Reason-based Belief Revision in Social Networks

Fenrong Liu

Tsinghua University and University of Amsterdam

Applied Logic Seminar, TU Delft, 6 Oct 2014

Outline

1. Review on belief revision in social networks
2. Introducing reasons: new models
3. Discussion and conclusion

Doxastic influence

Consider influence regarding a single proposition *p*. If I do not believe *p* and some significant number or proportion of my friends do believe it. How should I respond?

(1) ignore their opinions and remain doxastically unperturbed.
(2) If I am influenced to change my beliefs there are at least two ways: I may *revise* so that I too believe *p* (*Rp*) or (more cautiously) *contract*, removing $\neg p$ (*Cp*).

Notations: $[Rp]Bp$; $[Cp]\neg B\neg p$. (Assuming success conditions)

Doxastic influence

We draw a distinction between two kinds of influence:

(a) influence that leads to revision (strong influence): Sp.
(b) influence that leads to contraction (weak influence): Wp.

We define a general operation of social influence regarding p (Ip) as the program

$$\texttt{if } \mathsf{S}p \texttt{ then } Rp \texttt{ else if } \mathsf{W}p \texttt{ then } C\neg p;$$
$$\texttt{if } \mathsf{S}\neg p \texttt{ then } R\neg p \texttt{ else if } \mathsf{W}\neg p \texttt{ then } Cp$$

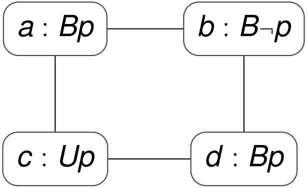
Strong and weak influence

- I am **strongly influenced** to believe p iff **ALL** (at least one) of my friends believe p **.**
- I am **weakly influenced** to contract by belief in $\neg p$ iff **NONE** of my friends believe $\neg p$.

Strong and weak influence captured with the following axioms:

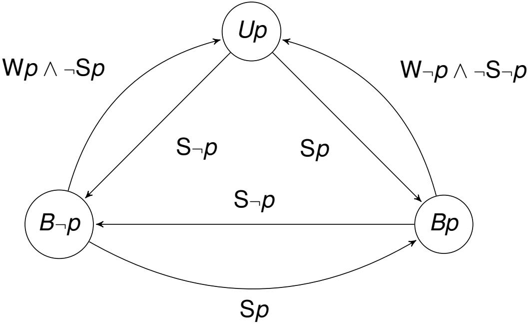
$$S\varphi \leftrightarrow (FB\varphi \wedge \langle F \rangle B\varphi)$$
$$W\varphi \leftrightarrow (F\neg B\neg\varphi \wedge \langle F \rangle B\varphi)$$

Distribution of doxastic states: Example



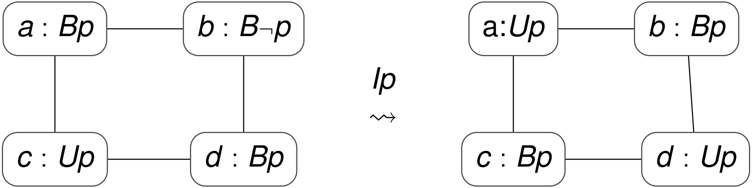
A network of four agents. Links between nodes indicate friendship (irreflexive and symmetric relation). Agent a believes p (written $a : Bp$) and has friends b and c ; agent b disbelieves p and has friends a and d; and so on.

Finite state of automaton



Example

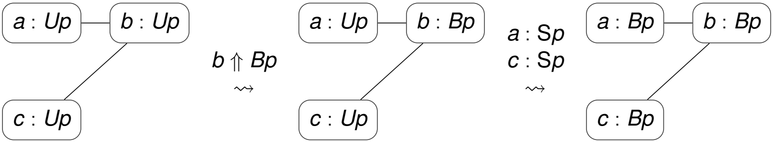
Example 1:



Observation: In the configuration on the right, *a* and *d* are both strongly influenced to believe *p*, and so a further application of *Ip* would cause them to revise their beliefs, returning them to their previous doxastic states.

Location is critical

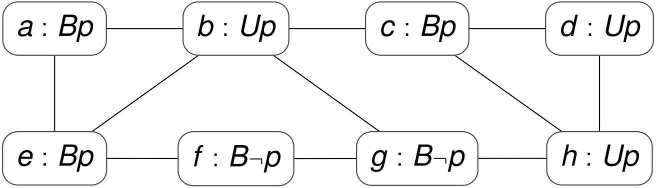
Example 2:



Stability and flux

A network is *stable* if the operator Ip has no effect on the doxastic states of any agent in the network. Unanimity within the community is sufficient for stability but not necessary:

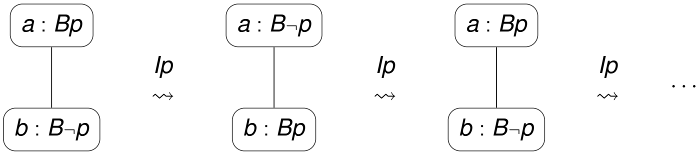
Example 3:



Stability and flux: one more example

Configurations that never become stable will be said to be *in flux*:

Example 4:



Characterizing stability

$$\neg(B\neg p \wedge \mathrm{W}p) \;\wedge\; \neg(Up \wedge \mathrm{S}p) \;\wedge\; \neg(Up \wedge \mathrm{S}\neg p) \;\wedge\; \neg(Bp \wedge \mathrm{W}\neg p)$$

Under the assumption of threshold influence, it is equivalent to

$$\begin{array}{l}
\neg(B\neg p \wedge F\neg B\neg p \wedge \langle F\rangle Bp) \wedge \\
\neg(\neg Bp \wedge \neg B\neg p \wedge FBp \wedge \langle F\rangle Bp) \wedge \\
\neg((\neg Bp \wedge \neg B\neg p \wedge FB\neg p \wedge \langle F\rangle B\neg p) \wedge \\
\neg(Bp \wedge F\neg Bp \wedge \langle F\rangle B\neg p)
\end{array}$$

A network is *stable* when every agent in the network satisfies this condition.

Main Reference

Fenrong Liu, Jeremy Seligman, and Patrick Girard, "Logical Dynamics of Belief Change in the Community", *Synthese*, Volume 191, Issue 11, pp 2403-2431, 2014.

Motivation of the new work

We change our beliefs for some reasons. Two aspects to be considered in this talk:

- Instead of changing beliefs due to pure pressure, we want to connect agent's **evidence** to her beliefs.
- Friends are not always equal, we may **trust** some friends more than others.

Ongoing joint work with Alexandru Baltag and Sonja Smets.

Definition (Weighting justification model)

A **weighting justification model** is a structure (S, E, w, V) where

- a finite set S of possible worlds.
- a family $E \subseteq \mathcal{P}(S)$ of non-empty subset $e \subseteq S$ ($\emptyset \notin E$), called evidence such that S is itself an evidence set ($S \in E$). A body of evidence is any consistent family of evidence sets, i.e. any $G \subseteq E$ such that $\cap G \neq \emptyset$. We denote $\mathcal{E} \subseteq \mathcal{P}(E)$ the family of all bodies of evidence.
- $w : E \to \mathbb{R}$.
- V a stardard valuation function.

This is taken from Fiutek, other forms of the definition appeared in various works by Baltag, Smets, Christoff, and Hansen.

Adding epistemic and social components

Definition (Epistemic weighting social network model)

An **epistemic weighting social network model** is a structure $(S, E, A, F, \sim_a, w, \tau, V)$ where

- S is a finite set of possible states; $\sim_a$ is a binary epistemic indistingishable relation; V is a valuation function.
- E, a set of evidence.
- A is a finite set of agents.
- $F : S \to (A \to \mathcal{P}(A))$.
- $w : S \times A \to (E \to \mathbb{R})$.
- $\tau : S \times A \to (A \to \mathbb{R}^+)$.

Some notations

- $F^s(a)$ for a's **friends** at s.
- $w^s_a(e)$ for a's **strength** of her evidence e at s.
- τ^s_{ab} for a's **trust** towards b at s.
- $s(a) = \{t : s \sim_a t\}$.

Conditions

(1) $\sim_a$ is an equivalence relation

(2) $a \in F^s(a)$

(3) $s \sim_a s' \Longrightarrow \mathrm{F}^s(a) = F^{s\prime}(a)$

(3) $e \cap s(a) = \emptyset \Longrightarrow w_a^s(e) = 0$

(4) $s \sim_a s' \Longrightarrow w_a^s = w_a^{s'}$

(5) $s \sim_a s' \Longrightarrow \tau_{ab}^s = \tau_{ab}^{s'}$

Induced plausibility relation

Definition (Induced plausibility relation)

We define the notion of **the largest body of evidence consistent with a given state** $s \in S$ and write it as $E_s := \{e \in E \mid s \in e\}$

We can induce a **plausbility relation** on states directly from the weight comparison on $\mathcal{E}$: for two states s, $t \in S$, we put

$$s \leq_a t \quad \text{iff} \quad s \sim_a t \quad \text{and } \widetilde{w}_a(s) \leq \widetilde{w}_a(t)$$

where $\widetilde{w}_a(s) = \sum\{w_a^s(e) : e \in E_s\}$.

Defining doxastic notions

Given an epistemic weighting social network model, we can define the useful notions by using the plausibility order $\leq_a$:

- $K_a p = \{s \in S \mid s(a) \subseteq P\}$.
- $Best_a P = \{s \in P \mid t \leq_a s \text{ for all } t \in P\}$.
- $B_a^Q P = \{s \in S \mid best_a(Q \cap s(a)) \subseteq P\}$.
- $\Box_a P = \{s \in S \mid \forall t(s \leq_a t \Longrightarrow t \in P)\}$.

Dynamic update

Definition (Updated model)

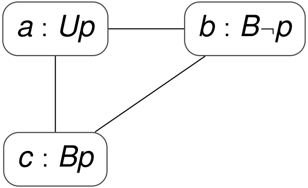
Given an epistemic weighting social network model $\mathcal{M}$= $(S, E, A, F, \sim_a, w, \tau, V)$, after one round of social communication, the updated model $\mathcal{M}'$ is defined as follows:

- S, E, A, F, V and τ remain the same.
- $w_a'^{s}(e) = \begin{cases} 0, & \text{if } e \cap s'(a) = \emptyset \\ \sum_{b \in F^s(a)} \tau^s_{ab} \cdot w^s_b(e), & \text{otherwise} \end{cases}$
- $s \sim_a' t \Leftrightarrow s \sim_a t$ and $w_b'^{s}(e) = w_b'^{t}(e)$ for all e, $b \in F^s(a)$

Note that friendship relation (F) and the trust weight τ can also change in a more complex setting.

Example

Example 5:

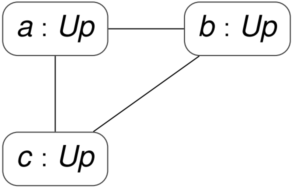


$$w_a(p) = 0, w_a(\neg p) = 0$$
$$w_b(p) = 0, w_b(\neg p) = 1$$
$$w_c(p) = 1, w_c(\neg p) = 0$$

Example continued: communication

Assuming $\tau_{aa} = \tau_{bb} = \tau_{cc} = \tau_{ab} = \tau_{ba} = \tau_{bc} = \tau_{cb} = \tau_{ac} = \tau_{ca} = 1$, after one-step communication, the updated model is (**stable**):

Example 6:

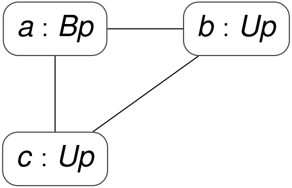


$$w'_a(p) = 1, w'_a(\neg p) = 1$$
$$w'_b(p) = 1, w'_b(\neg p) = 1$$
$$w'_c(p) = 1, w'_c(\neg p) = 1$$

With different trust weight

Assuming $\tau_{aa} = \tau_{bb} = \tau_{cc} = \tau_{ab} = \tau_{ba} = \tau_{bc} = \tau_{cb} = \tau_{ca} = 1$, and $\tau_{ac} = 2$, after communication, Example 5 changes into:

Example 7:

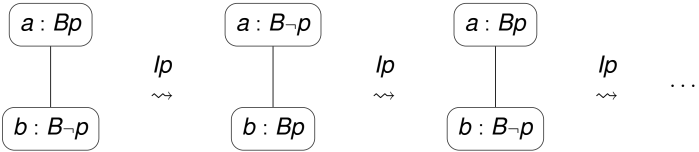


$$w'_a(p) = 2, w'_a(\neg p) = 1$$
$$w'_b(p) = 1, w'_b(\neg p) = 1$$
$$w'_c(p) = 1, w'_c(\neg p) = 1$$

In flux

Assuming $\tau_{ab} = \tau_{ba} = 2$, and $\tau_{aa} = \tau_{bb} = 1$, consider

Example 8:



?? $s \sim_a t \Longrightarrow \tau^s_{ba} = \tau^t_{ba}$ for all $b \in F^s(a)$

Does an agent know how much her friends trust her?

?? $s \sim_a t \Longrightarrow \tau^s_{bc} = \tau^t_{bc}$ for all $b, c \in F^s(a)$

Does an agent *a* always know how much her friend *b* trusts her friend *c*?

Example: different strategies

Example (Different strategies)

Assume $F^s(a) = \{b, c\}, \tau^s_{ab} = 10, \tau^s_{ac} = 1$. If a knows that and $\tau^s_{bc} = 10$, what would she do?

- keep updating as we proposed.
- optimistic: increase τ^s_{ac}
- pessimistic: lower τ^s_{ab}

Example

Example (Wikipedia: consider the further sources of evidence)

- Typically, it provides information (knowledge or belief, also supporting evidence)
- It has references, indicating the further sources of the information
- We can look at the names of references and change our trust weight towards the Wiki; use the old trust; or look it up ourselves, using our own trust weight to the author directly.

Some remarks

- We may need a set of update strategies, instead of one general rule.
- We use results from dynamic systems, to characterize the stability of the social network in the long term in this setting.
- Though our models are based on the weighting changes (non-AGM dynamics), but they can simulate AGM dynamics.

Future directions

- Social networks with different structure.
- Dynamics of trust change.
- Adding new friends and deleting olds friends.
- More qualitative approach.

Thank you for your attention!