HAVING THE LAST WORD: UNDERSTANDING HOW TO SAMPLE DISCUSSIONS ONLINE

A PREPRINT

Gioia Boschi  
Department of Mathematics  
King’s College London  
Strand, London WC2R 2LS  
gioia.boschi@kcl.ac.uk

Anthony P. Young  
Department of Informatics  
King’s College London  
Strand, London WC2R 2LS  
peter.young@kcl.ac.uk

Sagar Joglekar  
Department of Informatics  
King’s College London  
Strand, London WC2R 2LS  
sagar.joglekar@kcl.ac.uk

Chiara Cammarota  
Department of Mathematics  
King’s College London  
Strand, London WC2R 2LS  
chiara.cammarota@kcl.ac.uk

Nishanth Sastry  
Department of Informatics  
King’s College London  
Strand, London WC2R 2LS  
nishanth.sastry@kcl.ac.uk

June 11, 2019

ABSTRACT

In online debates, as in offline ones, individual utterances or arguments support or attack each other, leading to some subset of arguments winning (potentially from different sides of the debate). However, online conversations are much larger in scale than offline ones, with often hundreds of thousands of users weighing in, and so, readers are often forced to sample a subset of the arguments or arguments being put forth. Since such sampling is rarely done in a principled manner, users may not get all the relevant “winning” arguments to get a full picture of the debate from a sample. This paper is interested in answering the question of how users should sample online conversations to selectively favour winning arguments.

We apply techniques from argumentation theory and complex networks to build a model that predicts the probabilities of the normatively winning arguments given their location in idealised online discussions. Online discussions are modeled as reply networks where nodes represent the comments exchanged and directed edges represent replies that can either be supporting or attacking. Our model shows that the proportion of replies that are supporting in the network, the network’s in-degree distribution and the locations of unrebutted arguments (“the last words”) all determine the behaviour of the probability that a comment is a winning argument given its location. This is also verified with data scraped from the online debating platform Kialo. In predicting the locations of the winning arguments in reply networks, we can therefore suggest which arguments a reader should sample and read if he or she would like to grasp the winning opinions in such discussions. Our models have important implications for the design of future online discussion platforms.

Keywords  Argumentation Theory, Online Discussions, Probabilistic Analysis, Graph Sampling, Kialo

1 Introduction

1.1 Background and Research Question

The Internet has enabled large scale discussions on all sorts of topics, ranging from important news events to comments on user-posted videos, and they often have consequences in the offline world (e.g. [123456]). Early examples of
We are therefore motivated to ask the following question:

well.

which is hosted on an

As it stands the question is quite broad and our terminology is slightly vague. In this subsection, we present, at high level,

whether it is the portion that interests them or one that is filtered through a recommender system.

Further, comments exchanged over online discussions are often argumentative in nature, where replies to earlier

comments can be attacking or supporting. We are therefore interested in understanding which points being made in the
discussion can be deemed “winning”. The content of those portions shown to a reader may mislead them into thinking
their favoured arguments have won when those arguments actually have not won if one were to read and understand the
entire discussion, especially given that the multiple points of view present that can be attacking or supporting each other.
The presentation of such information will therefore shape people’s points of view and subsequent actions (e.g. [13]),
and pointing out where the winning arguments actually are located in a large discussion can help users negotiate “truth”
in the sense of grasping the prevailing positions.

We are therefore motivated to ask the following question:

Question 1. If one would like to read the “winning” comments from both (or all) sides of a large online discussion,
how should one sample the portion of the discussion to read?

To answer this question, we apply techniques from argumentation theory, data mining and complex networks to devise
a probabilistic model that gives insight into what fragments of a large-scale discussion a user should read if they are to
sample the winning arguments. We then verify the insights of this model on the online debating platform Kialo and
show that the assumptions and predictions of the model are consistent with the structure and the location of winning
arguments in Kialo reply networks.

1.2 Sharpening the Question

As it stands the question is quite broad and our terminology is slightly vague. In this subsection, we present, at high level,
our modelling choices regarding the terms mentioned in the question. Our object of study is the (online) discussion
which is hosted on an (online) discussion platform such as Reddit or Twitter. In a discussion, (Internet) users post
free text comments that should share their opinion or knowledge on some topic that is being discussed. Often such
comments have attributes such as the time of posting, the username of the entity that has posted it, and the number of
likes or dislikes attributed by readers of that comment. In some discussion networks, comments may only reply to the
main news article, such as in BBC News’ Have Your Say (see Footnote 1), or comments can reply to other comments as
well. [3]

Representation: We can formalise online discussions as a directed graph (digraph), which we call a reply network.
The nodes represent comments and contain all relevant attributes, and the directed edges represent the direction of reply
by a comment later in time to a comment earlier in time. Commonly, reply networks are directed trees (i.e., each node
or comment can only be a reply to one previous comment node); we use the term reply trees to refer to such reply
networks, and this paper focuses exclusively on reply trees. The root of a reply tree is (e.g.) the first thesis for debate or
the news article under discussion. The tree grows when the root node receives comments, and those comment nodes
may receive further comments, perpetuating a potentially infinite reply hierarchy. The leaves of a reply tree are the
comments that have not received any replies yet. [4] The depth of a reply network is the length of the longest path from
root to leaf, i.e., from the first thesis of the debate to the most nested reply; the allowed depth is a consequence of the
user interface (UI) of the discussion platform. In this terminology, depth zero reply networks are trivial as they consist
of one (e.g.) news article that disallows comments. Depth one reply networks, such as BBC News’ Have Your Say,

---

1 For example, the BBC News article reporting on former United Kingdom (UK) Prime Minister Tony Blair’s thoughts on Brexit
https://www.bbc.co.uk/news/uk-politics-38996179 (last accessed 7/4/2019) has attracted over $10^8$ comments. More
recent work concerning how UK Members of Parliament (MPs) and citizens engage with each other on Twitter has shown that the
579 MPs on Twitter have a total of 12.83 million followers that can reply to any Tweet posted by the MPs [12].

2We define terms using italics. Words in brackets are optional. In this case, for example, “online discussion” and “discussion” are
used interchangeably in this paper.

3 For example, consider the following news article from the Daily Mail, a UK newspaper: https://www.dailymail.co
.uk/news/article-6894539/Meghan-Markle-snubs-Queens-doctors-birth-doesnt-want-men-suits.html (last ac-
cessed 10/4/2019). This has attracted over $9 \times 10^7$ comments, and allows for comments to reply to other comments.

4 Graph-theoretically the root is a sink node (i.e. with zero out-degree) and the leaves are source nodes (i.e. with zero in-degree),
which is the opposite convention to how the terms “root” and “leaf” are used for directed trees.
allow for one reply, to the news article only (i.e., comments may not reply to another comment). Many reply networks such as those obtained from Twitter, Reddit or Kialo (see Section 1.3) allow for a more interesting nesting structure by permitting replies of arbitrary depth; we mainly focus on such discussions in this paper. Given a reply network, one can treat any node as the root of a new reply network, comprising of all replies to that node and all replies to such replies.

Example 1. Consider the following fragment of the discussion concerning the Daily Mail news article referenced in Footnote 3, where we only show the text of each comment and abbreviate each comment with the symbol $a_i$, for $i \in \{0, 1, 2\}$. In this paper, the symbol $:=$ is read (the right symbol is) defined as (the left information).

1. $a_0 :=$ The news article with headline Meghan Markle ‘snubs the Queen’s doctors for her birth because she doesn’t want “men in suits” delivering her baby’ and its content text.
2. $a_1 :=$ A fantastic role model for independent women
3. $a_2 :=$ Are you kidding me!

The reply network is represented in Figure 1.1

![Figure 1.1: The Reply Network from Example 1](image)

In Figure 1.1, the root node is $a_0$ and the (only) leaf node is $a_2$.

Sampling: By sampling the discussion, we mean selecting a subgraph of the corresponding reply network and treating this subgraph as its own reply network. In general, this subgraph may not be a full subgraph or even a connected subgraph. The available ways to sample a reply network is also affected by the UI. For example, Twitter uses hashtags to display comments regarding similar topics, and Disqus-based discussion networks allow for the user to sort the comments by depth first search according to the time of posting or from most liked to least liked. One can then sample by, e.g. reading the first $n$ consecutive comments from the top, where $n$ is much less than the total number of comments, or sample in some other suitable way. Understanding such sorting or comment-presentation mechanisms can help users (ranging from casual readers to active debaters) better understand what has already been said, and what to read in order to make most sense of the content and make better informed decisions.

Example 2. (Example 1 continued) One sample of the reply network in Figure 1.1 is the full subgraph from the nodes $a_0$ and $a_1$ only, represented in Figure 1.2

![Figure 1.2: The Reply Network from Example 2](image)

This means the user would be shown the news article (node $a_0$), and the single comment (node $a_1$).

Winning Comments: What does it mean to say that a comment has “won” in a discussion? In this paper, we apply ideas from argumentation theory, a branch of artificial intelligence (AI) that is concerned with the rational and transparent resolution of conflicting claims (e.g. [15]). We will represent reply networks as bipolar argumentation frameworks (BAFs) [16], where each comment is an argument and each reply edge is either an attack or a support. There are

---

6 Graph-theoretically this is the full subgraph induced by a given node and its predecessor set (“following the arrows backwards”) from that node.
6 e.g. $X := \{a, b, c\}$ means the symbol $X$ is defined as the set $\{a, b, c\}$. We may occasionally use $=:$, where definitions go the other way, e.g. $17 := d$ means the variable $d$ is defined to be the quantity 17.
6 e.g. See [https://www.quantamagazine.org/mapping-the-brain-to-build-better-machines-20160406/](https://www.quantamagazine.org/mapping-the-brain-to-build-better-machines-20160406/) last accessed 7/4/2019.
8 These are technical terms that will be defined in Sections 2.1 and 2.2. Intuitively, arguments are abstracted as entities that can attack or support other arguments. We treat each node in a reply tree as a free-standing argument, with the “reply to” relation inducing attack or support relations between arguments.
principled ways of combining supports into attacks such that we extract an argumentation framework \[17\] from each BAF; this enables one to define \textit{normatively} what the winning arguments are from the attacks. We will recap the formal definitions of “winning” in Section \[2.2\] Intuitively, an argument is \textit{winning} if either it is unrebutted (a leaf node in a reply tree, see Footnote \[4\]), or that every other argument that attacks it is losing. An argument is \textit{losing} if it is attacked by some winning argument.\[14\] This paper focuses on reply trees, and a result from argumentation theory states that there is always a unique, non-empty set of winning comments \[12\] Theorem 30\]. The winning comments can express a range of viewpoints (potentially from more than one side of a debate) as long as they are not fundamentally contradictory.

\textbf{Example 3.} (Example \[7\] continued) We can represent the reply network as a BAF where, judging from the text of each comment, \(a_1\) supports \(a_0\) and \(a_2\) attacks \(a_1\). From \[16\] (see Section \[2.2\] below), this means \(a_2\) attacks \(a_0\) as well. As \(a_2\) is unrebutted because nothing attacks it, \(a_2\) wins and both \(a_1\) and \(a_0\) are attacked by a winning argument, and hence are losing. The corresponding BAF is represented in Figure \[1.3\].

\begin{figure}[h]
  \centering
  \includegraphics[width=0.5\textwidth]{example3_baf.png}
  \caption{The BAF from Example 3. We adopt the convention that green (dotted) edges denote supporting edges, and red (solid) edges denote attacking edges.}
\end{figure}

In light of this background from Argumentation Theory, we can sharpen our question to:

\textbf{Question 2.} If one would like to read the normatively winning arguments in a reply network representing a large online discussion, as modelled by argumentation theory, then how should one sample the reply network?

This paper seeks to provide a probabilistic answer to the above question and compare it to data from an online discussion platform.

\subsection*{1.3 Assumptions behind Argumentation and Our Choice of Kialo}

In this subsection, we make explicit some assumptions from Section \[1.2\] in order to use argumentation theory for modelling online discussions, and how these assumptions can be problematic in general online discussion platforms. We will see how our choice of dataset will address these assumptions.

We use data from discussions hosted on \textit{Kialo}, an online debating platform.\[9\] Kialo discussions are well moderated, allowing us to freely assume that each reply is actually making a coherent and relevant point in relation to the argument that it is replying to; such high standards of moderation differentiates Kialo from other platforms such as Twitter and Reddit. Assuming that replies are relevant allows us to straightforwardly induce a reply tree from the structure of the Kialo discussion: each “reply” action induces an attack or support relation (as in Figure \[1.3\]) between two users’ statements within the discussion. Once a reply tree/network has been induced, we can then apply powerful results from argumentation theory (Section \[2.1\]). Furthermore, users taking part in a Kialo debate are asked to classify their arguments as attacking or supporting. Thus the reply networks induced are automatically bipolar argumentation frameworks.

It is easy to see how our assumptions may become invalid when the discussion is not moderated: For instance, trolls may seek to turn the discussion into an abusive exchange (e.g. \[19\]) and sidetrack a discussion away from its main topic, or bots could jump in and hijack a conversation to distribute spam (e.g. \[21\]). Note however, that there are automated methods to recognise relevance and the nature of the directional relations between pieces of text (e.g. using textual entailment \[22\] \[23\], sentiment shifts via psycholinguistics \[24\] \[25\] or novel machine learning architectures \[26\] \[27\]). Such methods can be used to apply bipolar argumentation theory to non-moderated or very noisy discussions.

In summary, our use of Kialo as a dataset is motivated by its clean moderated structure as well as the classification of arguments into “supporting” and “attacking”. These features allows us to abstract away from the content of the arguments and examine the relationships between different arguments. In other words, we are able to ignore the relatively difficult task of \textit{mining} arguments, and use real world data to support our theoretical findings.

\[9\] An argument can be \textit{undecided} if it is neither winning nor losing \[18\], but Section \[2.1\] will explain why we do not need to consider undecided arguments.

\[10\] The left sub-figure is taken from https://stackoverflow.com/questions/49854754/kialo-how-can-i-view-the-argument-topology-map-after-i-have-entered-an-argument last accessed 14/4/2019. The right sub-figure is taken from http://mycareacademy.org/all/a-new-digital-debating-tool-for-collaborators-kialo/ last accessed 15/4/2019.
Figure 1.4: Example of a Kialo discussion. In this example, the thesis states “The Earth is round”; it is represented by the grey rectangle on the right diagram and the grey, filled-in circle on the left diagram. In the left diagram, the concentric annuli represent the replies made at depth 1, depth 2... etc, where each slice of the annulus denotes a single replying comment. Each slice at a given depth is attached to none or many slices at the next depth, denoting the number of comments replying to that initial comment slice. Further, in each annulus, replies are coloured green if they are pro and red if they are con, and the shade of the colouring represents the likelihood, with higher likelihood denoted by a darker shade of red or green (or grey for the thesis). In the right diagram, we have two example replies highlighted: a pro reply at depth 2 and a con reply at depth 3. Each is one among several replies (denoted by the smaller red or green rectangles) given their depth.

1.4 Contributions of this Paper

Recall from Section 1.2 that our Question 2 is: If one would like to read the normatively winning arguments in a reply network representing a large online discussion, as modelled by argumentation theory, then how should one sample the reply network? This paper applies techniques from complex networks to articulate a probabilistic model that predicts the probability that an argument at a given distance from the root is winning. Specifically, it makes the following contributions:

1. In the case where the in-degree distribution of the reply tree is homogeneous, we identify three regimes of behaviour characterised by the probability $q \in [0, 1]$ of a replying edge to be a support:
   
   (a) If $q = \frac{1}{2}$, then the probability of an argument winning does not depend on the distance from the root.
   
   (b) If $q < \frac{1}{2}$, we get an oscillatory behaviour in the probability of a node winning, depending on its distance from the root.
   
   (c) If $q > \frac{1}{2}$, the probability of a node winning decreases monotonically the closer one approaches the root of the reply tree.

Since winning arguments are located in different parts of the reply tree in these three distinct regimes, online discussion platforms and their user interfaces can estimate the supportiveness of the discussion (i.e., $q$), and thereby guide users to parts of the tree that are richer in winning arguments.

2. In the case of scale-free graphs, simulations of ensembles of bipolar argumentation frameworks indicate that the best way to sample the graphs is from the arguments that are located farther from the root to the comments closer to it.

3. We then apply these insights to Kialo data. We find that Kialo reply trees are well-approximated by scale-free trees when the proportion of supporting replies in the reply tree is medium or high ($q > 0.4$) and by homogeneous in-degree graphs when the discussion is very aggressive ($q < 0.2$). This allows us to apply the model’s results to indicate to the user where he or she should read in the Kialo reply tree.

Overall, we find that across the models we consider, as well as in the empirical data from Kialo, the leaves of a discussion literally “have the last word”, i.e., unrebutted arguments at the leaves of reply trees have an enormous influence on the winning arguments: leaves win by default and thereby influence which other arguments win deeper in
We show that even a conservative position of simply not accepting arguments as winning until they have been supported or attacked by at least one other argument (i.e., only considering non-leaf nodes in the whole reply tree) is not sufficient. We then suggest new methods for calculating winning arguments, such that their effect is dampened.

To the best of the authors’ knowledge, this work is the first to combine argumentation theory with complex networks to analyse online discussions. While previous work has used argumentation theory to understand online discussions (e.g. [28] and [29]), previous research has mainly focused on mining arguments from natural language expressions. We are motivated by the complementary question of understanding where winning arguments might be, once they have been mined, and suggesting to the user where to look (sample) for such winning arguments.

Our contributions are theoretical in nature, but also provide important insights for the future design of platforms for online discussions, by answering Question [2] The “takeaway” for a system / UI designer is that a user should sample the reply network at the appropriate distances from the root where the probability of an argument winning is highest (this sampling can potentially supported by the platform or its UI, but can also be done manually by an interested and committed user). The sampling probability is calculated by our model, and depends on factors such as the in-degree distribution of the reply network and the proportion of replies that are supports. The good agreement between synthetic and real data reveals the appropriateness of the use of our probabilistic approach to answer a question that could in general depend on multiple factors of complex human behaviour.

1.5 Structure of this Paper

In Section 2, we summarise the concepts from argumentation theory needed to understand this paper and review related work in argumentation theory and social media analytics. In Section 3 we explain how we have scraped and cleaned our dataset of Kialo discussions for this work. In Section 4 we present our contributions. In Section 5 we articulate methods of dampening the effects of the leaf arguments. In Section 6 we outline some limitations of our analysis and conclude with future work.

2 Technical Background and Related Work

In this section we review the relevant technical background in argumentation theory and related previous research in social media analytics and argumentation-theoretic analyses of social media.

2.1 Abstract Argumentation Theory

Argumentation theory is the branch of AI concerned with the rational and transparent resolution of conflicting claims. Argumentation theory has roots in philosophy and linguistics, where its traditional emphasis has been on understanding what makes individual arguments deductively valid. However, the seminal paper [17] presented argumentation as concerned with how one should reason about different arguments that are individually compelling but draw opposing conclusions. The resulting theory is called abstract argumentation, as it abstracts away from the content and structure of each argument, as well as the nature of their disagreements. This allows one to represent arguments and their disagreements as nodes and directed edges of a digraph, and conflicts are resolved through an appropriate choice of subset of nodes that satisfy some normative requirements for what it means for arguments to “win” a debate. Its success has since been applied to multi-agent systems, legal informatics (see chapters in [15]) and medical informatics (e.g. [30]).

More formally, let $A$ be a set of arguments under consideration, where by treating each argument as a point means we have abstracted away from each argument’s structure and content. Let $R \subseteq A^2$ be the attack relation, such that for $a, b \in A$, $(a, b) \in R$ means the argument $a$ attacks or disagrees with the argument $b$. The digraph $\langle A, R \rangle$ is called an argumentation framework (AF).

Example 4. Consider the AF with $A = \{a, b, c\}$ and $R = \{(a, b), (b, a), (b, c)\}$. This is depicted in Figure 2.1.

---

12 The formal definition of what it means for an argument to win is recapped in Section 2.1.
13 This section follows the notation and terminology of [31].
Further, there are four basic definitions of what it means for a set of arguments to be winning, collectively referred to as the Dung semantics after their inventor [17]. More precisely, given an argumentation framework \( \langle A, R \rangle \), let \( S \subseteq A \) and \( a, b \in A \). We say \( S \) is conflict-free iff no two arguments in \( S \) attack each other. We say \( S \) defends or reinstates an argument \( a \) iff every attacker of \( a \) is in turn attacked by some argument in \( S \). We say \( S \) is self-defending iff for all \( a \in S \), \( a \) is defended by \( S \). We say \( S \) is admissible iff \( S \) is conflict-free and self-defending.

**Example 5.** (Example 4 continued) The set of arguments \( \{ a, c \} \) is conflict-free. As \( b \) attacks \( c \), \( \{ b, c \} \) is not conflict-free. Further, \( \{ b \} \) defends \( b \) and therefore \( \{ b \} \) is self-defending. By symmetry, \( \{ a \} \) is also self-defending. The sets \( \{ a \} \), \( \{ b \} \) and \( \{ a, c \} \) are admissible.

There are four basic definitions of what it means for a set of arguments to be winning, collectively referred to as the Dung semantics after their inventor [17].

- We say \( S \) is a complete extension iff it is admissible and all arguments that are defended by \( S \) also belong to \( S \); intuitively one should believe everything he or she can defend.
- We say \( S \) is a preferred extension iff it is a \( \subseteq \) -maximal complete extension; intuitively, one should believe as much as it is consistently possible to believe.
- We say \( S \) is the grounded extension, denoted \( G \), iff it is the \( \subseteq \) -least complete extension; intuitively, one should adopt a sceptical stance by only believing that which can be defended by unattacked arguments.
- Lastly, we say \( S \) is a stable extension iff it is a preferred extension that attacks all arguments outside of it; intuitively, stable extensions adopt the extreme position of “you’re either with us or against us”.

**Example 6.** (Example 5 continued) The grounded extension \( G \) is \( \emptyset \). The set of preferred extensions \( \text{PREF} = \{ \{ b \}, \{ a, c \} \} \subseteq \mathcal{P}(A) \). The preferred extensions are also stable because \( b \) attacks both \( a \) and \( c \), while the set \( \{ a, c \} \) attacks \( b \) via \( a \).

It follows from these definitions that the set of unattacked arguments, \( U \subseteq A \), is always contained in the grounded extension \( G \) [31 Corollary 7.45] [18]. As \( G \) is the \( \subseteq \) -smallest complete extension, \( U \) is contained in all complete extensions, and hence all preferred and stable extensions. This captures the principle from everyday reasoning that claims are usually taken to be true unless there are good reasons to believe otherwise, such as there being an explicit rebuttal to the claim. This is how argumentation theory formalises the idea that arguments having the last word should win.

Given that there are multiple solution concepts, are there situations where all of the Dung semantics coincide, therefore giving a single, well-defined notion of “winning”? Yes, as [17 Theorem 30] gives a sufficient condition on the topology of \( \langle A, R \rangle \) for this to be the case. A consequence of [17 Theorem 30] for this paper is that if \( \langle A, R \rangle \) is finite and acyclic, then all four Dung semantics coincide; this is certainly the case with AFs that model reply networks whose underlying digraph structure is a tree.

**Example 7.** (Example 6 continued) One can see from Figure 2.1 that there is a two-cycle involving the arguments \( a \) and \( b \). Therefore, it is possible that there are multiple sets of winning arguments. In this case, we have two preferred extensions.

---

14 Graph-theoretically, conflict-free sets are independent sets.

15 Vacuously, \( \emptyset \subseteq A \) is always admissible.

16 It follows that if every argument in \( \langle A, R \rangle \) is attacked (i.e. there are no source nodes), the grounded extension is empty - see [31 Corollary 6.9]. Further, as every set, the grounded extension exists and is unique for all argumentation frameworks.

17 If \( X \) is a set then \( \mathcal{P}(X) \) is its power set.

18 This is trivially the case if \( U = \emptyset \).

19 The converse to [17 Theorem 30] is not in general true, i.e. a finite AF can have cycles but still have one set of arguments that is winning, see, e.g. [31 Theorem 8.12].
Therefore, for AFs that are trees, we can focus on calculating the grounded extension $G$ - this is our set of winning arguments, and this calculation can be done in polynomial time [32]. Further, $G$ is stable, and hence for all arguments $a \notin G$, $a$ is attacked by some argument in $G$ and hence is losing, while all $b \in G$ are winning. Therefore, arguments are either winning or losing, and there are no undecided arguments [18].

The algorithm that calculates the winning arguments is as follows. Upon input of $\langle A, R \rangle$ it defines two sets $in, out \subseteq A$ such that all winning arguments will be placed in $in$, and all losing (i.e. not winning) arguments will be placed in $out$. The algorithm first places all unattacked arguments in $in$, all arguments attacked by the unattacked arguments in $out$, then restricts to the full subgraph of $\langle A, R \rangle$ induced by the nodes $A - out$. This creates a new set of unattacked arguments which are added to $in$ and a new set of arguments attacked by the unattacked arguments which are added to $out$. This is repeated until the assignment to $out$ is completed and all remaining arguments of the appropriate full subgraph is assigned to $in$. This set is returned as the grounded extension.

More formally, the algorithm is as follows [33]:

Algorithm 1

\begin{algorithm}
\begin{algorithmic}[1]
\Function{GROUNDEDEXTENSION}{$\langle A, R \rangle$}
\State $in \leftarrow \emptyset$
\State $out \leftarrow \emptyset$
\While{$in \neq A$}
\State $in \leftarrow \{a \in A \mid (\forall b \in A) \ (b, a) \notin R\}$
\State $out \leftarrow \{a \in A \mid (\exists b \in in) \ (b, a) \in R\}$
\State $A \leftarrow A - out$
\State $R \leftarrow R \cap A^2$
\EndWhile
\Return $A$
\EndFunction
\end{algorithmic}
\end{algorithm}

Example 8. Consider an argumentation framework with arguments $A = \{a, b, c, d, e\}$ and attacks $R = \{(e, b), (d, c), (c, b), (b, a)\}$. We can picture this in Figure 2.2

![Figure 2.2: The argumentation framework from Example 8](image)

Applying Algorithm\[7\] we have initial assignments $in = \emptyset$, $out = \emptyset$, $A$ and $R$.

1. On the first iteration of the while loop, $in = \{d, e\}$, $out = \{c, b\}$, $A = \{a, d, e\}$ and $R = \emptyset$.
2. On the second iteration of the while loop, $in = \{a, d, e\}$, $out = \emptyset$, $A = \{a, d, e\}$ and $R = \emptyset$.
3. The while loop does not have a third iteration as $in = A = \{a, d, e\}$.

Therefore, the grounded extension is $\{a, d, e\}$, which are the winning arguments. The set of losing arguments is $\{b, c\}$.

Now suppose $e$ were absent from this example. Notice from Figure 2.2 (without $e$) how the fact that $d$ is unattacked and therefore winning by default implies that $c$ is losing, $b$ is winning and $a$ is losing. Indeed, $d$ reinstates $b$ because it attacks all of $b$’s attackers, in this case $c$.

In summary, abstract argumentation theory represents arguments and their disagreements as the nodes and directed edges of a digraph. There are general ways to calculate a set of winning arguments, which satisfy the normative requirements that no arguments within such a set attack each other, and the set of arguments counter-attack all attacks from arguments not in the set. In the case where the underlying digraph is a tree, there is a straightforward algorithm to calculate the “winning” arguments (Algorithm\[1\]), where intuitively all arguments that are unattacked are winning, all arguments attacked by some winning arguments are losing, and all arguments that are attacked by only losing arguments are winning.
2.2 Bipolar Argumentation Theory

One criticism of abstract argumentation is that it assumes the only way arguments can interact with each other is through disagreement. A more flexible model should allow for more types of interactions, and the simplest such interaction to include is to allow for arguments to agree with and hence support each other; this is bipolar argumentation theory \cite{16}.

Formally, a bipolar argumentation framework (BAF) is a structure \( \langle A, R_{sup}, R_{att} \rangle \), where \( A \) is our set of arguments and \( R_{sup}, R_{att} \subseteq A^2 \) are two binary relations on \( A \) that respectively represent supporting replies and attacking replies, i.e. for \( a, b \in A \), \( (a, b) \in R_{sup} \) means \( a \) supports (agrees with) \( b \), and \( (a, b) \in R_{att} \) means \( a \) attacks (disagrees with) \( b \).

We require that \( R_{sup} \cap R_{att} = \emptyset \). One can therefore think of \( \langle A, R_{sup}, R_{att} \rangle \) as a directed graph (digraph) where each edge has one of two colours. Example 9 (page 3) can be modelled as a BAF.

Given a BAF, how can one determine what the winning arguments are? If \( R_{sup} = \emptyset \), then the BAF reduces to an AF and the winning arguments are calculated as described in Section 2.1. However, as \( R_s \neq \emptyset \) in general, we need principled ways to incorporate supporting edges into attacking edges before we can invoke the results summarised in Section 2.1. Following \cite{16}, there are two ways of absorbing supports into attacks. Given a BAF \( \langle A, R_{sup}, R_{att} \rangle \) and \( a, b \in A \), we say a support path is a path in the underlying digraph that only traverses support edges. Let \( a \rightarrow_{sup} b \) denote that there exists a support path from \( a \) to \( b \). We say a support-defeats \( b \) iff \( (\exists c \in A) \left[ a \rightarrow_{sup} c, R_{att}(c, b) \right] \), and \( a \) indirectly defeats \( b \) iff \( (\exists c \in A) \left[ R_{att}(a, c), c \rightarrow_{sup} b \right] \). We define the argumentation framework of \( \langle A, R_{sup}, R_{att} \rangle \) to be the digraph \( \langle A, R \rangle \) where \( (a, b) \in R \) iff either \( a \) support-defeats \( b \) or \( a \) indirectly defeats \( b \). As usual, we say \( a \) attacks \( b \) iff \( (a, b) \in R \).

Once we have transformed \( \langle A, R_{sup}, R_{att} \rangle \) to \( \langle A, R \rangle \), we can apply the usual calculations to find the sets of winning arguments. Further, if the underlying digraph of \( \langle A, R_{sup}, R_{att} \rangle \) is a tree, then \( \langle A, R \rangle \) will also be a tree.

Example 9. Consider a BAF with arguments \( A = \{a, b, c, d, e\} \), attacks \( R_{att} = \{(c, b)\} \) and supports \( R_{sup} = \{(d, c), (e, b), (b, a)\} \). We can picture this in Figure 2.3 where following Example 3 support edges are coloured green and attacking edges are coloured red.

Figure 2.3: The bipolar argumentation framework from Example 9 where green (dotted) edges denote supports and red (solid) edges denote attacks

The support paths of length 1 in this BAF are \((e, b), (b, a)\) and \((d, c)\), and the support paths of length 2 in this BAF are \((e, b, a)\). Therefore, \( e \rightarrow_{sup} b \rightarrow_{sup} a \), \( d \rightarrow_{sup} c \) and \( e \rightarrow_{sup} a \). As \( c \) attacks \( b \), we can see that \( c \) indirectly defeats \( a \), and \( d \) support-defeats \( b \). Therefore, the corresponding argumentation framework of this BAF has the same arguments and attack relation \( R = \{(c, b), (c, a), (d, b)\} \). This is pictured in Figure 2.4.

Figure 2.4: The corresponding argumentation framework of the BAF in Figure 2.3 from Example 9

By inspection, the winning arguments of this BAF are \( \{c, d, e\} \).

In summary, one can generalise abstract argumentation into bipolar argumentation frameworks, where edges are either attacks or supports. There are principled ways of absorbing supports into attacks, such that bipolar argumentation frameworks can be transformed into abstract argumentation frameworks, where one can evaluate which arguments are winning.

2.3 Social Media Analytics

As stated in Section 1, online discussions cover a vast range of topics and involve many users; this is not surprising due to the growth of access to the Internet, especially through smartphones \cite{11}. Indeed, 62% of American adults get their
news on social media in 2016 [1], increasing to 67% in 2017 [13]. In the UK, the accounts of UK MPs are collectively being followed by the equivalent of almost 20% of the UK population [12]. It is reasonable to ask how can we analyse online discussions at scale.

Engaging with large-scale online discussions often lead users to suffer from information overload. For example, UK MPs learn to reply strategically and selectively to citizens concerned with specific topics that are also of interest to the MPs [12]. While MPs are guided by political issues, many discussion platforms have UIs that allow for readers to sort the comments, say from most liked to least liked. This seems to rely on a “wisdom of the crowds” effect to have the best points float to the top as indicated by the number of likes, allowing for the user to read the top few points made [31]. [2] has argued that such comment sorting and structuring mechanisms, including flagging, moderation and ways of detecting relevancy and novelty, can help increase user participation on news comments, improve the quality of comments, and promote constructive discussions. This is what moderation on Kialo also seeks to achieve [20].

It is important to study online discussions because they can affect the offline world. For example, [8] has shown that large companies (as defined by Forbes) may actively censor critical comments. However, the magnitude of such effects is open for debate in some cases. For example, [1] has clarified the factors behind how much the spread of fake news on social media can be responsible for influencing the 2016 US Presidential Election from the perspective of welfare economics, and argued that there are good reasons to argue that the effects are both small and large. Also, Twitter allows for UK MPs to follow other MPs; which may facilitate cross-party discussions despite ideological differences [12]. But on the individual level, we seek to understand how information can be presented to them such that they read the most compelling of comments made in a discussion.

2.4 Argumentation Theory in Social Media Analytics

Given the inevitable diversity and conflicts of the many views that are expressed, it is important for us to understand how many of these views are consistent, and how various differences can be resolved argumentatively and at scale. This makes online discussions a natural arena for argumentation theory to study.

Argumentation theory has been applied to both mine structured arguments from natural language text (e.g. [34]) and to analyse specific online discussion platforms. For example, [29] has designed and tested a pipeline on Twitter. Due to the comparatively noisy messages exchanged on Twitter, the pipeline has to first identify which tweets can be interpreted as self-contained arguments (ignoring those that are not, e.g. tweets that consist of just a URL and no other accompanying text), and also infer which replies are attacking or supporting. Once the appropriate BAF is extracted, one can calculate the winning arguments as explained in Section 2.2. Our work is different as we are applying these ideas to Kialo, which is a less noisy discussion platform where individual comments are moderated to be self-contained arguments and can contain more than the Twitter character limit. This bypasses the step in [29] where tweets need to be identified as self-contained arguments.

Further, [35, 28] have applied techniques of argument mining and evaluation to Debatepedia, where attack and support are identified via textual entailment (e.g. [23]), a technique that aims to reproduce how humans would use common sense to judge whether one piece of text or its negation follow from another piece of text. As Kialo already requires users to classify their comments as supporting or attacking, we can also bypass such techniques.

3 Details of the Kialo dataset

In this section we outline how discussions are initiated in Kialo and the procedure by which we scraped and cleaned the discussions.

3.1 Discussions and Sub-discussions on Kialo

To start a discussion on Kialo, the user needs to create a thesis, along with a tag that indexes the discussion. Users can attach multiple tags to the same thesis, which increases its visibility to users. Additionally, a discussion can be created with an option to add multiple theses to debate. For example, a discussion could start with an open question like “Who is the ultimate fighting hero from any fandom?” and several theses could be proposed as debatable options under this overarching question. In such a situation, this one discussion thesis could spawn multiple sub-discussions, each

---

20 See [https://support.kialo.com/hc/en-us/articles/360000631852-Moderating-Discussions](https://support.kialo.com/hc/en-us/articles/360000631852-Moderating-Discussions) last accessed 4/12/2019.

21 [http://www.debatepedia.org/en/index.php/Welcome_to_Debatepedia](http://www.debatepedia.org/en/index.php/Welcome_to_Debatepedia) last accessed 13/4/2019.

22 See [https://www.kialo.com/who-is-the-ultimate-fighting-hero-from-any-fandom-8857](https://www.kialo.com/who-is-the-ultimate-fighting-hero-from-any-fandom-8857) last accessed 8/4/2019.
proposing a candidate fighting hero, which will give rise to a separate reply tree of their own. Formally, this will be the predecessor-closed subtree from a given node of a reply tree (see Footnote 5, page 3).

3.2 Scraping and Cleaning Kialo Discussions

To obtain the dataset, we reverse engineered the Kialo app API, which obtains all the available tags on the Kialo website. This is done by first bootstrapping the query with certain featured tags on Kialo and then progressively expanding the tags dataset by adding the co-occurring tags with the bootstrap set. At the end of the process, we were exhaustively able to get 1120 tags, which covers almost all of the discussions hosted on Kialo. To verify this claim, we scripted another utility that exploited Kialo’s scrolling API to go as far back in the list as possible to get the oldest thread, and we ended up with the same number of threads to view. Figure 3.1 shows the histogram of top fifty popular tags among the 1120 along with corresponding number of threads associated with a particular tag. Tags like Politics, Ethics, Society, Law are the most popular.

As the next step we obtained all the discussion threads associated with each of these 1120 tags. This was done by mimicking the tag-based search feature of Kialo and getting all the results that show up for a particular tag based search. As mentioned before, each thread can be associated to one or more tags. Through our data collection scripts, we are able to obtain 1100 discussion threads. Our manual verification gives us a high degree of confidence that this is almost all of the debate activity on the service. We progressively crawl each discussion thread to acquire the data about the tree structure, votes on each argument and the argument text. This also includes all the sub discussion trees resulting due to debates having multiple thesis, as described in Section 3.1. Across these 1100 top-level discussion theses/topics/threads, various sub theses are created, leading to a total of 49,454 distinct discussion trees. Before analysing the data, we cleaned them by removing all the trees with less than twenty nodes and removing all the discussions with comments that have empty text or deleted branches. We are left with a total 2,835 final trees to analyse.

As stated in Section 1.2, we will represent these discussions as directed trees, where the directed edges point in the direction of the reply. Each edge can have a positive or negative sign, respectively indicating support or attack; the representation of such discussions as bipolar argumentation frameworks is therefore straightforward (Section 2.2), and we have ready-made techniques for evaluating which arguments are winning (Section 2.1).

---

23 Application Program Interface

24 See https://www.kialo.com/explore/featured (last accessed 8/4/2019)

25 This is as of March 2019.
We also acquire other supplemental meta-data such as the time of posting, the time of editing (if any) and the author meta-data. To our knowledge, this is the most complete snapshot of Kialo, as of March 2019.26 All discussions that were crawled from Kialo have a tree structure with a root node that represents the main thesis and each other node is a reply to its parent. Each reply answers only to the argument of the parent, so an answer in favour to a node, does not necessarily represent a support to the main thesis.

4 A Probabilistic Model of Winning Arguments in Synthetic Bipolar Argumentation Frameworks

4.1 The Probability of Winning Given the Level

In this section we will analyse and model probabilistically the distribution of winning arguments (a.k.a. winners) in simulated bipolar argumentation frameworks (BAFs) based on various kinds of reply trees. Our idea is to find a model that can describe the distribution of winners in real online discussions platforms that are suitably moderated, like Kialo. This would allow us to understand how to sample arguments in a discussions such to have the best perception of the “truth” as expressed by the normatively winning arguments, namely to sample a finite number of arguments from amongst those that have the highest probability of winning.

Recall from Section 3 that Kialo discussions can be represented by a directed tree, where the root is the main thesis under debate and the direction of each edge is the direction of the corresponding reply. In our model, each node apart from the root can attack or support its parent, and we represent this by a negative or positive edge (see Equation 4.4 below). Let \( q \in [0, 1] \) denote the probability that a reply is supporting. As replies must either be attacking or supporting, it follows that \( 1 - q \) is the probability that a reply is attacking. The only nodes that are not attacked or supported by any other argument are leaves in the reply tree. For each reply tree, when the probability \( q \) is not available, it can be estimated as \( \hat{q} \) that is the number of replies that are supporting divided by the total number of replies.

For each reply tree, we say that the root is at level 0, the set of nodes replying to the root is at level 1, the set of nodes replying to the set of nodes replying to the root is at level 2... etc. Let \( \hat{N} \in \mathbb{N} \) denote the depth of a given tree. The set of nodes at level \( \hat{N} \) are certainly leaves, but leaves do not need to be at level \( \hat{N} \).

**Example 10.** Recall Example 9 (page 9). This represents a reply tree with depth \( \hat{N} = 3 \). Node \( a \) is the root, at level 0. Node \( b \) is at level 1. Nodes \( c \) and \( e \) are at level 2. Node \( d \) is at level 3. In this case, we have \( \hat{N} = 3 \), but \( e \), at level 2, is a leaf that is not at level 3.

For a given reply tree, let \( 0 \leq h \leq \hat{N} \) be an integer denoting an arbitrary level in the tree. Let \( p_h \in [0, 1] \) be the winning probability of the nodes at level \( h \), and let \( p(k|h) \) be the probability of a node at level \( h \) having \( k \) child nodes at level \( h+1 \) that reply to it (attacking or supporting it).

Recall from Section 2.1 that as unrebutted arguments win by default, the probability that an argument at level \( \hat{N} \), i.e. a leaf, wins is 1. Therefore,

\[
p_{\hat{N}} = 1. \tag{4.1}
\]

Given a generic node \( a \) at level \( h < \hat{N} \) how do we calculate its \( p_h \)? Recall from Section 2.2 that for such an node to be a winning argument, all of its child nodes (at level \( h+1 \)) that support it must be winning, and all the child nodes that attack it must be losing arguments. Otherwise, a loses.

An argument \( a \) (leaf or non-leaf) at level \( h \) will have \( k \in \mathbb{N} \) replies from arguments at level \( h+1 \) with probability \( p(k|h) \), with a probability \( q \) for a reply to be a support. Each argument at level \( h+1 \) has its own probability \( p_{h+1} \) of winning. Furthermore, since arguments that are not winning are losing arguments (cf. Section 2.1), arguments at level \( h+1 \) are losing with a probability \( 1 - p_{h+1} \). Therefore, the expression for \( p_h \) is

\[
p_h = \sum_{k=0}^{\infty} [q p_{h+1} + (1 - q)(1 - p_{h+1})]^k p(k|h). \tag{4.2}
\]

In order to assess \( p_h \) in synthetic BAFs and therefore model which arguments are winning in a given online discussion, we re-express Algorithm 1 in slightly different notation such that it can calculate the winners directly from both attacking and supporting edges, rather than first converting from a BAF to its AF (as described in Section 2.2).
Given our set of arguments $A$ and set of winning arguments $G$ (Section 2.1), define a state function $s : A \rightarrow \{\pm 1\}$ where $s(i) = s_i$ such that $s_i = 1$ means $i \in G$, i.e. argument $i$ is winning, while $s_i = -1$ means $i \notin G$, i.e. argument $i$ is losing. The value of $s_i$ will be assigned iteratively to all $i \in A$ starting from level $N$ via the following rule:

$$s_i = \begin{cases} 
1 & \min_{j \in A} J_{ij} s_j 
eq 0 \\
\min_{j \in A} J_{ij} s_j & \text{else}
\end{cases}$$

(4.3)

where $(J_{ij})_{i,j \in A}$ is a matrix of size $|A| \times |A|$ defined as

$$J_{ij} := \begin{cases} 
-1 & (i, j) \in R_{\text{att}} \\
1 & (i, j) \in R_{\text{sup}} \\
0 & \text{else}
\end{cases}$$

(4.4)

Intuitively, the first case of Equation 4.3 is the case of a leaf node $i \in A$, while the second case assigns to $i$ $s_i = 1$ if all supporting nodes are winning and all attacking nodes are losing, otherwise $s_i = -1$ if there is some supporting node that is losing or attacking node that is winning. This convenient product form is made possible by the definition of Equation 4.4.

We can now calculate the frequency of winning arguments at level $h$. This quantity, averaged over an ensemble of reply trees with the same degree distribution and the same level of support $q$, will be our estimator of the probability $p_h$:

$$\overline{p}_h := \left\langle \sum_{i_h \in A} s_i \right\rangle / \left\langle \sum_{i_h \in A} s_i^2 \right\rangle,$$

(4.5)

where $i_h$ are the nodes belonging to level $h$, $\langle \cdot \rangle$ is the average over the ensemble of trees and $\sum_{i_h \in A} s_i^2$ is the number of arguments at level $h$.

In the next two subsections we will focus our analysis of the distribution of winners on two kinds of graphs:

- trees with homogeneous in-degree distribution (Section 4.2) and
- scale-free trees (Section 4.3).

A digraph with homogeneous in-degree distribution is one where the degree distribution is the same for all the nodes, i.e. $(\forall 0 \leq h \leq N) p(k|h) = p(k)$. Scale-free trees ([36]) are a class of graphs where the degree distribution of nodes follows a power law. For scale free graphs the condition $p(k|h) = p(k)$ no longer holds ([37]).

In the first case, the homogeneity of the in-degree distribution will allow us to classify analytically three main behaviours of $p_h$ as a function of the probability that an edge is supporting, $q$. We will then compare these results with simulations done on synthetic trees with Poisson in-degree distribution.

In the second case, the graphs have a more complicated structure, but we will see that some of its features will better approximate the structure of Kialo discussions. In this case it will not be possible to find exact formulae for $p_h$, so we will focus on the simulations on synthetic trees only. The results coming from this analysis will be then compared with the analysis of real discussions crawled from Kialo (Section 3), both to validate the probabilistic approach we are using and to interpret Kialo discussions using the results from our models.

### 4.2 Homogeneous In-Degree Distribution Graphs

#### 4.2.1 Establishing Three Types of Behaviour for the Winning Probability

Let us start studying the behaviour of winning arguments in the simpler case of trees with homogeneous in-degree distribution. The probability for an argument at level $h$ (where $0 \leq h \leq N$) having $k \in \mathbb{N}$ replying arguments satisfy the following:

$$p(k|h) = \begin{cases} 
p(k) & 0 \leq h < N \\
1 & h = N
\end{cases},$$

(4.6)

where the second case incorporates the last level that is only populated by leaves, which have in-degree zero (i.e., $k = 0$) and are always winning (Eq. 4.1). In this case it is possible to identify three classes of behaviours for different values of $q$.
Figure 4.1: The figures show different values of $\tilde{p}_h$ when the support probability $q$ is changed (here, $q \in \{0.1, 0.5, 0.9\}$). The degree distribution of the trees is Poisson with rate $\lambda = 2$ for all levels. The red dots represent the theoretical prediction of the winning probability.

1. when $q = \frac{1}{2}$ the expected fraction of winners is the same for all the levels.
2. when $q < \frac{1}{2}$ there is a high probability of attacks and the expected fraction of winners at a given level is determined by the parity of the level. This means that the proportion of winners oscillates between levels – i.e., the fraction of winning arguments increase and decrease from one level to the next.
3. when $q > \frac{1}{2}$ the expected proportion of winners is determined by the depth of the level. The lower is the level (i.e. the farther it stands from the leaves), the higher is the probability that at least one of the replies of the node is winning and attacking it. In other words, the fraction of winning arguments increase monotonically as we go from the root to the leaves.

We will now show that these three classes are present in all homogeneous trees with a generic in-degree distribution $p(k|h) = p(k)$.

**Theorem 1.** Let $p_h$ be the following winning probability of a node at level $h$:

$$p_h = \sum_{k=0}^{\infty} [qp_{h+1} + (1-q)(1-p_{h+1})]^k p(k) \quad \text{for } h < N,$$

with initial condition $p_N = 1$. Then the following results hold:

1. If $q = \frac{1}{2}$ then
$$p_h = p_{h+1} \quad \text{for } h \in [0, N-1]$$

2. If $q < \frac{1}{2}$ then
$$p_{N-2m} > p_{N-2m+1} \quad \text{for } m \in [0, N/2]$$
$$p_{N-2m-1} < p_{N-2m} \quad \text{for } m \in [0, (N-1)/2]$$

3. If $q > \frac{1}{2}$ then
$$p_h < p_{h+1} \quad \text{for } h \in [0, N-1]$$

**Proof.** We prove each case in turn.

1. If $q = \frac{1}{2}$ then $qp_{h+1} + (1-q)(1-p_{h+1}) = \frac{1}{2}$. Therefore
$$p_h = \sum_{k=0}^{\infty} \frac{p(k)}{2^k} = p_{h+1} \quad \text{for } h \in [0, N-1]$$

So irrespective of the in-degree distribution $p(k)$, $p_h$ will not depend on $h$.

29By parity we mean whether the path length from the leaf arguments to a given argument is even or odd. In an argumentation framework ($q = 0$), arguments at odd-length paths from the set of unattacked arguments $U$ are being indirectly attacked by $U$, while even-length paths from the unattacked arguments are being indirectly defended by $U$. See Example 8 or [17, Page 332].
2. If \( q < \frac{1}{2} \), then

\[
\begin{align*}
    p_h &> p_{h+1} \quad \text{if } p_{h+2} > p_{h+1}, \quad \text{and} \quad (4.13) \\
    p_h &< p_{h+1} \quad \text{if } p_{h+2} < p_{h+1}. \quad \text{and} \quad (4.14)
\end{align*}
\]

If \( p_{h+2} > p_{h+1} \), then

\[
\begin{align*}
    p_h &= \sum_{k=0}^{\infty} [q p_{h+1} + (1-q)(1-p_{h+1})] k p(k) \\
    &> \sum_{k=0}^{\infty} [q p_{h+2} + (1-q)(1-p_{h+2})] k p(k) = p_{h+1}. \quad (4.15)
\end{align*}
\]

This is because

\[
\begin{align*}
    q p_{h+1} + (1-q)(1-p_{h+1}) &> q p_{h+2} + (1-q)(1-p_{h+2}), \quad (4.17) \\
    -q(1-p_{h+1}) + (1-q)(1-p_{h+1}) &> -q(1-p_{h+2}) + (1-q)(1-p_{h+2}) \quad \text{and} \quad (4.18) \\
    (1-2q)(p_{h+2} - p_{h+1}) &> 0, \quad (4.19)
\end{align*}
\]

since \( q < \frac{1}{2} \) and \( p_{h+2} > p_{h+1} \).

If instead \( p_{h+2} < p_{h+1} \), then reasoning identically to the above implies that \( p_h < p_{h+1} \). From our initial condition \( p_N = 1 \) and

\[
p_{N-1} = \sum_{k=0}^{\infty} q^k p(k) < p_N = 1
\]

we obtain an oscillating trend of \( p_h \) as a function of \( h \): \( p_{N-2} > p_{N-1} < p_N \), \( p_{N-3} < p_{N-2} \) as \( p_{N-2} > p_{N-1} \), and so on.

3. Finally, if \( q > \frac{1}{2} \), then for all \( 0 \leq h \leq N-1 \),

\[
p_h < p_{h+1}, \quad (4.20)
\]

because \( (1-2q) < 0 \), therefore the opposite of Equation 4.19 holds: if \( p_{h+1} < p_{h+2} \), which holds true for \( p_{N-1} < p_N = 1 \), then \( p_h < p_{h+1} \), and so on monotonically.

This shows the result. \( \Box \)

### 4.2.2 Instantiating to Reply Trees with Poisson In-Degree Distribution and a First Answer to Question 2

The three cases in Theorem 1 show that the probability of the replies to be supports \( q \) determines the behaviour of \( p_h \). In particular we can recognise a transition at \( q = \frac{1}{2} \) between winning probability driven by level parity (for \( q < \frac{1}{2} \)) and a winning probability driven by distance from the root (for \( q > \frac{1}{2} \)). This is visualised in Figure 4.1 which shows the estimated winning probability \( \hat{p}_h \) of synthetic discussions on randomly generated trees with a Poisson in-degree distribution

\[
p(k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (4.21)
\]

with \( \lambda = 2 \) and for \( q \in \{0.1, 0.5, 0.9\} \). Each of the three sub-figures in Figure 4.1 is obtained by averaging the number of winning arguments of an ensemble of one thousand trees with the same depth \( N = 8 \). In this case we can perform the sum over \( k \) of Equation 4.17 where \( p(k) \) is Equation 4.21. The result is the series expansion for the exponential function, which can be written as

\[
p_h = \exp \left[ -\lambda (1 - q p_{h+1} - (1-q)(1-p_{h+1})) \right]. \quad (4.22)
\]

A corollary of Equation 4.22 is that for \( q = \frac{1}{2} \), \( p_h = e^{-\frac{1}{2}} \), which is consistent with substituting Equation 4.21 into Equation 4.12. This further verifies that for \( q = \frac{1}{2} \), the probability of a node winning at a given level is independent of that level. For \( \lambda = 2 \), \( e^{-1} \approx 37\% \), which is consistent with the horizontal red line on upper right of Figure 4.1.

Equation 4.22 is the theoretical estimation of the probability as a function of the level, which appears to be in good agreement with the simulations.
As a consequence of the above analysis we can conclude that we expect that the amplitude of the oscillations and the decrease, as we have seen in Equations 4.13, 4.14 and 4.20. However the higher is the lower bound the smaller will be the amplitude of the oscillations and the decrease, as \( q < \frac{1}{2} \) and how steep is its decay when \( q > \frac{1}{2} \). In particular, we will answer these questions in relation to the size of probability \( p_{\text{leaf}} := p(0); \) this is the probability for a node being a leaf. We will see that this analysis will be of particular interest in order to understand the distribution of winning arguments in Kialo data.

Consider a (homogeneous in-degree) tree with a certain support probability \( q \). We can rewrite Equation 4.2 as follows:

\[
\begin{align*}
    p_N &= 1 \\
    p_h &= \sum_{k=0}^{\infty} [q p_{h+1} + (1-q)(1-p_{h+1})]^k p(k) \\
    &= p(0) + \sum_{k=1}^{\infty} [q p_{h+1} + (1-q)(1-p_{h+1})]^k p(k).
\end{align*}
\]

As \( [q p_{h+1} + (1-q)(1-p_{h+1})]^k \leq q p_{h+1} + (1-q)(1-p_{h+1}) \) for all \( k \geq 1 \), we have that

\[
    p_h \leq p(0) + \sum_{k=1}^{\infty} [q p_{h+1} + (1-q)(1-p_{h+1})] p(k).
\]

As probabilities add to one, we have:

\[
    p_h \leq p(0) + [q p_{h+1} + (1-q)(1-p_{h+1})](1-p(0)) =: p_h^\text{max}.
\]

\[
    p_h \geq p(0) =: p_h^\text{min}.
\]

Equations 4.25 and 4.26 provide an upper and a lower bound of the function \( p_h \). The upper bound in Equation 4.25 is composed by two terms: one is the probability \( p(0) \) of a node to be a leaf and another term that depends on the winning probability \( p_{h+1} \) of the nodes at the following level. This second term is responsible for the oscillations of the upper bound in function of the level for \( q < \frac{1}{2} \) and the decreasing of it in function of the level for \( q > \frac{1}{2} \) (similar to what we have seen in Equations 4.13, 4.14 and 4.20). However the higher is the lower bound the smaller will be the amplitude of the oscillations and the decrease, as \( p_h \) will be squeezed between a large \( p_h^\text{min} \) and 1. This is shown in Figures 4.2 and 4.3. The left of these figures shows systems with a relatively small \( p(0) = 0.1 \), indicated by the green dashed line. The right of these figures shows systems with a relatively large \( p(0) = 0.5 \). This will respectively determine large and small oscillations of \( p_h \) in Figure 4.2 and long and short decrease of \( p_h \) in Figure 4.3. The blue dots represent the iterative solution of the equation:

\[
    p_h^\text{max} = p(0) + [q p_h^\text{max} + (1-q)(1-p_h^\text{max})](1-p(0)),
\]

which was obtained by assuming that the inequality in Equation 4.25 is saturated, i.e. \( p_h = p_h^\text{max} \). The red line in the same figure represents the Equation 4.27. The iterative solutions of Equation 4.27 indicated by the blue dots is obtained by projecting the points on this line to the diagonal black line. Starting from \( p_N = 1 \), we obtain the blue point 1 in the left of Figure 4.2 which is a solution of Equation 4.27 with initial \( p_{h+1} = 1 \). The projection of this point on the diagonal is the starting point of the new iteration. This new starting point, corresponding to a new value of \( p_{h+1} \), leads to the new solution of Equation 4.27 and is indicated by the blue point 2. Note that the new \( p_{h+1} \) is much smaller than the initial one, and also much smaller than the next value in the iteration represented by the blue point 3. As shown in Figure 4.2 the oscillations thus produced are larger when the value of \( p(0) \) is small. This representation of iterative solutions is called cob-webbing [38]. The same cob-webbing procedure for \( q > \frac{1}{2} \) is shown in Figure 4.3. In this case we do not have oscillations but similarly we can see that when \( p(0) \) is large there is a less pronounced decrease of \( p_h \).

As a consequence of the above analysis we can conclude that we expect that the amplitude of the oscillations and the decrease of the solution of \( p_h \) depend on the size of \( p(0) \). This means that the number of unreplied comments in the graph have a large impact on the behaviour of the winning probability. We will apply this analysis and result when analysing Kialo discussions in Section 4.4.
4.3 Scale-Free Graphs

In this subsection we will analyse what happens to the winning probability per level $p_h$ in scale-free graphs where the condition $p(k|h) = p(k)$ (Equation 4.6) no longer holds [37]. This means we are no longer able to study the behaviour of $p_h$ analytically. We will therefore analyse this case using synthetic BAFs. Even if the condition $p(k|h) = p(k)$ does not hold anymore, we can still expect to recognise the same behaviours of $p_h$ if the probability $p(k|h)$ does not differ much between the levels $h$. As mentioned in Section 4.1 the study of these graphs is important because Kialo discussion trees can be approximated by scale-free trees.

A common way to generate scale-free graphs (or trees) is using preferential attachment [36]. Intuitively, this means that every time a new node is attached to the tree, it will be attached with higher probability to a node with high degree. Using preferential attachment, we generated a series of scale-free synthetic trees following the growing network method described in [39]. Starting from a couple of connected nodes at each step we added a new node and connect it to an existing node $i$ with probability:

$$\pi_i = \frac{w_i}{\sum_j w_j},$$  \hspace{1cm} (4.28)

where $w_i$ is the degree (in-degree plus out-degree) of the node $i$. For each simulation we generated 1000 random trees of size 50. The sign of the edges in the tree are generated at random with a probability $q$ to be a support and $1 - q$ to be an attack. The average of the observables has been done aligning the maximum level $N$ of each tree, such that levels
with the same distance from the maximum level of their tree are averaged together. For example looking at examples \[3\] and \[9\] the argument \(a_2\) would be aligned with the argument \(d\) of the second, the argument \(a_3\) with the arguments \(c\) and \(d\), and the argument \(a_0\) would be aligned with the argument \(h\). This is because we expect that the nodes at the same distance from the last level of their tree have a comparable in-degree distribution. In all the figures that we will show in this subsection, the number of levels on the horizontal axis will correspond to the depth of the highest tree generated. Given the scale-free nature of our graphs, we will have a number of short trees with many leaves and few long trees. For this reason we will have several trees to average on for the statistics on higher levels and relatively few trees to average over as the level decreases. To maintain statistical significance, we only report winning probabilities for levels which have at least ten trees with nodes at that level.

In Figure 4.4, we compare the estimated probability of winning \(\bar{p}_h\) (in blue) with the estimated probability \(\bar{p}(0|h)\) of being a leaf at level \(h\) (in orange) as a function of the level for synthetic trees generated with the method above. As discussed before, the probability of leaves per level and the winning probability have been estimated by averaging the fraction of leaves per level and the fraction of winners per level in each tree over our ensemble of trees. Figure 4.5 shows that in a scale-free graph, the fraction of leaves per level is very high and decreases as the level decreases. Moreover especially for the higher levels, \(\bar{p}_h\) and \(\bar{p}(0|h)\) are similar, which means that the winners are typically unrebuted comments. Given the great difference between the probabilities \(p(0|h)\) at different levels we are not able to see the three different behaviours for \(p_h\) described in Section 4.2. Instead, it seems that the behaviour of \(p_h\) as a function of the levels is determined by the in-degree distribution for the various levels and in particular by \(p(0|h)\).

Remark. The best way to sample scale-free graph in order to read most of the winning arguments is to follow the indications given by the distribution of leaves in the graph, i.e. starting from the last level and ending with the root. This appears to be valid whatever the proportion of supporting replies \(q\) is in the graph, given that this is determined by the number of arguments that have “the last word” (i.e., leaves) that does not depend on \(q\).

4.4 Analysis of Kialo data

In this subsection we will analyse the discussions crawled from Kialo (Section 3) and their relative distribution of the winning arguments by comparing them to the simulated discussions analysed in Sections 4.2 and 4.3.

Recall from Section 3 that we have 2,835 reply trees from Kialo, with an mean (median) of 59 (33) arguments each (standard deviation is 77; this distribution is pictured on the left of Figure 4.5). An interesting feature of Kialo discussions is that they appear to have on average a good balance between attacks and supports; this is shown in the right of Figure 4.5, where we can see the distribution of the estimated level of support in each Kialo discussion \(\hat{q}\), obtained as the fraction of supporting edges to the total number of edges in each graph. The distribution is peaked around a mean of \((\hat{q}) = 0.47, median of 0.46 and standard deviation \(\sigma_{\hat{q}} = 0.15\). However not all the graphs have \(\hat{q} \sim 0.5\) – in fact we can see several with \(\hat{q} \gg 0.5\) and \(\hat{q} \ll 0.5\). Given that the results from synthetic trees are \(q\) dependent, we will divide Kialo trees in three classes as a function of the proportion of supporting edges \(\hat{q}\). The three classes are \(\hat{q} \leq 0.2, 0.4 \leq \hat{q} \leq 0.6\) and \(\hat{q} \geq 0.8\). The boundaries of the categories has been chosen such to be as extreme as possible in order to highlight the different behaviour of \(p_h\), but still large enough to have enough reply trees in each class for us to make statistically meaningful statements.

Figure 4.6 shows the in-degree distribution of Kialo trees. This distribution can be fitted by a power law (red line in the plot) for degrees larger that \(k_{\text{min}} := 17\) and \(\alpha = 3.05\). The expression for the power law is

\[
p(k) = \frac{k^{-\alpha}}{\zeta(\alpha) - \sum_{k=1}^{k_{\text{min}}-1} k^{-\alpha}},
\]

(4.29)
what was expected from the study of the overall distribution of the in-degree probability. This observation, on the one
with a quite homogeneous probability
\( p \)
where
\( \tilde{p} \)
with exponent
\( \alpha \)
with
\( k \geq k_{min} \)
and
\( \zeta(\alpha) := \sum_{n=1}^{\infty} \frac{1}{n^\alpha} \)
is the Riemann zeta function, for \( \alpha \in \mathbb{R} \) and \( \alpha > 1 \).
The form of \( p(k) \) suggests that Kialo discussions can be represented as scale-free trees, as discussed in Section
\ref{scale_free_trees}. However we will see that in some cases, the results from Section
\ref{oscillations} will be more relevant.

Given that Kialo trees seem to be similar to scale-free trees, we will study them performing the same kind of analysis
done in Section
\ref{scale_free_trees}. Figure
\ref{histogram_results} compares \( p_h \) with \( \tilde{p}(0|h) \). For \( 0.4 < \tilde{q} < 0.6 \) and \( \tilde{q} > 0.8 \) we can see that \( \tilde{p}(0|h) \)
does not follow the same behaviour as in the other two cases. It must be noted that in this case the probability of having leaves does not vary much, making this term act more similar to the probability of leaves per level, \( p_{leaf} \), in the case of homogeneous trees. As a consequence of this, the probability of winning \( p_h \) per level in Kialo discussions appears to have similarities with both scale-free trees and homogeneous
in-degree distribution trees depending on \( \tilde{q} \).

For \( \tilde{q} > 0.8 \) we can see the same decrease in winning probabilities that we observed for the homogeneous in-degree
distribution trees (figure
\ref{histogram_results}) and the scale-free synthetic trees. For \( 0.4 < \tilde{q} < 0.6 \) the behaviour of \( p_h \) appears to be the
same decreasing behaviour observed for \( q = 0.5 \) in scale-free trees.

For \( \tilde{q} < 0.2 \) we can partly observe the oscillations as a function of the levels that we saw in Figure
\ref{oscillations} for homogeneous in-degree trees, but in this case the oscillations are smaller and irregular. The presence of oscillations is consistent with a quite homogeneous probability \( \tilde{p}(0|h) \) over the levels. For example in the levels 14, 15, 16 and 17 in the top
left of Figure
\ref{histogram_results} which have approximately the same small value of \( \tilde{p}(0|k) \), we can clearly recognise a more regular oscillatory percentage of winners. However in principle, being thus distorted, this non monotonicity could also be the result of some noise. In Section
\ref{oscillations} we will answer to this question clearly showing that these oscillations have the same
origin as the oscillations seen in the homogeneous case.

Interestingly, the results for \( \tilde{q} < 0.2 \) suggests that in Kialo discussions, there is a correlation between the level of
support \( \tilde{q} \) and the form of the graph, which in this case seems to be less scale-free like and more homogeneous than
what was expected from the study of the overall distribution of the in-degree probability. This observation, on the one
hand suggests that it could be interesting an in-depth characterisation of the structure of Kialo discussions with different
levels of support. On the other hand, it confirms the validity of our overall probabilistic approach to the analysis of winning arguments. It seems in fact that the percentage of winners is fairly described by synthetic BAFs (where attacks and supports are just i.i.d. random variables), provided that they are shaped into structures with appropriate in-degree distributions that are representative of the real BAFs for different average levels of support $\hat{q}$ in the discussions.

**Remark.** As in the case for scale-free trees, by considering Kialo discussions for all the three categories of $\hat{q}$, not only do we have the number of leaves per level being one major contribution to the total number of winners, but also the behaviour of $\hat{p}(0|\hat{h})$ appears to greatly determine the trend of the winners as a function of the level $h$ especially for discussions where the level of support is medium or high, i.e. $\hat{q} > 0.4$. For this reason, the best way to sample from this discussions is sampling the arguments starting from the level with the highest percentage of leaves and ending with the level with the lowest. In graphs where the proportion of supporting replies is very low, $\hat{q} < 0.2$ the best way to sample the graph is to still follow the higher percentage of leaves, but, as the results of the next section will confirm, between two subsequent levels with the same percentage of leaves it is always better to sample levels at an even distance from the last level of their reply tree.

5 Removing leaves from winners statistics

Our study of synthetic reply networks and comparison with Kialo data (Section 4) seems to highlight that comments that have the last word (i.e. the leaf comments in discussion trees) represent a determining factor in establishing the rest of the winners. This is consistent with argumentation theory, which assumes that arguments that have the last word win by default (Section 2.1). It can be argued that comments that have been answered, and so have been evaluated positively or negatively by others, are more representative of the truth and should have a greater importance than those that have not yet been replied to. For this reason we propose a different way to count the number of winning arguments in the graph, still using Equation 4.3 to determine the winners. The difference from the method used in the previous section is that when calculating the fraction of winners per level, we consider only the nodes that are not leaves. In other words, we only consider those arguments that have had a chance to be supported or attacked by at least one other argument.

**Example 11.** If in a given level of a reply tree there are five losing arguments and ten winning arguments, and among the ten winning arguments five of them are leaves, then the fraction of winners in this level is $\frac{10 - 5}{5 + 10 - 5} = \frac{5}{10} = \frac{1}{2}$. However, if we are to ignore the leaves, then the fraction of winners is $\frac{10 - 5}{5 + 10 - 5} - \frac{5}{10} = \frac{1}{2}$.

In other words given a reply tree, defining

$$\hat{p}_h := \frac{\# \text{winners at level } h}{n_h},$$  \hspace{1cm} (5.1)

$$\hat{p}(0|h) := \frac{\# \text{leaves at level } h}{n_h},$$  \hspace{1cm} (5.2)

where $n_h$ is the number of nodes in level $h$, we will have

$$\hat{p}_h^{nl} = \frac{\# \text{winners at level } h - \# \text{leaves at level } h}{n_h - \# \text{leaves at level } h} = \frac{\hat{p}_h - \hat{p}(0|h)}{1 - \hat{p}(0|h)}.$$
Therefore, the estimated probability of winning arguments per level is

\[ \hat{p}_h = \left( \frac{\hat{p}_h - \hat{p}(0|h)}{1 - \hat{p}(0|h)} \right), \]

where the average \( \langle \cdot \rangle \) is over an ensemble of graphs with the same degree probability and the same level of support \( q \).

### 5.1 Predicting the Fraction of Rebutted Winning Arguments

For reply trees with homogeneous in-degree distributions, we expect that the shape of the winning probability of the non-leaf arguments \( p_h^{\text{nl}} \) as a function of the level would not differ much from the old probability distribution \( p_h \), given that for all \( 0 \leq h < N \), \( p(0|h) = p(0) \) and the estimated probability of rebutted winning comments per level will be:

\[ \tilde{p}_h^{\text{nl}} = \left( \frac{\hat{p}_h - \hat{p}(0)}{1 - \hat{p}(0)} \right). \]

In scale-free graphs, the situation becomes more interesting since the estimated probability of leaves per level has a non-trivial dependence on the level \( h \), therefore we can obtain that \( \tilde{p}_h^{\text{nl}} \) behaves differently from \( \tilde{p}_h \). In general, we can predict the behaviour of \( \tilde{p}_h \) by separating the contribution of leaves from that of validated comments in Eq. (4.2) in the following way:

\[ p_h^{\text{nl}} = \sum_{k=1}^{\infty} \left[ q \left( p(0|h + 1) + (1 - p(0|h + 1))p_h^{\text{nl}} \right) + (1 - q)(1 - p(0|h + 1))(1 - p_h^{\text{nl}}) \right] k p(k|h), \]

where the first term is the probability of being supported by a leaf, or being supported by a non-leaf that is winning. The second term is the probability of being attacked by a non-leaf that is losing. Note that this time the sum over all replies to the node at level \( h \) starts from \( k = 1 \) in order to exclude the leaves (which would correspond to \( k = 0 \)).

Given that we do not have an analytical formula for \( p(k|h) \), and in this case we do not even have the predominant contribution of leaves to determine the shape of \( p_h \), we need to find another way to estimate it. One possible solution is to approximate Equation 5.5 using the fraction of replies per level of a single synthetic graph. We define

\[ \hat{k}_h = \sum_{i_h \in \mathbb{N}} k_{i_h}, \]

where \( i_h \) indexes the nodes belonging to level \( h \), and \( k_{i_h} \in \mathbb{N} \) is the in-degree of node \( i_h \). Therefore, Equation 5.6 is the average in-degree of level \( h \). Using Equations 5.1, 5.2 and 5.6, we can estimate the new probability of winning per level approximating Eq. 5.5 as follows:

\[ p_h^{\text{nl}} \approx \left( q \left( \hat{p}(0|h + 1) + (1 - \hat{p}(0|h + 1))p_h^{\text{nl}} \right) + (1 - q)(1 - \hat{p}(0|h + 1))(1 - p_h^{\text{nl}}) \right)^\hat{k}_h. \]

For \( q = \frac{1}{2} \), we can see that the formula simplifies and we have the following expression that determines the probability of winning:

\[ p_h^{\text{nl}} \approx \left( q \hat{k}_h \right). \]

In this case the number of replies to each level is the only relevant ingredient to determine the winning probability of that level.

For \( q \neq 1/2 \) the situation is different because in order to determine \( p_h^{\text{nl}} \) we should also consider the fraction of leaves present in level \( h + 1 \). The percentage of validated winners per level will thus be a non-trivial combination of the contribution of the number of replies and their probability to be leaves.

### 5.2 Analysis of Kialo Data and Discussion

Figure 5.1 shows the comparison between the estimated \( \tilde{p}_h^{\text{nl}} \) and the theoretical \( p_h^{\text{nl}} \) in synthetic scale-free trees (left column) and Kialo discussions (right column). In the Kialo discussions, \( p_h^{\text{nl}} \) has been estimated using Equation 5.7 for
Figure 5.1: In the figure we compare the estimated winning probability of validated arguments \( \tilde{p}_{nl}^{h} \) with its theoretical prediction \( p_{nl}^{h} \) for Kialo discussions and scale-free synthetic graphs and different levels of support. The scale-free graphs have been generated as usual and the quantities are averaged over 1000 trees of size 50.

5.1.1 Medium Level of Support: Balanced Discussions

As we discussed in the previous section, for \( q = 0.5 \), the number of replies to the arguments at a certain level is the only factor that determines the winning probability for that level (see Equation 5.8). A general result about scale-free graphs \[37\] is that the levels with the highest number of nodes are expected to be in the middle of the graph. This explains why in Figure 5.1 for scale-free graphs we see a minimum of the winning probability for central levels, while we know that for synthetic graphs \( \hat{k}_{h} \) does not depend on \( q \). The behaviour of \( p_{nl}^{h} \) of Kialo discussions for \( \hat{q} > 0.8 \) seems again to be the same as the one of scale-free graphs for high levels.

5.1.2 High Level of Support: High Agreeability or Supportive Discussions

For \( q = 0.9 \), we know that not only the number of replies is important for \( p_{nl}^{h} \), but also the fraction of leaves per level. This is also confirmed by the observation that in scale-free graphs the trend of \( p_{nl}^{h} \) when \( q = 0.9 \) is different from the trend when \( q = 0.5 \), while we know that for synthetic graphs \( \hat{k}_{h} \) does not depend on \( q \). The behaviour of \( p_{nl}^{h} \) of Kialo discussions for \( \hat{q} > 0.8 \) seems again to be the same as the corresponding behaviour of scale-free graphs.
5.2.3 Low Level of Support: Controversial or Attacking Discussions

On the contrary, for $q = 0.1$ the behaviour of scale-free graphs appears to be totally different from the behaviour of Kialo graphs when $\tilde{q} < 0.2$. In fact, $\tilde{p}_h$ in Kialo discussions shows a clear oscillatory behaviour typical of homogeneous trees. This may suggest again that when the discussions are particularly controversial (as captured by small values of $\tilde{q}$), their structure is very different from the one of discussions with higher values of $q$. At variance with the oscillatory behaviour of $\tilde{p}_h$ for $\tilde{q} < 0.2$, oscillations are particularly evident in $\tilde{p}_h$. These amplification of the oscillation behaviour can be understood by looking at Equation 4.27 that was approximately describing the oscillation of $p_{\text{max}}$ in homogeneous trees. Their amplitude was directly dependent on $p(0)$ as we discussed earlier. However if we subtract $p(0)$ from $p_{\text{max}}$ and we divide by $(1 - p(0))$ (for $p(0) < 1$), we are left with the full-amplitude oscillating part only, i.e. more formally

$$\frac{p_{\text{max}} - p(0)}{1 - p(0)} = q p_{h+1}^{\text{max}} + (1 - q) \left(1 - p_{h+1}^{\text{max}}\right). \tag{5.9}$$

In fact the analysis of no-leaves winner statistics formalized in Equation 5.3 is in this case effectively zooming on (and making it finally apparent beyond any possible doubt) the oscillating part of the winning statistics that is expected in a homogeneous tree rather than a scale-free tree. As a consequence we can also conclude that the uneven behaviour previously observed for $\tilde{p}_h$ in the first panel of Figure 4.7 was indeed a largely squeezed and slightly deformed oscillatory behaviour as well.

Remark. 
In conclusion, when looking at the statistics of non-leaf arguments, the method of sampling the discussion to get as many of the winning arguments as possible at first is highly non trivial. For balanced graphs, where the amount of attack and support is comparable ($q \approx 0.5$), we only need the number of nodes per level in order to predict the percentage of winners and we can estimate it using Eq. 5.8. When there is a high level of agreeability or support ($q > 0.5$), we can follow the same indications that we gave both for scale-free graphs and homogeneous in-degree graphs. For graphs in which the discussions is very controversial ($q$ small) the method of sampling the graph is the same as for homogeneous graphs, i.e. sampling levels according to their parity starting from the last level of the tree.

5.3 Reducing the Importance of Unanswered Comments

In the previous subsection, we have shown that the percentage of leaves per level, $p(0|h)$, is a very important factor even when leaves are removed from the statistics. In fact, even if they are not present in the statistics, they can still influence the winning probability of the arguments at the inner levels of the tree. This is potentially problematic since leaf nodes are arguments that may not have been examined fully (as an extreme example, a comment that has just been authored will not have been examined at all by any of the other participants in the discussion). According to the rules of argumentation theory, one attack from such an unrebutted argument is enough to make the target of this attack lose. In this sub-section, we depart from traditional argumentation frameworks and suggest different methods to establish winning arguments that dampen the high degree of influence that leaf nodes have.

5.3.1 A Different Rule for Nodes with only Leaves as Replies

This method proposes that we treat nodes that have only leaves as replies differently from the others. Given a generic node the method implies that the leaves that are replying to it will not influence its winning status. However, if all the replies of the node are leaves, the winning rule will consider them as important as usual. We will therefore expect the leaves to have a diminished influence compared to the first case. More formally:

$$s_i = \begin{cases} 1 & (\forall j \in A) J_{ij} = 0 \\ \min_{j \in A} J_{ij} s_j, & (\forall j \in A) k_j = 0 \\ \min_{j \in A} J_{ij} (1 - \delta_{k_j,0}), & \text{else} \end{cases}, \tag{5.10}$$

5.3.2 Majority Rule

A different possibility to reduce the effect of the leaves is to apply a majority rule in order to reduce the effect of the leaves. Recall from Section 2.2 that a node is losing if at least one of its attacking replies is winning, or at least one of its supporting replies is losing. In this way it is sufficient that a unique unrebutted argument attacks a node in order for it to lose. Implementing a majority rule means that we can consider an argument to be losing if it has the majority of its replies in one of the two situation just described. In this way the contribution of attacking unreplied arguments is less

$$s_i = \begin{cases} 1 & (\forall j \in A) J_{ij} = 0 \\ \min_{j \in A} J_{ij} s_j, & (\forall j \in A) k_j = 0 \\ \min_{j \in A} J_{ij} (1 - \delta_{k_j,0}), & \text{else} \end{cases}, \tag{5.10}$$

23
important because we need more than just one attack from a leaf to make it a losing argument. Given this, the new rule to establish if a node is winning will be:

\[
s_i = \begin{cases} 
1 & \text{if } (\forall j \in A) J_{ij} = 0 \\
\text{sign}(J_{ij} s_j) & \text{else.}
\end{cases}
\] (5.11)

### 5.3.3 Generalised Majority Rule

This more generalised majority rule is similar to the majority rule above, but gives different weights to arguments that have different degrees and different opinions. The idea is that when two arguments \(a\) and \(b\) are replying to the same argument \(c\), if \(a\) is winning and \(b\) is losing, the opinion of \(a\) should count more than \(b\) in the majority computation.

Furthermore, we wish to give more importance to winning nodes that have been evaluated by (i.e., have received replies from) many other arguments. Intuitively, if an argument is winning and has a lot of replies, this means that many arguments agree with it. Nodes that have a lot of replies have therefore passed a “harder” test in order to win, in comparison to nodes with fewer replies. The opposite happens with losing nodes. If two nodes are losing, the one with lower in-degree is the more influential – If a node has been validated by many other arguments and it is found losing, then many argument disagree with it, so its loss is more “severe”. This translates in the following rule:

\[
s_i = \begin{cases} 
1 & \text{if } (\forall j \in A) J_{ij} = 0 \\
\text{sign}(J_{ij} s_j (k_j + 1)^{s_i \alpha}) & \text{else.}
\end{cases}
\] (5.12)

where \(k_j\) is the degree of node \(j\) and \(\alpha\) is a parameter that controls the importance of the weight \((k_j + 1)^{s_i \alpha}\). When \(\alpha = 0\) we obtain the simple majority rule of the previous case.

In summary, these three methods represent possible ways to relax the rules of argumentation theory in order to give less importance to arguments that have the last word in the discussion. More work and analysis are needed in order to assess what are the effects of these methods on the winning probability \(p_h\).

### 6 Conclusions and Future Work

#### 6.1 Summary of Results

In this paper we have applied ideas from bipolar argumentation theory and complex networks to simulate an ensemble of synthetic reply trees where the nodes are arguments and the directed edges are attacking or supporting replies. We then built a model that calculates the probability that an argument will win given the distance that argument is from the last level of the tree, subject to some assumptions on the in-degree distribution of the reply trees (e.g., being scale-free, or homogeneous). This model indicates how one should sample such a reply tree so as to choose the levels where arguments are winning with the highest probability.

The first result is that this probabilistic approach appears to be a good way to tackle the problem because it can predict the winning probability of online discussions, when its results are compared to real data that we scraped from Kialo, an Internet debating platform. The probabilistic approach also reveals three different schemes of behaviour for the winning probability as a function of two global parameters of the reply tree: (1) the percentage of attacking and supporting arguments in the graph and (2) the structure of the discussion tree, i.e., how comments are organised in it. Data from Kialo indicates that very controversial discussions (i.e., reply trees with a high proportion of attacks) are organised in a different way, specifically that the number of unreplied arguments per level is quite homogeneous amongst the levels and the behaviour of the winning probability follows the model for reply trees with a homogeneous degree distribution. When the graph is not as controversial (i.e., when the conversation is more supportive), the Kialo reply trees seem to behave similar to trees with a scale-free degree distribution, both in the probability of unreplied comments per level and in the location of the winning arguments.

From our analysis, we can thus classify online discussions by the proportion of supporting replies. When the proportion of supporting replies is medium or high (i.e., the discussion points largely agree with each other), the best way to read the comments is to start from the last level (i.e., most recent comments first) and arriving to the root comment in reverse order of the level in the reply tree. In this case, our model suggests that a temporal ordering of new user utterances, with the most recent comments appearing first (a common sorting method employed on platforms such as Disqus, and available as an option on other platforms such as Twitter or Facebook) may show a higher proportion of winning arguments than other sorting methods, but pure temporal ordering is not in itself sufficient – it needs to be tweaked, allowing comments to be read based on the level in the reply tree rather than just the time stamp of each post. In contrast, when discussions are combative or aggressive, (i.e., individual discussion points have low levels of support),

\[
\text{...}
\]
the best way to read the comments is still to start from the last level, but following the parity of the levels (i.e., read every other level, since each winning level of comments attacks and invalidates the previous level).

We also observed that when the number of attacks and support is balanced in the graph, the new probability of winning per level is guided only by the number of replies to comments at that level. Today’s platforms support this (very crudely) by sorting based on overall level of support (e.g., sort by number of likes and comments, or numbers of upvotes and downvotes). However when the level of support is high or low, the new winning probability is guided both by the number of replies and by the percentage of leaves per level; thus current methods of sorting posts in online discussions can miss many of the winning arguments, especially given the combative and aggressive nature of many online discussions.

An important result that appears from our analysis is that arguments that have the “last word” (unreplied comments) have a great impact on the winning probability of all the other arguments. This is due to a fundamental assumption in argumentation theory where all unrebutted arguments, and hence arguments that are not replied to, win by default. We may overcome this problem by removing the leaves from the winning statistics. The analysis of the behaviour of the winning probability of the non-leaf arguments highlight again the difference between very aggressive discussions and less aggressive discussions, confirming what was observed in the previous case.

We also found that removing leaves (comments without replies) from the statistics is not sufficient to remove their influence on the structure of the winning arguments. In Section 5, we suggested possible ways to overcome this problem by define new rules for determining the winning arguments such that the role of the leaves becomes less important.

6.2 Directions for Future Work

We conclude by suggesting interesting directions for future work to consider.

6.2.1 More Sophisticated Argumentation Models

This work has applied relatively basic frameworks from argumentation theory. If one were to take more information about the comments into account, e.g. the information about how compelling an argument is in Kialo as represented by the horizontal bar (Figure 1.4, page 5), then can this information be represented in more sophisticated argumentation models, such as the quantitative frameworks described in [41].

6.2.2 Other Online Discussions

We can expand our analysis to online discussions other than Kialo in order to better investigate the correlation between the proportion of supporting replies, \( \hat{q} \), of a reply tree, and its in-degree distribution. We can integrate various argument cleaning techniques, such as those applied to Twitter [29], to see whether the patterns we observe in Kialo are replicable across other online discussions. We can also apply the methods outlined in Section 5.3 to investigate how the winning probability changes and to confirm that the unrebutted comments have a more minor role in Kialo and other datasets.

6.2.3 More Ways to Sample

Our probability model can guide readers in where to read in a Kialo discussion in a “top-down” way, but many discussion platforms have sorting and comment structural options already built in, and comparisons of sorting and sampling methods across platforms can be instructive. For example, in BBC’s Have Your Say (recall Footnote [1] page 2), there are ways to sort the comments by their number of likes or by their time of posting. Following the suggestion in [31], suppose a user sorts the comments from most liked to least liked, then how many comments should a reader sample in order to get, say, 10% of the winning arguments? How about 50% of the winning arguments? What if the reader were to sort but reads every fifth argument, or just randomly samples by scrolling down and stopping to read at random points. What is the probability that such a random sample will get all of the losing arguments from the odd levels of a tree? If we use as a metric of how many winning arguments are sampled, then which method of sorting is the best, and how does Kialo’s UI of displaying the pro and con arguments in two columns compare with these other methods of sampling?

6.2.4 The Problem with Normativity

One criticism that can be leveled at this work is that the normative notion of a winning argument is irrelevant to helping people see the winning arguments. This is because people can be irrational, and may not be convinced by winning arguments that do not support their point of view. For example, a reader of The Guardian (which is considered to be left of centre) may disagree with the winning comments of the comment networks in Breitbart News (a right leaning
news website), and vice versa. There have been research relating how people perceive how arguments disagree and when they are justified (e.g. [42, 43]), which have discovered that if people have a choice between the grounded and preferred extensions (see Section 2.1), they tend to go for the preferred extensions. Therefore, we would like to conduct experiments to test the recommendations of our model, e.g. are people more convinced by the winning arguments that we suggest them to sample, or not? What if we “unroll” the argumentation framework by displaying the counter-arguments and how they are rebutted as well?

References

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

[2] Nicholas Diakopoulos and Mor Naaman. Towards quality discourse in online news comments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pages 133–142. ACM, 2011.

[3] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. How Useful are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings. In *Proceedings of the 19th international conference on World wide web*, pages 891–900. ACM, 2010.

[4] Manos Tsagkias, Wouter Weerkamp, and Maarten De Rijke. News comments: Exploring, modeling, and online prediction. In *European Conference on Information Retrieval*, pages 191–203. Springer, 2010.

[5] Carole Cadwalladr and E Graham-Harrison. The cambridge analytica files. *The Guardian*, 21:6–7, 2018.

[6] Jaram Park, Meeyoung Cha, Hoh Kim, and Jaeseung Jeong. Managing bad news in social media: A case study on domino’s pizza crisis. In *Sixth International AAAI Conference on Weblogs and Social Media*, 2012.

[7] Sandra. L. Emerson. Usenet - A Bulletin Board for Unix Users. *Byte*, 8(10):219, 1983.

[8] Sam H. Dekay. How large companies react to negative Facebook comments. *Corporate Communications: An International Journal*, 17(3):289–299, 2012.

[9] Sherice Gearhart and Seok Kang. Social Media in Television News: The Effects of Twitter and Facebook Comments on Journalism. *Electronic News*, 8(4):243–259, 2014.

[10] Damian Trilling. Two Different Debates? Investigating the Relationship Between a Political Debate on TV and Simultaneous Comments on Twitter. *Social science computer review*, 33(3):259–276, 2015.

[11] Julia Murphy and Max Roser. Internet. *Our World in Data*, 2019. [https://ourworldindata.org/internet](https://ourworldindata.org/internet), last accessed 7/4/2019.

[12] Pushkal Agarwal, Nishanth Sastry, and Edward Wood. Tweeting MPs: Digital Engagement between Citizens and Members of Parliament in the UK. *ArXiv preprint arXiv:1904.00004*, 2019.

[13] Jeffery Gottfried and Elisa Shearer. News use across social media platforms 2017, 2017. [http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/](http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/) last accessed 7/4/2019.

[14] Anthony P. Young, Sagar Joglekar, Kiran Garimella, and Nishanth Sastry. Approximations to Truth in Online Comment Networks. 2018. Available from [http://comma2018.argdiap.pl/wp-content/uploads/Young.pdf](http://comma2018.argdiap.pl/wp-content/uploads/Young.pdf) last accessed 10/4/2019.

[15] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*, volume 47. Springer, 2009.

[16] Claudette Cayrol and Marie-Christine 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.

[17] Phan Minh 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.

[18] Martin Caminada. On the Issue of Reinstatement in Argumentation. In *European Workshop on Logics in Artificial Intelligence*, pages 111–123. Springer, 2006.

[19] Erin E Buckels, Paul D Trapnell, and Delroy L Paulhus. Trolls just want to have fun. *Personality and individual Differences*, 67:97–102, 2014.

[20] Whitney Phillips. Loling at tragedy: Facebook trolls, memorial pages and resistance to grief online. *First Monday*, 16(12), 2011.
[21] Steven Gianvecchio, Mengjun Xie, Zhenyu Wu, and Haining Wang. Humans and bots in internet chat: measurement, analysis, and automated classification. IEEE/ACM Transactions On Networking, 19(5):1557–1571, 2011.

[22] Ion Androutsopoulos and Prodromos Malakasiotis. A survey of paraphrasing and textual entailment methods. Journal of Artificial Intelligence Research, 38:135–187, 2010.

[23] Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto. Recognizing textual entailment: Models and applications. Synthesis Lectures on Human Language Technologies, 6(4):1–220, 2013.

[24] James W. Pennebaker, Martha E. Francis, and Roger J. Booth. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 71(2001):2001, 2001.

[25] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology, 29(1):24–54, 2010.

[26] Oana Cocarascu and Francesca 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.

[27] Oana Cocarascu and Francesca Toni. Mining bipolar argumentation frameworks from natural language text. 2017. Available from https://spiral.imperial.ac.uk/bitstream/10044/1/56643/2/final.pdf, last accessed 8/4/2019.

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

[29] Tom Bosc, Elena Cabrio, and Serena Villata. Tweeties Squabbling: Positive and Negative Results in Applying Argument Mining on Social Media. In COMMA, pages 21–32, 2016.

[30] Nadin Kökciyan, Isabel Sassoon, Anthony P. Young, Martin Chapman, Talya Porat, Mark Ashworth, Vasa Curcin, Sanjay Modgil, Simon Parsons, and Elizabeth Sklar. Towards an Argumentation System for Supporting Patients in Self-Managing their Chronic Conditions. In Proceedings of the AAAI Joint Workshop on Health Intelligence (W3PHIAI 2018), 2018.

[31] Anthony P. Young. Notes on Abstract Argumentation Theory. ArXiv preprint arXiv:1806.07709, 2018.

[32] Paul E. Dunne and Michael Wooldridge. Complexity of Abstract Argumentation. In Argumentation in Artificial Intelligence, pages 85–104. Springer, 2009.

[33] Michael Wooldridge. An Introduction to Multiagent Systems. John Wiley & Sons, 2009.

[34] Marco Lippi and Paolo Torroni. Argumentation Mining: State of the Art and Emerging Trends. ACM Transactions on Internet Technology (TOIT), 16(2):10, 2016.

[35] Elena Cabrio and Serena Villata. Combining Textual Entailment and Argumentation Theory for Supporting Online Debates Interactions. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 208–212, 2012.

[36] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. science, 286(5439):509–512, 1999.

[37] Zsolt Katona. Levels of a scale-free tree. Random Structures & Algorithms, 29(2):194–207, 2006.

[38] Johannes Müller and Christina Kuttler. Methods and models in mathematical biology. Springer, 2015.

[39] Paul L. Krapivsky and Sidney Redner. Organization of growing random networks. Phys. Rev. E, 63:066123, May 2001.

[40] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. SIAM review, 51(4):661–703, 2009.

[41] Joao Leite and Joao Martins. Social abstract argumentation. In Twenty-Second International Joint Conference on Artificial Intelligence, 2011.

[42] Marcos Cramer and Mathieu Guillaume. Directionality of attacks in natural language argumentation. In CEUR Workshop Proceedings. RWTH Aachen University, 2018.

[43] Iyad Rahwan, Mohammed I Madakakatel, Jean-François Bonnefon, Ruqiyabi N Awan, and Sherief Abdallah. Behavioral experiments for assessing the abstract argumentation semantics of reinstatement. Cognitive Science, 34(8):1483–1502, 2010.