*September 2018*

# Approximations to Truth in Online Comment Networks (Extended Abstract)

Anthony P. YOUNG [a,1], Sagar JOGLEKAR [a], Kiran GARIMELLA [b] and Nishanth SASTRY [a]

[a] *Department of Informatics, King's College London*

[b] *Data Science Laboratory, Ecole Polytechnique Federale de Lausanne*

**Abstract.** We motivate and outline an analysis pipeline used to measure approximations to "truth" in online comment networks. After structuring such networks as bipolar argumentation frameworks, we investigate whether reading only part of the comments will allow the reader to get a "fair" approximation of the "truth" of the network; this is formalised as comparing the grounded extension of what has been read (as an induced sub-framework) with the grounded extension of the entire network (the "truth"). We close by outlining future work, in particular by considering factors would affect the rate of such an approximation to the "truth" with respect to which portion of the comments are read.

**Keywords.** Argument mining, bipolar argumentation frameworks, media and communication studies

## 1. Introduction

Discussions that take place online have effects in the real world, such as having possibly influenced the course of the 2016 US Presidential Election [1]. Indeed, many people now have their opinions about a vast range of topics shaped by reading such discussions online [8]. Websites that host these discussions tend to reveal only a part of the whole discussion according to some policy, such as offering the option to display the comments from most to least liked.[2] Our broader question is *how do such policies influence whether readers can get the "truth" of the discussion?* But we cannot answer this unless we make precise the idea of what it means for readers to get this "truth". We therefore outline an analysis pipeline based in argumentation theory to formalise and measure this quantity of "getting the truth" depending on how many of the comments the reader would have read. Section 2 recaps the appropriate background, Section 3 outlines this pipeline, and Section 4 discusses future work.

## 2. Background

Recall that a **bipolar argumentation framework** (BAF) is a structure $\langle A, R_{att}, R_{sup} \rangle$ where $A$ is the set of arguments, $R_{att} \subseteq A^2$ denotes when two arguments disagree, and

[1] Corresponding Author: Department of Informatics, King's College London, Bush House, Strand Campus, 30 Aldwych, WC2B 4BG, London, United Kingdom. E-mail: peter.young@kcl.ac.uk.

[2] See, for example, the comments in `https://www.bbc.co.uk/news/uk-politics-38996179`, last accessed 15/6/2018.

$R_{sup} \subseteq A^2$ denotes when two arguments agree [3]. There are principled ways to absorb $R_{sup}$ into $R_{att}$ [4], resulting in an **abstract argumentation framework** (AF) $\langle A, R \rangle$, from which we can calculate various **extensions** (normatively winning arguments) [6]. Given an online comment network, each comment can be treated as an argument, and each reply can be classified as an attack and support based on the texts of the comments using various methods (e.g. deep learning [5]); here, support relations are necessary because we are working with natural language rather than logic [9]. This allows us to represent a comment network as a BAF, such that the extensions serve as a notion of the "truth", which guide the reader into what he or she should believe.

## 3. Formalism and Analysis Pipeline

Suppose our comment network has $N \in \mathbb{N}^+$ comments and has been structured into the AF $\langle A, R \rangle$. Suppose a reader reads $0 \le n \le N$ such comments in a given order that is specified by the policy of the platform hosting the comment network. The comment network that has been read is represented by the AF $\langle A', R' \rangle$; this is an induced sub-framework of $\langle A, R \rangle$ w.r.t. the $n$ comments read. As a starting point, we choose to compare the grounded extensions because it always exists, is unique [6], and is tractable [7]. Let $G$ and $G'$ denote the grounded extensions of $\langle A, R \rangle$ and $\langle A', R' \rangle$ respectively. We measure how much this reader "gets" amongst the winning comments with the **Jaccard coefficient**, i.e. $J(G, G') := \frac{|G \cap G'|}{|G \cup G'|} \in [0, 1]$ and $J(\varnothing, \varnothing) := 1$. Given a comment network, we can plot $J$ vs. $n$ to get a qualitative idea of how $J$ varies w.r.t. $n$. Our data analysis pipeline is illustrated as follows:

$$\begin{tikzcd}
\text{Entire comment network} \arrow[r] & \text{BAF} \arrow[r] & \text{AF} \arrow[r] & G \arrow[dr] & \\
 & & & & J(G, G') \text{ given } n \\
\text{Comment network, } n \text{ comments} \arrow[r] & \text{Induced BAF} \arrow[r] & \text{AF'} \arrow[r] & G' \arrow[ur] &
\end{tikzcd}$$

We do not expect $J$ to monotonically increase w.r.t. $n$. For example, in **simple reinstatement** [10], the AF is $A = \{a, b, c\}$, $R = \{(c, b), (b, a)\}$, and $G = \{a, c\}$. Suppose we read the AF in the order of $a$, $b$ then $c$. If $n = 0$, $G' = \varnothing$ so $J = 0$. If $n = 1$, $G' = \{a\}$ so $J = 0.5$. If $n = 2$, $G' = \{b\}$ so $J = 0$. If $n = 3$, $G' = G$ so $J = 1$.

## 4. Conclusions with Future Work

We have outlined a pipeline designed to measure the speed of "getting" the "truth" when reading online comment networks w.r.t. to the number of comments, structured as an AF with "truth" being the grounded extension. As discussed in Section 1, this will allow us to investigate how the various policies the websites hosting such comment networks would affect this speed, such as which of sorting the comments by upvotes or by chronological order is faster. Further, there are many modelling choices that need to be better motivated, such as whether it is fair to treat all comments as arguments when [2] argues that one should not, at least in the context of Twitter, or that how replies which are neither attacks nor supports can be treated. Addressing these problems will be a step towards providing a data-driven understanding of online discourse.

*September 2018*

## References


[1] H. Allcott and M. Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2):211–36, 2017.

[2] T. Bosc, E. Cabrio, and S. Villata. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. In *COMMA*, pages 21–32, 2016.

[3] C. Cayrol and M.-C. Lagasquie-Schiex. On the Acceptability of Arguments in Bipolar Argumentation Frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 378–389. Springer, 2005.

[4] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolarity in Argumentation Graphs: Towards a Better Understanding. *International Journal of Approximate Reasoning*, 54(7):876–899, 2013.

[5] O. Cocarascu and F. Toni. Identifying Attack and Support Argumentative Relations using Deep Learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1374–1379, 2017.

[6] 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.

[7] P. E. Dunne and M. Wooldridge. Complexity of Abstract Argumentation. In *Argumentation in Artificial Intelligence*, pages 85–104. Springer, 2009.

[8] J. Gottfried and E. Shearer. News use across social media platforms 2017, 2017. `http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/`, last accessed 15/6/2018.

[9] S. Modgil. Revisiting Abstract Argumentation Frameworks. In *International Workshop on Theory and Applications of Formal Argumentation*, pages 1–15. Springer, 2013.

[10] I. Rahwan, M. I. Madakkatel, J.-F. Bonnefon, R. N. Awan, and S. Abdallah. Behavioral Experiments for Assessing the Abstract Argumentation Semantics of Reinstatement. *Cognitive Science*, 34(8):1483–1502, 2010.