Using Matrix Exponentials for Abstract Argumentation

Carl COREA [a] and Matthias THIMM [a]
[a] *University of Koblenz-Landau, Germany*

Abstract. We investigate the relationship between semantics for formal argumentation and measures from social networking theory. In particular, we consider using *matrix exponentials*, which are measures used for link prediction and recommendation in social networks, as a way to measure acceptability of arguments in abstract argumentation frameworks. We reformulate the approach of matrix exponentials to adhere for the fact that, compared to the social network setting, edges in argumentation frameworks have a negative connotation, arguments linked by edges should not be accepted together, and empirically evaluate this approach on benchmark graphs from ICCMA'15. Moreover, matrix exponentials can also be used for prediction in so-called signed social networks, which have both positive and negative edges denoting friend and foe relationships. As these networks bear a close resemblance to *bipolar* argumentation frameworks, we extend our framework and investigate the applicability of matrix exponentials from signed networks to be used in bipolar argumentation frameworks as well. Finally, we evaluate postulates for ranking-based argumentation semantics for our approach.

Keywords. Abstract Argumentation, Bipolar Argumentation, Ranking Semantics, Network Theory, Matrix Exponential

1. Introduction

In the field of formal argumentation [9,3], one can entail the validity of defeasible claims by the analysis of arguments supporting these claims. Other than actual proofs for claims, arguments are defeasible, meaning that the validity of their conclusions can be challenged by other arguments. Reasoning about a claim therefore does not only adhere to the sole existence of arguments supporting this claim, but is also interconnected to other counterfactual arguments. In order to evaluate which sets of arguments are acceptable, one can represent the conflicting relationships of arguments as a directed graph, constituting an argumentation framework where arguments may be seen as vertices and the attack of one argument to another as a directed edge [9]. There have been major efforts directed towards the development of semantics to derive which arguments are to be accepted. This aim mainly comprises finding subsets of arguments in a framework that are compatible with each other and therefore promote the acceptance of this subset. In other words, defining the argumentation semantics relates to answering which arguments can be jointly accepted. There have been different proposals to this matter—see [3] for an excellent overview—, still there are some behaviors that can be commonly observed in mainstream semantics. For example, following Rahwan et al. [14], we find that most of the classical argumentation semantics agree on the *principle of reinstatement*, i. e., that arguments which are defended by acceptable arguments should be deemed acceptable.

Mathematically speaking, studies in abstract argumentation semantics are concerned with graph-theoretic measures on (directed) graphs. Another discipline that studies the very same mathematical object is (social) network theory [11]. Here, graphs are used to model e. g. social networks where nodes are people and edges between nodes can be interpreted by a *friend* relationship. A particular problem in this area is *link prediction* or *friend recommendation*, i. e., measures that aim at predicting whether a new relationship will be established in the future or to recommend possible friends to the users of an online social network such as Facebook[1] or Twitter[2]. These approaches can analyze the relations of a person to directly related friends and can then calculate friend recommendations. This also means a system can rank the recommendations by a score, which can be seen as the system's certainty of this recommendation. Methodologically, these approaches also bear resemblance to *ranking semantics* [1] for abstract argumentation, which aim at ordering or assigning numerical values quantifying acceptability to arguments. Moreover, some social networks are *signed* [13,12], meaning that both *friend* relationships as well as *foe* relationships are present[3]. Conceptually, these networks are very similar to *bipolar argumentation frameworks* [2,7,6] which allow the representation of both an *attack* between arguments as well as *support*.

We believe that because of the close methodological similarities of the research disciplines of formal argumentation and social networking theory, a thorough investigation of the applicability of the methods from one field to the other may be beneficial. Similar observations have already been made by some other researchers in the field of formal argumentation, see e. g. [4,15,17,8,5]. In this paper, we focus on investigating exponentials of adjacency matrices of graphs—a simple measure for link prediction [13,12]—to be used for abstract argumentation and bipolar argumentation. More precisely, the contributions of this paper are as follows:

- We investigate the concept of matrix exponentials for both abstract and bipolar argumentation and define measures that aim at assessing acceptability of arguments (Section 4).
- We formally compare our approach to *ranking semantics* and investigate its compliance with rationality postulates (also in Section 4).
- We conduct some experiments with our new measures and benchmark graphs from the *First International Competition on Computational Models of Argumentation (ICCMA'15)*[4] in order to obtain some empirical evidence on the hypothesized relationships of abstract argumentation and matrix exponentials (Section 5).

We introduce necessary preliminaries on abstract and bipolar argumentation frameworks in Section 2, preliminaries on network theory in Section 3, and conclude in Section 6.

2. Abstract and Bipolar Argumentation Frameworks

An abstract argumentation framework $\mathfrak{A}$ is defined as a pair $\mathfrak{A} = (Arg, R_{Att})$, where Arg is a finite set of arguments and $R_{Att} \subseteq Arg \times Arg$. For two arguments $A,B \in Arg$, we

[1] `http://facebook.com`
[2] `http://www.twitter.com`
[3] See e. g. `http://www.slashdot.org`
[4] `http://argumentationcompetition.org/`

say that A *attacks* B, iff $(A,B) \in R_{Att}$, which we denote as $A \to B$ in the following. An argument A *defends* C against B, iff $B \to C$ and $A \to B$. So, through $\mathfrak{A}$ we can formalize arguments and their relations.

Let $Att_{\mathfrak{A}}(X)$ for a set X of arguments be the set of its attackers, i. e., $Att_{\mathfrak{A}}(X) = \{y \in Arg \mid \exists x \in X, y \to x\}$. Semantics are given to an argumentation framework $\mathfrak{A} = (Arg, R_{Att})$ through *extensions* [9,3], i. e., sets $S \subseteq Arg$. Some important notions on extensions are as follows:

- $S \subseteq Arg$ is *conflict-free* iff there are no arguments $A, B \in S$, such that $A \to B$.
- $S \subseteq Arg$ *defends* an argument $A \in S$ iff for all $B \notin S$, if $B \to A$ then there exists an argument $C \in S$, such that $C \to B$.

The function $\mathfrak{F} : 2^{Arg} \to 2^{Arg}$ is defined as $\mathfrak{F}(B) = \{A \mid B \textit{ defends } A\}$ and is also called the characteristic function of $\mathfrak{A}$.

- $S \subseteq Arg$ is *admissible* iff S is conflict-free and S defends all of its elements.
- $S \subseteq Arg$ is *preferred* iff S is a maximal set, with respect to set inclusion, among the admissible sets of $\mathfrak{A}$.
- $S \subseteq Arg$ is *stable* iff S is conflict-free and for all $B \notin S$ there exist an argument $A \in S$ such that $A \to B$.
- $S \subseteq Arg$ is *complete* iff it is an admissible set and all acceptable arguments, in respect to S, also belong to S.
- $S \subseteq Arg$ is *grounded* iff S is the least fixed point of $\mathfrak{F}$.

The set of admissible/preferred/stable/complete/grounded extensions establish the *semantics* of an argumentation framework, cf. [9,3]. Note the grounded extension is always uniquely defined while stable extensions may not exist [9].

Bipolar argumentation frameworks [2,7,6] extend abstract argumentation frameworks by introducing an additional relation between argumentations to denote *support*. In the following, we focus on the framework of [7]. Formally, a bipolar argumentation framework $\mathfrak{B}$ is a tuple (Arg, R_{Att}, R_{Supp}), where (Arg, R_{Att}) is an abstract argumentation framework and $R_{Supp} \subseteq Arg \times Arg$ represents the support relation. In addition to the notions for abstract argumentation, for two elements $A, B \in Arg$, we say A supports B iff $(A,B) \in R_{Supp}$, also denoted by $A \rightsquigarrow B$. Notions of acceptability and semantics are extended as follows.

- Given arguments $A, B \in Arg$, a *supported defeat* is a sequence $AR_1 \ldots R_{n-1}B$ (for $n \geq 3$), such that for all i=1...n-2, $R_i \in R_{Supp}$ and $R_{n-1} \in R_{Att}$.
- Given arguments $A, B \in Arg$, an *indirect defeat* is a sequence $AR_1 \ldots R_{n-1}B$ (for $n \geq 3$), such that for all i=2...n-1, $R_i \in R_{Supp}$ and $R_1 \in R_{Att}$.

We say that S set-defeats A, iff there exists an argument $B \in S$, such that there is some directed path from B to A that is a (supported or indirect) defeat. Similarly, we say that S set-supports A, iff there exists a sequence of the form $A_1R_1 \ldots R_{n-1}A$ (for $n \geq 2$), such that all $R_i \in R_{Supp}$ and $A_1 \in S$.

In bipolar argumentation frameworks, the concept of conflict-freeness has to be extended by defining *internal* and *external* conflict-freeness.

- $S \subseteq Arg$ is *internally conflict-free* iff S does not set-defeat any of its elements.
- $S \subseteq Arg$ is *externally conflict-free* iff S does not set-defeat and set-support the same argument.

The above notions allow us to define the two important properties of *conflict-freeness* and *safeness* in bipolar frameworks.

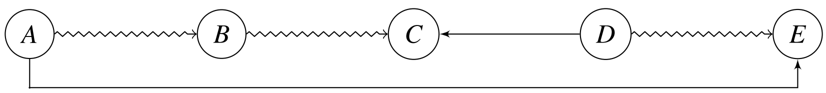


Figure 1. Bipolar argumentation framework from Example 1

- $S \subseteq Arg$ is *conflict-free* iff $\nexists$ A, $B \in S$, such that $\{A\}$ set-defeats B.
- $S \subseteq Arg$ is *safe* iff $\nexists$ $B \in S$, such that S set-defeats and set-supports B.

However, following Cayrol et al. [7], admissibility can be further distinguished into so-called *d-admissibility* and *s-admissibility*.

- *D-admissibility (D in the sense of Dung):* Let $S \subseteq Arg$. S is a d-admissible set, iff S is conflict-free and S defends all its elements.
- *S-admissibility (S in the sense of safe):* Let $S \subseteq Arg$. S is an s-admissible set, iff S is safe and S defends all its elements.

Building on the notions of admissibility above, semantics of bipolar argumentation frameworks are analogously defined as for abstract argumentation frameworks.

Example 1. *Consider the bipolar argumentation framework in Figure 1. There,* A,B,D *is the d-preferred extension, where* A,B *and* D *are the s-preferred extensions respectively.*

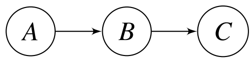
3. Network Theory and Matrix Exponentials

Much attention in the field of network theory has been directed towards investigating social network structures [11,10] to understand the interaction and processes between humans. In this context, graphs are used as a mathematical model of these network structures. For example, a graph could be utilized to depict users through nodes and friendship relations through directed or undirected edges. There, a graph G =(V,E) us usually represented by its adjacency matrix $A \in \{0,1\}^{|V|\times|V|}$. Every matrix component A_{ij} is defined as:

$$A_{ij} = \begin{cases} 1, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

It is important to realize, that the adjacency matrix of a directed graph, such as the graph of an argumentation framework, is not symmetric, as an edge from *i to j* does not imply an edge from *j to i* and vice versa. In this work, we will assume that graphs have directed edges. Note that [17] also makes use of adjacency matrices of argumentation frameworks to characterize classical semantics.

In the abovementioned social network example, edges have a positive connotation as they indicate friendship relations. However, real social structures are also subject to negative effects, e. g. not only friendly but also antagonistic relationships between entities. To model these different relationships, signed networks [13,12] introduce edges, that are annotated with positive or negative signs. Positive edges represent friendship, while negative edges represent antagonism. A signed network G is defined as a triple G=(V, E, σ), where V is a finite set of vertices, E is the set of edges and σ : E $\rightarrow$ {-1,+1}



(i)

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

(ii) Adjacency matrix

$$\begin{bmatrix} 1 & 1 & 0.5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

(iii) Matrix exponential

Figure 2. (i) Friendship graph, (ii) corresponding adjacency matrix, and (iii) exponential scores from Example 2

is a function that assigns a sign to the edges. Given this signed network G, its adjacency matrix $A \in \{\text{-1,0,1}\}^{|V|\times|V|}$ is defined as:

$$A_{ij} = \begin{cases} \sigma(\{i,j\}), & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The adjacency matrices of graphs enable us to do graph analysis by the means of algebraic theory. This allows for approaches like *link prediction* or *recommender systems* [10], which try to analyze the social structure and predict/recommend the creation of new edges to users. An important principle in this field is the so-called *friend-of-a-friend* principle [13], which basically recommends to a user the friends of his own friends. As the adjacency matrix represents the existing friendship relations, i. e. paths of length 1, the simple friend-of-a-friend recommendation can be calculated by A^2, which represents paths of length 2 between users. In practice, the friend-of-a-friend recommendation also considers other users and edges, e. g. users that are connected by longer paths, leading to the *matrix exponential* of A.

Definition 1. *Let* A *be some matrix. The* matrix exponential $exp(A)$ *of* A *is defined as*

$$\exp(\mathrm{A}) = \sum_{i=0}^{\infty} \frac{A^i}{i!} \tag{3}$$

In other words, $exp(A)$ sums paths of any length between users, but weights these by the inverse factorial path length. The result is a $|V|\times|V|$ matrix that contains a so-called *predication* or *recommendation score* (simply called *exponential score* in the following) for connecting to new users. The higher this score, the more likely it is that a new edge will be established in the future.

Example 2. *Figure 2 shows the adjacency matrix of the depicted friendship-graph, as well as the exponential scores computed with the matrix exponential. So we can see for example, that the system recommends* A *to befriend user* C *with a score of 0.5, as* C *is also a friend of* B. *Since Figure 2 shows a directed graph and* C *has no outgoing edges yet, there can be no recommendation made.*

We will discuss the exponential score more closely and investigate its relationship to acceptability of arguments in the next section.

4. Matrix Exponentials for Abstract Argumentation

When comparing the graph representation of argumentation frameworks to the introduced friendship networks, a fundamental difference that can be observed is the different connotations of edges. In friendship networks, edges symbolize a positive friendship relation, whereas the directed edges of argumentation frameworks represent attack relations. A further difference can be identified regarding edges. Where in friendship networks, the path length does not change the semantics of this path, the path length does have to be considered in argumentation frameworks, due to the concept of defense relations. So for example, in an argumentation framework, a path of length 1 has a negative connotation, i. e. attack, but continuing on this path via a further edge changes this to a positive connotation, i. e. a defense. Given that we have defined the semantics of a recommendation score in such a way, that a higher score value is superior to a lower value, when trying to compute a matrix exponential for a graph representing an argumentation framework, we have to integrate the different connotations of path length in order to adhere to the semantics of the recommendation score. As a result of this, paths of odd length should contribute negatively to the exponential score and paths of even length should contribute positively to the exponential score. We therefore represent attacks as negative entries in the adjacency matrix, in order to account for these mentioned factors regarding the semantics of paths in the graph.

Definition 2. *Let $\mathfrak{A} = (Arg, R_{Att})$ be an abstract argumentation framework with $Arg = \{a_1, \ldots, a_n\}$. The matrix $\hat{A}^{\mathfrak{A}} \in \{\text{-}1,0\}^{\,|\,Arg\,|\,\times\,|\,Arg\,|}$ with*

$$\hat{A}^{\mathfrak{A}}_{ij} = \begin{cases} -1, & \textit{if } (a_i, a_j) \in R_{Att} \\ 0, & \textit{otherwise} \end{cases} \tag{4}$$

for all $i, j = 1, \ldots, n$ is called adjusted adjacency matrix *of* $\mathfrak{A}$.

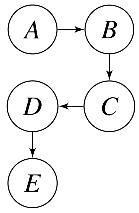
Now we simply apply the matrix exponential to the adjusted adjacency matrix of argumentation frameworks to obtain exponential scores for arguments.

Definition 3. *Let $\mathfrak{A} = (Arg, R_{Att})$ be an abstract argumentation framework with $Arg = \{a_1, \ldots, a_n\}$. The* matrix exponential *$exp(\mathfrak{A})$ is the $|\,Arg\,| \times |\,Arg\,|$ real-valued matrix defined via $exp(\mathfrak{A}) = exp(\hat{A}^{\mathfrak{A}})$. For $a_i, a_j \in Arg$ the entry $exp(\mathfrak{A})_{ij} \in \mathbb{R}$ is called* acceptability assessment of a_j wrt. a_i.

An entry $exp(\mathfrak{A})_{ij}$ is thus the accumulated and weighted sum of paths between a_i and a_j where paths of odd length contribute negatively and paths of even length contribute positively. Thus, the interpretation of a *positive* acceptability assessment of a_j wrt. a_i is that a_i *supports* a_j or "if a_i is accepted then so should a_j". On the other hand, a negative acceptability assessment of a_j wrt. a_i indicates some contradiction between the arguments and "if a_i is accepted then a_j should not be accepted". Furthermore, the i^{th} column in $exp(\mathfrak{A})$ gives an overview on how argument a_i is assessed by all arguments in the framework.

Example 3. *Figure 3 shows (i) an argumentation framework, (ii) its adjusted adjacency matrix, and (iii) its matrix exponential. The acceptability assessment of* B *wrt.* A *is* -1*, the acceptability assessment of* C *wrt.* A *is* 0.5 *and so on. We can observe a correlation between the acceptability assessment and acceptability as proposed by traditional argumentation semantics. An admissible set in this framework is {A,C,E}, and as we can see, the assessments for all pairs of arguments within this set are non-negative. Take the*

(i)



$$\begin{bmatrix} 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

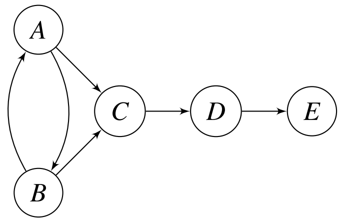
(ii) Adjacency matrix

$$\begin{bmatrix} 1.00 & -1.00 & 0.5000 & -0.1667 & 0.0416 \\ 0.00 & 1.00 & -1.00 & 0.5000 & -0.1667 \\ 0.00 & 0.00 & 1.00 & -1.00 & 0.5000 \\ 0.00 & 0.00 & 0.00 & 1.00 & -1.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

(iii) Exponential acceptability

Figure 3. Abstract argumentation framework (i), corresponding adjacency matrix (ii) and acceptance scores computed with the modified matrix exponential(iii).

(i)



$$\begin{bmatrix} 1.54 & -1.17 & -0.63 & 0.36 & -0.13 \\ -1.17 & 1.54 & -0.63 & 0.36 & -0.13 \\ 0.00 & 0.00 & 1.00 & -1.00 & 0.50 \\ 0.00 & 0.00 & 0.00 & 1.00 & -1.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

(ii) Exponential acceptability

Figure 4. Abstract argumentation framework (i) and acceptance scores (ii).

constellation A→B→C*. From the viewpoint of* A*, the explicit attack to* B *is weighted more significantly than the defense relation from* A *to* C*, resulting in the acceptability assessments* -1 *and* 0.5 *respectively. In our opinion, this resembles characteristics similar to the results observed in the experiments by Rahwan et al. [14]. If argument* A *explicitly attacks* B*, we can be certain* A *does not accept* B*. But in a human context, would this really mean that* A *would also accept* C*? The acceptability assessment of* C *wrt.* A *is a reasonable prediction for the outcome of an argumentation process [14].*

Example 4. *Figure 4 depicts (i) an argumentation framework with cycles and (ii) its matrix exponential. Looking at the acceptability assessments for argument* A *(first column), the fact that* B *is attacked by* A*, as well as* B *also attacking* A*, has increased the assessment that* B *is to be resented. When observing the triangle between* A*,* B *and* C*, we find that* A *attacks* C *but is also involved in a defense relation via* A→B→C*. The acceptability assessment of* C *wrt. A is still negative. This underlines, that explicit attacks against a node are considered more important than a defense relation.*

In order to make use of the assessments in the matrix $exp(\mathfrak{A})$ we now introduce a simple way to aggregate them and obtain a single value for each argument. In a first step, we consider a relative assessment wrt. some given set of arguments.

Definition 4. *Let $E \subseteq Arg = \{a_1, \ldots, a_n\}$ be a set of arguments and $a_j \in Arg$. The* relative acceptance score $\mathsf{score}_E(a_i)$ *of a_i wrt. E is defined via*

$$\mathsf{score}_E(a_j) = \sum_{a_i \in E} exp(\mathfrak{A})_{ij}$$

The value $\mathsf{score}_E(a_j)$ aggregates the acceptability assessments of arguments in E for argument a_j. A positive value $\mathsf{score}_E(a_j)$ indicates that a_j is assessed as acceptable. In particular, if E is any classical extension (such as a preferred extension), we would expect that $\mathsf{score}_E(a_j)$ for every $a_j \in E$ is positive, indicating joint acceptability of arguments within an extension. We will come back to this issue in Section 5.

A more general acceptance score can be defined by considering the accumulated acceptability assessments wrt. to all arguments.

Definition 5. *Let $a_i \in Arg$. The* absolute acceptance score $\mathsf{score}(a_i)$ *of a_i is defined via*

$$\mathsf{score}(a_i) = \mathsf{score}_{Arg}(a_i)$$

The value $\mathsf{score}(a_i)$ indicates the general acceptance of a_i within the argumentation framework. As for the relative acceptance score from above, we will have a closer look at the relationship of this score with classical argumentation semantics in Section 5. Note that the above method for aggregating an acceptance score for an argument is simple summation. Other aggregation methods are imaginable as well but left for future work.

Formally, our approach of acceptance scores resembles *ranking semantics* for argumentation frameworks, which recently have gained some attention in the community, see e. g. the work by Amgoud et al. [1]. In general, a ranking semantics S is a function that orders the arguments of an argumentation framework $\mathfrak{A}$ = *(A,R)* in an order from most acceptable to least acceptable.

Definition 6. *Let $\mathfrak{A} = (Arg, R_{Att})$ be an abstract argumentation framework. Define the relation $S^{\mathfrak{A}}_{exp} \subseteq Arg \times Arg$ through $(a,b) \in S^{\mathfrak{A}}_{exp}$ iff* $\mathsf{score}(a) \geq \mathsf{score}(b)$.

For sets of arguments $X, Y \subseteq Arg$ define furthermore $X \leq_{S^{\mathfrak{A}}_{exp}} Y$ if for all $y \in Y$ there is a $x \in X$ with $(x,y) \in S^{\mathfrak{A}}_{exp}$. Define also $X <_{S^{\mathfrak{A}}_{exp}} Y$ if for all $y \in Y$ there is a $x \in X$ with $(x,y) \in S^{\mathfrak{A}}_{exp}$ and $(y,x) \notin S^{\mathfrak{A}}_{exp}$.

In [1] several rationality postulates have been proposed which should be satisfied by any argumentation semantics based on ranking. Although the motivation of our approach originates in a different field than argumentation theory, it still satisfies many of these postulates as our next result shows.

Theorem 1. *The ranking S_{exp} satisfies the following postulates:*

Abstraction *Let* $\mathfrak{A}$ = (A,R) *and* $\mathfrak{A}$' = (A',R') *be two argumentation frameworks. If* $\mathfrak{A}$ *and* $\mathfrak{A}'$ *are isomorphic, i. e., there is a bijection $f : A \to A'$ with* $\forall$a,b $\in$ A, aRb *iff* f(a)R'f(b)*, then $(a,b) \in S^{\mathfrak{A}}_{exp}$ whenever $(f(a), f(b)) \in S^{\mathfrak{A}'}_{exp}$.*

Independence *For all simply-connected components $\mathfrak{B} = (A', R')$ of $\mathfrak{A} = (A, R)$,* $\forall$a,b $\in$ *A', $(a,b) \in S^{\mathfrak{B}}_{exp}$ whenever $(a,b) \in S^{\mathfrak{A}}_{exp}$.*

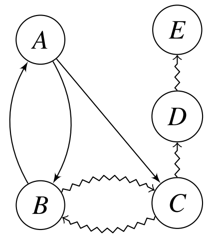
Counter-Transitivity *If $Att_{\mathfrak{A}}(b) \leq_{S^{\mathfrak{A}}_{exp}} Att_{\mathfrak{A}}(a)$ then $(a,b) \in S^{\mathfrak{A}}_{exp}$.*

Strict Counter-Transitivity *If $Att_{\mathfrak{A}}(b) <_{S^{\mathfrak{A}}_{exp}} Att_{\mathfrak{A}}(a)$ then $(a,b) \in S^{\mathfrak{A}}_{exp}$.*

Quality Precedence *If* $\exists$c $\in$ *$Att_{\mathfrak{A}}(b)$ such that* $\forall$d $\in$ *$Att_{\mathfrak{A}}(a)$, $(d,c) \notin S^{\mathfrak{A}}_{exp}$, then $(b,a) \notin S^{\mathfrak{A}}_{exp}$.*

The proof of the above theorem is omitted due to space limitations but straightforward. For a detailed discussion of these postulates see [1].

Before discussing some experiments in Section 5 we will have a look at the case of bipolar argumentation frameworks first. In particular, the matrix exponential can be



(i)

$$\begin{bmatrix} 0 & -1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1.78 & -2.01 & -2.01 & -0.78 & -0.23 \\ -1.41 & 2.38 & 2.01 & 0.78 & 0.23 \\ -0.59 & 1.41 & 1.78 & 1.23 & 1.78 \\ 0.00 & 0.00 & 0.00 & 1.00 & 1.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix}$$

(ii) Adjacency matrix (iii) Exponential acceptability

Figure 5. Bipolar argumentation framework (i), corresponding adjacency matrix (ii) and acceptance scores computed with the matrix exponential(iii).

applied to bipolar argumentation frameworks in more direct way by explicitly utilizing signed networks as in Equation (2).

Definition 7. *Let* $\mathfrak{B}$ *=(*Arg, R$_{Att}$, R$_{Supp}$*) be a bipolar argumentation framework with* $Arg = \{a_1, \ldots, a_n\}$*. The matrix* $\hat{B}^{\mathfrak{B}} \in$ *{-1,0,1}* $^{|\,Arg\,| \times |\,Arg\,|}$ *with*

$$\hat{A}^{\mathfrak{A}}_{ij} = \begin{cases} -1, & \mathit{if}\ (a_i, a_j) \in R_{Att} \\ 1, & \mathit{if}\ (a_i, a_j) \in R_{Supp} \\ 0, & \mathit{otherwise} \end{cases} \tag{5}$$

for all $i, j = 1, \ldots, n$ *is called* adjusted adjacency matrix *of* $\mathfrak{B}$.

Definition 8. *Let* $\mathfrak{B}$ *=(*Arg, R$_{Att}$, R$_{Supp}$*) be a bipolar argumentation framework with* $Arg = \{a_1, \ldots, a_n\}$*. The* matrix exponential $exp(\mathfrak{B})$ *is the* $|\,Arg\,| \times |\,Arg\,|$ *real-valued matrix defined via* $exp(\mathfrak{B}) = exp(\hat{B}^{\mathfrak{B}})$*. For* $a_i, a_j \in Arg$ *the entry* $exp(\mathfrak{B})_{ij} \in \mathbb{R}$ *is called* acceptability assessment of a_j wrt. a_i.

Relative and absolute acceptance scores can be defined for bipolar frameworks in the same way as for classical argumentation frameworks.

Example 5. *Figure 5 shows (i) a bipolar argumentation framework, (ii) its adjacency matrix, and (iii) the corresponding acceptance scores. From the point of view of argument* A*, both* B *and* C *have an acceptability assessment of roughly* -2*, indicating a strong case that these two arguments are not to be accepted by* A*. This low score can be attributed to the explicit attack relations. Furthermore, we see that argument* C *supports* D*, which itself supports* E*. Again from the view of* A*, the acceptability assessments are negative for both* D *and* E*. Yet, the assessment for* E *is slightly higher than for* D*. In our opinion, this showcases the successful implementation of our underlying motivation.*

5. Experimental Evaluation

In this section, we report on some experiments we conducted in order to test the applicability of the relative and absolute acceptance scores introduced above. The experiments have been conducted on graphs from the *First International Competition on Computational Models of Argumentation*[5] (ICCMA'15) [16], which will be described in Sec-

[5] http://argumentationcompetition.org

tion 5.1. We will describe our experiment goals and setup in more detail in Section 5.2 and present our results in Section 5.3.

5.1. Benchmark graphs

Our study has been conducted on all argumentation frameworks from ICCMA'15. However, for all graphs we ignored empty extensions, as the absolute- and relative acceptance scores are only properly defined for extensions with at least one element via Definition 4, respectively Definition 5.

5.2. Experiment Goal and Setup

We computed the absolute- and relative acceptance score for all arguments in all frameworks wrt. complete, preferred, grounded, and stable semantics. If e. g. multiple stable extensions can be defined for a single framework, we computed the scores for all of these extensions and then took the average. So for every framework, we were able to assign a relative- and absolute acceptance score to each argument wrt. one of the mentioned semantics. Our experiment was implemented in GNU Octave. For the largest ICCMA data-sets, computing mentioned scores took under a minute. In our opinion, this shows that it is feasible to compute our proposed semantics in practice.

Regarding relative acceptance scores, i. e. the assessment of an argument from the viewpoint of the respective extension, we can distinguish between the relative acceptance score of extension members and non-extension members. As a result, we aggregated the individual relative acceptance scores into the three subgroups of positive (> 0), neutral ($= 0$) and negative scores (< 0) for these two types of arguments. Furthermore, we computed the average values for the relative- and absolute acceptance scores. Subsequently, the aim of our experiment was to validate these values against the following hypotheses.

Hypothesis 1. *If E is a complete, grounded, stable or preferred extension, the relative acceptance scores of all members of E are strictly positive.*

Hypothesis 2. *If E is a complete, grounded, stable or preferred extension, the relative acceptance scores of all non-members of E are strictly negative.*

Hypothesis 3. *If E is a complete, grounded, stable or preferred extension, the absolute acceptance scores of all members of E are on average higher than the absolute acceptance score of all non-members of E.*

5.3. Results

As a first result, we found that the relative acceptance scores of extension members are strictly positive. This conforms with the definition of admissible extensions as there may be no attacks within extensions. In our opinion, this shows that our approach manifests a ranking-based semantics that does not contradict extension-based argumentation semantics. Figure 6 shows the distribution of the average relative acceptance score of *extension members* under the grounded semantics for all considered graphs. The distributions for the complete, stable and preferred extension are very similar and were therefore omitted. As all relative acceptance scores of extension members are strictly positive, we accept hypothesis 1.

Figure 6 also shows the average distribution of relative acceptance scores for *non*-extension elements. As can be observed, there are some graphs where the relative accep-

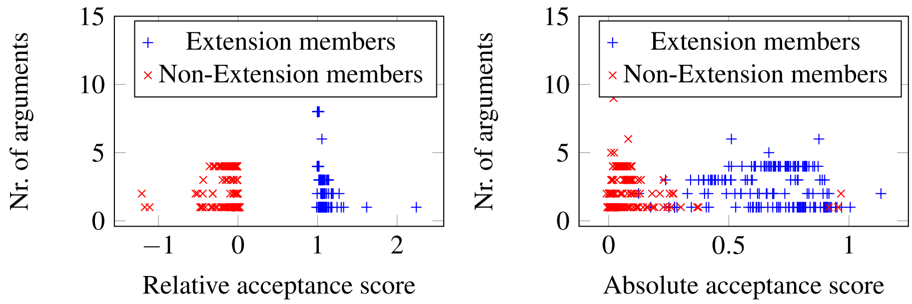


Figure 6. Distribution of acceptance scores of ICCMA graphs wrt. grounded extensions. The x-axis delineates the relative acceptance score (left), respectively the absolute acceptance score (right). The y-axis corresponds to the number of arguments with the respective score.

tance score of non-extension elements is greater than `0`. We therefore reject hypothesis 2. However, it is noticeable that the scores of non-extension members are on average much lower than the scores of extension members. To put this into perspective, the average relative score for the extension members was `1.029`, clearly distinguishable from the average relative acceptance score of non-extension members, which was `-0.103`.

When considering absolute acceptance scores, we incorporate assessments of arguments from the viewpoint of non-extension members. Here, the average absolute acceptance score of extension members was `0.733`. So even when incorporating all non-extension members, i. e. attacks that should lower the individual acceptance scores, the average acceptance score of extension members did not decrease significantly. Regarding non-extension members, the average absolute acceptance score was `0.065`. Non-extension members can form components that conform to each other. These components might not be admissible, still they can promote each other internally. This support will therefore increase the absolute acceptance score. Yet, we can accept hypothesis 3, as we find that admissible arguments wrt. the complete, grounded, preferred or stable extension of all graphs are ranked as more acceptable on average based on our approach.

6. Summary

We have proposed an approach that computes acceptance scores of arguments using matrix exponentials from network theory. We applied this approach to both classical and bipolar argumentation frameworks and investigated its properties. Moreover, we conducted experiments to evaluate the applicability of our approach and its relationships to classical semantics. The empirical evaluation yielded two results. First, the relative acceptance score of extension members is positive in 100% of the cases. This conforms with classical semantics and shows that our approach is able to compute scores that adhere to these semantics. Secondly, for absolute acceptance scores the relative scores from before did not change significantly. In our opinion, this is a further point indicating that the semantics we assigned to the score, namely a higher score relating to a higher acceptability of an argument, is plausible. Figure 6 clearly shows, that the relative- and absolute acceptance scores of extension arguments are on average higher than the respective scores of non-extension arguments.

Our work contributes both from the conceptual as well as the computational point of view to research in computational argumentation. First, our ranking-based semantics gives a new perspective on fine-grained assessments of acceptability due to the use of matrix exponentials and its links to network theory. Second, as the (approximations of) the matrix exponentials are feasible to compute and due to the strong relationships between our semantics and classical semantics one could exploit this for computational purposes and develop a new solver for abstract argumentation based on our approach. We leave this topic for future work.

Further future work could be directed towards investigating comparable scores based on other recommendation measures from network theory. It is also worth noting that it might prove as beneficial to apply abstract argumentation theory to network theory, e. g. to consider whether certain argumentation semantics might be utilized to promote an understanding of groups in social networks.

References

[1] L. Amgoud and J. Ben-Naim. Ranking-based semantics for argumentation frameworks. In *Scalable Uncertainty Management*, pages 134–147. Springer, 2013.

[2] L. Amgoud, C. Cayrol, M.-C. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23(10):1062–1093, 2008.

[3] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *The Knowledge Engineering Review*, 26(4):365–410, 2011.

[4] S. Bistarelli, D. Pirolandi, and F. Santini. Solving weighted argumentation frameworks with soft constraints. In *Proceedings of CSCLP'09*, pages 1–18, Berlin, Heidelberg, 2011.

[5] J. Butterworth and P. Dunne. Spectral techniques in argumentation framework analysis. In *Proceedings of the 6th International Conference of Computational Models of Argument*, 2016.

[6] C. Cayrol and M. Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning*, 54(7):876–899, 2013.

[7] C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *Proceedings of ECSQARU'05*, pages 378–389, 2005.

[8] M. Correia, J. Cruz, and J. Leite. On the efficient implementation of social abstract argumentation. In *Proceedings of ECAI'04*, pages 225–230, 2014.

[9] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial intelligence*, 77(2):321–357, 1995.

[10] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[11] E. Estrada. *The Structure of Complex Networks: Theory and Applications*. Oxford University Press, 2011.

[12] J. Kunegis. Applications of structural balance in signed social networks. *arXiv preprint arXiv:1402.6865*, 2014.

[13] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *Proceedings of SDM'10*, pages 559–559, 2010.

[14] I. Rahwan, J.-F. Bonnefon, M. I. Madakkatel, R. N. Awan, and S. Abdallah. Experiments for assessing floating reinstatement in argument-based reasoning. *Conference of the Cognitive Science Society*, 2010.

[15] Y. Tang, K. Cai, E. Sklar, P. McBurney, and S. Parsons. A system of argumentation for reasoning about trust. In *Proceedings of the 8th European Workshop on Multi-Agent Systems*, 2010.

[16] M. Thimm, S. Villata, F. Cerutti, N. Oren, H. Strass, and M. Vallati. Summary report of the first international competition on computational models of argumentation. *AI Magazine*, 37(1):102–104, 2016.

[17] Y. Xu and C. Cayrol. The matrix approach for abstract argumentation frameworks. In *Proceedings of the 2015 International Workshop on Theory and Applications of Formal Argument (TAFA'15)*, July 2015.