Explanation Semantics for Abstract Argumentation

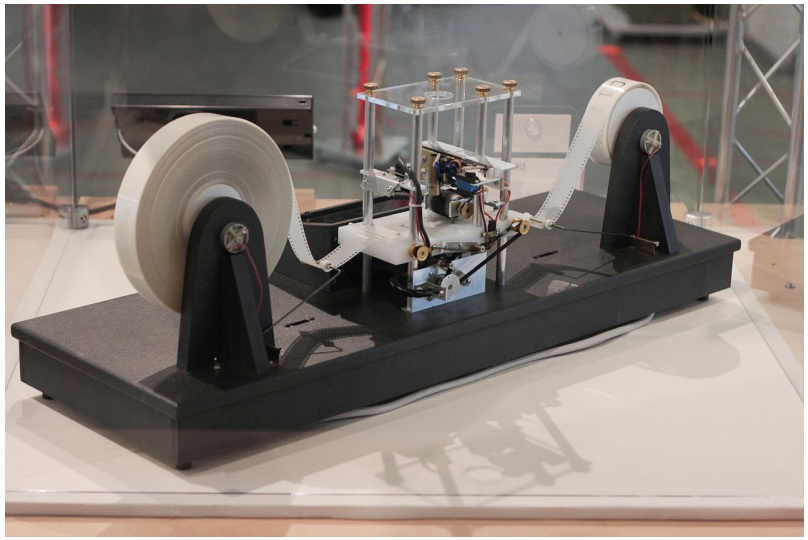
Beishui LIAO [a], Leendert VAN DER TORRE [b,a]

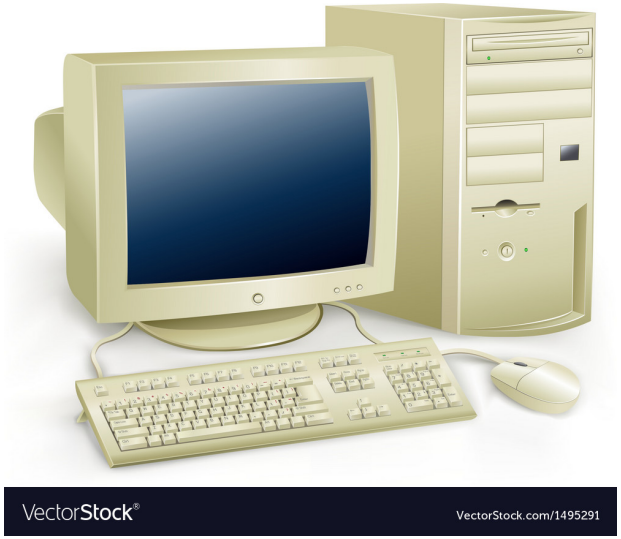
[a] *Zhejiang University*

[b] *University of Luxembourg*

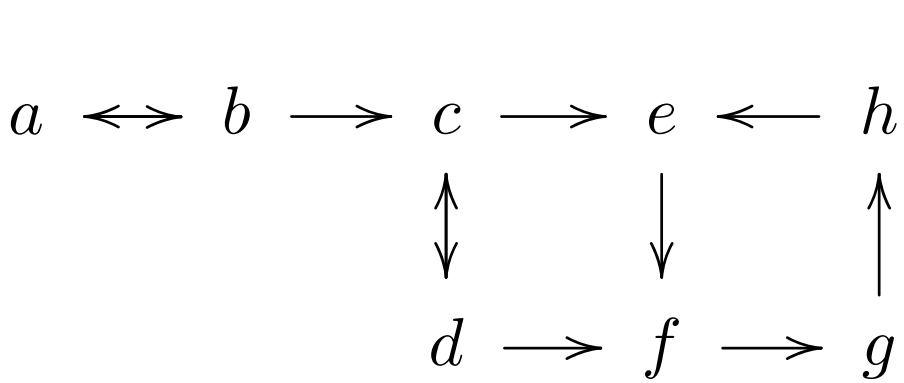
At sabbatical leave at TUWIEN

Is Dung's graph the Turing machine of reasoning?



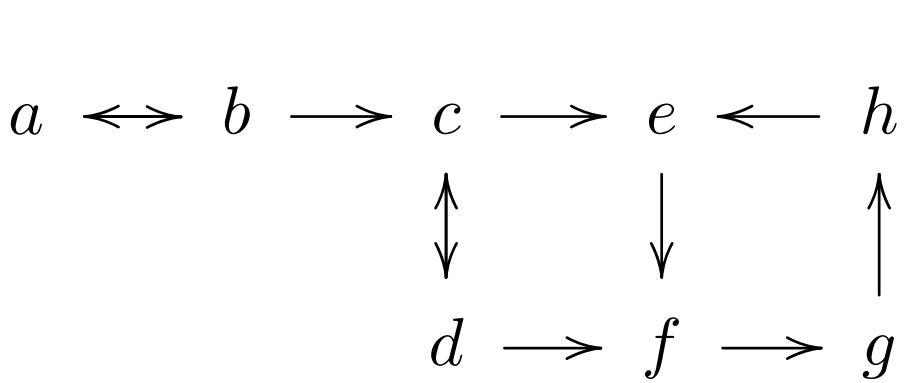


Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

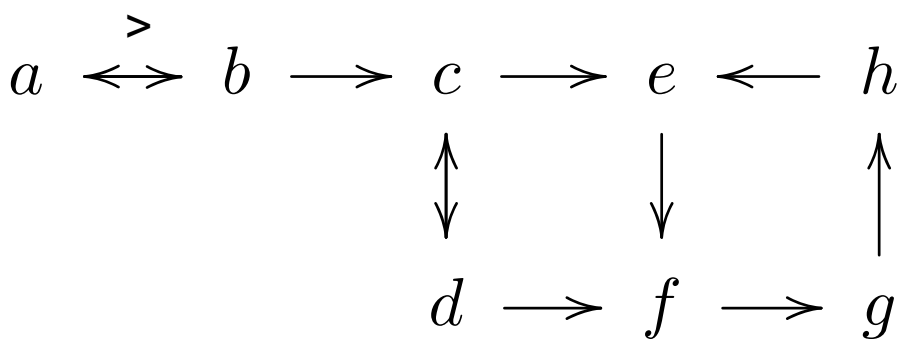
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph

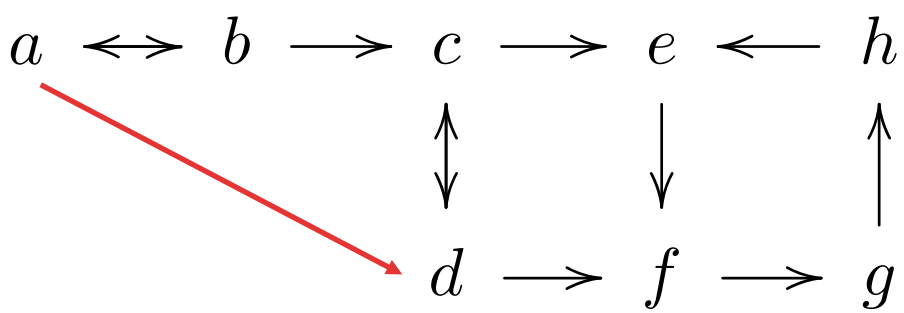
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$

1. Adding ingredients to the graph
    1. Preference

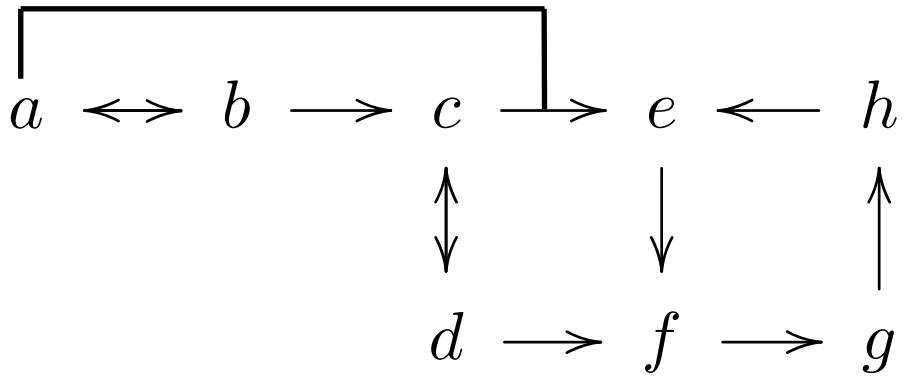
Generalising Dung's abstract argumentation



$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
    1. Preference
    2. Support

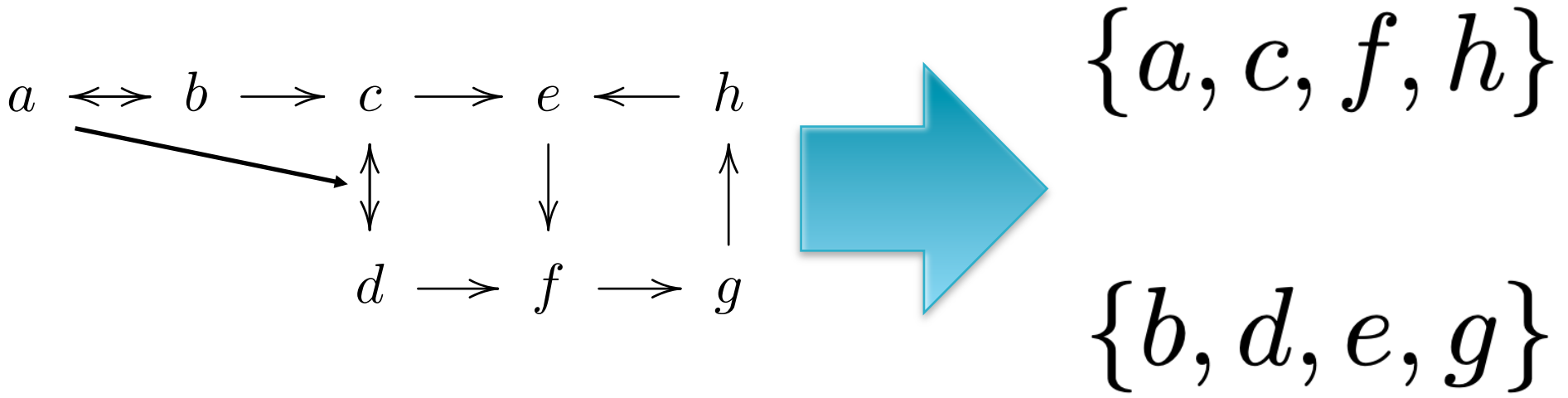
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

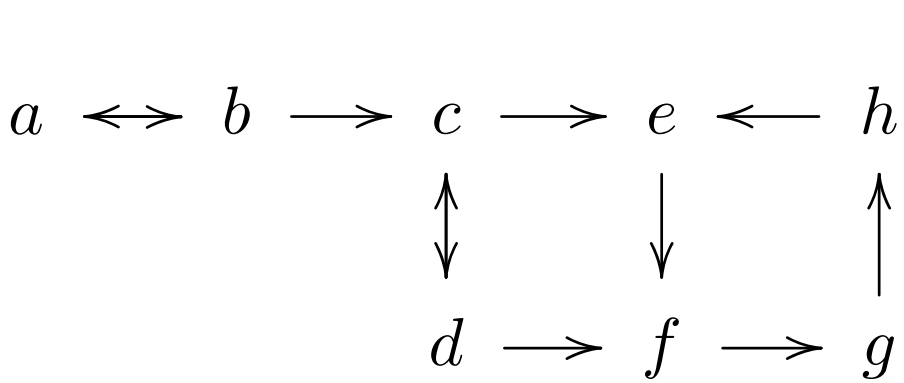
1. Adding ingredients to the graph
    1. Preference
    2. Support
    3. Collective attack/support

Generalising Dung's abstract argumentation



1. Adding ingredients to the graph
    1. Preference
    2. Support
    3. Collective attack/support
    4. Higher-order attack / support

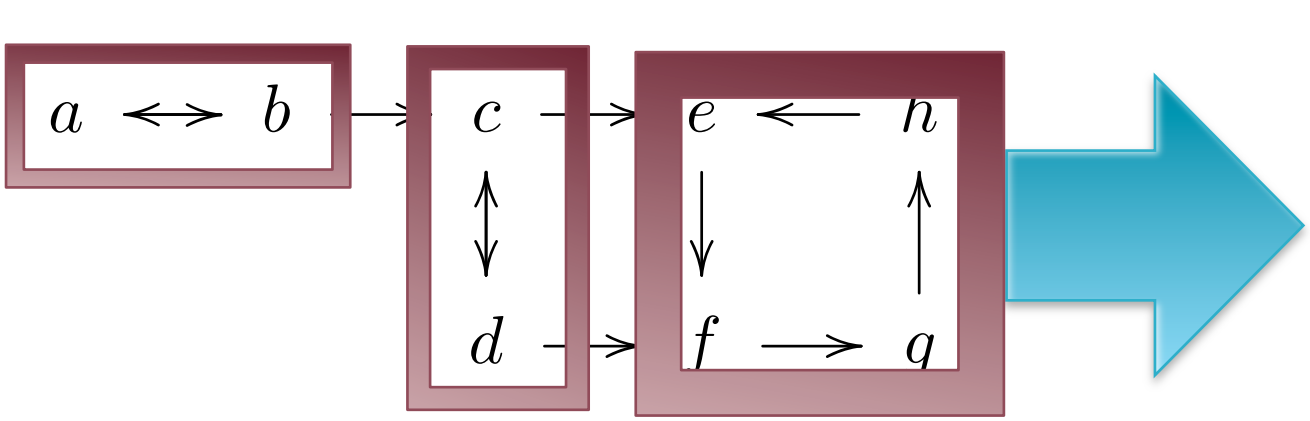
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
    1. Preference
    2. Support
    3. Collective attack/support
    4. Higher-order attack / support
    5. ADF

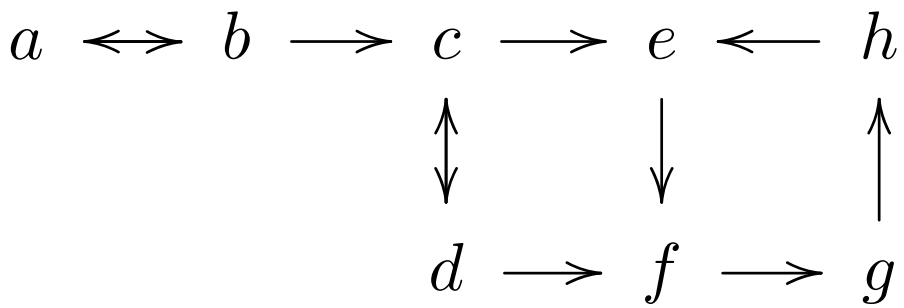
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
    1. Preference
    2. Support
    3. Collective attack/support
    4. Higher-order attack / support
    5. ADF
    6. Input/output (multi-sorted)

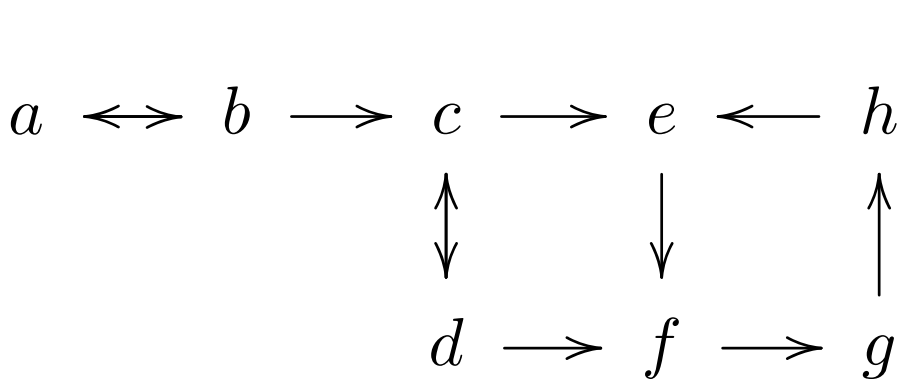
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
    1. Preference
    2. Support
    3. Collective attack/support
    4. Higher-order attack / support
    5. ADF
    6. Input/output (multi-sorted)
    7. …

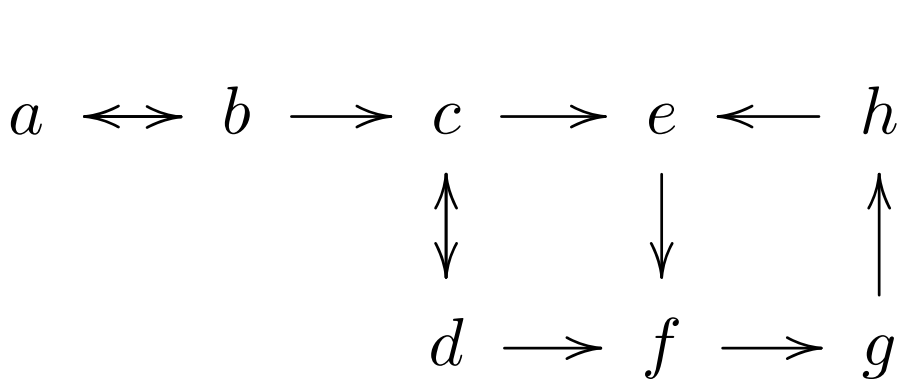
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
2. Introducing new semantics

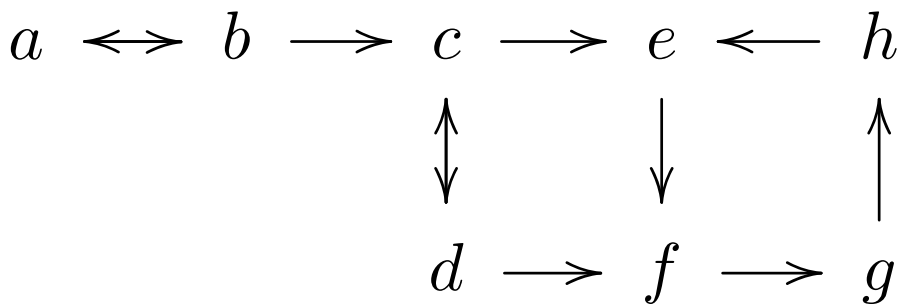
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
2. Introducing new semantics
    1. Admissibility-based: semi-stable, …

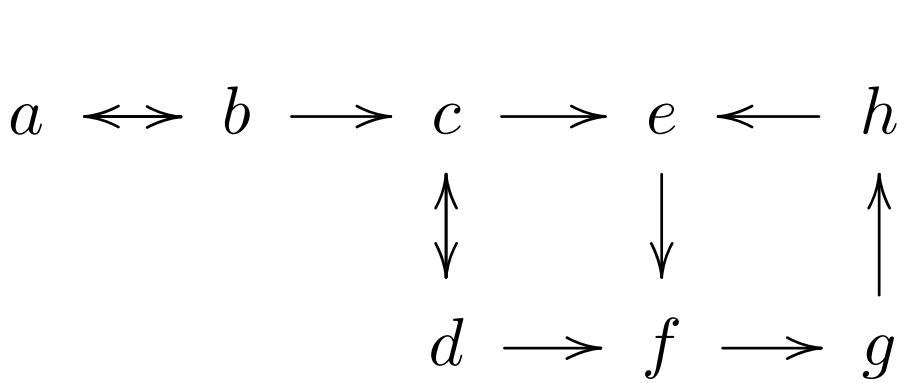
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
2. Introducing new semantics
   1. Admissibility-based: semi-stable, …
   2. Naïve-based: CF2, Stage2, SCF2, …

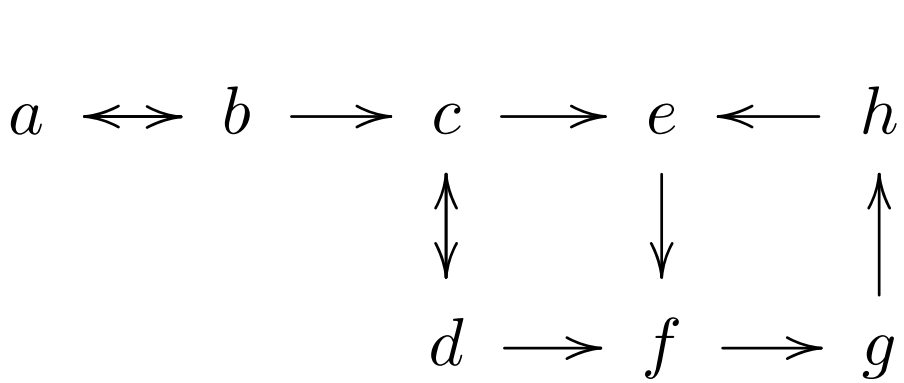
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
2. Introducing new semantics
   1. Admissibility-based: semi-stable, …
   2. Naïve-based: CF2, Stage2, SCF2, …
   3. Weak admissibility based: …

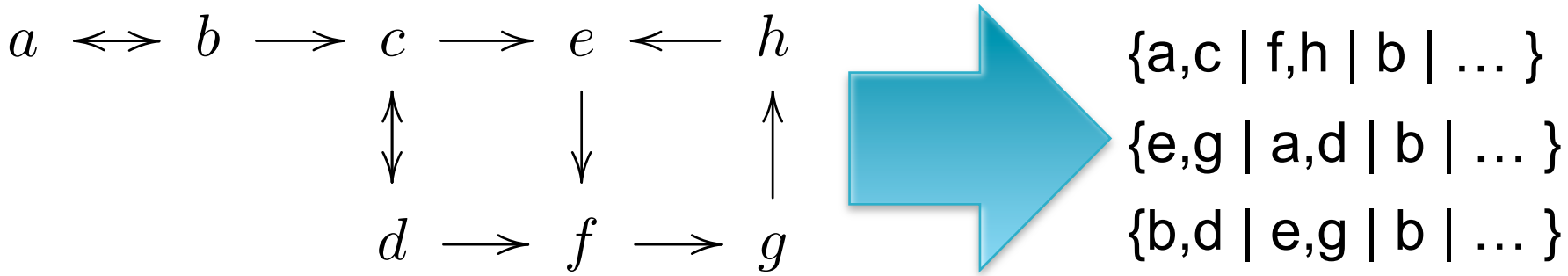
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

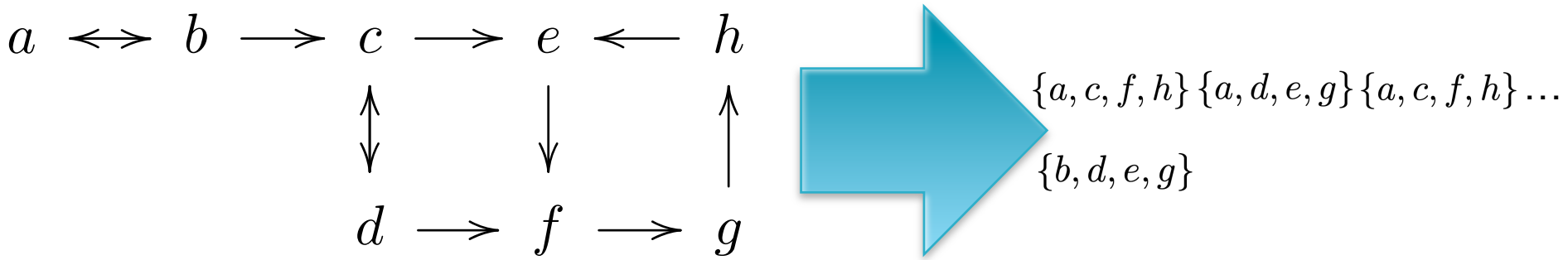
1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions

Generalising Dung's abstract argumentation



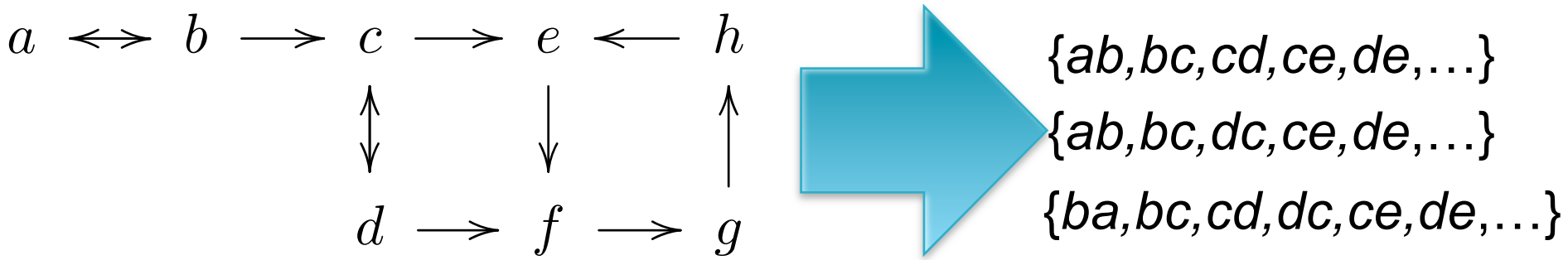
1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions
    1. Ranked / contrary-to-duty semantics

Generalising Dung's abstract argumentation



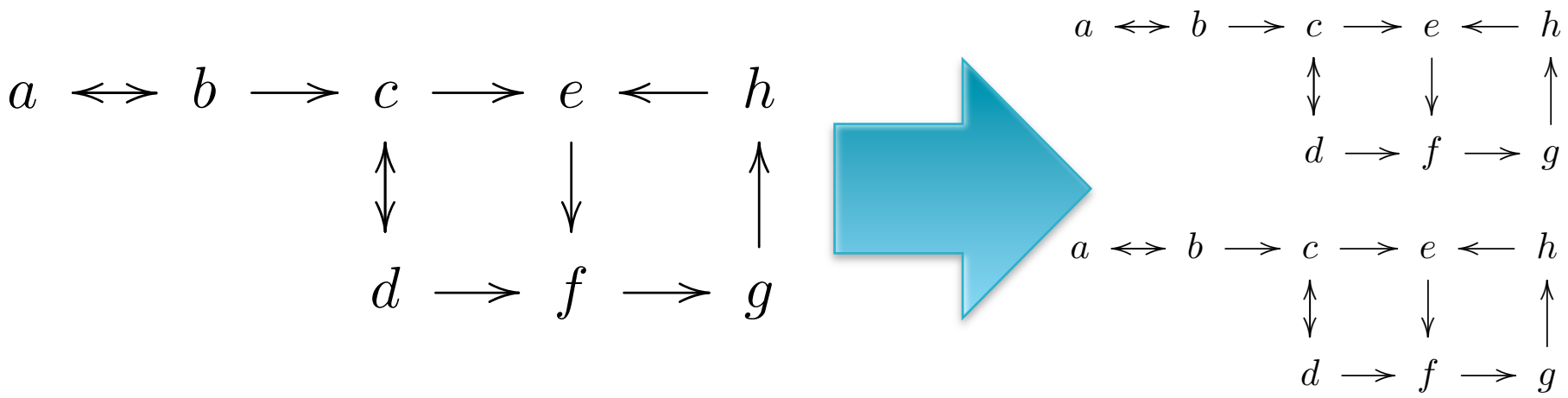
1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions
    1. Ranked / contrary-to-duty semantics
    2. Probabilistic / sequence semantics

Generalising Dung's abstract argumentation



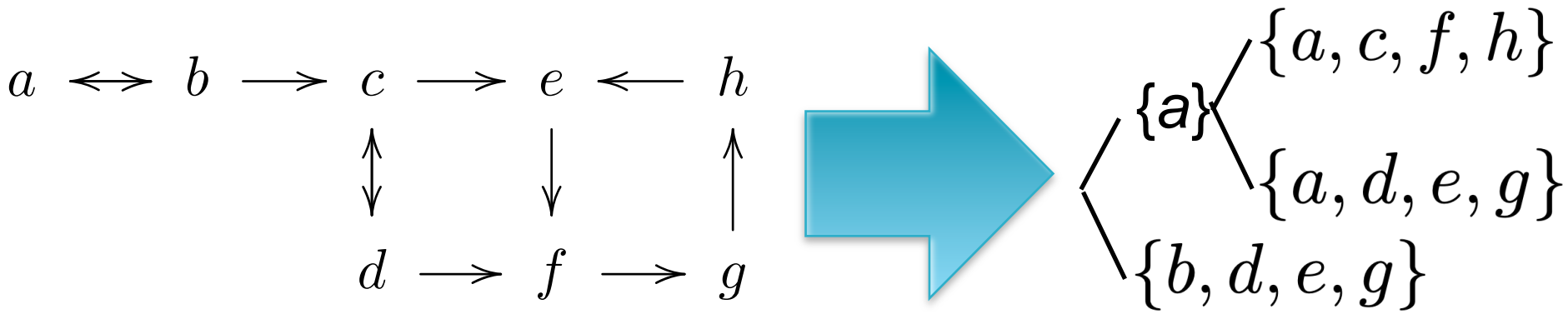
1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions
    1. Ranked / contrary-to-duty semantics
    2. Probabilistic / sequence semantics
    3. Attack semantics (IJCAI11)

Generalising Dung's abstract argumentation



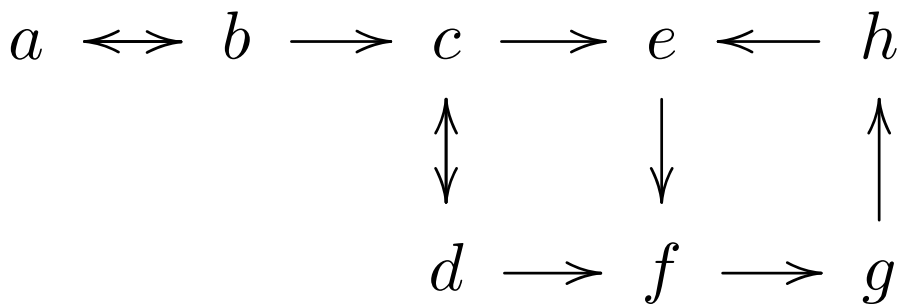
1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions
   1. Ranked / contrary-to-duty semantics
   2. Probabilistic / sequence semantics
   3. Attack semantics (IJCAI11)
   4. Sub-framework semantics (AICOL17)

Generalising Dung's abstract argumentation



1. Adding ingredients to the graph
2. Introducing new semantics
3. Adding ingredients to the extensions
    1. Ranked / contrary-to-duty semantics
    2. Probabilistic / sequence semantics
    3. Attack semantics (IJCAI11)
    4. Sub-framework semantics (AICOL17)
    5. Decision-graph semantics (COMMA18)

Generalising Dung's abstract argumentation

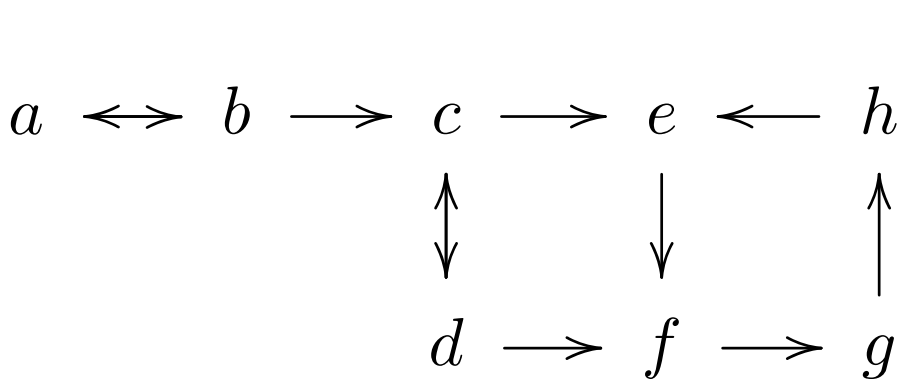


$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
   - How to instantiate this? Already VERY difficult for simple graphs

Beishui Liao, Nir Oren, Leender van der Torre, Serena Villata:
Prioritized norms in formal argumentation. J. Log. Comput. 29(2): 215-240 (2019)

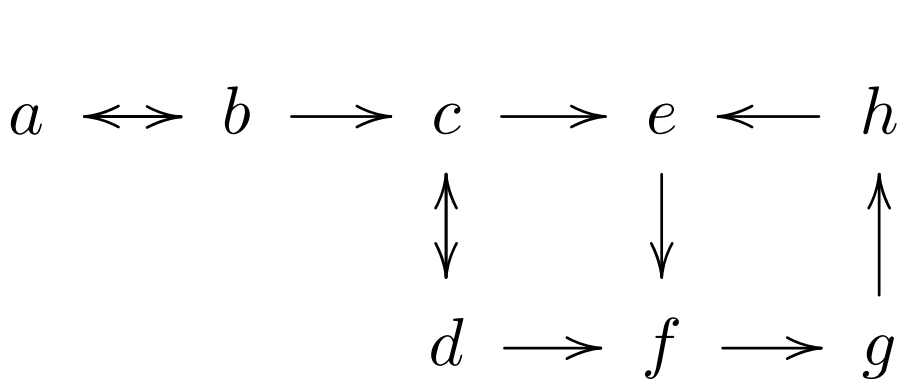
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

1. Adding ingredients to the graph
   - How to instantiate this? Already VERY difficult for simple graphs
2. Introducing new semantics
   - Does this satisfy the rationality postulates? Easily violated by semantics

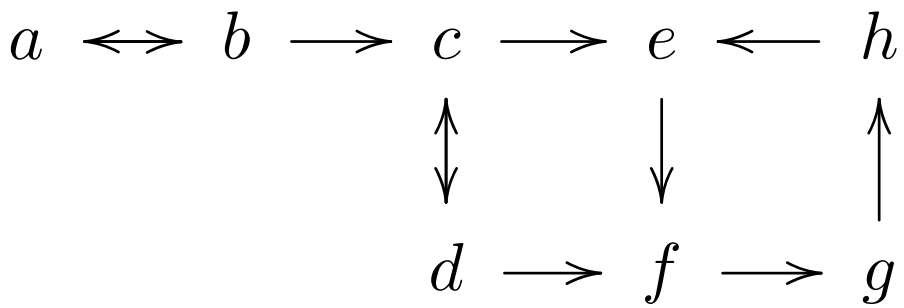
Generalising Dung's abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

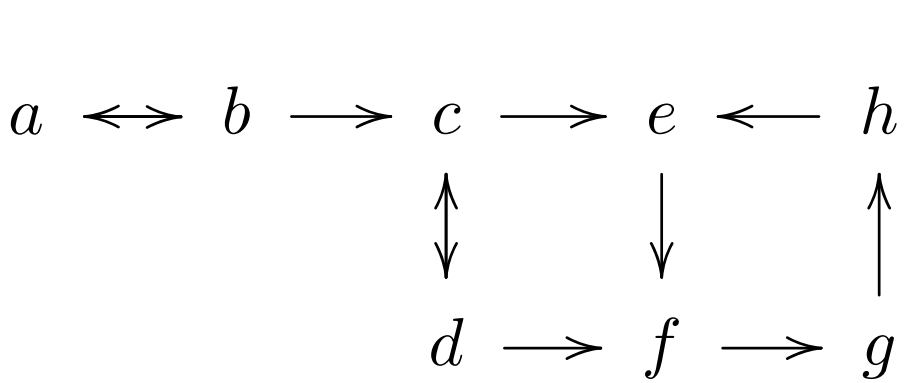
1. Adding ingredients to the graph
   - How to instantiate this? Already VERY difficult for simple graphs
2. Introducing new semantics
   - Does this satisfy the rationality postulates? Easily violated by semantics
3. Adding ingredients to the extensions
   - We can do it conservatively, extracting more information from the graph
   - We can obtain more satisfactory notions of equivalence

Explanation semantics for abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

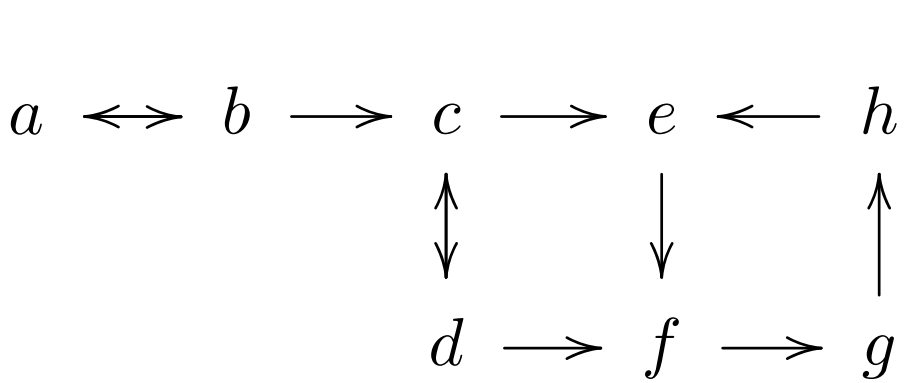
Explanation semantics for abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

- Explanation in psychology and cognitive science
  - Experimental concept: discriminative, minimal, ...

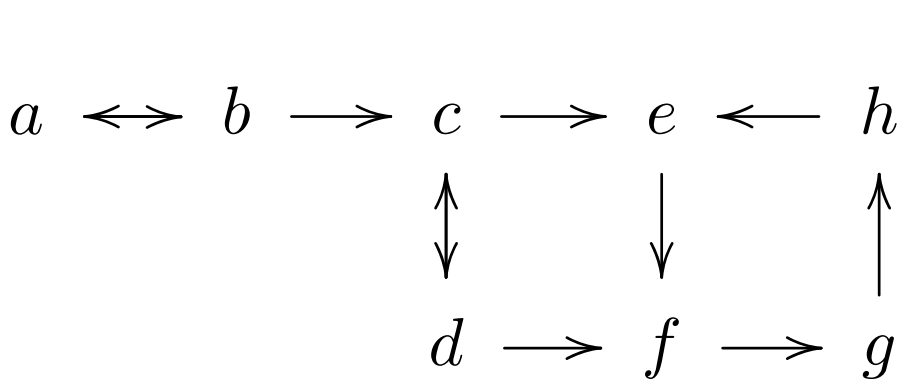
Explanation semantics for abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

- Explanation in psychology and cognitive science
- Explanation in machine learning
  - Tweaking the parameters

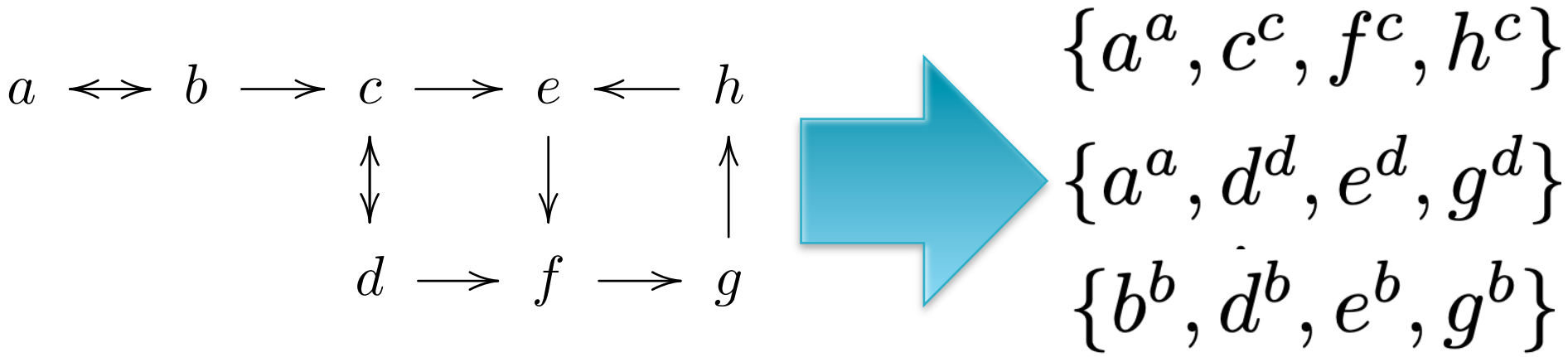
Explanation semantics for abstract argumentation



$$\{a, c, f, h\}$$
$$\{a, d, e, g\}$$
$$\{b, d, e, g\}$$

- Explanation in psychology and cognitive science
- Explanation in machine learning
- Explanation in knowledge representation & reasoning
  - Self explanatory (e.g. decision variables, parameters)
  - Everything else must be explained

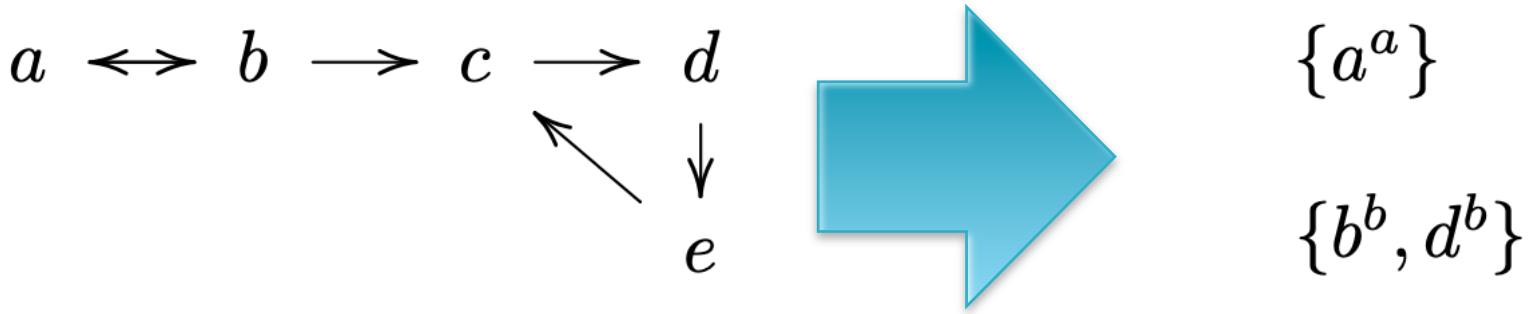
Explanation semantics for abstract argumentation



- Explanation in psychology and cognitive science
- Explanation in machine learning
- Explanation in knowledge representation & reasoning
  - Self-explanatory (e.g. decision variables, parameters)
  - Everything else must be explained

A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
4. Direct defense: the explanation defends the explained argument



A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
4. Direct defense: the explanation defends the explained argument
5. Minimality: expanation is subset minimal

A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
4. Direct defense: the explanation defends the explained argument
5. Minimality: expanation is subset minimal

$$a \longrightarrow b \longrightarrow c \longrightarrow d \longrightarrow e$$

$$\{a^{\{\}}, c^{\{a\}}, e^{\{c\}}\}$$

$$\{a^{\{\}}, c^{\{\}}, e^{\{\}}\}$$

A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
4. Direct defense: the explanation defends the explained argument
5. Minimality: expanation is subset minimal

- Dung semantics not representable if we accept 1-5

A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
5. Minimality: expanation is subset minimal
6. Transitivity: if R explains a, S explains b, and b in R, then S subset R

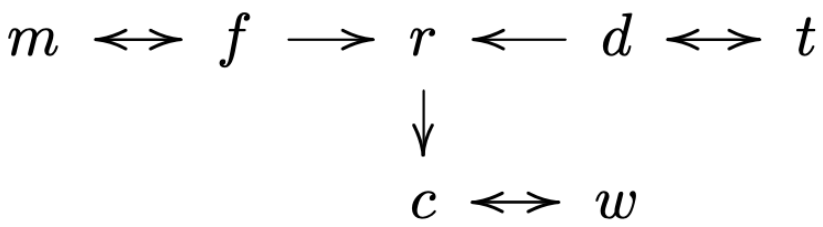
$$a \leftrightarrow b \leftarrow c \leftrightarrow d$$

$\{a^a, c^c\}$
$\{a^a, d^d\}$
$\{b^{\{b,d\}}, d^d\}$

- If argument is explained by one argument, then it is self explanatory

A principle-based analysis

1. Uniqueness: every accepted argument is explained by one set
2. Acceptance: the explanation arguments are themselves accepted
3. Indirect defense: iteratively applying the characteristic function on the explanation will give us the explained argument
5. Minimality: expanation is subset minimal
6. Transitivity: if R explains a, S explains b, and b in R, then S subset R
7. Explanation inheritance



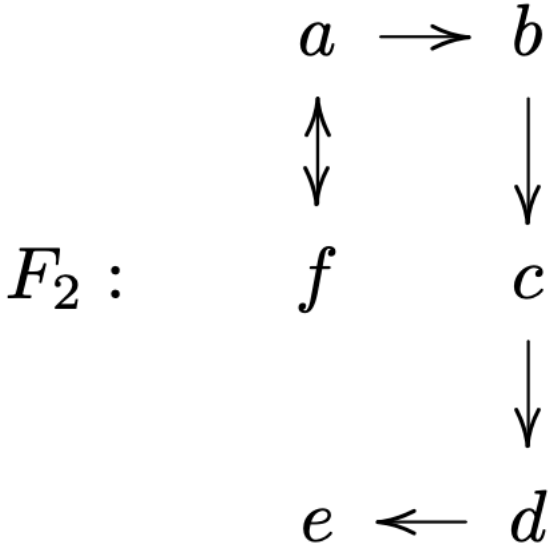
Many extensions
7 excludes e.g.:

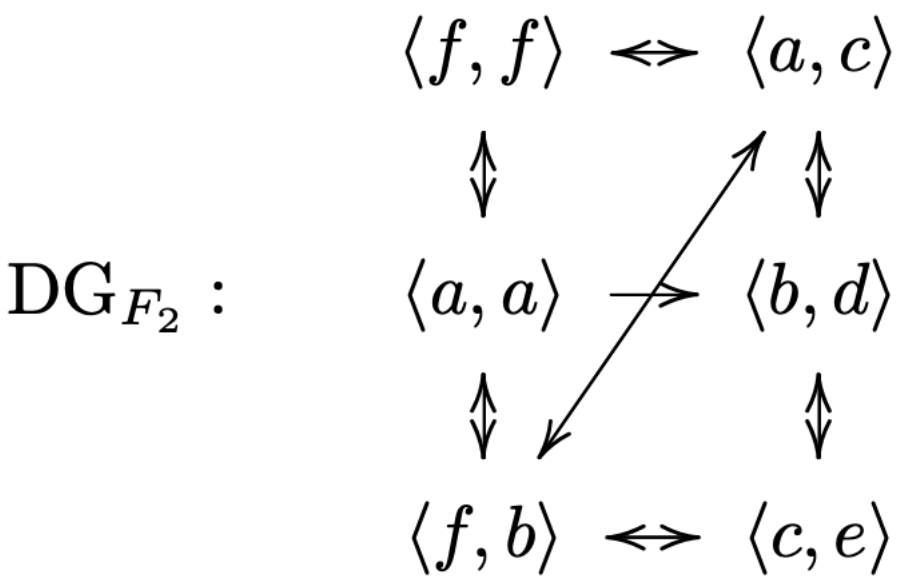
$\{m^m, t^t, r^m, w^t\}$

- From Rienstra et al, KR 2018

What else is in the paper?

- Dung semantics is representable by all except direct defense
- Concrete explanation-based semantics
  - Derived from Dung semantics and the principles
- Explanation based on weak defense graphs





Summary and further research

- Don’t extend the graphs, but extract more information from them
  - See also our COMMA18 paper on representation equivalences
- Direct defense versus minimal defense

Further work

- Explanation in psychology, cognition, informal argumentation, …
- Explaining rejection (and undecided)
- Ranked explanations, numerical explanations, …
- Structured explanations: evidence, ethical & legal principles, …
- Dialogical explanations

Jiàoshī jié kuài lè! 教师节快乐