prop. 2</td><td>Lem. 2, Lem. 7</td><td>Alg_{Def}</td></tr>
<tr><td>$Reins_1$</td><td>Def. 13</td><td rowspan="2">Th. 3, 4, Cor. 1</td><td>Lem. 3, Lem. 8</td><td>Alg_{Reins_1}</td></tr>
<tr><td>$Reins_2$</td><td>Def. 14</td><td>Lem. 4, Lem. 9</td><td>Alg_{Reins_2}</td></tr>
<tr><td>CA</td><td>Def. 15</td><td>Th. 5, Prop. 3</td><td>Lem. 5, Lem. 10</td><td>Alg_{CA}</td></tr>
</table>

- Th. 1 to 5 and Prop. 2 to 4 establish links between our approach and the approach described in [BDDLS22].
- Lem. 1 to 5 are dedicated to the characterisation of maximal explanations. Lem. 6 to 10 are dedicated to the characterisation of minimal explanations.
- Th. 6 is about emptyness. Th. 7 is about emptyness and uniqueness. Th. 8 is about maximality and uniqueness. Th. 9 is about the link between maximality and minimality.
- Prop. 4 is about the computation of the maximal explanations (using a link with some results given in [BDDLS22]). Prop. 5 is about the computation of the minimal explanations.

Table 1: Summary of the results

attacks), but not a correct and complete explanation; determining whether there exists such an explanation, and providing it, would ensure that an explanation that suits best the agent would be provided. Classes of explanations, by the wideness of the explanations that they propose, encourage the feasibility of such an approach. In order to do so, a deeper investigation of the inner structure of our classes of explanation, and more specifically of the links that we think it could have with lattices, may be of help.

[BDDLS22] also proposed explanations to contrastive questions: generalising them to classes of explanations, following the approach presented in the current paper, could be addressed. Extending XVer to additional semantics (especially preferred and grounded) may also be considered.

In some cases, our notion of minimal explanations could be refined. Consider for instance Fig. 1 and the question "Why is $\{b\}$ not admissible?". In this case, when the explanation for the defence principle is built, all the attackers of b in the initial framework must be considered (so the arguments c, d and the attacks (c, b) and (d, b) belong to this explanation) whereas only one would be useful for explaining that b is non defended, since neither c, nor d are attacked. Thus, if we want to obtain such a minimal explanation, we could consider either lifting some conditions in the definitions of our explanations, or defining new classes dedicated to the case where a principle is not respected by a set.

Finally, we plan to investigate more notions of Graph Theory in order to provide other kinds of visual explanations. In particular, the notion of graph isomorphism seems of great interest to us, especially to provide ways of reasoning by association (explain a result via a structurally identical argumentation framework that the user already accepted).

References

[BB21a] AnneMarie Borg and Floris Bex. A basic framework for explanations in argumentation. *IEEE Intelligent Systems*, 36(2):25–35, 2021.

[BB21b] AnneMarie Borg and Floris Bex. Necessary and sufficient explanations for argumentation-based conclusions. In *Proc. of ECSQARU*, volume 12897 of *LNCS*, pages 45–58, Prague, Czech Republic, 2021. Springer.

[BDDLS22] Philippe Besnard, Sylvie Doutre, Théo Duchatelle, and Marie-Christine Lagasquie-Schiex. Explaining semantics and extension membership in abstract argumentation. *Intelligent Systems with Applications*, 16:200118, 2022.

[BU21] Ringo Baumann and Markus Ulbricht. Choices and their consequences - explaining acceptable sets in abstract argumentation frameworks. In *Proc. of KR*, pages 110–119, Online event, 2021. IJCAI Organization.

[CČRT18] Oana Cocarascu, Kristijonas Čyras, Antonio Rago, and Francesca Toni. Explaining with argumentation frameworks mined from data. In *Proc. of DEXAHAI*, Southampton, United Kingdom, 2018.

[ČRA+21a] Kristijonas Čyras, Antonio Rago, Emanuele Albini, Pietro Baroni, and Francesca Toni. Argumentative XAI: A survey. In *Proc. of IJCAI*, pages 4392–4399, Online Event / Montreal, Canada, 2021. IJCAI Organization.

[ČRA+21b] Kristijonas Čyras, Antonio Rago, Emanuele Albini, Pietro Baroni, and Francesca Toni. Argumentative XAI: A survey. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence IJCAI*, pages 4392–4399, Online Event / Montreal, Canada, August 2021. IJCAI Organization.

[DD18] Wolfgang Dvorák and Paul E Dunne. Computational problems in formal argumentation and their complexity. *Handbook of formal argumentation*, 4:631–688, 2018.

[DM16] Sylvie Doutre and Jean-Guy Mailly. Quantifying the difference between argumentation semantics. In *Proc. of COMMA*, volume 287, pages 255–262, Potsdam, Germany, 2016. IOS Press.

[Dun95] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.

[FT15] Xiuyi Fan and Francesca Toni. On computing explanations in argumentation. In *Proc. of AAAI*, pages 1496–1502, Austin, Texas, USA, 2015. AAAI Press.

[GTWX16] Yang Gao, Francesca Toni, Hao Wang, and Fanjiang Xu. Argumentation-based multi-agent decision making with privacy preserved. In *Proc. of AAMAS*, pages 1153–1161, Singapore, 2016. ACM.

[LvdT20] Beishui Liao and Leendert van der Torre. Explanation semantics for abstract argumentation. In *Proc. of COMMA*, volume 326, pages 271–282, Perugia, Italy, 2020. IOS Press.

[NJ20] Andreas Niskanen and Matti Järvisalo. Smallest explanations and diagnoses of rejection in abstract argumentation. In *Proc. of KR*, pages 667–671, Rhodes, Greece, 2020. IJCAI Organization.

[RT21] Teeradaj Racharak and Satoshi Tojo. On explanation of propositional logic-based argumentation system. In *Proc. of ICAART*, volume 2, pages 323–332, Online Streaming, 2021. SCITEPRESS.

[SWW20] Zeynep Gozen Saribatur, Johannes Peter Wallner, and Stefan Woltran. Explaining non-acceptability in abstract argumentation. In *Proc. of ECAI*, volume 325, pages 881–888, Santiago de Compostela, Spain, 2020. IOS Press.

[UW21] Markus Ulbricht and Johannes Peter Wallner. Strong explanations in abstract argumentation. In *Proc. of AAAI*, pages 6496–6504, Online event, 2021. AAAI Press.

[VYT22] Srdjan Vesic, Bruno Yun, and Predrag Teovanovic. Graphical representation enhances human compliance with principles for

graded argumentation semantics. In *Proc. of AAMAS*, pages 1319–1327, Auckland, New Zealand, 2022. IFAAMAS.