# Smallest Explanations and Diagnoses of Rejection in Abstract Argumentation

Andreas Niskanen  Matti Järvisalo

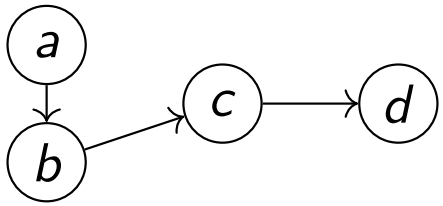
HIIT, Department of Computer Science, University of Helsinki, Finland

September 16, 2020 @ KR 2020, Online

# Motivation

## Argumentation in AI

- Active and vibrant area of modern AI research
- Central KR formalism for reasoning in abstract argumentation: **argumentation frameworks** (AFs) [Dung, 1995]




## Explaining and Diagnosing in Abstract Argumentation

- Understanding **reasons for rejection** important and nontrivial
- **Diagnosing** why no argument is accepted [Ulbricht and Baumann, 2019]
- **Explaining** credulous rejection of an argument [Saribatur et al., 2020]

# Contributions

## What?

- **Provide complexity results** for computing smallest explanations and diagnoses of credulous rejection of a given argument
- **Design declarative algorithms** for practical computation
  - both **argument-based** and **attack-based** explanations and diagnoses

## How?

**Identify correspondences** between

- minimal (smallest) explanations and (smallest) MUSes
- minimal (smallest) diagnoses and (smallest) MCSes

of propositional formulas in CNF

MUS = minimal unsatisfiable subset
MCS = minimal correction set

# Argument-Based Explanations and Diagnoses

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$.

## Definition

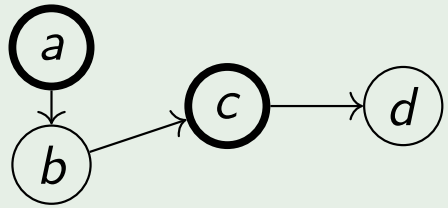
A set $A' \subseteq A$ of arguments is an **explanation** for rejecting $q$:
$q$ remains rejected in any sub-AF containing $A'$

## Definition

A set $A' \subseteq A$ of arguments is a **diagnosis** of rejecting $q$:
$q$ becomes accepted in sub-AF where $A'$ is removed

## Example

$\{a, c\}$ is an explanation for rejecting $d$



# Argument-Based Explanations and Diagnoses

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$.

## Definition

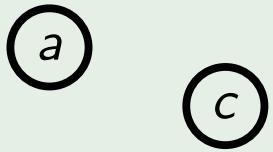
A set $A' \subseteq A$ of arguments is an **explanation** for rejecting $q$:
$q$ remains rejected in any sub-AF containing $A'$

## Definition

A set $A' \subseteq A$ of arguments is a **diagnosis** of rejecting $q$:
$q$ becomes accepted in sub-AF where $A'$ is removed

## Example

$\{a, c\}$ is an explanation for rejecting $d$



# Argument-Based Explanations and Diagnoses

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$.

## Definition

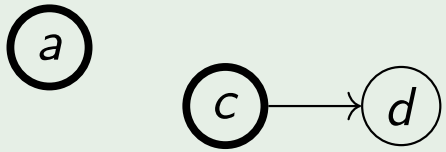
A set $A' \subseteq A$ of arguments is an **explanation** for rejecting $q$:
$q$ remains rejected in any sub-AF containing $A'$

## Definition

A set $A' \subseteq A$ of arguments is a **diagnosis** of rejecting $q$:
$q$ becomes accepted in sub-AF where $A'$ is removed

## Example

$\{a, c\}$ is an explanation for rejecting $d$



# Argument-Based Explanations and Diagnoses

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$.

## Definition

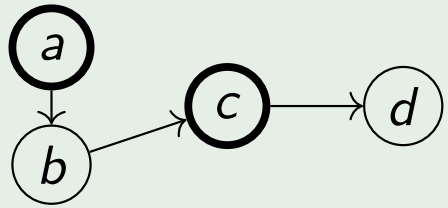
A set $A' \subseteq A$ of arguments is an **explanation** for rejecting $q$:
$q$ remains rejected in any sub-AF containing $A'$

## Definition

A set $A' \subseteq A$ of arguments is a **diagnosis** of rejecting $q$:
$q$ becomes accepted in sub-AF where $A'$ is removed

## Example

$\{a, c\}$ is an explanation for rejecting $d$



# Complexity Results

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$, and an integer $k \geq 0$.

**Theorem**

*Deciding whether there exists an* **explanation** $A' \subseteq A$ *with* $|A'| \leq k$ *for rejecting* $q$ *in* $F$ *under* $\sigma$ *is* $\mathbf{\Sigma_2^p}$**-complete**.

Consider the standard reduction from CNF to AFs. Reduce from deciding whether there is an unsatisfiable subset of size at most $k$. [Liberatore, 2005]

**Theorem**

*Deciding whether there exists a* **diagnosis** $A' \subseteq A$ *with* $|A'| \leq k$ *of rejecting* $q$ *in* $F$ *under* $\sigma$ *is* **NP-complete**.

Reduce from credulous acceptance under $\sigma$.

# Declarative Algorithms

Given an AF $F = (A, R)$, $q \in A$, $\sigma \in \{adm, stb\}$.

$\Rightarrow$ Propositional formulas (with hard and soft clauses) for which

- an **MUS** corresponds to a minimal **explanation**,
- an **MCS** corresponds to a minimal **diagnosis**.

## Computation of Smallest Explanations and Diagnoses

Declaratively via computing smallest MUS/MCS using

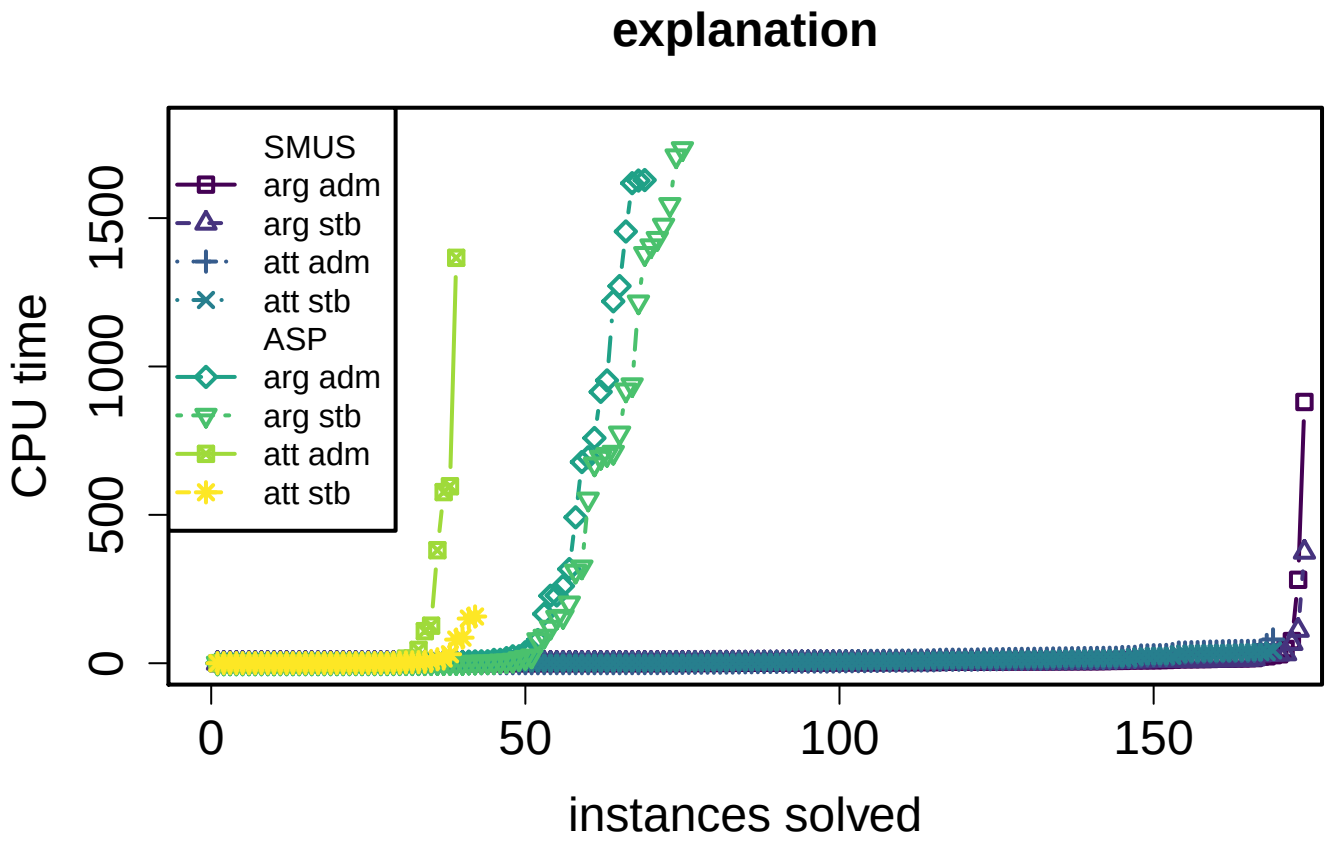
- system for extracting smallest MUS [Ignatiev et al., 2015]
- MaxSAT solver for computing smallest MCS [Ignatiev et al., 2019]

Implementation available online in open source:
`https://bitbucket.org/andreasniskanen/selitae`

# Experiments: Smallest Explanations

Comparison to recent ASP-based approach
for computing smallest explanations

[Saribatur et al., 2020]



# Conclusions

## Paper Summary

- **Complexity results** for deciding small explanations and diagnoses
  - $\Sigma_2^p$-completeness and NP-completeness
- **Algorithms** for computing smallest explanations and diagnoses
  - employing smallest MUS extractors and MaxSAT solvers

## Future Outlook

- Complexity of **attack-based** explanations and diagnoses open
- Dually: explaining and diagnosing **skeptical acceptance**


Dung, P. M. (1995).
On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.
*Artif. Intell.*, 77(2):321–358.


Ignatiev, A., Morgado, A., and Marques-Silva, J. (2019).
RC2: An efficient MaxSAT solver.
*J. Satisf. Boolean Model. Comput.*, 11(1):53–64.


Ignatiev, A., Previti, A., Liffiton, M. H., and Marques-Silva, J. (2015).
Smallest MUS extraction with minimal hitting set dualization.
In *CP*, volume 9255 of *LNCS*, pages 173–182. Springer.


Liberatore, P. (2005).
Redundancy in logic I: CNF propositional formulae.
*Artif. Intell.*, 163(2):203–232.


Saribatur, Z. G., Wallner, J. P., and Woltran, S. (2020).
Explaining non-acceptability in abstract argumentation.
In *ECAI*, volume 325 of *FAIA*, pages 881–888. IOS Press.


Ulbricht, M. and Baumann, R. (2019).
If nothing is accepted - repairing argumentation frameworks.
*J. Artif. Intell. Res.*, 66:1099–1145.