Weighted Argumentation for Analysis of Discussions in Twitter

Teresa Alsinet, Josep Argelich, Ramón Béjar, Cèsar Fernández,
Carles Mateu, Jordi Planes

Department of Computer Science, Universitat de Lleida
C/Jaume II 69, 25001 Lleida, Spain
{tracy,jargelich,ramon,cesar,carlesm,jplanes}@diei.udl.cat

Abstract

Twitter has become a widely used social network to discuss ideas about many domains. This leads to a growing interest in understanding what are the major accepted or rejected opinions in different domains by social network users. At the same time, checking what are the topics that produce the most controversial discussions among users can be a good tool to discover topics that can be divisive, what can be useful, e.g., for policy makers. With the aim to automatically discover such information from Twitter discussions, we present an analysis system based on Valued Abstract Argumentation to model and reason about the accepted and rejected opinions. We consider different schemes to weight the opinions of Twitter users, such that we can tune the relevance of opinions considering different information sources from the social network. Towards having a fully automatic system, we also design a relation labeling system for discovering the relation between opinions. Regarding the underlying acceptability semantics, we use ideal semantics to compute accepted/rejected opinions. We define two measures over sets of accepted and rejected opinions to quantify the most controversial discussions. In order to validate our system, we analyze different real Twitter discussions from the political domain. The results show that different weighting schemes produce different sets of socially accepted opinions and that the controversy measures can reveal significant differences between discussions.

Key words: Abstract argumentation, weighted arguments, semantic attacks, discussions in Twitter.

1 Introduction

The analysis of opinions on general and specialized social networks, has recently received a lot of attention on many application fields. For example, there is a vivid interest in the analysis of tourists' opinions about destinations and facilities, aimed at getting insight on tourist's behavior and preferences for improvement and investment policy planning [40,48,51], and similar efforts are being done on marketing [9], customer engagement [53], and related fields [35,16]. Less numerous are the contributions centered around analyzing, not individual opinions, but debates and conversations where the structural relations between opinions are a key component to be able to pinpoint the accepted, or winning, opinions in a discussion. In this line, key contributions are the works of Atkinson et al. about using argumentation for tools to support e-participation in deliberation processes [4,11,12,50].

Although there exist many specialized and generalist social networks, nowadays Twitter is one of the most widely ones when it comes to share and criticize relevant news, and the citizens response to news and events in Twitter is frequently taken as an indicator of the social interest for that topic. For example, this can be observed in the efforts of researchers in social sciences and artificial intelligence to use automatic tools to analyze trends in the responses of Twitter users to certain events [44,38] and even the use of Twitter data to sample the most prevailing opinions in the political landscape [14], or even to predict results regarding a government election [47]. The election of Twitter as a research target and tool when considering social networking sites is not an arbitrary decision [54], but relies on two major facts: size and openness. As for size, Twitter has an **active** audience of over 320 Million users per month (as of December 2015, [33]), with up to 1.3 billion registered users [49], tweeting an estimated 6000 tweets per second [45]. When it comes to openness, Twitter is an open social network, easy to use in research thanks to two design decisions, first, all tweets, unless stated the contrary, are open to be read by anyone, and second, since its very first days, Twitter has an open extensive developer API [34] that allows developers to interact with Twitter servers, extracting huge amounts of data easily.

In this paper we present a system for analysis of discussions in Twitter that is based on valued abstract argumentation. Discussions in Twitter are represented as weighted argumentation problems, where arguments are tweets and argument values are the weights used to model the relative social relevance of tweets, considering different ways to measure such relevance from data obtained from Twitter. To move from natural language inputs (the Twitter data) to valued abstract argumentation, we develop an automatic labeling system, based on support vector machines, to discover the semantic relation between tweets in a discussion and we associate weights with tweets from its social relevance taking into account three different attributes of a tweet: the number of followers of the author, the number of retweets and the number of favorites. We use ideal semantics as the argumentation model to capture the set of socially accepted tweets in our weighted argumentation problems. We also define measures for the controversy analysis between accepted tweets and the rest of tweets in the discussion, with the goal of providing tools to analyze which topics may be causing a major controversy (and so possible division) between users of the social network. Finally we analyze some Twitter discussions as case studies for our system. The results show that our system allows us to discover subsets of opinions in Twitter that are widely accepted and defended by others, under different weighting schemes, and that the controversy analysis shows that even for discussions with similar solution sizes, as the ones we analyze as case studies, the controversy measures can reveal that some discussions may be more critical than others with respect to social division.

The rest of the paper is organized as follows. In the next subsection we summarize the more relevant related work. In Section 2 we define the formal structures for our system, the weighted labeled discussion graphs. In Section 3 we present the weighted argumentation framework we use for computing the accepted tweets in a solution. In Section 4 we present the architecture of our discussion analysis system and in Section 5 we define the controversy measures we compute for the solution of a discussion. Finally, in Section 6 we present some case studies and analyze them with our system. We end the paper with some conclusions and a discussion of future work.

1.1 Related work

In the literature we can find several extensions to Dung's work considering the weighting of arguments and relations between them. Prakken and Sartor [43] attach priorities to arguments defining a partial order between them. Arguments are expressed in a logic-programming language with both weak and strong negation, conflicts between arguments are decided with the help of priorities on the rules.

Other approaches have formalized the role of preferences at an abstract level. In Amgoud and Cayrol's [3,2] Dung's framework is augmented with a preference ordering on the set of arguments, so that an attack by an argument X on an argument Y is successful only if Y is not preferred to X. On the other hand, Modgil [41] extends Dung's theory to accommodate arguments that claim preferences between other arguments, thus incorporating meta-level argumentation-based reasoning about preferences at the object level. Other extensions are those of Cayrol and Lagasquie-Schiex [13], which introduce priorities in the selection of arguments in order to represent different levels of selection in the solution where the value of each argument is a function of the values of the arguments that are related to it, not taking into account any weight assigned a priori to the arguments.

The main difference between our system and these abstract semantic extensions is that our approach considers different weighting functions for computing both weights attached to arguments and relations between arguments and weighting functions are dynamically evaluated from attributes representing social relevance.

Recently, departing from Dung's model of argument systems, Matt and Toni [39] formalize a measure of argument strength by applying the concept of value of a game, as defined in Game Theory, and Dunne et al. [23] associate attacks with weights indicating the relative strength of the attack. Bench-Capon [6] defines valued-based abstract argumentation by attaching to each argument the social values that it promotes, and making the semantics dependent on a particular preference order over values, representing a particular audience. This valued-based abstract argumentation framework has been used as underlying semantics for defining the notion of argument acceptability in our system.

Our acceptability semantics draws from the so-called "ideal semantics" promoted by Dung, Mancarella and Toni [18] as an alternative basis for skeptical reasoning within abstract argumentation settings. Informally, ideal acceptance not only requires an argument to be skeptically accepted in the traditional sense but further insists that the argument is in an admissible set all of whose arguments are also skeptically accepted. While the original proposal was couched in terms of the so-called preferred semantics for abstract argumentation, in [24] the notion of "ideal acceptability" has been extended to arbitrary semantics, showing that standard properties of classical ideal semantics, e.g. unique status, continue to hold in some extension-based semantics (see also [22] for an analysis of the computational complexity of the ideal semantics within abstract argumentation frameworks and assumption-based argumentation frameworks). In our system the solution for a set of tweets is computed by extending the algorithm for computing the ideal extension for an argumentation framework presented in [21], but adapting it to work with weights.

Regarding argumentation models for social context, Leite and Martins [37,26] propose a semantical extension of Dung's Abstract Argumentation Framework [17] that incorporates social voting by adding votes to arguments and to relations between them. Votes are assumed to be extracted from an online debating system. The exploitation of Twitter by means of argumentation frameworks has also been explored by Grosse et al. [30,31], who created a framework which allows opinion mining from incrementally generated Twitter queries, triggering the construction of argument trees such as those found in classical dialogue-based argumentation [7]. In their approach, an argument is a set of tweets for a given query (mainly a set of hashtags), and a tree is a hierarchical relation between them, with subsumption and conflict relations. The trees obtained resemblance dialectical trees used in their previous work on Defeasible Logic Programming (DeLP) [46], although no argumentation algorithm is defined to extract the most relevant arguments from trees.

Our system is close to the argumentation framework developed by Cabrio and Villata [10]. The authors use bipolar argumentation algorithms and semantics to evaluate the set of accepted arguments, given the support and the attack relations among them. The arguments and the relations among them are detected by an automated framework by applying natural language techniques, since the system is focused on online debate, such as Debatepedia.

One key difference between our system and the one proposed by Cabrio and Villata is that we incorporate weighted arguments, by means of different weighting schemes, and define attacks between them by means of preference relations over the weights. We believe that the incorporation of weights to get the relative relevance of arguments, considering information taken from the social network, is an important aspect if we want to finally build tools that are useful for analyzing discussions considering different sources of information for socially accepted arguments. Yet our argumentation system can be utilized to analyze discussions in different social networks, in this work we have focused on the analysis of Twitter discussions. The discussions extracted from Twitter are characterized by: limited number of characters by tweet, use of emoticons and jargon, and social relevance attributes. From these elements, we compute weighted arguments and relations between them by means of an automatic labeling system based on Support Vector Machines. It is worth noticing that in the case of social networks where discussions tend to have more lengthy texts (like for example Reddit), we could consider the use of argumentation frameworks that use more structured arguments, like DeLP [29] or its more recent extension: the weighted argumentation framework RP-DeLP [1], that adds several defeasibility levels in the propositional logic knowledge base and it is based on a recursive ideal semantics. To transform the natural language texts to structured propositional logic knowledge bases, we could consider for example the recent approach followed in [52] to transform English natural language sentences to propositional logic sentences. However, with respect to Twitter, as we have commented above, the usually short length of tweets makes the extraction of more complex structures from single tweets unfeasible, so in the particular case of Twitter the approach of considering tweets as atomic arguments seems the most natural.

2 Weighted Labeled Discussion Graphs

As already pointed out, in this paper we consider the discussions in the social network Twitter. So, in this section we define Twitter discussions and formalize the computational structures used in our system to represent such discussions and reason about them.

Let t_1 and t_2 be two different tweets. We say that t_1 *answers* t_2 whenever t_1 is a reply to the tweet t_2 or t_1 mentions (refers to) tweet t_2. Observe that a tweet can answer many tweets.

Definition 1 *(Discussion Graph) Let Γ be a non-empty set of tweets. The Discussion Graph* (DisG) *for 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 2 *(Labeled Discussion Graph) A labeled discussion graph* (LDisG) *is a tuple (T, E, L), where (T, E) is a discussion graph and L is a labeling function for edges $L : E \to S$, where S represents a set of possible semantic relations for any directed edge (t_1, t_2).*

In this work, the set of *semantic relations* S for a directed edge (t_1, t_2) we have considered is $\{$*criticizes*, *supports*, *none*$\}$, *criticizes* meaning that tweet t_1 does not agree with the claim expressed in tweet t_2, *supports* that tweet t_1 agrees with the claim expressed in tweet t_2 and *none* if the relation is none of the previous two.

It is worth noticing that one could also consider the labeling with the *criticizes* relation for tweet pairs that are not direct answers, by using the information from the *supports* relation; i.e. if t_1 answers t_2, t_3 answers t_1 and t_3 does not answer t_2 then, when t_1 supports t_2 and t_3 criticizes t_1, we could also consider that t_3 criticizes t_2. This approach was proposed in the system defined by Cabrio and Villata [10] and the authors used Bipolar Argumentation Frameworks to extract such indirect relations from online debates where arguments and relations between them are extracted from much more textual information than typical tweets. In our system, when constructing relations between tweets written with natural language and with other attributes such as emoticons, jargon, onomatopoeia and abbreviations, it is often evident that there is uncertainty about whether some of the criticisms and supports hold. So, since our classification system on relations between tweets handles a lot of uncertainty, we have found that the information contained in a typical tweet almost never allows us to consider a sound indirect criticism relation between tweets t_3 and t_2, if t_3 is not an answer for t_2, and thus, indirect relations between tweets have not been considered yet neither in our labeling model nor in the implementation of the system. However, our implementation allows us to incorporate indirect relations extending the labeling model and without affecting the other components of the system, in case we consider further social networks where these indirect relations could make sense. Therefore, with such future goal in mind, when computing semantic relations between tweets, we compute both the *criticizes* relation and the *supports* relation between tweets, even if with the current reasoning algorithm that we will describe later, the *supports* relation is not used for Twitter discussions.

Definition 3 *(Weighted Labeled Discussion Graph) A weighted labeled discussion graph* (WLDisG)*, is a tuple $\langle T, E, L, R, W\rangle$, where (T, E, L) is a* LDisG*, R is a nonempty set of ordered values and W is a weighting function $W : T \rightarrow R$ that assigns a weight value in R to each tweet in T, representing the social relevance of the tweet.*

When working with discussion graphs that we obtain from Twitter, in order to weight tweets considering different notions for social relevance, we have considered the following three weighting schemes that use different sources of information from Twitter:

Followers-weighted The weighting function $W : T \rightarrow R$ maps the number of followers of the author of a tweet to a value in R. Graphs weighted with this function will be referenced as *followers-weighted labeled discussion graphs* (FoWLDisG).

Retweets-weighted The weighting function $W : T \rightarrow R$ maps the number of retweets of a tweet to a value in R. Graphs weighted with this function will be referenced as *retweets-weighted labeled discussion graphs* (ReWLDisG).

Favorite-weighted The weighting function $W : T \rightarrow R$ maps the number of favorites of a tweet to a value in R. Graphs weighted with this function will be referenced as *favorite-weighted labeled discussion graphs* (FaWLDisG).

3 Weighted Abstract Argumentation

Once we have introduced the formal representation for discussions graphs obtained from Twitter discussions, the next key ingredient is the reasoning model used to obtain the set of socially accepted tweets. As we have said in the introduction, we use valued abstract argumentation (VAF) [6] for modeling the weighted argumentation problem associated with discussion graphs and ideal semantics [19] for defining the set of socially accepted tweets.

Valued abstract argumentation is based on the extension of abstract argumentation with a valuation function *Val* on a set of values R for arguments and a (possible partial) preference relation *Valpref* between values in R. For our work, we use the valued argumentation framework introduced by Bench-Capon in [5], also called *audience-specific value-based argumentation framework* in [6].

Definition 4 *(Valued argumentation framework) A valued argumentation framework (VAF) is a tuple* $\langle A, attacks, R, Val, Valpref\rangle$ *where* A *is a set of arguments,* $attacks$ *is an irreflexive binary relation on* A*,* R *is a nonempty set of values,* Val *is a valuation function* $Val\colon A \to R$ *that assigns to each argument in* A *a weight value in* R*, and* $Valpref \subseteq R \times R$ *is a preference relation on* R *(transitive, irreflexive and asymmetric), reflecting the value preferences of arguments.*

Definition 5 *(Defeat) Given a VAF* $\langle A, attacks, R, Val, Valpref\rangle$ *and arguments* a *and* b *in* A*, we say that* a *defeats* b *(written* $defeats(a, b)$*) iff* $(a, b) \in attacks$ *and* $(Val(b), Val(a)) \notin Valpref$*. We also say that* a *effectively attacks* b*.*

We now define the ideal semantics we use in this work for the set of accepted arguments for a given argumentation problem.

Definition 6 *(Conflict-free) Given a VAF* $\langle A, attacks, R, Val, Valpref\rangle$ *and a set* $S \subseteq A$ *of arguments we say that* S *is conflict-free iff* $\forall\ a, b \in S, (a, b) \notin attacks$ *or* $(Val(b), Val(a)) \in Valpref$*; i.e.* $\neg defeats(a, b)$*.*

We note that there might be an attack between two arguments in a conflict-free set, if this relation of attack is not effective; i.e. given a conflict-free set S and arguments $a, b \in S$, it can be the case that a attacks b whenever b is preferred than a according to the preference relation *Valpref*. For instance, if we instantiate the set of ordered values R to the natural numbers $\mathbb{N}$, the valuation function *Val* to a mapping from arguments to $\mathbb{N}$ and the preference relation *Valpref* to the total order relation on $\mathbb{N}$, then it could be that a attacks b whenever $Val(b) > Val(a)$.

Definition 7 *(Acceptable argument) Given a VAF* $\langle A, attacks, R, Val, Valpref\rangle$*, a set* $S \subseteq A$ *of arguments and an argument* $a \in A$ *we say that* a *is acceptable with respect to* S *iff* $\forall\ b \in A, defeats(b, a)$ *implies that* $\exists\ c \in S, defeats(c, b)$*.*

Definition 8 *(Preferred extension) Given a VAF* $\langle A, attacks, R, Val, Valpref\rangle$ *and a set* $S \subseteq A$ *of conflict-free arguments we say that* S *is an admissible extension if for any* $a \in S$*,* a *is acceptable with respect to* S*. We say that* S *is a preferred extension if* S *is a maximally admissible extension; i.e. for any argument* $b \in A$ *if* b *is acceptable with respect to* S*,* $b \in S$*.*

The ideal semantics for a VAF is defined trough the ideal extension. Following [19], the ideal extension is defined as follows.

Definition 9 *(Ideal extension) Given a VAF* $\langle A, attacks, R, Val, Valpref\rangle$ *the ideal extension is the largest admissible extension contained in every preferred* *extension.*

In [19] the authors prove that this extension is unique.

At this point we are ready to formalize the valued argumentation framework for a given weighted labeled discussion graph.

Definition 10 *(VAF for a* WLDisG*) Let* $G = \langle T, E, L, R, W \rangle$ *be a* WLDisG*. The VAF for* G *is* $\mathcal{F} = \langle T, attacks, R, W, Valpref \rangle$*, where*

- *the set of arguments is the set of nodes (or tweets)* T*,*
- *attacks between arguments are defined as follows:*

$$attacks = \{(t_1, t_2) \mid (t_1, t_2) \in E \text{ and } L(t_1, t_2) = criticizes\}$$

- R *is the non-empty set of ordered values that models the social relevance or weight of tweets,*
- *the valuation function for arguments is the weighting function* $W : T \to R$ *for tweets and*
- *the preference relation* $Valpref \subseteq R \times R$ *is the ordering relation over* R*.*

Finally the ideal extension $S \subseteq T$ of $\mathcal{F}$ is the *accepted set of tweets* based on the weighting scheme W and we refer to it as the *solution* of G.

4 Weighted Discussion Analysis in Twitter

The system we have developed allows us to analyze any set of tweets. However, in this work we deal with discussions where a tweet only answers previous tweets, so the graphs that we obtain are always acyclic, and we refer to each of them as *discussion on a tweet*. Formally, a discussion on a tweet rt is a set of tweets Γ such that $rt \in \Gamma$ and the discussion graph for Γ is connected and acyclic and the out degree of rt is zero. We refer to the tweet rt as *the root tweet of the discussion*.

The architecture of our system has two main components: a Discussion Retrieval and a Reasoning System (Figure 1). The first component takes a discussion on a tweet and outputs its WLDisG by managing a tweet collections database, and the second component takes a WLDisG and outputs the set of accepted tweets of the discussion. The Discussion Retrieval component has three phases: the Discussion Tweets Retrieval, which outputs a DisG; the Labeling, which outputs a LDisG; and, finally, the Weighting, which outputs a WLDisG. The first phase inserts data in the tweet collections database, and the second and third phases use data from the database. This three phases and the reasoning component are described in the following subsections.

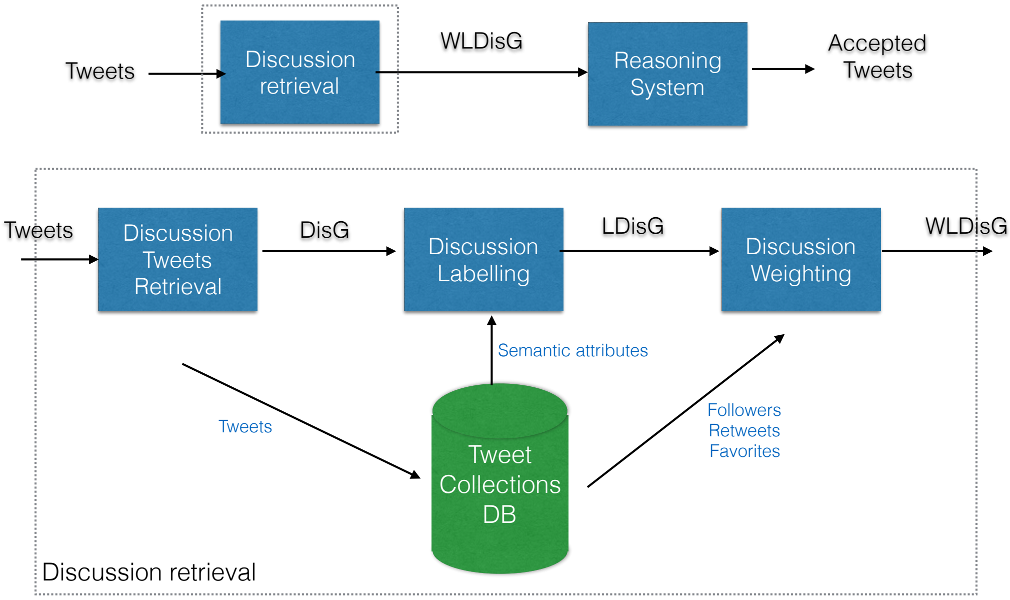

Fig. 1. Architecture of the analysis discussion system.

4.1 Discussion Tweets Retrieval

We have developed an automatic discussion crawler [1] that given a root tweet interacts with Twitter to find the largest discussion on the root tweet, and downloads the tweets of the discussion using the Twitter API. The information that we receive from the Twitter API it is not only the text of the tweets, we also receive additional information like the author, if it is a reply, number of retweets, favorite count, number of followers, etc. This additional information is crucial to be able to build WLDisG instances, as we describe in the next subsections.

The tweets of each discussion are stored in a collection with a NoSQL Database Management System (DBMS), that in our case is MongoDB. [2] To generate a DisG instance from a discussion stored in the collection, we create a graph (T, E) where a vertex t is added to T for each tweet of the collection, and an edge (t_1, t_2) is added to E whenever tweet t_1 answers tweet t_2.

[1] The crawler uses the python `selenium` web driver to interact with the web page associated with a Twitter discussion.
[2] `https://www.mongodb.com/`

4.2 Discussion Labeling

In the second phase for discussion retrieval we have to generate labels, in the set $\{criticises, supports, none\}$, for all the answers in a DisG instance (T, E) for obtaining a LDisG (T, E, L) instance.

With the aim of being able to work with discussions in several languages, we have chosen an automatic labeling system based on Support Vector Machines (SVM). Our SVM model for labeling relations between tweets considers different attributes obtained from the tweets of an answer (t_1, t_2). On the one hand, we have attributes that count the number of occurrences of relevant words in the tweets t_1 and t_2. We have considered two kinds of words: regular words and stopwords. We have considered the inclusion of stopwords as attributes because the typical tweet is very short and the fraction of stopwords that can be giving information about the kind of answer could be relevant. For example, in the next tweet taken from one of the case study discussions of Section 6:

`@ponpimpampum @LL_Sosa Jajajaja...!!! muy bueno.`

we can observe that the stopwords `...` and `!!!` give information about the sentiment associated to the tweet.

On the other hand, we also consider attributes that have to be computed from the text and from the additional information that comes with the tweets. In particular, for each tweet these attributes are the number of images and the number of URLs mentioned in the tweet, the number of positive and negative emoticons and the sentiment expressed by the tweet. Our labeling system incorporates a sentiment analysis computation module [32,42] that given the set of words in a tweet it provides a sentiment value in the range $[-5, +5]$, where -5 is the most negative sentiment and +5 is the most positive sentiment. Finally, the sentiment value is incremented (or decremented) considering every positive (negative) emoticon.

Since SVM follows a supervised learning approach, we first have to train a model from an already labeled data set of answers. To this aim, we have collected a set of several Twitter discussions, on the Spanish language, and we have manually labeled the answers in the discussions to be able to train a SVM labeling model for Spanish discussions.

The training collection contains 12 discussions and a total of 582 pairs of tweets (answers). We have considered the creation of SVM models with different number of regular words (w) and different number of stopwords (s). The words in the collection are sorted by number of occurrences, and for an SVM model with w regular words, we select the first w most frequent regular words. The

SVM model	# w	# s	Accuracy
1	150	50	60%
2	175	25	48%
3	200	50	52%
4	225	25	48%

Table 1
Accuracy: 582 pairs of Spanish tweets and four SVM models.

stopwords have been obtained from the natural language toolkit (NLTK) [8] and have been also ordered by number of occurrences, so we also select the s most frequent stopwords.

Using this training collection, we have trained four models with different values for w and s, in order to get a good labeling model. In order to compute the sentiment for the tweets in this collection, we have taken the AFINN data [3] used in [32,42] and we have translated the words to Spanish. Table 1 shows the accuracy results obtained with our four different models, where each one has a different number of regular words (w) and stopwords (s). The accuracy is the percentage of correctly classified answers from a test set not included in the training set.

As we observe, the best results are obtained for models 1 and 3 where the ratio of stopwords to regular words is higher. We believe that a better performance for the automatic labeling system is hard to obtain given the text characteristics in Twitter: limited number of characters by tweet and use of emoticons, jargon and sarcasm. Another option that we considered was to use the same labeling system used in the argumentation framework of Calabria et al. [10] which is based on the EDITS library [36]. The EDITS library is a textual entailment discovery system that works with some specific modules for the English language. So, using it for other languages, as it is the goal of our system, implies adapting such modules with specialized knowledge in natural language processing. By contrast, our supervised learning process for a labeling model can be used for databases of tweets in any language without any special adaptation.

4.3 Discussion Weighting

Next, from a given LDisG instance we generate a WLDisG instance incorporating weights to each tweet. As we said before, we have considered three weighting schemes, based on followers, retweets, and favorite count, that give place

[3] `http://www2.compute.dtu.dk/~faan/data/AFINN.zip`

to three different weighted graphs: FoWLDisG, ReWLDisG and FaWLDisG. The information for computing the weights with such schemes is obtained from the data stored in the collection of tweets for the discussion.

As we have defined in Section 2, for each weighting scheme we use a weighting function $W : T \to R$ that maps information from a tweet to a value in R, a set of ordered values that models the weight or the social relevance of tweets. In our implementation we have instantiated the set of ordered values R to the natural numbers $\mathbb{N}$ and we have used the same weighting function for the three weighting schemes that use different sources of information from Twitter. Thus, for each weighting scheme, we have considered a function $W : T \to \mathbb{N}$ that maps the number of followers, retweets or favorites of a tweet t, denoted as n_t, to a logarithmic scale that in our case is $W(t) = \lfloor \log_{10}(n_t+1) \rfloor$, allowing us to quantify the orders of magnitude of the number of followers, retweets or favorites of tweets and assigning different weight values (natural numbers) to tweets only if the difference in their social support is significant. Finally, we have instantiated the preference relation *Valpref* to the total order relation on $\mathbb{N}$. So, for each weighting scheme, in our current implementation a tweet t_1 is preferred to a tweet t_2 iff $\lfloor \log_{10}(n_{t_1} + 1) \rfloor > \lfloor \log_{10}(n_{t_2} + 1) \rfloor$.

4.4 Reasoning System

Finally, given the WLDisG associated with a discussion for a root tweet, we compute the set of accepted tweets in the solution by computing the ideal extension for the associated VAF.

To this end we have extended the algorithm for computing the ideal extension for an argumentation framework presented in [21], but adapting it to work with valued arguments. An *argumentation framework* (AF) is a tuple $\langle A, attacks \rangle$ where A is a set of arguments and *attacks* is an irreflexive binary relation on A.

Regarding the implementation we have used an approach based on Answer Set Programming (ASP) available in the argumentation system ASPARTIX [27] but we have extended it to work with VAFs, as the current implementation in ASPARTIX only works with non-valued arguments. To develop such extension we have modified the manifold ASP program explained in [28] incorporating:

- the *valuation function* for arguments,
- the *preference relation* between argument valuations and
- the *effective attack* relation between arguments (defeat relation).

For the sake of completeness, we present here a high-level description of the algorithm, but we refer the reader to the work [28] to know the details of the manifold ASP program for non-valued arguments. The algorithm is based on the following characterization of the ideal extension for an AF. According to [21] an argument a is in the ideal extension for an AF if and only if: (i) no argument that attacks a belongs to an admissible extension and (ii) for any argument b that attacks a there is at least one argument c in the ideal extension that attacks b.

Admissible and ideal extensions for AF are only based on the attack relation between arguments, since no valuation function is considered. So, in order to compute the ideal extension of a VAF we follow the algorithm proposed by Dunne but considering the notion of effective attack based on valuation functions.

Given a VAF $\mathcal{F}$=$\langle A, \mathit{attacks}, R, \mathit{Val}, \mathit{Valpref}\rangle$ the algorithm reduces the problem of computing its ideal extension to the problem of computing an extension for the AF $\mathcal{F}^* = (A_{\mathcal{F}^+} \cup A_{\mathcal{F}^-}, \mathit{attacks}^*)$, where [4]

- $A_{\mathcal{F}^+}$ is the set of arguments that are in some admissible extension for $\mathcal{F}$ but are not defeated by arguments in an admissible extension for $\mathcal{F}$ and do not defeat arguments in an admissible extension for $\mathcal{F}$.
- $A_{\mathcal{F}^-}$ is the set of arguments that do not belong to any admissible extension for $\mathcal{F}$.
- $\mathit{attacks}^*$ is the attack relation with pairs $(a, b) \in \mathit{attacks}$ such that a defeats b and either $a \in A_{\mathcal{F}^+}$ and $b \in A_{\mathcal{F}^-}$, or $a \in A_{\mathcal{F}^-}$ and $b \in A_{\mathcal{F}^+}$.

The *ideal extension for the VAF* $\mathcal{F}$ is the extension (set of arguments) for $\mathcal{F}^*$ that satisfies:

$$\{a \in A_{\mathcal{F}^+} \mid a \in \mathit{admissible}(\mathcal{F}^*)\}, \tag{1}$$

where $\mathit{admissibe}(\mathcal{F}^*)$ is the union of all admissible extensions for $\mathcal{F}^*$.

Our manifold ASP program works by computing all the arguments that belong to some admissible extension for the VAF $\mathcal{F}$, then from them it builds both sets $A_{\mathcal{F}^+}$ and $A_{\mathcal{F}^-}$, it builds the AF $\mathcal{F}^*$ described above and, finally, it iteratively computes the arguments that satisfy Condition 1 for $\mathcal{F}^*$ with the algorithm for bipartite AFs described in [20].

In the worst-case, this algorithm decides if each argument of $\mathcal{F}$ is admissible solving one NP query per each argument in A, but all these NP queries can be executed non-adaptively in parallel. Then, finding the set of arguments that satisfy Condition 1 for the bipartite framework $\mathcal{F}^*$ can be performed with polynomial time. Overall, this gives an algorithm that performs a polynomial number of non-adaptative NP queries. This was shown in [22], where it was also shown that this complexity upper bound is probably tight, as computing

[4] Observe that the underlying graph for the AF $\mathcal{F}^*$ is bipartite.

the ideal extension of a general AF was shown to be hard for the complexity class $FP_{||}^{NP}$. However, there are special cases where computing the ideal extension can be performed more efficiently, like for example when the underlying graph is bipartite or acyclic [20]. A more recently studied case, that also allows polynomial time computation of the ideal extension, is when the tree-width of the underlying graph is bounded by a constant [25]. So, there are argumentation algorithms that are able to efficiently solve big discussions in our system, at least for the above mentioned cases.

Our aim in this work is to develop a system that is able to solve general discussions, without restrictions on the structure of their underlying graphs. To achieve this objective, we have considered an algorithm for finding the ideal extension of a general VAF. However, so far, with our current implementation of the labeling component for Twitter discussions, the underlying graphs that we compute are always acyclic, and so, in future extensions of the system, we can always switch to specific efficient algorithms when scalability to big acyclic discussions is needed.

5 Controversy Analysis

As we have said in the introduction, we want to develop a system to not only get the set of accepted tweets with weighting schemes oriented to social relevance, but also to analyze how much controversy there is among the accepted tweets and the rejected ones. Controversy analysis can be performed taking into account different aspects of opposed opinions between two or more participants in a discussion. So, we define two measures that may capture different aspects related with controversy of opposed views in discussions.

Definition 11 *(Controversy degree) Let $\langle T, E, L, R, W\rangle$ be a* WLDisG *and let $S \subseteq T$ be its solution. We define the controversy degree of S as the number of tweets in $T \setminus S$ that attack the solution S:*

$$\mid \{t_1 \mid (t_1, t_2) \in E \textit{ and } L(t_1, t_2) = \textit{criticizes and } t_1 \in T \setminus S \textit{ and } t_2 \in S\} \mid .$$

Note that the controversy degree measures the number of rejected tweets that criticize some accepted tweets. So a zero controversy degree means that accepted tweets effectively attack those that have been rejected but the rejected ones do not attack the solution. Observe that a high controversy degree indicates that many tweets not in the solution attack some tweets in the solution.

Although the controversy degree offers relevant information about opposing views in and out of the solution, it does not give us information on the structure of the relations between them. There can be many different discussions with a high controversy degree that however have a quite different structure in their solutions. For example, a high controversy degree may be due to a high number of tweets attacking a single tweet in the solution. With the aim of analyzing not only the number of attacked tweets in the solution but also the structure of consecutive attacks we propose what we call a controversial path in the solution.

Definition 12 *(Controversial path) Given a* WLDisG $G = \langle T, E, L, R, W \rangle$ *and its solution* $S \subseteq T$*, a controversial path is a directed path* $t_1, t_2, \ldots, t_n$ *in* G *with* $t_1 \in S$ *and that any two consecutive tweets* t_i, t_{i+1} *in the path satisfy that* $(t_i, t_{i+1}) \in E$ *with* $L(t_i, t_{i+1}) = criticizes$ *and* $t_i \in S \leftrightarrow t_{i+1} \in T \setminus S$.

Definition 13 *(Controversy depth) Given a* WLDisG $G = \langle T, E, L, R, W \rangle$ *and its solution* $S \subseteq T$*, the controversy depth of* S *is defined as the length of the longest controversial path.*

Note that the controversy depth measures the number of tweets in a discussion with chained opposed opinions, starting attacks from the solution and returning the attacks from outside of the solution to the solution. So a zero controversy depth means that all tweets of the discussion are in the solution and a high controversy depth indicates that many tweets in the discussion, in and out the solution, are involved in the acceptability of some opinions. In the next section where we analyze the solution of several Twitter discussions, we also evaluate their controversy by means of these measures.

6 Case Studies

In this section we analyze three Twitter discussions obtained from the political domain by computing the solution of the discussion graphs (the accepted tweets) and also analyzing the controversy of the solution as defined in the previous section. Table 2 shows the main structural characteristics of the discussion graphs for the examples selected. For each discussion graph, we show its size (number of tweets), the number of attack edges, the number of leaves (number of tweets without any attacks) and its depth (the length of the longest path in the graph). We consider the attack edges only because these are the ones used when computing the set of accepted tweets in the solution. From the number of leaves and the depth of the graph, we get an upper bound on the number of different controversial paths and on their depth respectively. So, observe that the three discussions show different structural characteristics, which is reflected also in their solutions and in their controversy.

The first discussion is about the accusation of tax fraud over a former Spanish

Discussion	size	attacks	leaves	depth
1	24	9	21	2
2	38	36	23	11
3	27	24	13	8

Table 2
Structural characteristics of the Twitter discussions analyzed.

minister of economy. [5] The discussion has a very simple structure, possibly the most common in Twitter. A root tweet starts a discussion, wherein the majority of tweets support the root tweet, some replies criticize it, and almost no replies between non-root tweets.

The second discussion is about the proposal from a Spanish politician for the possible legalization of marijuana in Spain. [6] It is a discussion where although some users argue that it could be good, there is discussion about the true intentions of the politician making such a proposal. In this second discussion most of the tweets attack the root tweet, but there are also some attacks between secondary tweets.

The third discussion is about the real significance of a public campaign to protest against a governmental law. [7] This discussion differs from the previous one, since the attacks are distributed in a more uniform way between tweets.

6.1 Followers-weighted Labeled Discussion Graphs

We first analyze the FoWLDisG instances for the three discussions. That is, when the weights are computed from the followers-count of the authors of the tweets.

Figure 2 shows the FoWLDisG instance for the first discussion and its solution. Each tweet is represented as a vertex, where the root tweet of the discussion is labeled with 0 and the other vertices are labeled with consecutive identifiers according to the temporal generation order of the tweets in the social network. A directed edge from tweet A to tweet B indicates that A attacks B. Support relations are not shown in the figure. Then, the set of vertices colored in *red scale* are the vertices in the solution, where the darkness of the color is directly

[5] The discussion can be found at `https://twitter.com/iescolar/status/588727284061863936`
[6] The discussion can be found at `https://twitter.com/ElHuffPost/status/588629431306059776`
[7] The discussion can be found at `https://twitter.com/europapress/status/587042571534360577`

proportional to its weight. Vertices colored in *gray scale* are those not in the solution, and its darkness is also directly proportional to its weight.

For the first discussion, the solution contains all the tweets except four ones. This is because almost all the answers (edges) in this discussion are labeled as *supports*, and from the attack edges only four are effective, that are the ones that cause four tweets to be outside of the solution. As already stated in Section 3, we can find attacks between tweets in the solution but only if they are not effective, as it happens in this discussion.

It is interesting to note that here the high weight of the root tweet is due to the high number of users interested in the opinions of the user, that is a well known journalist in his country. In this case the root tweet contains a mixture of objective information and personal opinion, and the high followers count of the user can be, to some extent, related to the support level for his ideas by other users. So, using the followers count as a weight can be a good indication of the level of acceptance of the ideas of that user by others. However, in many cases a follower can simply be a person interested in reading the opinions of the journalist without being supportive of all his opinions.

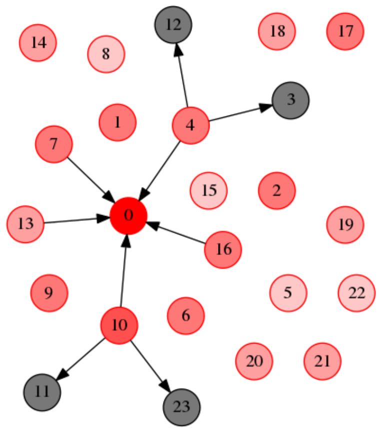

Fig. 2. FoWLDisG instance for Discussion 1: Edges represent attack relations and red vertices accepted tweets in the solution.

For the second discussion, the solution contains 32 of the 38 tweets. In this discussion there are many more *attack* edges than in the previous one, but the number of tweets in the solution is still quite large because many of the attacks are not effective. Figure 3 shows the FoWLDisG instance for this discussion and its solution. Observe that differences in the structure of the discussion graph with respect to the previous discussion clearly affect the structure of the solution. The discussion graph of the first discussion is more *flat*, with no deep paths, in contrast with the discussion graph of the second discussion, that contains a path with 11 tweets. This will be discussed with more detail when we analyze the solution controversy.

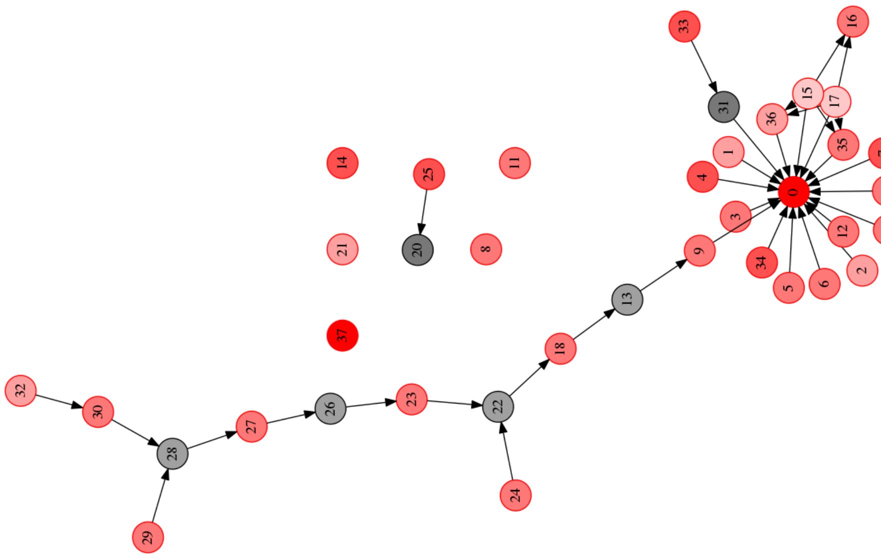

Fig. 3. FoWLDisG instance for Discussion 2.

In this discussion, the root tweet contains only a news headline with only objective information, because the author of the root tweet is a news agency. So, a high number of followers for the author of this root tweet may only indicate that many people is interested in following the information provided by that news agency, but not necessarily that all such people is supporting the information provided by the root tweet. In such cases, it could be more fair to adjust the weight of the root tweet to reflect better its social support.

For the third discussion, the solution contains 21 of the 27 tweets. As in the previous discussion, this high number of tweets in the solution is mainly due to the low number of effective attacks. Figure 4 shows the FoWLDisG instance for this discussion and its solution. Compared with the previous discussion, here the longest path in the discussion graph is slightly smaller (8 tweets), although we also observe some other paths with different lengths.

We next analyze the controversy of the solutions for the FoWLDisG instance of the three example discussions, to further detect differences between their solutions. Table 3 shows the different characteristics we have measured from the solutions for the three discussion graphs. For each discussion graph, we show the size of its solution, its controversy degree and its controversy depth. To facilitate their comparison, for each value we also show in parenthesis their values normalized in the scale $[0, 1]$. That is, we show the size of the solution ($\mid S \mid$) divided by the size of the graph ($\mid T \mid$), the controversy degree divided by the number of tweets that are not in the solution ($\mid T \setminus S \mid$) and the controversy depth divided by the length of the longest controversial path that could be potentially build if there were the appropriate attack edges $((2 \cdot min(\mid S \mid, \mid T \setminus S \mid) + 1)$.

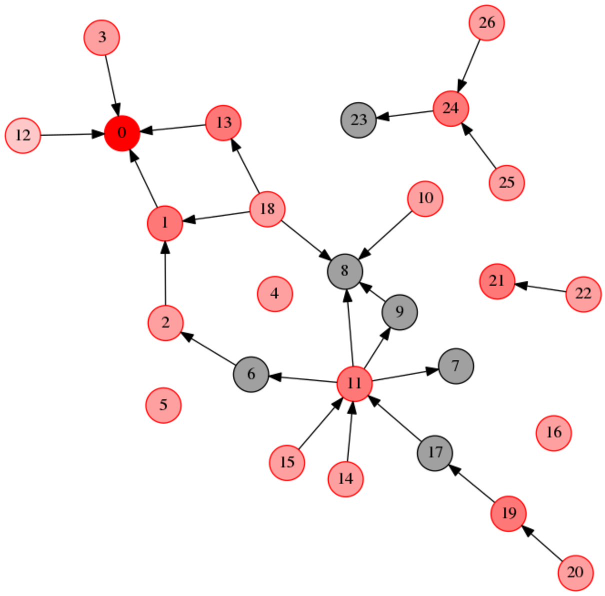

Fig. 4. FoWLDisG instance for Discussion 3.

We observe that the normalized solution sizes are very similar, being only slightly larger the one for the second discussion graph. However, the controversy measures show significant differences that indicate quantitative differences in the characteristics of the solutions. Remind that the first discussion graph has a more *flat* structure, and this is reflexed in a low controversy degree (few tweets attacking the solution) and also in a low controversy depth (low depth in the controversial paths). Remember that it is possible to have a high controversy degree but low controversy depth, so in this case having a low value for both measures indicates that there is almost no controversy between the accepted and rejected tweets and no depth discussions between them. By contrast, the situation in the other two discussion graphs is quite different. The second discussion shows a higher value for both the controversy degree (5) and the controversy depth (9), being both values correlated. Thus, most of the attacks from non-solution tweets to solution tweets belong to the longest controversial path. This situation also occurs in the third discussion, but with lower controversy degree (2) and lower controversy depth (5). In this case all attacks from non-solution tweets belong to the longest controversial path.

6.2 Retweets-weighted Labeled Discussion Graphs

We next analyze the ReWLDisG instances for the discussions. For the first discussion graph, the solution does not change when we consider this weighting scheme, because even if the actual weights are different, the actual effective attacks remain the same. But for the other two discussion graphs the so-

		controversy	
Discussion	solution size	degree	depth
1	20 (0.83)	0 (0)	2 (0.22)
2	32 (0.84)	5 (0.83)	9 (0.69)
3	21 (0.77)	2 (0.33)	5 (0.38)

Table 3
Structural characteristics of the solutions for the FoWLDisG instances.

lutions differ with respect to the FoWLDisG instances. Figure 5 shows the ReWLDisG instance for the second discussion and Figure 6 for the third discussion.

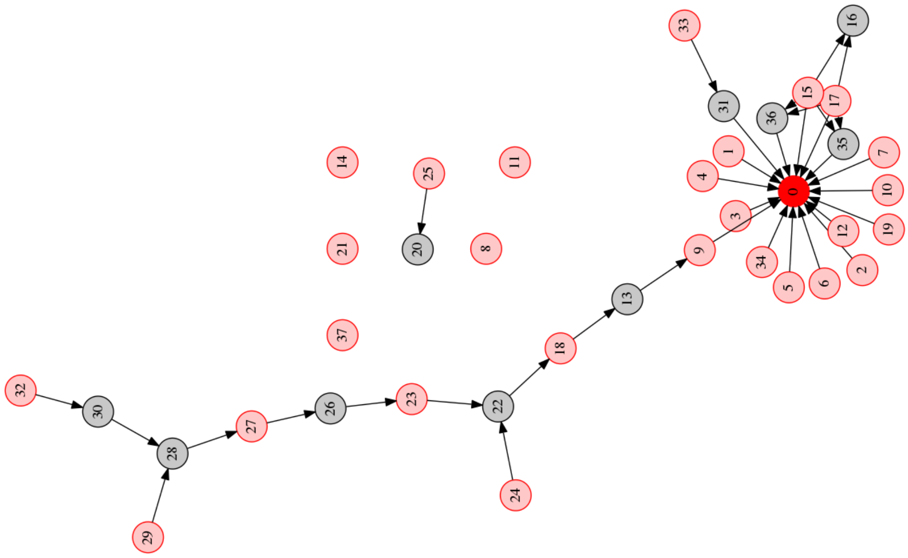

Fig. 5. ReWLDisG instance for Discussion 2.

We next analyze their controversy and the difference in their solutions with respect to the FoWLDisG instances. Table 4 shows their solution size, their controversy measures and also the symmetric difference between the solution for the ReWLDisG instance and the FoWLDisG instance ($| Fa \Delta Re |$).

For the second discussion graph, we have a slight decrease in the (normalized) controversy degree and depth (because we have less tweets in the solution). The symmetric difference contains four tweets, although from these four tweets in the symmetric difference only one appears in a longest controversial path in the FoWLDisG instance but not in a longest one for the ReWLDisG instance. These are the *two* longest controversial paths in the FoWLDisG instance (tweets

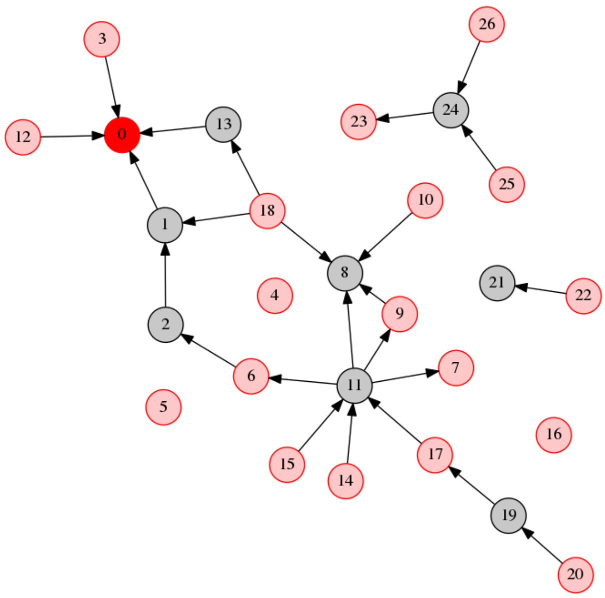

Fig. 6. ReWLDisG instance for Discussion 3.

in the solution are shown in bold):

$$\mathbf{29} \to 28 \to \mathbf{27} \to 26 \to \mathbf{23} \to 22 \to \mathbf{18} \to 13 \to \mathbf{9}$$

$$\mathbf{30} \to 28 \to \mathbf{27} \to 26 \to \mathbf{23} \to 22 \to \mathbf{18} \to 13 \to \mathbf{9}$$

and from these two paths, only the first one remains a longest controversial path in the ReWLDisG instance. So, the small changes in the solutions do not modify the length of the longest controversial path, although decrease slightly the normalized value.

For the third discussion, we observe an increase in the controversy degree although the controversy depth is almost the same as before, indicating *more* controversy but not deeper controversial discussion paths. Regarding the solution, the symmetric difference with the FoWLDisG instance is quite significant (12 tweets). This symmetric difference modifies the actual longest controversial paths, although the length is almost the same. For the FoWLDisG instance the longest controversial path is:

$$\mathbf{19} \to 17 \to \mathbf{11} \to 6 \to \mathbf{2}$$

but for the ReWLDisG instance it is:

$$\mathbf{20} \to 19 \to \mathbf{17} \to 11 \to \mathbf{6} \to 2$$

So, observe that what happens is that, since the tweet 19 is effectively attacked by the tweet 20 in the ReWLDisG instance, the role of the tweets in the path are reversed, so we end up with a controversial path with only one more tweet, but with many different tweets in the solution.

	controversy			
Discussion	solution size	degree	depth	$\mid Re\Delta Fo \mid$
1	20 (0.83)	0 (0)	2 (0.22)	0
2	28 (0.73)	7 (0.70)	9 (0.42)	4
3	19 (0.70)	5 (0.62)	6 (0.35)	12

Table 4
Structural characteristics of the solutions for the ReWLDisG instances.

6.3 Favorite-weighted Labeled Discussion Graphs

Finally, we analyze the FaWLDisG instances for the discussions. In this case, the solutions are the same to the ones of the ReWLDisG instances, so we do not show any figures with the discussion graphs. This is because, although the weighting schemes are based on different social network attributes (the retweets-count and the favorites-count), the computed weights tend to be correlated. That is, if a tweet is marked as favorite by a user it is very likely that it will also be retweeted by that user, although the converse implication is not probably as solid. So, because we work with a logarithmic rescaling of such values, the effective attacks do not change.

The results of the controversy analysis are shown in Table 5, together with the comparison of the solution sizes between these discussion graphs and the previous ones (FoWLDisG and ReWLDisG). Because the solutions are equal to the solutions for the ReWLDisG instances, we obtain the same controversy results of the ReWLDisG instances and the same symmetric differences with the FoWLDisG instances that we have between ReWLDisG and FoWLDisG instances.

	controversy				
Discussion	solution size	degree	depth	$\mid Fa\Delta Fo \mid$	$\mid Fa\Delta Re \mid$
1	20 (0.83)	0 (0)	2 (0.22)	0	0
2	28 (0.73)	7 (0.70)	9 (0.42)	4	0
3	19 (0.70)	5 (0.62)	6 (0.35)	12	0

Table 5
Structural characteristics of the solutions for the FaWLDisG instances.

Overall, we believe that the analysis of these examples, with the three classes of discussion graphs, shows that our framework allows to finely discover subsets of opinions in Twitter that are widely accepted and defended by others, considering different ways to weight the opinions that may be relevant in different social contexts. Also, the controversy analysis gives a way to discover discussions that may be arousing a major division or deep and dynamic discussions between users of the social network. Identifying such discussions may be important towards having tools to help policy makers to pay attention to the more critical topics that concern theirs citizens.

7 Conclusions and future work

In this paper we present a novel approach to apply argumentation reasoning to social network discussions. Although our reasoning system can be used for understanding what are the major accepted or rejected opinions in different domains by social network users, in this work we consider Twitter, one of the most widely used social networks to discuss ideas about many domains.

With the aim to automatically discover such information from Twitter, we develop an analysis system that moves from tweet collections to weighted labeled discussion graphs, where weights represent the social relevance of tweets and labeled edges represent the semantic relation between tweets. In particular we associate weights with tweets from its social relevance taking into account three different attributes of a tweet: the number of followers, the number of retweets and the number of favorites. The automatic labeling system is based on Support Vector Machines and considers different attributes obtained from a tweet as regular and stopwords words, the number of images and the number of URLs mentioned in the tweet, the number of positive and negative emoticons and the sentiment expressed by the tweet. The reasoning model is based on Valued Abstract Argumentation and accepted tweets are those defined by ideal semantics.

We also define measures for controversy analysis with the goal of quantifying the divisions between accepted and rejected tweets. Finally, we analyze three case studies. The results show that our system allows us to discover subsets of opinions in Twitter that are widely accepted and defended by others, under different weighting schemes. The controversy analysis shows that even for discussions with similar solution sizes, as our case studies, the controversy measures can reveal that some discussions may be more critical than others with respect to social division. As far as we know these are the first experimental results related with Twitter discussion analysis based on valued argumentation semantics.

Future work will be addressed in three main directions. First, we will analyze social networks with different characteristics to those of Twitter for which the evaluation of support relations between arguments can be particularly relevant or useful.

We have developed a reasoning system able to solve discussions without restrictions on the structure of their underlying graphs. However, as we have explained before in the paper, there are argumentation algorithms for particular tractable cases of discussions, such those with acyclic discussion graphs. So, given that in our current implementation of the discussion retrieval component we get discussions with this particular structure, our second aim will be to design specialized argumentation algorithms taking as starting point the work of Charwat and Dvorák [15] about bounded tree-width argumentation algorithms.

In our analysis system relations between arguments are crisp in the sense that given a pair of tweets either there is an attack or a support relation, or there is no relation between them. Finally, our future interest will be to extend our system in order to model degrees of attacks between arguments and then to incorporate these degrees into the acceptability semantics. In order to combine weighted arguments with graded attacks we will consider the approach proposed by Dunne [23]. We will also consider graded acceptability semantics where attacks inside the solution could lead to a stratified hierarchy of accepted tweets as proposed in the weighted argumentation framework RP-DeLP where accepted conclusions are rank-ordered sets from stratified knowledge bases.

Acknowledgments

This work has been partially funded by the Spanish MICINN Projects TIN2014-53234-C2-2-R, TIN2015-71799-C2-2-P and ENE2015-64117-C5-1-R. The authors would like to thank anonymous reviewers for providing helpful comments to improve the paper.

References

[1] Teresa Alsinet, Ramón Béjar, Lluis Godo, and Francesc Guitart. RP-DeLP: a weighted defeasible argumentation framework based on a recursive semantics. *J. Log. Comput.*, 26(4):1315–1360, 2016.

[2] Leila Amgoud and Claudette Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *J. Autom. Reasoning*, 29(2):125–169, 2002.

[3] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1–3):197–215, 2002.

[4] Katie Atkinson, Trevor J. M. Bench-Capon, and Peter McBurney. PARMENIDES: facilitating deliberation in democracies. *Artif. Intell. Law*, 14(4):261–275, 2006.

[5] Trevor J. M. Bench-Capon. Value-based argumentation frameworks. In *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning (NMR 2002)*, pages 443–454, 2002.

[6] Trevor J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[7] Trevor J. M. Bench-Capon and Paul E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.

[8] Steven Bird. NLTK: the natural language toolkit. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics, ACL 2006*, pages 17–21, 2006.

[9] Suzan Burton and Alena Soboleva. Interactive or reactive? marketing with Twitter. *Journal of Consumer Marketing*, 28(7):491–499, 2011.

[10] Elena Cabrio and Serena Villata. A natural language bipolar argumentation approach to support users in online debate interactions. *Argument and Computation*, 4(3):209–230, 2013.

[11] Dan Cartwright and Katie Atkinson. Political engagement through tools for argumentation. In *Proceedings of Computational Models of Argument COMMA 2008*, pages 116–127, 2008.

[12] Dan Cartwright and Katie Atkinson. Using computational argumentation to support e-participation. *IEEE Intelligent Systems*, 24(5):42–52, 2009.

[13] Claudette Cayrol and Marie-Christine Lagasquie-Schiex. Graduality in argumentation. *J. Artif. Intell. Res. (JAIR)*, 23:245–297, 2005.

[14] Andrea Ceron, Luigi Curini, Stefano M Iacus, and Giuseppe Porro. Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens? political preferences with an application to Italy and France. *New Media & Society*, 16(2):340–358, 2014.

[15] Günther Charwat and Wolfgang Dvorák. DYNPARTIX 2.0 - dynamic programming argumentation reasoning tool. In *Proceedings of Computational Models of Argument, COMMA 2012*, pages 507–508, 2012.

[16] Shu-Chuan Chu and Yoojung Kim. Determinants of consumer engagement in electronic word-of-mouth (ewom) in social networking sites. *International Journal of Advertising*, 30(1):47–75, 2011.

[17] Phan Minh Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, 77(2):321 – 357, 1995.

[18] Phan Minh Dung, Paolo Mancarella, and Francesca Toni. Computing ideal sceptical argumentation. *Artif. Intell.*, 171(10-15):642–674, 2007.

[19] Phan Minh Dung, Paolo Mancarella, and Francesca Toni. Computing ideal sceptical argumentation. *Artif. Intell.*, 171(10-15):642–674, 2007.

[20] Paul E. Dunne. Computational properties of argument systems satisfying graph-theoretic constraints. *Artif. Intell.*, 171(10-15):701–729, 2007.

[21] Paul E. Dunne. The computational complexity of ideal semantics I: abstract argumentation frameworks. In *Proceedings of Computational Models of Argument, COMMA 2008*, pages 147–158, 2008.

[22] Paul E. Dunne. The computational complexity of ideal semantics. *Artif. Intell.*, 173(18):1559–1591, 2009.

[23] Paul E. Dunne, Anthony Hunter, Peter McBurney, Simon Parsons, and Michael Wooldridge. Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artif. Intell.*, 175(2):457–486, 2011.

[24] Wolfgang Dvorák, Paul E. Dunne, and Stefan Woltran. Parametric properties of ideal semantics. In *IJCAI*, pages 851–856, 2011.

[25] Wolfgang Dvorák, Reinhard Pichler, and Stefan Woltran. Towards fixed-parameter tractable algorithms for abstract argumentation. *Artif. Intell.*, 186:1–37, 2012.

[26] Sinan Egilmez, João Martins, and João Leite. Extending social abstract argumentation with votes on attacks. In *Theory and Applications of Formal Argumentation - Second International Workshop, TAFA 2013, Beijing, China*, pages 16–31, 2013.

[27] Uwe Egly, Sarah Alice Gaggl, and Stefan Woltran. Aspartix: Implementing argumentation frameworks using answer-set programming. In Maria Garcia de la Banda and Enrico Pontelli, editors, *Proceedings of the 24th International Conference on Logic Programming, ICLP 2008, Udine, Italy*, volume 5366 of *Lecture Notes in Computer Science*. Springer, 2008.

[28] Wolfgang Faber and Stefan Woltran. Manifold answer-set programs for meta-reasoning. In *Proceedings of Logic Programming and Nonmonotonic Reasoning, LPNMR 2009*, volume 5753 of *Lecture Notes in Computer Science*, pages 115–128. Springer, 2009.

[29] Alejandro Javier García and Guillermo Ricardo Simari. Defeasible logic programming: An argumentative approach. *TPLP*, 4(1-2):95–138, 2004.

[30] Kathrin Grosse, Carlos Iván Chesñevar, and Ana Gabriela Maguitman. An argument-based approach to mining opinions from Twitter. In *Proceedings of the First International Conference on Agreement Technologies, AT 2012*, volume 918 of *CEUR Workshop Proceedings*, pages 408–422. CEUR-WS.org, 2012.

[31] Kathrin Grosse, María Paula González, Carlos Iván Chesñevar, and Ana Gabriela Maguitman. Integrating argumentation and sentiment analysis for mining opinions from Twitter. *AI Commun.*, 28(3):387–401, 2015.

[32] Lars Kai Hansen, Adam Arvidsson, Finn Arup Nielsen, Elanor Colleoni, and Michael Etter. Good friends, bad news-affect and virality in Twitter. In *International Workshop on Social Computing, Network, and Services (SocialComNet 2011)*, 2011.

[33] Twitter Inc. Twitter about company. `http://web.archive.org/web/20160222014218/https://about.twitter.com/company/`, 2016. [Online; accessed 22-February-2016].

[34] Twitter Inc. Twitter developers. `https://dev.twitter.com//`, 2016. [Online; accessed 22-February-2016].

[35] Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11):2169–2188, 2009.

[36] Milen Kouylekov and Matteo Negri. An open-source package for recognizing textual entailment. In *Proceedings of the ACL 2010 System Demonstrations*, ACLDemos 2010, pages 42–47, 2010.

[37] João Leite and João Martins. Social abstract argumentation. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, IJCAI 2011, pages 2287–2292. AAAI Press, 2011.

[38] Clare Llewellyn, Claire Grover, Jon Oberlander, and Ewan Klein. *Re-using an Argument Corpus to Aid in the Curation of Social Media Collections*, pages 462–468. European Language Resources Association (ELRA), 5 2014.

[39] Paul-Amaury Matt and Francesca Toni. A game-theoretic measure of argument strength for abstract argumentation. In *Proceedings of the 11th European Conference on Logics in Artificial Intelligence, JELIA 2008*, pages 285–297, 2008.

[40] Laura McCarthy and Debra Stock. How travelers use online and social media channels to make hotel-choice decisions. *Cornell Hospitality Reports*, 10(18):6–18, 2010.

[41] Sanjay Modgil. Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, 173(9-10):901–934, 2009.

[42] Finn Arup Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts'*, pages 93–98, 2011.

[43] Henry Prakken and Giovanni Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics*, 7(1):25–75, 1997.

[44] Rob Procter, Farida Vis, and Alex Voss. Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, 16(3):197–214, 2013.

[45] Real Time Statistics Project. Twitter usage statistics. `http://www.internetlivestats.com/twitter-statistics/`, 2016. [Online; accessed 22-February-2016].

[46] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, 1st edition, 2009.

[47] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *International AAAI Conference on Web and Social Media*, 2010.

[48] Daniel Villatoro, Jetzabel Serna, Vctor Rodrguez, and Marc Torrent-Moreno. The tweetbeat of the city: Microblogging used for discovering behavioural patterns during the mwc2012. In Jordi Nin and Daniel Villatoro, editors, *Citizen in Sensor Networks*, volume 7685 of *Lecture Notes in Computer Science*, pages 43–56. Springer Berlin Heidelberg, 2013.

[49] Adrianus Wagemakers. There is a possibility that the quality of Twitter's users is deteriorating. `http://web.archive.org/web/20160222090004/http://www.businessinsider.com/twitter-monthly-active-users-2015-7?r=UK&IR=T/`, 2015. [Online; accessed 22-February-2016].

[50] Maya Wardeh, Adam Wyner, Katie Atkinson, and Trevor J. M. Bench-Capon. Argumentation based tools for policy-making. In *International Conference on Artificial Intelligence and Law, ICAIL '13, Rome, Italy, June 10-14, 2013*, pages 249–250, 2013.

[51] Nigel L. Williams, Alessandro Inversini, Dimitrios Buhalis, and Nicole Ferdinand. Community crosstalk: an exploratory analysis of destination and festival Ewom on Twitter. *Journal of Marketing Management*, 31(9-10):1113–1140, 2015.

[52] Adam Z. Wyner, Tom M. van Engers, and Anthony Hunter. Working on the argument pipeline: Through flow issues between natural language argument, instantiated arguments, and argumentation frameworks. *Argument & Computation*, 7(1):69–89, 2016.

[53] Mimi Zhang, Bernard J. Jansen, and Abdur Chowdhury. Business engagement on Twitter: a path analysis. *Electronic Markets*, 21(3):161–175, 2011.

[54] Michael Zimmer and Nicholas John Proferes. A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66(3):250–261, 2014.