Playing with Argumentation

Nicolas Maudet
nicolas.maudet@lip6.fr

Université Pierre et Marie Curie

EPCL-BTC: 19th of Novembre 2013

Outline of the talk

Motivation: Online Debate Platforms

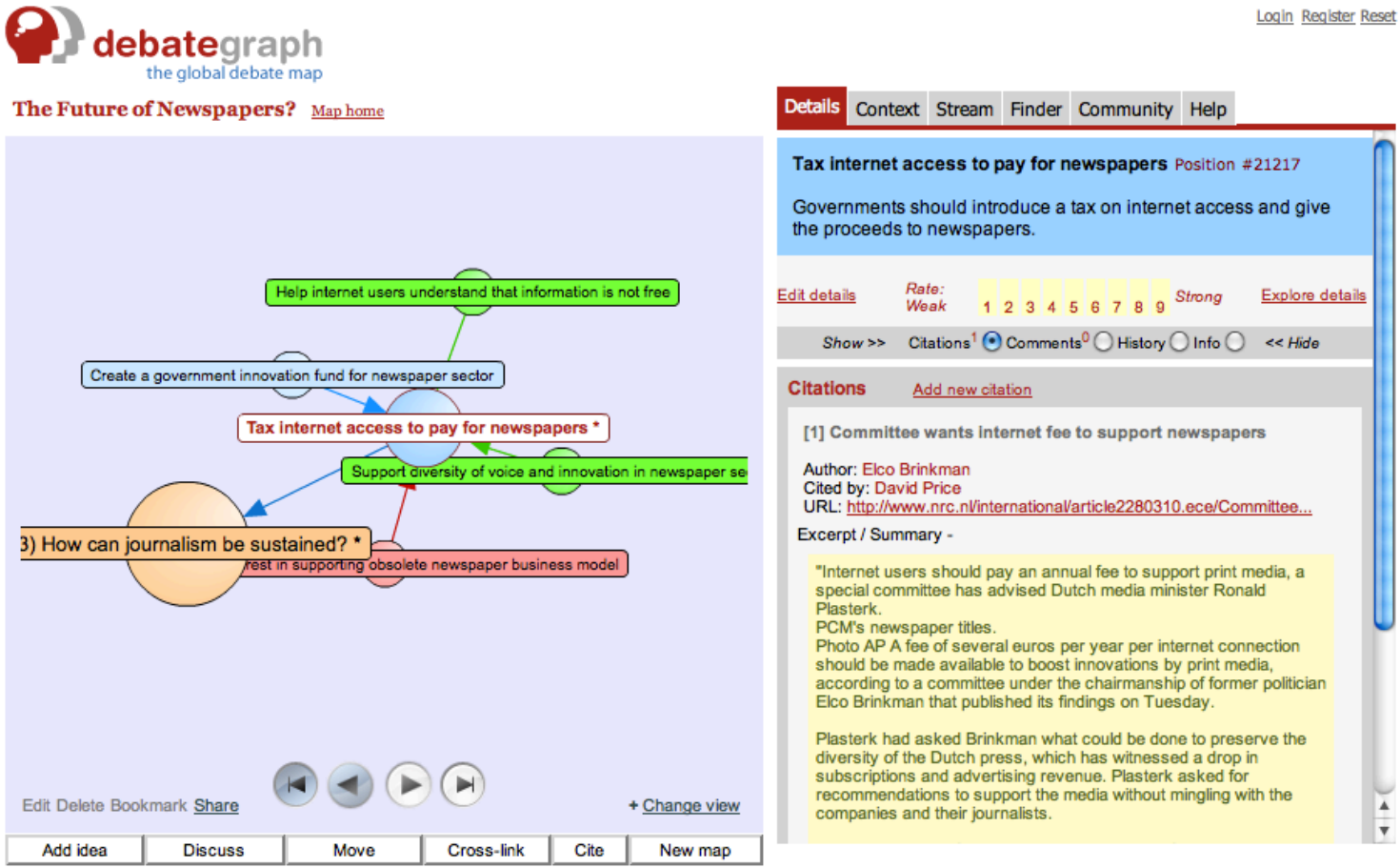


Figure: debategraph.org

Motivation: Online Debate Platforms

Some practical problems with these systems:

- the number of agents and arguments put forward give rise to unfocused debates, difficult to follow and interpret
- agents may have unequal access to the debate platform
- agents cannot be assumed to fully cooperative

One approach is to rely on crowdsourcing.
Here we take a more normative view and regulate the debate by means of a protocol. The resulting setting may be approached with game theoretical tools.

Argumentation and Game-Theory

There are existing works relating argumentations and games, in particular:

- proof-theoretical procedures involving an agent trying to defend an argument (PRO) and an agent trying to destroy this argument (CON).
  winning strategy = provability = x belongs to some/all extensions

There also are existing works relating argumentation and non cooperative game-theoretical notions. But here we need to define the components of a game

- what are the goals of agents?
- what are the permitted moves?
- what is the winning criterion?

A two-agent example

Example: *the argumentative battle of the sexes*
Suppose the arguments of a a given argumentation system are partitioned among two agents (blue and red):

- x_1: we should go to a football match
- x_2: we should go to a ballet
- x_3: I can't stay outside for a long time
- x_4: my worst enemy goes to the ballet as well...

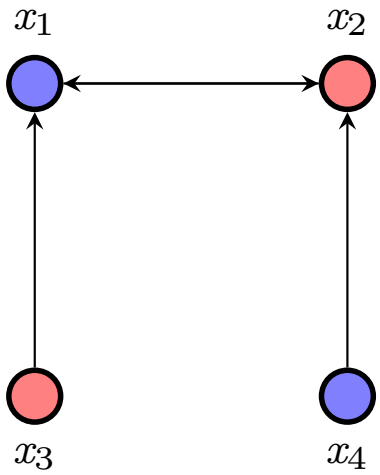
What are the goals of agents? Assume agents have preferences on the possible extensions, possibly based on some goals they have. Take :

- blue : football $\succ$ ballet $\succ \emptyset$ (say, 2,1,0 as utility)
- red : ballet $\succ$ football $\succ \emptyset$ (say, 2,1,0 as utility)

Rahwan and Larson. *Argumentation and Game-Theory*. Argumentation in AI.

A two-agent example

- strategies: subsets of (his own) arguments to put forward

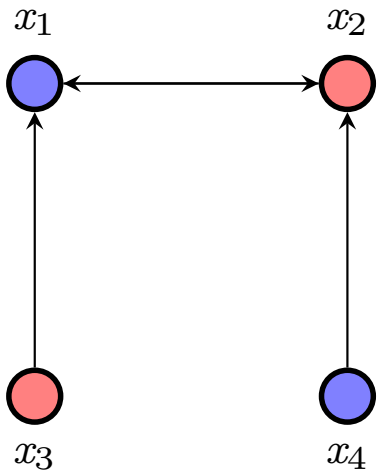


	$\{x_1\}$	$\{x_4\}$	$\{x_1, x_4\}$	$\{\}$
$\{x_2\}$	0,0	0,0	1,2	2,1
$\{x_3\}$	0,0	0,0	0,0	0,0
$\{x_2, x_3\}$	2,1	0,0	0,0	2,1
$\{\}$	1,2	0,0	1,2	0,0

- $\{x_3\}$ and $\{x_4\}$ are weakly dominated.
- Nash equilibrium = no agent has an incentive to unilaterally deviate.
  Can you spot NE?

A two-agent example

- strategies: subsets of (his own) arguments to put forward



	$\{x_1\}$	$\{x_4\}$	$\{x_1, x_4\}$	$\{\}$
$\{x_2\}$	0,0	0,0	1,2	2,1
$\{x_3\}$	0,0	0,0	0,0	0,0
$\{x_2, x_3\}$	2,1	0,0	0,0	2,1
$\{\}$	1,2	0,0	1,2	0,0

- $\{x_3\}$ and $\{x_4\}$ are weakly dominated.
- Nash equilibrium = no agent has an incentive to unilaterally deviate.
  Can you spot NE?
  $\langle\{x_2\}, \{x_1, x_4\}\rangle, \langle\{x_2, x_3\}, \{x_1\}\rangle, \langle\{\}, \{x_1, x_4\}\rangle, \langle\{x_2, x_3\}, \{\}\rangle, \langle\{\}, \{\}\rangle$

A two-agent setting

However, this type of analysis is not highly relevant for the types of debates that we envision:

- moves are not played simultaneously
- and even a sequential analysis would be poorly relevant:
  - it is difficult to assume a predefined sequence of turns (round-robin protocol) when there are many agents
  - it is difficult to assume full knowledge of the utility functions of the other agents

An analysis in terms of dynamics of better/best responses is more appropriate.

Bonzon and Maudet. *On the outcomes of multiparty argumentation*. AAMAS-11.

Outline of the talk

Argumentation systems

Abstract view of argumentation (no specification of the actual content of arguments), due to [Dung, 95]

- Argumentation system AS defined as a pair
  - A : set of arguments
  - R : attack relation ($\subseteq A \times A$)

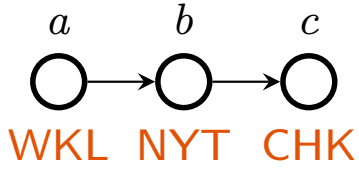
  The meaning of the attack relation is that an argument "defeats" another argument.
- Argumentation graph

$$AS = \{a, b, c\}, \{(b, c), (a, b), (a, c)\}$$

P. M. Dung. *On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games*. AIJ, 1995.

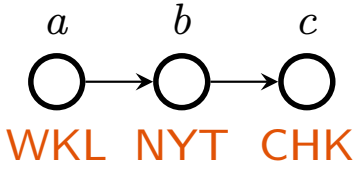
Argumentation systems + Preferences

1. Label each argument according to the value it puts forward, or the source it comes from, etc.

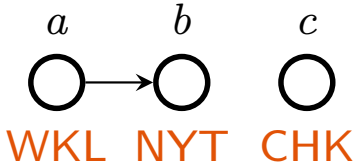


Argumentation systems + Preferences

1. Label each argument according to the value it puts forward, or the source it comes from, etc.



2. Discard attacks coming from less preferred values/sources ("the attack is not strong enough to defeat the argument")



$$\text{WKL} \succ \text{CHK} \succ \text{NYT}$$

T. M. Bench-Capon. *Value-based argumentation frameworks*. NMR, 2002.

A Simple Example

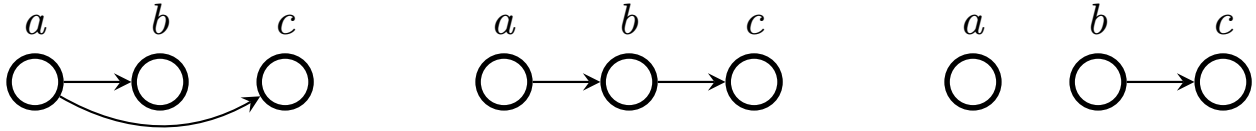
Consider the following arguments:

(c) The US army is preparing a secret plan to retreat from Afghanistan (source: Wikileaks)

(b) Our (informed) sources say the documents are fake. (source: NYT)

(a) The media cannot be trusted on military issues (source: N. Chomsky)

Why systems may differ



- a_1 thinks CHK is the more credible source, and sees WKL as a media (more credible than the NYT).
- a_2 thinks NYT is more credible than WKL, but that CHK is more credible than NYT. But he believes WKL cannot be seen as a media.
- a_3 thinks the NYT is the more credible source, and that CHK always says stupid things.

Acceptability

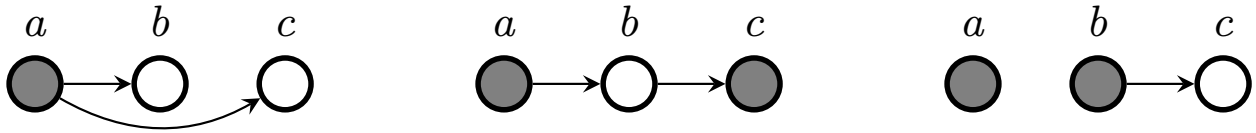
Now we need to define what (sets of) arguments should be considered as "justified" point of view.
Many different ways to define this! ($\Rightarrow$ argumentation semantics)

- S collectively defends $a \in A$ iff $\forall b \in A$ such that bRa, $\exists c \in S$ such that cRb
- S is a grounded extension iff S is the least fixed point of the characteristic function F of AS

$$F(S) = \{a \text{ such that } S \text{ collectively defends } a\}$$

- always exists a unique grounded extension, denoted by $\mathcal{E}(\text{AS})$
- can be computed in polynomial time

Acceptablity: example



- a_1 thinks CHK is the more credible source, and sees WKL as a media (more credible than the NYT).

$$a_1 : \mathcal{E}(AS_1) = \{a\}$$

- a_2 thinks NYT is more credible than WKL, but that CHK is more credible than NYT. But he believes WKL cannot be seen as a media.

$$a_2 : \mathcal{E}(AS_2) = \{a, c\}$$

- a_3 thinks the NYT is the more credible source, and that CHK always say rubbish.

$$a_3 : \mathcal{E}(AS_3) = \{a, b\}$$

Assumptions

Now suppose that these agents want to exchange their different points of view. How shall they proceed?

The following is taken for granted:

- many agents debating
- a single issue is under discussion (focused agents)
- no coordination takes place among agents
- agents agree on the set of (potential) arguments
- agents may disagree on the attack relations among these arguments

And we stick to the Dung's semantics (other approaches are possible).

Martins and Leite. *Social abstract argumentation*. IJCAI-13.

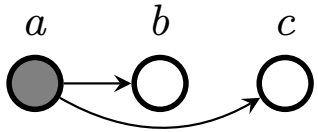
Merged Argumentation System

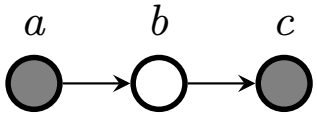
Introduced by [Coste-Marquis *et al.*, AIJ'07]

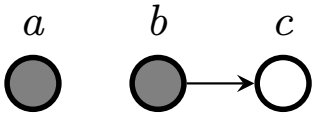
- n agents holding an argumentation system AS_i
- Majority Argumentation System
  - Attacks supported by a majority of agents
  - Ties broken in favour of the absence of an attack
- $MAS_N = \langle A, M \rangle$ where
  - $M \subseteq A \times A$
  - xMy when $\#\{i \in N | xR_i y\} > \#\{i \in N | xR_i y\}$
- Merged outcome: $\mathcal{E}(MAS_N)$

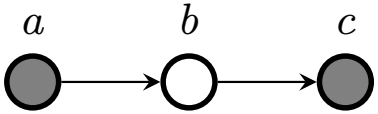
Coste-Marquis *et al.*. *Merging Argumentation Systems*. AIJ, 2007.

Merged argumentation system: example









$\mathcal{E}(MAS) = \{a, c\}$

But requires agents to send their full argumentation system, from which a very small part is likely to be relevant in the end!
$\Rightarrow$ incremental protocols closer to the way debates naturally evolve.

Outline of the talk

Basic ideas of the protocol

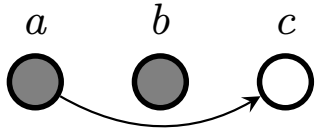
- agents prefer state of the debate where the issue has the same acceptability status as in their individual system.
- Convenient to distinguish two sets of agents ("sides"):
  - $CON = \{a_i \in N | d \notin \mathcal{E}(\mathsf{AS}_i)\}$
  - $PRO = \{a_i \in N | d \in \mathcal{E}(\mathsf{AS}_i)\}$
- the proposed protocol is based on the principle of direct relevance [Prakken, 2005]: a move is valid iff it changes the current status (accepted or not) of the issue (here: in the way which satisfies the agent)

H. Prakken. *Coherence and Flexibility in Dialogue Games for Argumentation*. JLC, 2005.

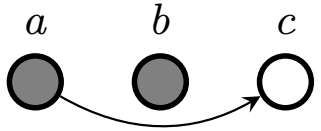
A relevance-based multiparty protocol

1. Agents report their individual view on the issue to the authority, which then assign (privately) each agent to PRO or CON.
2. The first round starts with the issue on the gameboard and the turn given to CON.
3. Until a group of agents cannot move :
   - 3.1 agents independently propose moves to the central authority;
   - 3.2 the central authority
     (i) picks the first (or at random) relevant move from the group of agents whose turn is active,
     (ii) update the gameboard, and
     (iii) passes the turn to the other group

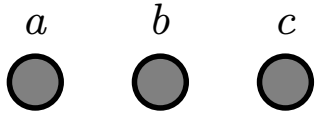
$t = 1$ - a_1 plays for CON: $RP_1^1 = \{(a, c)\}$



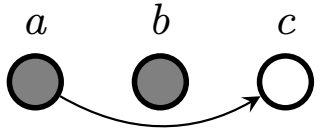
$t = 1$ - a_1 plays for CON: $RP_1^1 = \{(a, c)\}$



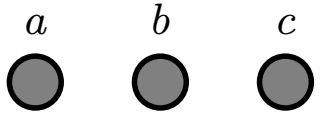
$t = 2$ - a_2: $RP_2^2 = \{(a, c)\}$



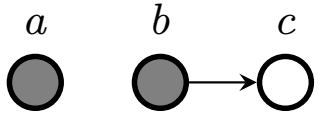
$t = 1$ - a_1 plays for CON: $RP_1^1 = \{(a, c)\}$



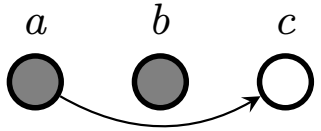
$t = 2$ - a_2: $RP_2^2 = \{(a, c)\}$



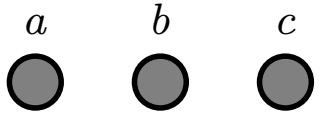
$t = 3$ - a_3 plays for CON: $RP_3^3 = \{(b, c)\}$



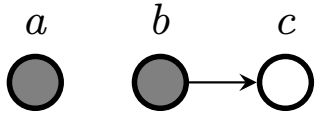
$t = 1$ - a_1 plays for CON: $RP_1^1 = \{(a, c)\}$



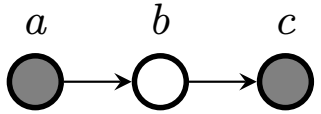
$t = 2$ - a_2: $RP_2^2 = \{(a, c)\}$



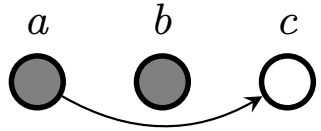
$t = 3$ - a_3 plays for CON: $RP_3^3 = \{(b, c)\}$



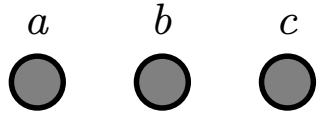
$t = 4$ - a_2: $RP_2^4 = \{(a, b), (a, c)\}$



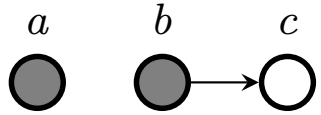
$t = 1$ - a_1 plays for CON: $RP_1^1 = \{(a, c)\}$



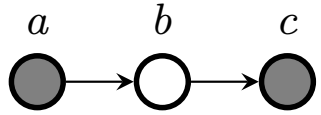
$t = 2$ - a_2: $RP_2^2 = \{(a, c)\}$



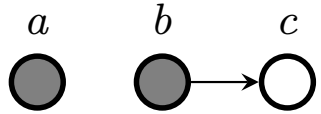
$t = 3$ - a_3 plays for CON: $RP_3^3 = \{(b, c)\}$



$t = 4$ - a_2: $RP_2^4 = \{(a, b), (a, c)\}$



$t = 5$ - a_3: $RP_3^5 = \{(b, c), (a, b)\}$



$t = 6$ - a_2 cannot add c in the extension

The status of the issue does not coincide with the merged outcome.
Agent 1 has no incentive to say that aRb ($\Rightarrow$ manipulation).

Outline of the talk

What are the outcomes with such a protocol?

Let $X \in \{CON, PRO\}$.
If $X = PRO$ (resp. CON), $\overline{X} = CON$ (resp. PRO).

- The issue of the debate is a possible outcome for a group X if this group has a possibility to set the acceptability status of this argument coincide in the debate and in their individual system.
- The issue is a necessary outcome for X if this issue is not a possible outcome for $\overline{X}$.

We now ask the following questions:

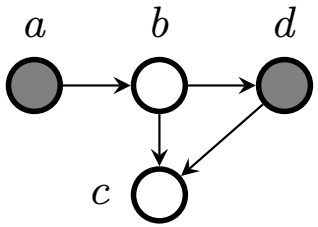
1. is the outcome pre-determined from the initial situation?
2. does the outcome coincide with the merged outcome?
3. is it useful to allow reinforcement moves?

Control of an edge of the global arguments-control graph

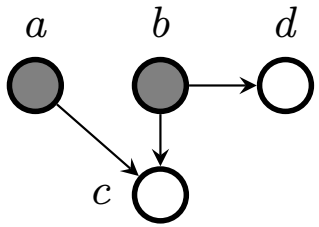
We collect the relevant information under the form of an arguments-control graph.

- constructive control: $X^{+}_{(a,b)}$ iff $|add_{(a,b)} \cap X| > |rem_{(a,b)} \cap \overline{X}|$
  the number of agents in X who can add (a, b) is greater than the number of agents in $\overline{X}$ who can remove it.
- destructive control: $X^{-}_{(a,b)}$ iff $|rem_{(a,b)} \cap X| \geq |add_{(a,b)} \cap \overline{X}|$
  the number of agents in X who can remove (a, b) is greater or equal than the number of agents in $\overline{X}$ who can add it.
- playability: $X_{\bullet}$ iff $|add_{(a,b)} \cap X| > 0$
  the move can be played.

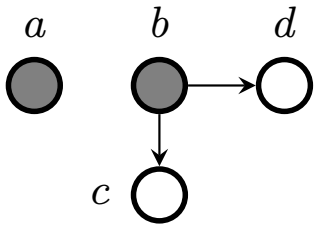
Example (ctd.)



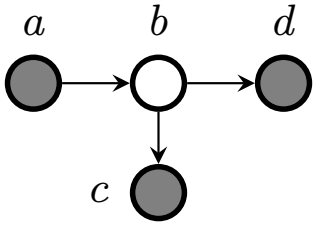
agent 1



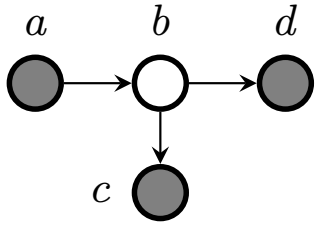
agent 2



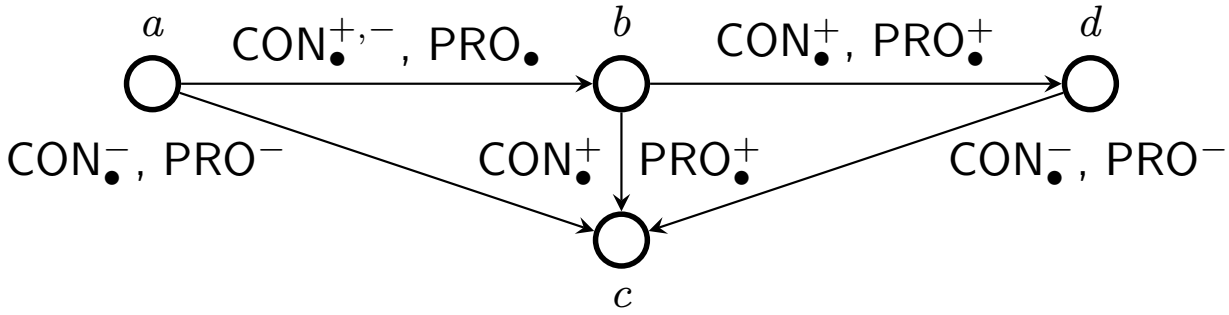
agent 3



agent 4



agent 5



Who wins the debate?

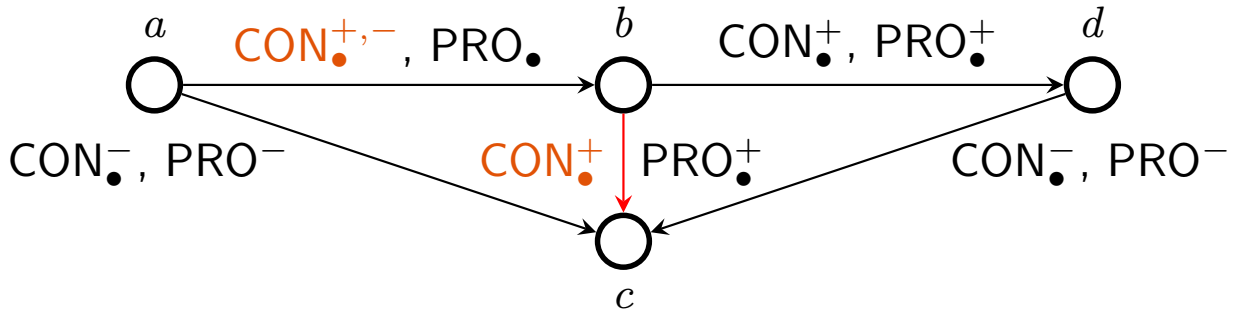
An edge (a, b) is an attack (resp. defense) edge for d if there is an even (resp. odd) length path from b to d. Note that an edge may be both attack and defense.

Definition

A **path for** d **controlled by** CON is an odd-length path from x to d such that
(i) CON has constructive control on all the attack edges for d, and
(ii) CON has destructive control on all the defense edges for d attacking x.

Possible outcome for CON

The issue d is a possible outcome for CON if there exists a tree for d controlled by CON.



A tree controlled by CON $\Rightarrow d$ is a possible outcome for CON.
Is it a necessary outcome?

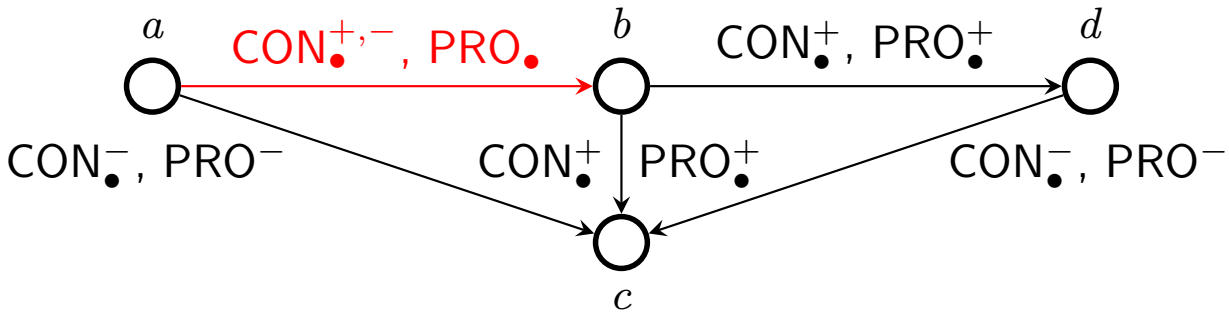
Who wins the debate?

However the issue may be possible for PRO even when such paths exist. Intuitively, some moves that are both attack/defense moves can block the paths controlled by CON.

Definition

An edge (x, y) on a path P is a switch for d if (i) it is a defense for d on P, (ii) it is playable by CON, (iii) there exists a even-length path from y to d such that all the attack edges are playable by CON and all the defense edges are playable by PRO, and (iv) PRO has the destructive control on at least one attack edge from this path.

Example (ctd.)



The edge aRb is a switch for the issue d.

$\Rightarrow$ debates may be "open".

Coincidence with the merged outcome?

In general, no guarantee that the outcome will be similar to the one obtained via merging. Can we find sufficient conditions for this to hold?

Reachability of the merged outcome

When the ACG does not contain edge (a, b) such that $X^{+,-}(a, b)$, then reachability of the merged outcome is guaranteed.

Extended protocol with reinforcement moves?

Idea: allow agents to reinforce (resp. weaken) moves that would change the status of the issue if it was to be deleted.

Ineffiency of reinforcement moves

It is never beneficial for an agent to play reinforcement moves (if weakening are symmetrically allowed). Furthermore, it may be the case that the issue is no longer a possible outcome if X uses reinforcement moves.

The key observation is to note that by playing reinforcement moves a group of agent may lose some destructive control.

Outline of the talk

Some challenges

- how can we specify rules for argumentation games and let agent play with different strategies?
- computational tasks faced by agents in these debates?
- can we quantify more precisely the distance to the merged outcome?