A probabilistic author-centered model for Twitter discussions*

Teresa Alsinet[1][0000−0003−4594−5502], Josep Argelich[1][0000−0003−4089−6422], Ramón Béjar[1][0000−0002−5208−685X], Francesc Esteva[2][0000−0003−4466−3298], and Lluis Godo[2][0000−0002−6929−3126]

[1] INSPIRES Research Center – University of Lleida
Jaume II, 69 – 25001 Lleida, SPAIN
{tracy,jargelich,ramon}@diei.udl.cat
[2] AI Research Institute, IIIA-CSIC
Bellaterra, SPAIN
{godo,esteva}@iiia.csic.es

Abstract. In a recent work some of the authors have developed an argumentative approach for discovering relevant opinions in Twitter discussions with probabilistic valued relationships. Given a Twitter discussion, the system builds an argument graph where each node denotes a tweet and each edge denotes a criticism relationship between a pair of tweets of the discussion. Relationships between tweets are associated with a probability value, indicating the uncertainty on whether they actually hold. In this work we introduce and investigate a natural extension of the representation model, referred as probabilistic author-centered model. In this model, tweets by a same author are grouped, describing his/her opinion in the discussion, and are represented with a single node in the graph, while edges stand for criticism relationships or controversies between opinions of Twitter users in the discussion. In this new model, interactions between authors can give rise to circular criticism relationships, and the probability of one opinion criticizing another are evaluated from the probabilities of criticism among the individual tweets that compose both opinions.

1 Introduction

In a recent work [2], an argumentative approach has been proposed for discovering relevant opinions in Twitter with probabilistic valued relationships.

Argumentation-based reasoning models aim at reflecting how humans make use of conflicting information to construct and analyze arguments. An argument is an entity that represents some grounds to believe in a certain statement and that can be in conflict with arguments establishing contradictory claims. The most commonly used general argumentation framework is Dung's abstract argumentation model [8].

In abstract argumentation, a graph is used to represent a set of arguments and counterarguments. Each node is an argument and each edge denotes an attack between arguments. Different kinds of semantics for abstract argumentation frameworks

* This work was partially funded by the Spanish MINECO/FEDER Projects TIN2015-71799-C2-1-P and TIN2015-71799-C2-2-P, by the European H2020 Grant Agreement 723596, and by the 2017 SGR 1537 and 172.

have been proposed that highlight different aspects of argumentation (for reviews see e.g. [4,5,16]). Usually, semantics for abstract argumentation frameworks are given in terms of sets of extensions, which are suitable consistent sets of arguments. For a specific extension, an argument is either accepted or rejected and, usually, there is a set of extensions that is consistent with the semantic context.

The analysis of Twitter by means of argumentation frameworks has also been explored by Grosse et al. [13] with the aim of detecting conflicting elements in an opinion tree to avoid potentially inconsistent information. Moreover, in order to mine arguments from Twitter, Bosc et al. [6] proposed a binary classification mechanism (argument-tweet vs. non argument) and Dusmanu et al. [10] applied supervised classification to identify arguments on Twitter and evaluated facts recognition and source identification for argument mining.

Given a Twitter discussion, i.e. a set of tweets generated from a root tweet, the system developed in [2] builds a weighted argument graph where each node denotes a tweet, each edge denotes a criticism relationship between a pair of tweets of the discussion and the weight of nodes models the social relevance of tweets from data obtained from Twitter. In Twitter, a tweet always answers or refers to previous tweets in the discussion, so the obtained underlying argument graph is acyclic. Moreover, when constructing relationships between tweets from informal descriptions expressed in natural language with other attributes such as emoticons, jargon, onomatopoeia and abbreviations, it is often evident that there is uncertainty about whether some of the criticism relationships actually hold. So, to take into account this fact in the model, each edge of an argument graph is associated with a probability value, quantifying such uncertainty on criticism relationships between pairs of tweets. The solution of a weighted argument graph for a Twitter discussion is computed by means of the reasoning system we developed in [1], where the graph is mapped to a valued abstract argumentation framework (VAF) [3] and the so-called ideal semantics [9] is used to evaluate the set of socially accepted tweets in a discussion from the weights assigned to the tweets and the criticism relationships between them.

In this work we introduce a natural extension of our previous representation model for Twitter discussions [2], that will be called *probabilistic author-centered model*. In this new model, tweets within a discussion are grouped by authors, such that tweets of a same author describe his/her opinion in the discussion that is represented by a single node in the graph, and criticism relationships denote controversies between the opinions of Twitter users in the discussion. In this model, the interactions between authors can give rise to circular criticism relationships, and the probability of one opinion criticizing another is evaluated from the individual probabilities of criticism among the tweets that compose both opinions. So, the underlying argument graph can contain cycles and a model for the aggregation of probabilities has to be proposed. Moreover, to compute the set of accepted authors' opinions in a discussion, we also extend our previous reasoning system [1] which is based on the acceptance of tweets of a discussion and not on its authors. This new representation and reasoning model can be of special relevance for assessing Twitter discussions in fields where identifying groups of authors whose opinions are globally compatible or consistent is of particular interest.

The rest of the paper is organized as follows. In Section 2, we recall from [2] the formal graph structure to model Twitter discussions. Then, in Section 3, we describe the author-centered model for representing discussions in Twitter and, in Section 4, we formalize the probabilistic weighting scheme of criticism relationships between authors' opinions. Finally, in Section 5 we define the reasoning system to compute the sets of accepted and rejected opinions and, in Section 6, we conclude.

2 Twitter discussion graph

In this section, we introduce a simplified computational structure of the one proposed in [2] to represent a Twitter discussion with probabilistic valued relationships, that will be called *probabilistic discussion graph*. In such a graph, each node will denote a tweet, each edge will denote an answer relationship between a pair of tweets of the discussion, and each edge will be attached a probability value, indicating the probability that a criticism relationships between the pair of tweets actually holds. We provide more formal definitions next.

Definition 1. *(Twitter Discussion) A Twitter discussion Γ is a non-empty set of tweets. A tweet $t \in \Gamma$ is a triple $t = (m, a, f)$, where m is the up to 140 characters long message of the tweet, a is the author's identifier of the tweet and $f \in \mathbb{N}$ is the number of followers of the author, according to its temporal instant generation during the discussion. Moreover, if t_1 and t_2 are tweets from different authors, We say that t_1 answers t_2 iff t_1 is a reply to the tweet t_2 or t_1 mentions (refers to) tweet t_2.*

Definition 2. *(Discussion Graph) The Discussion Graph* (DisG) *for a Twitter discussion Γ is the directed graph (T, E) such that for every tweet in Γ there is a node in T and if tweet t_1 answers tweet t_2 there is a directed edge (t_1, t_2) in E. Only the nodes and edges obtained by applying this process belong to T and E, respectively.*

Definition 3. *(Probabilistic Discussion Graph) A probabilistic discussion graph* (PDisG) *for a Twitter discussion Γ is a triple $\langle T, E, P\rangle$, where*

- *(T, E) is the* DisG *graph for Γ and*
- *P is a labeling function $P : E \to [0, 1]$ that attaches a probability value $p \in [0, 1]$ to every edge $(t_1, t_2) \in E$, meant as the degree of belief that tweet t_1 is a criticism to tweet t_2, i.e. that the message of t_1 does not agree with the claim expressed in the message of t_2. So, $p = 1$ means that it is fully believed that tweet t_1 disagrees with the claim expressed in tweet t_2, while $p = 0$ means that it is fully believed that tweet t_1 agrees with the claim expressed in tweet t_2.*

Given a PDisG $\langle T, E, P\rangle$ for a Twitter discussion Γ and two tweets $t_1, t_2 \in \Gamma$, we will say that t_1 *criticizes* t_2, written $t_1 \rightsquigarrow t_2$, iff t_1 answers t_2 and the degree of belief that the message of tweet t_1 is a criticism to the message of tweet t_2 is greater than zero. In other words, $t_1 \rightsquigarrow t_2$ iff $(t_1, t_2) \in E$ and $P(t_1, t_2) > 0$.

In Twitter, every tweet in a discussion can reply at most one tweet, but can mention many tweets, and all of them are prior in the discussion. So, every tweet can answer (and criticize) many prior tweets, either from a same author or from different ones. Given a

tweet t_1, we consider the set of tweets $\{t_{1_{a_1}}, \ldots, t_{1_{a_n}}\}$ that t_1 is answering to as those tweets including (i) the tweet that t_1 is replying to, and (ii) all the other previous tweets in the discussion by authors mentioned by t_1.

To check whether a tweet t_1 does not agree with the claim expressed in one of its answered tweets $t_{1_{a_i}}$, the system uses an automatic labeling system based on Support Vector Machines (SVM). The description of the method we used to train the SVM can be found in [2]. The SVM model is built from a set of 582 pairs of tweets (answers) obtained from a discussion set on Spanish politics, and manually labeled with the most probable label: criticism or not criticism. To build the SVM model, for each pair of tweets $(t_1, t_{1_{a_i}})$ we consider different attributes from the tweets of the pair: attributes that count the number of occurrences of relevant words in the tweets and attributes that have to be computed from the message. In particular, for each tweet, we have considered regular words and stop-words, the number of images, the number of URLs mentioned in the tweet, the number of positive and negative emoticons and the sentiment expressed by the tweet. We use LibSVM [7] to train a probabilistic SVM model, that is, a labeling function that assigns a probability value p for each possible label to each answer $(t_1, t_{1_{a_i}})$. The probability estimates can be obtained by using Platt's likelihood method [15]. LibSVM uses the same Platt's method but algorithmically improved [14]. With our SVM model for Spanish politics discussions, we obtain an accuracy of 75% over our training set of tweet pairs. This SVM model, obtained from such small data set, may not be good enough to be used in a final system, but one can always consider training a SVM model with a larger data set.

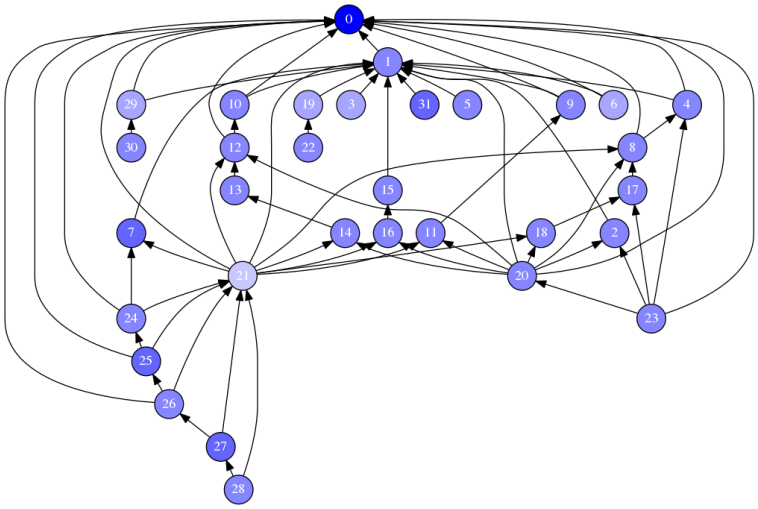

Fig. 1. Tweet-based model for a Twitter discussion.

In Figure 1 we show the PDisG for a Twitter discussion[3] from the political domain obtained by our discussion retrieval system. Each tweet is represented as a node and each criticism relationship between tweets is represented as an edge (answers with

[3] The discussion URL is
`https://twitter.com/jordievole/status/574324656905281538`

probability values greater than zero). The root tweet of the discussion is labeled with 0 and the other tweets are labeled with consecutive identifiers according to their generation order. The discussion has a simple structure. The root tweet starts the discussion (node 0), the reply (node 1) criticizes the root tweet and the rest of tweets within the discussion criticize mainly node 0 and node 1. The discussion contains 32 tweets of 14 different authors, and 81 criticizes relations between tweets. Nodes are colored in *blue scale*, where the darkness of the color is directly proportional to the number of followers of the authors of the tweets with respect to the maximum value in the discussion. Notice that the graph does not contain cycles, since a tweet only answers previous tweets in the discussion.

3 Author-centered model

As we have already pointed out, our goal is to introduce and investigate an author-centered model of Twitter discussions with probabilistic valued relationships. To this end, we group tweets by authors and we consider that criticism relationships between tweets denote controversies at the level of authors.

In this work we consider discussions in which every author's opinion is consistent, discussions in which authors are not self-referenced and do not contradict themselves. That is, for each author a_i and each pair of tweets $t_1 = (m_1, a_i, f_1)$ and $t_2 = (m_2, a_i, f_2)$, we assume that messages m_1 and m_2 do not express neither conflicting nor inconsistent information. Next we define what we will understand by the opinion and the number of followers of an author in a Twitter discussion Γ (with authors' identifiers $\{a_1, \ldots, a_n\}$):

- The opinion of an author a_i in the discussion Γ, denoted T_{a_i}, is the set of tweets of a_i in Γ, i.e. $T_{a_i} = \{(m, a_i, f) \in \Gamma\}$.
- The number of followers of an author a_i in Γ, denoted $f_{a_i} \in \mathbb{N}$, is the mode of the set $\{f \mid (m, a_i, f) \in \Gamma\}$, which provides us with the most frequent number of followers of the author during the discussion.

Given a Twitter discussion, we notice that, in fact, every author a_i can be uniquely represented by his/her opinion T_{a_i}. So, we shall refer to both terms indistinctly. Next we define the probabilistic author graph for a given discussion.

Definition 4. *(Probabilistic Author Graph) Let Γ be a Twitter discussion with authors' identifiers $\{a_1, \ldots, a_n\}$ and let $\langle T, E, P\rangle$ be the* PDisG *for Γ. The probabilistic author graph (*ADisG*) for Γ is a triple $\langle \mathcal{T}, \mathcal{E}, \mathcal{P}\rangle$, where*

- *the set of nodes $\mathcal{T}$ is the set of authors' opinions $\{T_{a_1}, \ldots, T_{a_n}\}$, i.e. a node for each author.*
- *the set of edges $\mathcal{E}$ is the set of answers between different authors in the discussion; i.e. there is an edge $(T_{a_i}, T_{a_j}) \in \mathcal{E}$, with $a_i \neq a_j$, iff there is $(t_1, t_2) \in E$ such that $t_1 \in T_{a_i}$ and $t_2 \in T_{a_j}$.*
- *$\mathcal{P}$ is a probabilistic weighting scheme, i.e. a map $\mathcal{P} : \mathcal{E} \to [0, 1]$ assigning to every edge $(T_{a_i}, T_{a_j}) \in \mathcal{E}$ a probability value in $[0, 1]$, that expresses a degree of*

belief with which the author a_i actually criticizes the author a_j. For each edge $(T_{a_i}, T_{a_j}) \in \mathcal{E}$, the value $\mathcal{P}(T_{a_i}, T_{a_j})$ is meant to be computed from the set of individual probabilities that tweets in T_{a_i} criticize tweets in T_{a_j}, i.e. from the set

$$\{P(t_1, t_2) \mid (t_1, t_2) \in E,\ t_1 \in T_{a_i} \text{ and } t_2 \in T_{a_j}\}.$$

Notice that an author can answer several authors in a discussion, and thus, can criticize several authors. However, if an author criticizes the opinion of another through several tweets, the set of discrepancies is represented with a single edge in $\mathcal{E}$ and with a single probability value, which is meant to denote the global belief that one opinion criticizes the other.

The ADisG graph shows discrepancies between authors only if there is some (explicit) criticism relationship between the tweets of the authors, and thus, indirect criticism relations between authors have not been considered yet in our model. For instance, consider a Twitter discussion with tweets $t_1 = (m_1, a_1, f_1)$, $t_2 = (m_2, a_2, f_2)$ and $t_3 = (m_3, a_3, f_3)$, with $a_1 \neq a_2 \neq a_3$. Suppose that $t_1 \rightsquigarrow t_2$ and $t_3 \rightsquigarrow t_1$ i.e. $\{(t_1, t_2), (t_3, t_1)\} \subseteq E$, $P(t_1, t_2) > 0$ and $P(t_3, t_1) > 0$. In our current approach, we restrict ourselves to consider that $t_3 \rightsquigarrow t_2$ iff t_3 answers (replies or mentions) t_2. The reason is that the information contained in a typical tweet, written in natural language and with possibly other attributes, almost never allows us to consider a sound way to assess an indirect criticism relation between two tweets t and t' if t' does not directly reply or mention t.

In the next section we introduce three different probabilistic weighting schemes, depending on the semantics assumed for the criticism relation between two authors' opinions.

4 Probabilistic weighting schemes

In our approach, each node of an ADisG graph denotes an author's opinion, and relationships between nodes are mined from the prevailing sentiment among the aggregated tweets of the opinions.

To be more precise, let Γ be a Twitter discussion and let $\langle \mathcal{T}, \mathcal{E}, \mathcal{P} \rangle$ be the probabilistic author graph (ADisG) for Γ. Suppose further we have two authors' opinions or sets of authors' tweets $T_a, T_b \in \mathcal{T}$, with $(T_a, T_b) \in \mathcal{E}$. Our aim is to define a probabilistic weighting scheme $\mathcal{P} : \mathcal{E} \to [0, 1]$ for edges in $\mathcal{E}$, by combining in an appropriate form the individual probabilities values $\{P(t_1, t_2) \mid t_1 \in T_a \text{ and } t_2 \in T_b\}$, where we consider $P(t_1, t_2) = 0$ for pairs of tweets such that $(t_1, t_2) \notin E$. As we will see, the addition of zero values to this set will be harmless.

In the rest of this section we define three possible probabilistic weighting schemes $\mathcal{P}$, depending on the semantics assumed for the criticism relationship between the authors' opinions T_a and T_b.

4.1 Skeptical scheme

A skeptical notion of criticism between T_a and T_b can be defined as follows: T_a criticizes T_b, written $T_a \rightsquigarrow T_b$, when every tweet in T_b is attacked by some tweet in T_a, i.e. for all $t \in T_b$, there is $t' \in T_a$ such that $t' \rightsquigarrow t$.

In logical terms, we can define $T_a \rightsquigarrow T_b$ by the following clause:

$$T_a \rightsquigarrow T_b := \bigwedge_{t \in T_b} \left(\bigvee_{t' \in T_a} t' \rightsquigarrow t \right)$$

Assuming independence of all the $t' \rightsquigarrow t$'s, which is a reasonable assumption in our context,[4] we can easily compute the probability of $T_a \rightsquigarrow T_b$ as

$$\mathcal{P}(T_a \rightsquigarrow T_b) = \prod_{t \in T_b} \left(\bigoplus_{t' \in T_a} P(t', t) \right),$$

where $\oplus$ corresponds to the probabilistic sum operation $x \oplus y = x + y - x \cdot y$. Observe that 0 is a neutral element for $\oplus$ (i.e. $x \oplus 0 = x$), and so having probability values such that $P(t', t) = 0$ does not affect the computation of $\mathcal{P}(T_a \rightsquigarrow T_b)$. Analogously for the next schemes.

4.2 Credulous scheme

On the other hand, a credulous notion of criticism between T_a and T_b can be defined as follows: T_a criticizes T_b, written $T_a \rightsquigarrow^c T_b$, when there is at least one tweet $t \in T_b$ that is attacked by a tweet $t' \in T_a$, i.e. when there are $t \in T_b$ and $t' \in T_a$ such that $t' \rightsquigarrow t$.

In logical terms, $T_a \rightsquigarrow^c T_b$ can be now expressed by the following clause:

$$T_a \rightsquigarrow^c T_b := \bigvee_{t \in T_b} \left(\bigvee_{t' \in T_a} t' \rightsquigarrow t \right).$$

Again, assuming independence of all the $t' \rightsquigarrow t$'s, we can easily compute the probability of $T_a \rightsquigarrow T_b$ as

$$\mathcal{P}(T_a \rightsquigarrow^c T_b) = \bigoplus_{t' \in T_a, t \in T_b} P(t', t).$$

4.3 Intermediate scheme

A more flexible definition of when T_a criticizes T_b is to stipulate that this holds when for *most* of the tweets $t \in T_b$ there is a tweet $t' \in T_a$ such that $t \rightsquigarrow t'$. We denote this notion of attack as $T_a \rightsquigarrow_{most} T_b$.

The question is how we interpret the quantifier *most*. A first option is to understand *most* as a proportion of at least r, for some $r \geq 0.5$ to be chosen. For any set X, let us define $most(X) = \{S \subseteq X \mid \frac{|S|}{|X|} \geq r\}$. Then we can express $T_a \rightsquigarrow_{most} T_b$ as follows:

$$T_a \rightsquigarrow_{most} T_b := \bigvee_{S \in most(T_b)} T_a \rightsquigarrow S.$$

[4] This is because in our probabilistic model the label $P(t_1, t_2)$ assigned to an edge (t_1, t_2) is based only on the information inside the tweets t_1 and t_2 and not on other answers from the same authors.

But we can simplify a bit this expression. Indeed, since if $S \subset R$ then $(T_a \rightsquigarrow S) \vee (T_a \rightsquigarrow R) = T_a \rightsquigarrow S$, we can write

$$T_a \rightsquigarrow_{most} T_b := \bigvee_{S \in Min(most(T_b))} T_a \rightsquigarrow S,$$

where $Min(most(X))$ denotes the minimal subsets of X with a proportion of at least r. Then, we can compute:

$$\mathcal{P}(T_a \rightsquigarrow_{most} T_b) = \mathcal{P}(\bigvee\{T_a \rightsquigarrow S : S \in Min(most(T_b))\}).$$

This can be computationally expensive. However, we can provide a lower approximation taking into account that for any probability we have $P(A \cup B) \geq \max(P(A), P(B))$:

$$\mathcal{P}_*(T_a \rightsquigarrow_{most} T_b) = \max\{\mathcal{P}(T_a \rightsquigarrow S) : S \in Min(most(T_b))\}.$$

Interestingly enough, there is a simple procedure to compute P_*:

(i) compute, for all $t \in T_b$, the probabilities $\mathcal{P}(T_a \rightsquigarrow t) = \bigoplus_{t' \in T_a} P(t', t)$;
(ii) rank them, from higher to lower: $P(T_a \rightsquigarrow t_1) \geq P(T_a \rightsquigarrow t_2) \geq \ldots$;
(iii) let k be the smallest index such that $\frac{k}{|T_b|} \geq r$.

Then, we have $\mathcal{P}_*(T_a \rightsquigarrow_{most} T_b) = \prod_{i=1}^{k} P(T_a \rightsquigarrow t_i)$.

5 Mining the set of consistent opinions

Once we have introduced the author-centered model of discussions in Twitter, the next key component is the definition of the reasoning system to compute the set of accepted authors' opinions. To this end, we have extended the reasoning system developed in [1] to deal here with ADisG graphs. The approach, described in the rest of the section, consists of mapping an ADisG graph, with a particular probabilistic weighting scheme, to a valued abstract argumentation framework (VAF) and considering the ideal semantics to compute the (unique) set of consistent authors' opinions of the discussion. Bench-Capon's valued abstract argumentation [3] is an extension of abstract argumentation with a valuation function *Val* for arguments taking values on a set R equipped with a (possibly partial) preference relation *Valpref*. Ideal semantics [9] guarantees that all opinions in the solution are consistent and that the solution is maximal in the sense that it contains all acceptable arguments.

5.1 The argumentation-based reasoning system

Given an ADisG for a Twitter discussion with a given probabilistic weighting scheme, we build a corresponding VAF where arguments represent authors' opinions and attacks between arguments represent discrepancies between authors' opinions according to an uncertainty threshold α, which characterizes how much uncertainty on probability values we are ready to tolerate.

Definition 5. *(VAF for an* ADisG*) Let Γ be a Twitter discussion with authors identifiers $\{a_1, \ldots, a_n\}$ and let $\alpha \in [0,1]$ be a threshold on the probability values. If $G = \langle \mathcal{T}, \mathcal{E}, \mathcal{P} \rangle$ is the* ADisG *graph for Γ with probabilistic weighting scheme $\mathcal{P}$, the Valued Argumentation Framework for G relative to the threshold α, written VAF(G, α), is the tuple VAF(G, α) = $\langle \mathcal{T}, \texttt{attacks}, R, Val, Valpref \rangle$, where*

- *each node (or author's opinion) T_{a_i} in $\mathcal{T}$ results in an argument,*
- *$\texttt{attacks}$ is an irreflexive binary relation on $\mathcal{T}$ and it is defined according to the threshold α as follows: $\texttt{attacks} = \{(T_{a_i}, T_{a_j}) \in \mathcal{E} \mid \mathcal{P}(T_{a_i}, T_{a_j}) \geq \alpha\}$,*
- *R is a non-empty set of relevance values,*
- *Valpref $\subseteq R \times R$ is an order relation (transitive, irreflexive and asymmetric) on the set of relevance values R.*
- *Val $: \mathcal{T} \rightarrow R$ is a valuation function that assigns relevance values to authors' opinions or arguments,*

An important element of our approach is the use of an uncertainty threshold α. It represents the maximum probability value under which we would be prepared to disregard criticism relationships between authors' opinions. So, the $\texttt{attacks}$ relation is interpreted as follows: the opinion of the author a_i is in disagreement with the opinion of the author a_j with at least a probability value α, according to the probabilistic weighting scheme $\mathcal{P}$.

Given such a VAF(G, α) = $\langle \mathcal{T}, \texttt{attacks}, R, Val, Valpref \rangle$, a *defeat* relation (or effective attack relation) between arguments (authors' opinions) is defined according to the valuation function *Val* and the preference relation *Valpref* as follows:

$$defeats = \{(T_{a_i}, T_{a_j}) \in \texttt{attacks} \mid (Val(T_{a_j}), Val(T_{a_i})) \notin Valpref\}.$$

As we have already pointed out, we consider the ideal semantics for computing the set of consistent authors' opinions of a discussion. The ideal semantics for valued argumentation is defined through the ideal extension (solution) which guarantees that the set of tweets in the solution is the maximal set of tweets that is consistent, in the sense that there are no defeaters among them, and all the tweets outside the solution are defeated by a tweet within the solution. That is, if a tweet outside the solution defeats a tweet within the solution, it is, in turn, defeated by another tweet within the solution. In other words, the solution is the biggest consistent set of tweets that defeats any defeater outside the solution. In [9] the authors prove that the ideal extension is unique.

Formally, given a VAF(G, α) = $\langle \mathcal{T}, \texttt{attacks}, R, Val, Valpref \rangle$, a set of arguments $S \subseteq \mathcal{T}$ is *conflict-free* iff for all $T_{a_i}, T_{a_j} \in S, (T_{a_i}, T_{a_j}) \notin defeats$. Given a conflict-free set of arguments $S \subseteq \mathcal{T}$, S is *maximally admissible* iff

(i) for all $T_{a_1} \notin S$, $S \cup \{T_{a_1}\}$ is not conflict-free and
(ii) for all $T_{a_1} \notin S$ and $T_{a_2} \in S$, if $(T_{a_1}, T_{a_2}) \in$ *defeats*, there exists $T_{a_3} \in S$ such that $(T_{a_3}, T_{a_1}) \in$ *defeats*.

Accordingly, we define what the solution of a discussion Γ is as follows.

Definition 6. *(Solution of a discussion) Given the* ADisG *graph $G = \langle \mathcal{T}, \mathcal{E}, \mathcal{P} \rangle$ for a discussion Γ and a probabilistic weighting scheme $\mathcal{P}$, the* set of accepted authors' opinions *of Γ for given a threshold α, or* solution *of Γ, is the largest admissible conflict-free set of authors' opinions $S \subseteq \{T_{a_1}, \ldots, T_{a_n}\}$ in the intersection of all maximally admissible conflict-free sets in the valued argumentation framework VAF(G, α).*

5.2 Implementation and analysis of results

As for the implementation purposes, we have instantiated the set of relevance values R to the set of natural numbers $\mathbb{N}$, and the preference relation *Valpref* to the natural order on $\mathbb{N}$. We have also instantiated the valuation function Val to the function $\texttt{followers} : \mathcal{T} \rightarrow \mathbb{N}$, with $\texttt{followers}(T_{a_i}) = \lfloor \log_{10}(f_{a_i} + 1) \rfloor$, where $f_{a_i} \in \mathbb{N}$ is the number of followers of the author a_i computed as the mode of the set $\{f \mid (m, a_i, f) \in T_{a_i}\}$ (i.e. the most frequent number of followers of the author during the discussion). This function allows us to quantify authors' relevance from the orders of magnitude of authors' followers, since we want to consider that one author is more relevant than another only if the number of followers is at least ten times bigger for the first author.

To implement the reasoning system, we have used the Answer Set Programming (ASP) approach of the argumentation system ASPARTIX [11]. Actually, we have extended ASPARTIX to deal with VAFs, as the current implementation only works with non-valued arguments. To develop such extension we have modified the manifold ASP program described in [12] to incorporate the valuation function for arguments and the preference relation.

The author-centered approach allows us to perform an analysis of results different from the tweet-based approach proposed in [1]. Aggregating the information by author allows us to identify the set of authors whose opinions are consistent or in agreement in the discussion, the authors involved in a circular argumentative discussion, and the most controversial authors. That is, for instance, we can look for the authors who receive the greatest number of criticisms, the authors who participate in the greatest number of cycles, or the authors that generate the longest argumentative chains.

Figure 2 shows the solution for an ADisG graph instance for the discussion of Figure 1. To build the ADisG graph, we have used the intermediate probabilistic weighting scheme $\mathcal{P}_*(T_{a_i} \rightsquigarrow_{most} T_{a_j})$ with the proportion parameter $r = 0.6$.[5] To find the solution for the ADisG graph (the set of accepted opinions of the discussion according to Definition 6), we have used the uncertainty threshold $\alpha = 0.6$ and the above `followers` valuation function for estimating the authors' relevance in Twitter. According to it, the authors of the discussion are stratified in five levels denoting their relevance, namely: level 0 (lowest level): $\{11\}$, level 1: $\{5, 6, 7, 13\}$, level 2: $\{0, 1, 3, 4, 8, 9, 10\}$, level 3: $\{12\}$ and level 4: $\{2\}$.

The nodes colored in blue are the accepted authors (authors' opinions in the solution) and the nodes colored in gray are the rejected ones, where the darkness of the color is directly proportional to the value of the `followers` function of each author. The edges colored in black are the answers between authors that cannot be classified as attacks, since the criticism probabilities are below the threshold $\alpha = 0.6$, while the edges colored in red are attacks between authors; i.e. answers with a criticism probability of at least the threshold $\alpha = 0.6$. For attack edges, the darkness of the color is directly proportional to the criticism probability with respect to the maximum value. With $r = 0.6$ and $\alpha = 0.6$, 11 answers between authors do not give rise to attacks. The ADisG graph has 13 cycles considering all answers among authors, and Authors 8 and 2 seem to be the most controversial ones.

[5] We plan to implement the other weighting schemes in the near future.

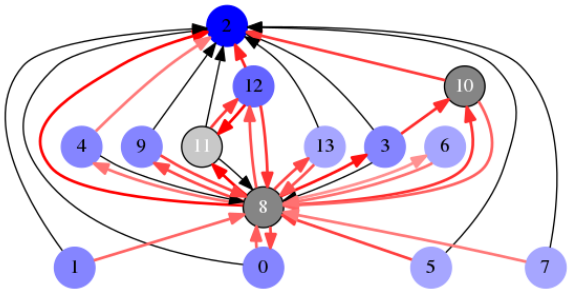

Fig. 2. Author-centered model and its solution.

The *solution* contains 11 of the 14 authors and only 3 are rejected (Authors 8, 10 and 11). On the one hand, Author 2 is the owner of the root tweet of the conversation (node 0 in the tweet-based model of Figure 1), and a total of four other authors (4, 9, 10 and 12) attack him, but he in turn does not reply later to the rest of tweets of the conversation. So, Author 2 is not involved in any cycle in the ADisG graph. Because the weight of Author 2 is greater than the one of any of his attacking authors, Author 2 belongs to the solution of the graph. With respect to the four attackers of Author 2, two of them (4 and 12) are also in the solution, since Author 12 does not defeat Author 2 and his weight is greater than the one of any of his attacking authors. On the other hand, Author 12 defeats Author 8 and this allows Authors 3, 4 and 9 to be in the solution, while in turn, accepting Author 3 causes Author 10 to be rejected. When analyzing the cycles of the graph, we obtain that Author 8 is involved in a total of 8 cycles, considering only attacks answers among authors, and almost all authors involved in cycles with Author 8 are in the solution (0, 6, 13 , 9 and 12). Thus, Author 8 produces a lot of circular discussions, but the weight of Author 12 is high enough to make Author 8 lose the discussion. Observe that in the ideal semantics, authors with a same weight that form a cycle are not accepted if none of the authors in the cycle is attacked by other authors outside of the cycle and accepted in the solution. Hence, in this discussion with high controversy around Author 8 (with a high number of cycles), we end up accepting many of these authors' opinions. Finally, as Authors 1, 5 and 7 only attack Author 8, all of them are also in the solution, while Author 11 is rejected, since it is defeated by Author 12.

6 Conclusions and future work

In this paper we have introduced first ideas on a probabilistic author-centered approach to analyze the set of accepted authors' opinions in Twitter discussions. We model discussions with a graph, where nodes represent whole sets of tweets of a single author, and thus representing his opinion, and edges between nodes represent criticism relationships between authors. Then, using valued abstract argumentation and ideal semantics, we compute the set of winning authors in the discussion. By comparing the set of accepted opinions with the rejected ones, we can detect the degree of polarization between both sets.

As future work, we plan to extend the author-centered model to also consider support relationships between tweets and also to explore more credulous acceptability semantics.

References

1. Alsinet, T., Argelich, J., Béjar, R., Fernández, C., Mateu, C., Planes, J.: Weighted argumentation for analysis of discussions in Twitter. Int. J. Approx. Reasoning 85, 21–35, doi: 10.1016/j.ijar.2017.02.004 (2017)
2. Alsinet, T., Argelich, J., Béjar, R., Fernández, C., Mateu, C., Planes, J.: An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships. Pattern Recognition Letters (2018), doi: 10.1016/j.patrec.2017.07.004, in press
3. Bench-Capon, T.J.M.: Value-based argumentation frameworks. In: Proceedings of 9th International Workshop on Non-Monotonic Reasoning, NMR 2002. pp. 443–454 (2002)
4. Bench-Capon, T.J.M., Dunne, P.E.: Argumentation in Artificial Intelligence. Artif. Intell. 171(10-15), 619–641, doi: 10.1016/j.artint.2007.05.001 (2007)
5. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. Artif. Intell. 128(1-2), 203–235, doi: 10.1016/S0004-3702(01)00071-6 (2001)
6. Bosc, T., Cabrio, E., Villata, S.: Tweeties squabbling: Positive and negative results in applying argument mining on social media. In: Computational Models of Argument - Proceedings of COMMA 2016. pp. 21–32, doi: 10.3233/978-1-61499-686-6-21 (2016)
7. Chang, C., Lin, C.: LIBSVM: A library for support vector machines. ACM TIST 2(3), 27:1–27:27, doi: 10.1145/1961189.1961199 (2011)
8. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77(2), 321 – 357, doi: 10.1016/0004-3702(94)00041-X (1995)
9. Dung, P.M., Mancarella, P., Toni, F.: Computing ideal sceptical argumentation. Artif. Intell. 171(10-15), 642–674, doi: 10.1016/j.artint.2007.05.003 (2007)
10. Dusmanu, M., Cabrio, E., Villata, S.: Argument mining on twitter: Arguments, facts and sources. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017. pp. 2317–2322 (2017)
11. Egly, U., Gaggl, S.A., Woltran, S.: ASPARTIX: Implementing argumentation frameworks using answer-set programming. In: Proceedings of the 24th International Conference on Logic Programming, ICLP 2008. pp. 734–738 (2008)
12. Faber, W., Woltran, S.: Manifold answer-set programs for meta-reasoning. In: Proceedings of Logic Programming and Nonmonotonic Reasoning, LPNMR 2009. pp. 115–128 (2009)
13. Grosse, K., González, M.P., Chesñevar, C.I., Maguitman, A.G.: Integrating argumentation and sentiment analysis for mining opinions from Twitter. AI Commun. 28(3), 387–401, doi: 10.3233/AIC-140627 (2015)
14. Lin, H.T., Lin, C.J., Weng, R.C.: A note on Platt's probabilistic outputs for support vector machines. Machine Learning 68(3), 267–276, doi: 10.1007/s10994-007-5018-6 (2007)
15. Platt, J.C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in Large Margin Classifiers. pp. 61–74. MIT Press (1999)
16. Simari, G.R., Rahwan, I.: Argumentation in Artificial Intelligence. Springer Publishing Company, (2009)