UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL AND APPLIED SCIENCES

Electronics and Computer Science

An Investigation into Dialectic and Eristic Argumentation on the Social Web

by

Tom Blount

Thesis for the degree of Doctor of Philosophy

August 2015

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
Electronics and Computer Science

Doctor of Philosophy

AN INVESTIGATION INTO DIALECTIC AND ERISTIC ARGUMENTATION ON THE SOCIAL WEB

by Tom Blount

Argumentation, debate and discussion are key facets of human communication, shaping the way people form, share and promote ideas, hypotheses and solutions to problems. Argumentation can broadly be broken down into collaborative problem solving or truth-seeking (known as dialectic argumentation) and quarrelling without hope for a resolution, either aggressively or for the purpose of recreation, catharsis or entertainment (known as eristic argumentation). Techniques used within argumentation can likewise be classified as primarily fact-based (logical), or emotion/audience-based (rhetorical).

The social web, consisting of the people, tools and communities that form over the world wide web, is a growing way in which individuals, social groups and even corporations share content, ideas and information, as well as hold discussions and debates. As the social web becomes more widely used, the potential for using it as a means to study how people communicate and collaborate on an enormous scale dramatically increases. Current models of argumentation often focus on formal argumentation techniques, in which participants are expected to abide by a stringent set of rules or practices. However, on the social web there is no such code of conduct. Antisocial behaviour, which often stems from argumentation, can have a negative impact on online communities, driving away new users and stifling participation.

Case-studies were carried out on three different areas of the social web to determine the strengths and weaknesses of modelling social, eristic argument on the web when using current models and ontologies. This preliminary work indicates that existing techniques for modelling argumentation are insufficient to capture the structure and dynamic of argumentation taking place on the social web.

Following this, augmentations were made to current modelling ontologies for the purpose of capturing a sub-set of rhetorical tactics. These were then used as part of an investigation re-examining the previous case studies to determine the prevalence of rhetorical tactics in argumentation within areas of the social web as well as investigating correlations that may be drawn between the use of these tactics and the machine-readable characteristic of the post (e.g. length, word-complexity, etc.). It was found that even this small sub-set of rhetorical tactics was regularly employed throughout each case study. Correlations between tactics and post features were also found, although these were not conclusive due to the discrete and binary nature of the features examined.

Based on these observations, future work will focus on extending the argumentation model further to capture additional rhetorical information, due to the importance of these types of interactions. This will be used in an experiment to analyse how different rhetorical and eristic features impacts users of the social web participating in discussion, and how this affects their perceptions of the topic and engagement with the argument. The results of this can then be used to further supplement the model of eristic argumentation on the web.

Contents

List of Figures

List of Tables

Acknowledgements

Thanks to both my supervisors David Millard and Mark Weal for their guidance, advice and support throughout this work. My thanks also go to my partner, for her love and support, and to my friends, who have encouraged and assisted me on my journey. Last, but not least, thanks go to the EPSRC for providing the funding for this work.

Chapter 1

Introduction

1.1 Problem Space and Motivation

Argumentation is fundamental to human communication – it is how people share new information and new ideas, and propose courses of action that see them carried out (**??**). As a result, there is a large amount of research on argumentation from a wide variety of disciplines and topics, including: philosophy, and the nature of fallacies and how they may be critically appraised (**?**); sociology, and the need to differentiate between classical logic and social argumentation due to the need for the capability to reason using only partial knowledge (**?**); law, and the need for measures of certainty and belief when modelling and reasoning over assertions (**?**); and artificial intelligence, and the use of agent-based systems such as dialogue games, as methods for reasoning over argument (**??**).

Argumentation can be (broadly) separated into two categories based on the goals and intended outcome. Firstly, dialectic argument, in which the participants are engaged in rational discourse with the aim of either discovering the particular truth behind a matter, or formulating a solution or resolution for a set of circumstances (**?**). Secondly, eristic argument, in which there is no clear goal and the participants are not trying to come to a resolution but are quarrelling with the aim of being seen to win, either in the eyes of their opponent or, more often, in the eyes of spectators (**??**). Arguments can shift between these two forms, or contain "pockets" of one form within the other. Orthogonally to this, there are the notions of logic and rhetoric. While often used in modern parlance as a pejorative term, rhetoric is simply the art of discourse, and convincing an audience to one's point of view based on one's knowledge of the topic at hand and, crucially, one's knowledge of the audience themselves (which clearly lends itself to the eristic form) whereas logic deals with reasoning between established facts (which lends itself to the dialectic form).

However, as **?** note, *"perhaps out of fear of metaphysics or of 'psychologizing,' present-day logicians tend to concentrate exclusively on formalized arguments that lack any direct relation*

with how argumentation is conducted in practice." Social argumentation, or the way people argue day-to-day, often has a very different structure to formalised models. In these instances, the aim of a proponent is not to prove themselves right through irrefutable logic, but simply to make others believe that they have proved themselves right. Schopenhauer, in his satirical work *The Art of Always Being Right*, emphasises that *"A man may be objectively in the right, and nevertheless in the eyes of by-standers, and sometimes in his own, he may come off worst"* (**?**).

This is particularly relevant when applied to the social web. As a network of social relationships that are created, formed and maintained through the world wide web, the social web (and the social media presented across it) are rife with discussion, debate, and argumentation (**?**). As the web (and in particular the number of people, tools and communities that make up the social web) grows and becomes totally ubiquitous (**?**, p. 559), the potential for using it to investigate how truly massive communities interact, communicate and argue increases dramatically. However, the social web presents a number of challenges for extracting and analysing arguments, particularly due to the lack of clear indicators of argument structure. This problem is compounded by the type of language used; often highly informal, incorporating slang and irregular punctuation and grammar (**?**), and by the number of distinct ecosystems on the social web, each with their own constraints and cultures (**?**).

There are a number of challenges when considering maintaining the social web as an inclusive platform for diverse and vibrant content, especially debate and discussion. There is a tendency for users to interact and associate with others who are similar in terms of traits, (such as race, age, or education) and beliefs (such as religion or politics), known as homophily (**?**) and is compounded by the introduction of "filter bubbles", the effect of content providers tailoring search results or default displays towards the preferences of individual users (**?**). This can lead to sites becoming "echo chambers" in which well-known views and opinions are repeated, little original content is produced and there is virtually no dissent or debate (**?**). This can be further exacerbated by reputation systems, enforcing which views are acceptable in a given community by rewarding users who agree and punishing those who disagree, or those considered "outsiders" of the accepted group or culture. At the opposite end of the spectrum, where there is constant and stimulated debate, there is equal (if not greater) potential for conflict. While critical and reasonable debate, and even (respectful) recreational quarrels, are things to be encouraged, there is a visible tendency to "shout down" the opposition, including attempts to silence dissenting opinions through abuse and threats. As a result online communities can become incredibly hostile spaces, culminating in anti-social behaviour, including vulgar abuse and, at the most extreme, threats of sexual violence, and death threats (**??**).

However, in this document the case is made that disregarding these interactions from argumentation models is a mistake. Accurately modelling them is the first step towards understanding exactly how argumentation is applied across the social web, and the ways in which creators and consumers of social media engage with argumentation. This information can then be applied towards creating tools and environments that discourage these types of abuse to facilitate more social argumentation.

1.2 Aim & Research Contribution

One key feature of social argumentation is the notion of the (presence of) an audience (**??**). The audience's perception of the argument is something that is often overlooked in formal models of argument, despite evidence that perception of argument can be altered through multiple means such as cultural associations (**?**), pre-existing biases (**?**) or peripheral information(**?**). The ultimate aim of this research is to explore how perception of argumentation specifically on the social web can be altered based on the types of tactics used, and how this can be used to develop more thorough models of argumentation. To achieve this, it is first important to be able to correctly model and represent the arguments that occur socially. In this way, the key features of informal arguments can be identified and categorised. This can then be used to determine exactly which features of argumentation are considered most important by users, and those that they are most likely to engage, reply to, critique, and how these features shape users' overall interpretation of an argument. The work described in this report examines how formal models currently map arguments, and applies a combination of these models to an argument (or arguments) on the social web to determine which features are well captured, and those that are not. This has led to the formulation of a hypothesis that the presence of particular rhetorical tactics affects both a user's perception of an argument, and the way in which they engage with it (described further in Section 5.2.1), which will be the focus of future work.

1.3 Report Structure

Background information on the topic area, both in argumentation and online behaviour, as well as the state of the research field at present, is discussed in Chapter 2. A preliminary investigation into the capabilities of current models of social argumentation, and an analysis of the results, is detailed in Chapter 3. This is expanded in Chapter 4, in which these models are developed and adapted to encompass further social and rhetorical information, and are used to examine the prevalence of a subset of rhetorical tactics in web-based argumentation and their correlation with machine readable features (such as post length, language, etc.). Finally, Chapter 5 sums up the findings and outlines a plan for a final thesis and future work.

Chapter 2

Background and Literature Review

2.1 Rhetoric and Argumentation

Rhetoric is often used in modern parlance as a pejorative to describe persuasive language that lacks substance, or containing empty or insincere promises; formally, however, it refers to the art of persuasion, whether spoken or written. In particular, it focuses on the act tailoring one's argument to the situation at hand based on knowledge of events and, crucially, knowledge of the audience (**?**).

2.1.1 Modes of Persuasion

Aristotle, in his treatise on rhetoric, described three "persuasive modes" that can be employed in an attempt to sway an audience: through the words that are used (*logos*), through the character of the rhetor or their opponent (*ethos*), and through the emotions of the audience (*pathos*) (**?**). These modes may be applied individually, or in conjunction with one another. *Logos* describes an appeal to logic or reason. This is the method by which one might rationalise a position, often backing it up with evidence or statistics. It is important to note that, when enacting *logos*, it is not strictly necessary for the logic to be sound, or the evidence provided to be factual – it can be warped to fit a particular purpose, or even outright fabricated (however, this will usually also invoke another of the modes described below). The key element is that it appears to be reasonable and thus, appeals to an audience's sense of reason (**??**). *Ethos* is an appeal a person's character or sense of ethics and morals. This can be used in an attempt to strengthen the position of the rhetor's argument or to weaken their opponent's position. For example, if a rhetor can state that they are an expert in the field that they are debating then it is likely their audience will lend their argument more credence than if they were a novice. This specific case is known as an argument from authority, or *argumentum ab auctoritate* (**??**). Similarly, an argument can be made that attacks an opponents position indirectly, by attacking their credentials rather than refuting their claims (*argumentum ad hominem*). Although such an argument is not logically

sound (and constitutes a fallacy), it is still often used in practice and in certain circumstances is a viable (and often effective, if somewhat underhand) means of persuading an audience (**?**). Finally, *pathos* is an appeal to emotion, whereby an attempt is made to evoke a particular feeling in an audience in the hope that this will influence their opinion on a position. This can be done in both positive and negative terms. For example, flattering an audience, or promising them a boon, can shift them towards accepting a particular course of action. On the other hand, threatening them with the potentially undesirable consequences of their actions can cause them to reconsider even if these consequences are unlikely or, indeed, impossible. A classic example is the appeal to fear (*argumentum ad metum*) (**??**).

2.1.2 Dialectic and Eristic Argument

The terms dialectic and eristic were coined in Ancient Greece to describe modes of argumentation with different goals and were popularised in Plato's *Republic* (**?**). A dialectic argument takes the form of two or more parties engaged in rational discourse with the aim of either discovering the particular truth behind a matter, or formulating a solution or resolution for a set of circumstances (**?**). For example, an academic presenting their findings and rationalising that they are indeed valid, given the rigorous methodology they have used and the weight of evidence this has provided is an example of a dialectic argument. Likewise, a peer reviewer that disagrees with the findings by pointing out a specific flaw in the experimental methodology and explaining how this should be resolved, is another example. The arguments tend to rely heavily on the mode of *logos*. In contrast, an eristic argument is an argument in which there is no clear resolution in the minds of the participants: they are not motivated by solving a problem, or convincing their opponent, but to be victorious (**?**). There may be different reasons for arguing in this vein, from quarrelling for its own sake as a form of catharsis (**?**), to being seen to "win" the argument in the eyes of spectators (**?**). As a result, these arguments chiefly favour the modes of *ethos* and *pathos*.

2.1.3 Modelling Argument

There are many different models and frameworks used to capture particular aspects of argumentation. These aspects include notions of trust (**?**, p. 752), focus on argument topic or chronology (**?**) and the ability to demonstrate support for or refutation of other points in the structure (**?**). Some examples are discussed below, with respect to their technical structure, their influence in the field and their practical applications.

2.1.3.1 Toulmin Model

Toulmin developed his model from the school of philosophy in the 1950s as a means of demonstrating an approach to practical (rather than theoretical) argumentation, by attempting to show

the internal structure (and thus, consistency) of an argument (**?**). The general form of Toulmin's argument, shown in Figure 2.1, follows the structure of a *claim*, or conclusion, that is backed up with generally agreed upon facts (the *data*). The *claim* can be *qualified* ("definitely", "maybe", "probably", etc.) and any potential *rebuttals* accounted for. Then, key to the Toulmin model, the *claim* and *data* are connected using either an implicit or explicit *warrant*, or justification – this can then be supported by a particular *backing* (**?**, p. 347-350). A specific example can be seen in Figure 2.2, which shows an argument reasoning that Alice is a British citizen. Toulmin's model has been a particularly influential piece of work and has had an impact of decades of argumentation research in fields as far ranging as law, rhetoric and education (**?**, p. 8-10; **?**, p. 5, 12). However, there has been discussion as to the effectiveness of different aspects of the framework. In its favour, the means of explicitly stating the connecting warrant (and associated backing) can improve cross domain discourse. On the other hand, because models themselves are focused towards internal structure, there is no criteria for modelling overall structure (such as a group of arguments that refutes or support one another's claims). There is also no concept of resolving an argument (for example, on the grounds of logic or value); although this may have been by design, it negates the possibility of evaluating the strength of a given argument (**?**, p. 349-350; **?**, p. 5, 12). Among other applications, the Toulmin model has been incorporated into the Argument Model Ontology[1], an OWL ontology to allow classification of academic arguments. This is used in conjunction with CiTO, an ontology for factually and rhetorically categorising citations (**?**, p. 8).

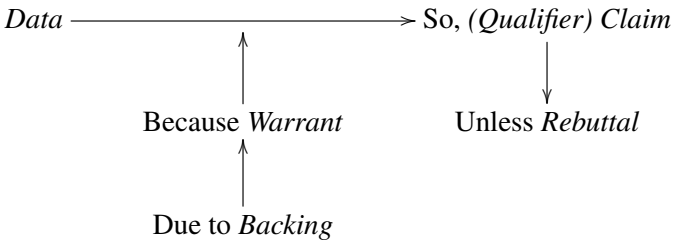


Figure 2.1: General form of Toulmin's diagram (**?**, p. 104)

[1] http://www.essepuntato.it/2011/02/argumentmodel

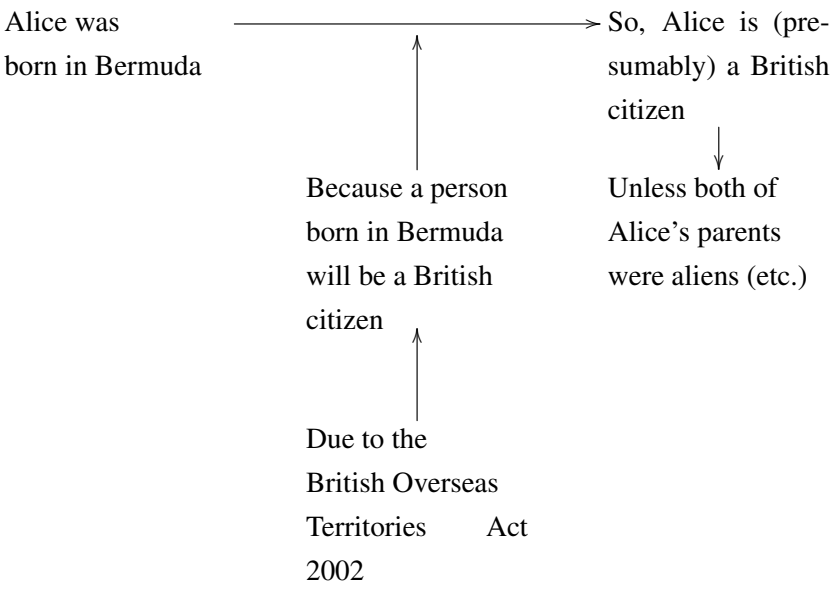


Figure 2.2: Example usage of Toulmin's diagram (**?**, p. 105), examining whether Alice is a British citizen

2.1.3.2 Information-Based Issue Systems

Information-Based Issue System (IBIS) models are a particular type of dialectic process originally designed to aid in solving so-called "wicked problems" (**?**) – problems of social policy to which there is no clear definition, methodology or even end-goal (**?**). IBIS models are represented as trees, made up of four different types of node. Firstly, *Issues* represent the problems that need to be solved, or questions that must be answered. Generally, there is one "root" *Issue* to be deliberated, but other sub-*Issues* can be created as necessary during the reasoning process. *Ideas* are proposed solutions or answers to these *Issues*, and each *Idea* can then be weighted positively or negatively using *Arguments For* and *Arguments Against*. IBIS models have seen wide usage in the field of design rationale and cognitive ergonomics where the assimilation of collective knowledge is required to solve problems (**??**). An example usage of an IBIS model is shown in Figure 2.3. Because of its dialectic context, the application of IBIS models is ideal when two or more parties are trying to resolve a complex problem, especially if they have differing (or even opposing) stakes. As might be expected, there are many IBIS-like systems used in system-design and knowledge aggregation. Delibatorium[2] is a tool, developed by **?** of MIT, that uses an IBIS approach to solving challenging problems such as *"Is carbon offsetting a good idea?"*. The IBIS approach is invoked to aid the collaboration of large amounts of people separated across space and time by preserving a topic-centric (rather than time-centric) structure.

[2] http://deliberatorium.mit.edu

IBIS structures have also been included in an extension to the SIOC ontology devoted to representing argumentation[3]. This ontology uses the IBIS notation of *Issues* and *Ideas* to formalise the process of solving a problem over social media (**?**).

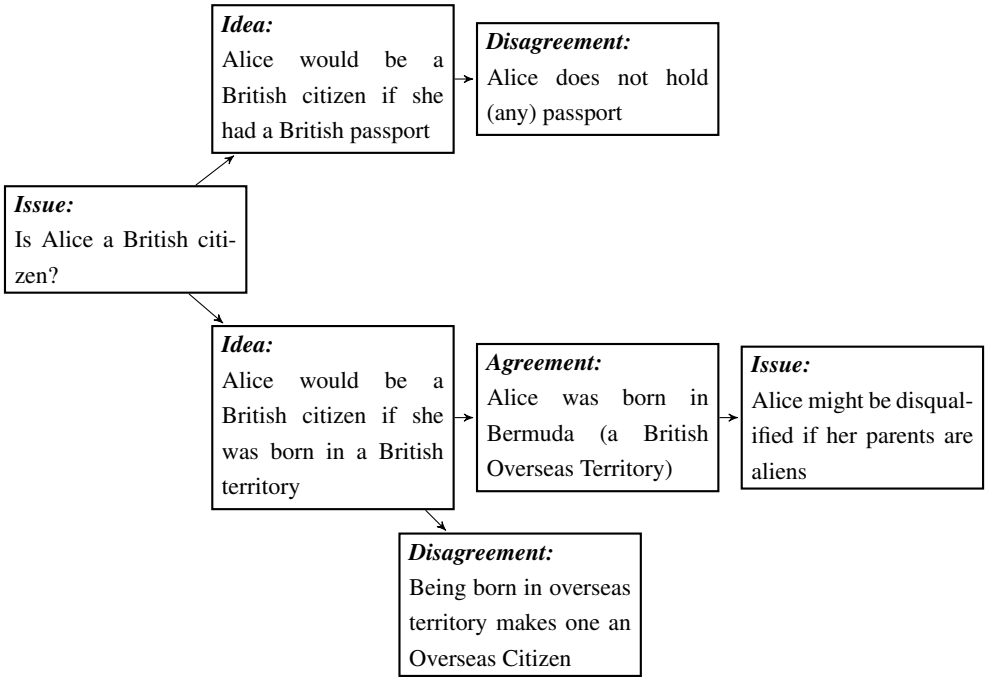


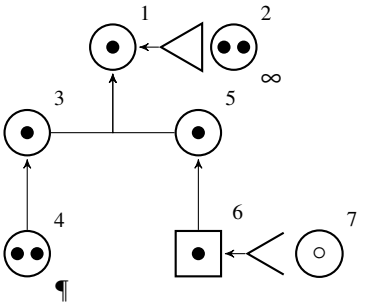
Figure 2.3: Example usage of an IBIS model, examining whether Alice is a British citizen

2.1.3.3 Wigmore's Chart

"Wigmore's chart", conceived in 1913, is a means of recording argumentation originally devised for use in legal trials. The chart models the chain of interactions between competing arguments from both participants and can be used to evaluate the overall conclusion that should be drawn (**?**, p. 751). It takes the form of a directed graph where each node represents a particular fact. The shape of each node relates to the nature of the assertion; squares represent testimony given under oath; a triangle represent an explanation of or support for the node it "points" to; an open angle refutes the argument it points to and all other assertions (such as claims, physical evidence or related legal statutes) are represented by circles. These can additionally be marked to denote arguments by the defence or prosecution, but are not discussed here for clarity (**??**). Symbols relate further information about the nature of these assertions: an infinity symbol (∞) states that a node denotes sensory evidence that may be (re)produced in court; a pilcrow (¶) denotes an assertion that can be taken as fact with no further evidence (such as a precedence case); a lack of a symbol shows that the claim is implied from further reasoning in the graph. In addition, Wigmorean analysis can incorporate the notions of *strong belief* (●●), *belief* (●) *doubt*

[3] http://rdfs.org/sioc/argument

(?) *disbelief* (○) and *strong disbelief* (○○) (**?**, p. 751-756; **?**). Little is known about precisely how often this type of analysis is used manually, although it is thought that it is carried out in courthouses around the world (**?**). However, efforts are being made to automate the process by parsing the natural language propositions made in court and transforming these into a Wigmore diagram to aid judges, barristers and juries in their deliberations (**?**).



1 Alice is a British citizen

2 Alice has a British passport

3 A person born in a British territory will be a British citizen

4 British Overseas Territories Act 2002

5 Alice was born in Bermuda

6 Alice's parents testify that she was born in Bermuda

7 Alice's parents' testimony could be biased in her favour

Figure 2.4: Example Wigmore graph, examining whether Alice is a British citizen

2.1.3.4 Dung's Framework

Similar to Wigmore's method, Dung's framework (which uses the format of set theory) focuses on the aspect of arguments attacking, (implicitly) supporting and, ultimately, defeating one another (**?**). Dung defines an *Argument Framework* as a pair such that $AF = \langle AR, attacks \rangle$ where AR is a set of arguments $\{a_1, a_2, ..., a_n\}$ and $attacks$ is a binary relation such that $attacks \subseteq AR \times AR$. $attacks$ describes which arguments are "defeated" by one another: for example, if a_1 is the argument "Alice is not a British citizen" and a_2 is the argument "Alice has a British passport" then $(a_2, a_1) \in attacks$. The set of *conflict free* arguments is a maximal set of arguments that do not attack each other. An argument a_1 is *acceptable* with regard to a set of arguments S if there is no argument a_2 that attacks a_1 that is not itself attacked by an argument in S. A set of arguments is *admissible* if each argument is considered *acceptable* with respect to the set. The maximal *admissible* set is known as a *preferred extension* (**?**). **??** has extended this framework to incorporate the idea of "value" or principle to arguments. When circumstances arise such that two possible resolutions to a dispute are equally (logically) valid, different audiences will have differing preferences based on the principles they feel are most important. For example, say that two solutions for combating crime are put forward: reading the general public's private correspondence or an expensive social program of education and rehabilitation. If each has been proven to be equally effective, audiences that value minimisation of cost may favour the former whereas audiences that value individual privacy might choose the latter.

2.1.3.5 The Argument Interchange Format

The Argument Interchange Format (AIF) is a framework for representing argumentation as a directed graph (**?**). Created as part of the Argument Web project (**?**), which aims to link the concepts of natural language argumentation with abstract mathematical modelling (including capturing *"linguistically sophisticated manoeuvres"* (**?**)), the AIF is primarily a description, with specifications in a number of languages including RDF and SQL.

At its highest level, the AIF can be conceptually divided into an "upper" ontology and a "forms" ontology. The upper ontology consists of the building blocks of the argument structure, while the forms ontology applies context, for example, by differentiating between logical attacks based on faulty evidence, witness bias, or appeals to authority. The data, claims and conclusions that make up the argument are modelled by Information nodes (I-nodes). There can be no direct relationship between I-nodes. Instead, there must be an intermediary Scheme node (S-nodes). These S-nodes are subdivided into three applications: Rule of Inference Applications (RA-nodes), Conflict Applications (CA-nodes) and Preference Applications (PA-nodes). RA-nodes and CA-nodes simply denote an inference or conflict (logical or otherwise) between one or more pieces of information. PA-nodes, however, denote a preference of one piece of information over another. For example when discussing economics, while it may be difficult to logically prove the superiority of a regulated market over a free market, or vice-versa, the personal beliefs and preferences of proponent and opponent will feature heavily in their reasoning on such issues (**?**). This structure is displayed in Figure 2.5.

In their work on an extension to the AIF, dubbed the AIF+, Reed et al. build on the work of O'Keefe to differentiate between two separate notions of argumentation (**??**): the first, which they term argument$_1$, is a logically constructed set of claims and evidence used to back these claims (or attack other claims), as in *"Alice put forward her argument"*. The second, termed argument$_2$, refers to a dialogue – the exchange of ideas and opinions between two or more people, as in *"Alice and Bob were having an argument*. A result of this work was to introduce a new set of nodes. The first, a subset of I-nodes dubbed Locutions (L-nodes), model locutionary acts (or utterances) in an argument$_2$. That is, they record precisely what was said. The second, a subset of S-nodes dubbed Transition Applications (TA-nodes), represent transitions between L-nodes (with associated forms such as a challenge or response). Thirdly Illocutionary Applications (YA-nodes), also a subset of S-nodes, represent the "illocutionary force" and serve to link each argument$_1$ to the overall argument$_2$. Figure 2.6 shows how this structure can be visualised. Consider the locution *"All men are mortal, and Socrates is a man. Therefore, Socrates is mortal."* The statement itself is modelled using the L-node on the rightmost side of the diagram. On the leftmost side is the core AIF structure, which show the premises formed as two I-nodes (*"Socrates is a man"* and *"All men are mortal"*), linked to the conclusive I-node (*"Socrates is mortal"*) by way of an RA-node. The L-node is connected to this argument$_1$ by way of the YA-node, shown in the middle.

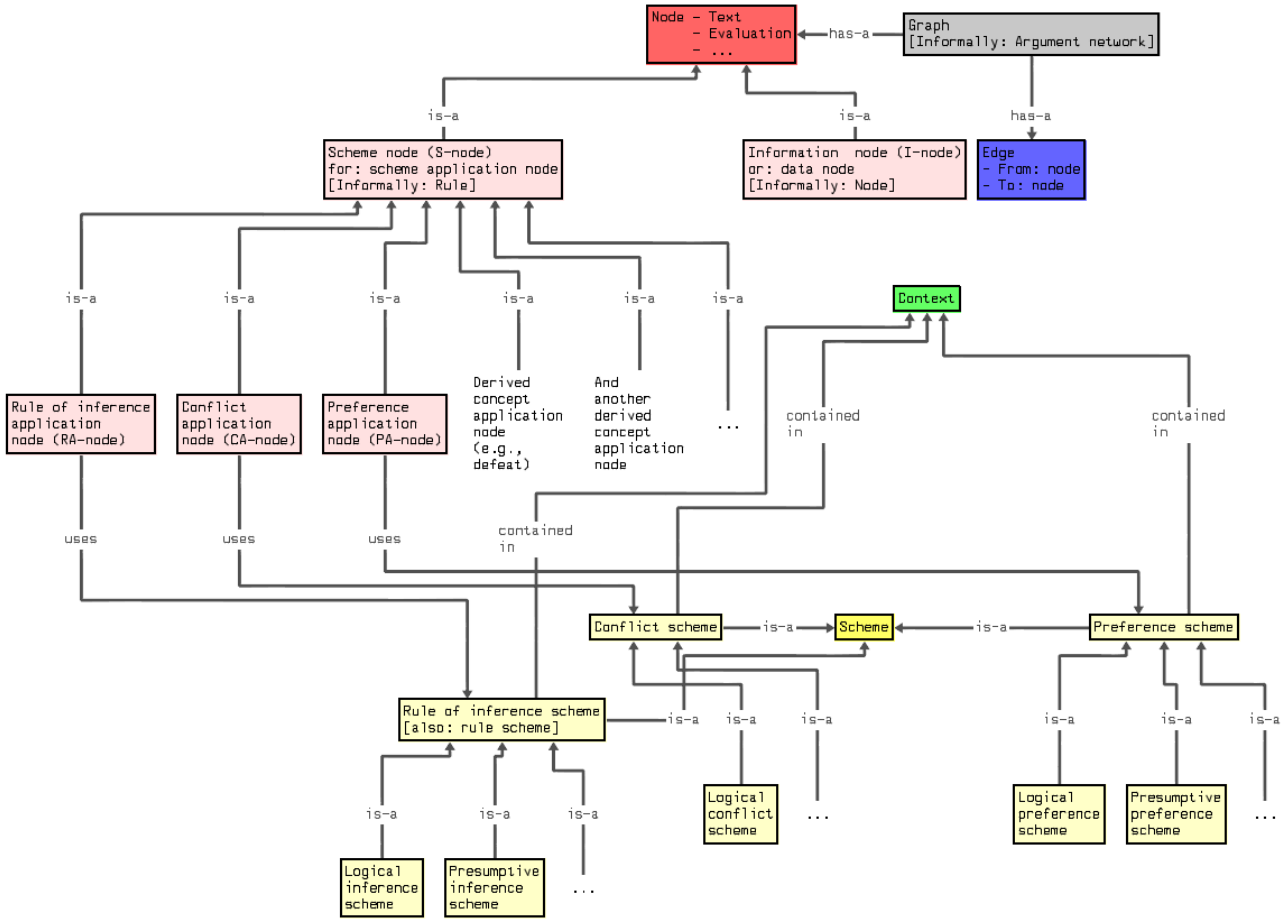


Figure 2.5: An overview of the AIF Ontology (**?**)

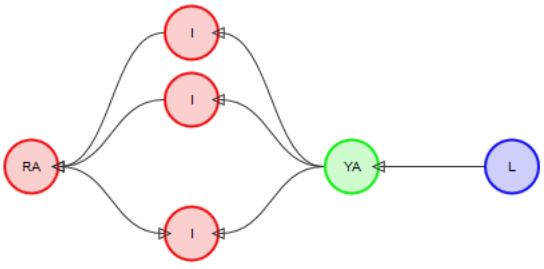


Figure 2.6: Visualisation of a simple AIF+ graph

2.2 Online Communication and Interaction

2.2.1 Social Media and the Social Web

The social web consists of the people, tools and communities that form over the world wide web, and is a way for individuals to share content, ideas and information. The social web presents a number of challenges for extracting and analysing arguments, particularly due to the lack of clear "indicators" of argument or structure. This problem is compounded by the type of language used; often highly informal, incorporating slang and irregular punctuation and grammar (**?**). As the social web becomes more and more ubiquitous, the potential for using it to investigate how truly massive communities interact, communicate and argue increases dramatically.

Many theoretical models of argumentation are based on the assumption of a dialectic argument, as their purpose is to aid the participants with the process of understanding the information discussed, or to reason over the model and draw conclusions regarding the outcome. However, in social media there is a clear proliferation of eristic argumentation (**?**). This makes the role of audience an important feature to consider: when an individual responds to a post on the social web their post is often seen not just by the author of the post they reply to, but by many other users as well. In fact, many posts may be directed at this wider audience to seek approval, voice dissent, or provoke other emotions (**?**). Consider the analogy of a political hustings: neither candidate believe they can change the mind of their opponent, but instead are debating with a view to sway their audience. Schneider et al. note though, that currently it is difficult to model the value of eristic arguments as participants are free to *"sling propositions that they would not commit to under other circumstances"* as a means of catharsis, recreation or entertainment (**?**).

? classify six distinct categories of social media: collaborative projects, blogs, content communities, social networking sites, virtual game worlds and virtual social worlds. Collaborative projects allow many different users to create, maintain and often discuss content. This category includes sites such as the online encyclopaedia *Wikipedia*[4], which allow users to write and edit articles and *Urban Dictionary*[5], a user generated dictionary of slang and internet culture. **?** compare blogs (web-logs) to personal websites, in that they allow users to post information about the subject of their choice – these posts are often timestamped and presented reverse-chronologically. *Wordpress*[6] and *Blogger*[7] are two social media sites specialised for this purpose. "Micro"-blogging sites that pose limits on the amount of content that can be shared in a single post, such as *Twitter*[8], also fall into this category. Content communities revolve around the concept of publishing (and ultimately sharing) different forms of media. These include

[4] https://en.wikipedia.org/
[5] http://urbandictionary.com/
[6] http://wordpress.com/
[7] http://blogger.com
[8] http://twitter.com/

sites for publishing video (such as *Vimeo*[9]), images (such as *Flickr*[10]), audio (such as *SoundCloud*[11]) and many other different types of media. Social networking sites allow users to create a profile detailing information about themselves (such as home town, or music preferences) and then connect their profiles with the profiles of others on the site. Examples include *Facebook*[12] and *Google+*[13]. Virtual game worlds (such as *World of Warcraft*[14]) encompass online games in which a user controls a digital avatar to accomplish certain tasks (such as slaying a virtual dragon, or defeating another player's avatar). Similarly, virtual social worlds (such as *Second Life*[15]) encompass virtual spaces in which users have an avatar, but there is no specified aim or end-goal – the medium exists solely to facilitate social interaction. In this work, less focus is afforded to these latter two areas of the social web due to the the issue that as participants are controlling a virtual avatar, and may be playing a particular "role" rather than their real self, this can affect their behaviour and engagement in a discussion **?**. There is also the tendency for discussions to centre on the mechanics of the game world itself (**?**).

2.2.2 Anti-Social Behaviour

Anti-social behaviour is a growing problem on the social web, and often arises from debates or discussions that get out of hand (**???**). This behaviour can arise from simple misunderstandings due to the difficulty in conveying tone through text, or as a deliberate act by individuals lashing out at other participants in a discussion. Incidents include flaming, in which a user simply hurls emotional abuse (**?**, p. 13); spamming, in which a user floods the medium with content, often unrelated to the topic in hand, in the hope of drowning out other participants or as a means of advertising a commercial product (**?**); trolling, in which a user posts seemingly innocuous but deliberately fallacious argument to provoke other members of the group into becoming outraged (although there is debate as to whether this term refers to the bridge-dwelling monster of myths, or the fishing term for dangling a baited line behind a boat) (**?**); and much more serious incidents of directed threats and stalking (**???**).

As a result, there is a concerted research effort into the best way to tackle these issues before they cause serious harm to individuals, or the field as a whole. **?** discuss a wide variety of approaches (specifically in regard to the virtual social world *The Palace*[16], but these could be applied to other online spaces as well). The simplest solution is to moderate users' interactions and dispense warnings, "mutes" (where a user may observe, but not contribute) or, in extreme cases, bans as and when the situation warrants. While effective for dealing with small or close-knit communities, this approach does not scale when considering the social web.

[9] http://vimeo.com/
[10] http://flickr.com/
[11] http://soundcloud.com/
[12] http://facebook.com
[13] http://plus.google.com/
[14] http://battle.net/wow/
[15] http://secondlife.com
[16] http://thepalace.com

A different approach is to allow the community a degree of self-moderation. Reputation systems, for example, allow users within a community to assign "votes" to a particular account, or post, to show its trustworthiness. This allows new users to make judgements on whether to take a comment seriously, for example, or to purchase something from a particular seller in an online auction (**??**). However, this can also lead to a feedback loop in which communities become self-reinforcing; if users always vote for posts of similar sentiment (or against those that disagree), then gradually these sentiments will become dominant. Over time only users who hold these views will contribute to the site (further reinforcing the disparity) and the community as a whole will stagnate or worse, become distrustful or outright hostile to new members or "outsiders".

In another example of direct self-moderation, the popular online game *League of Legends*[17] implements a "tribunal" system in which players that are reported for poor behaviour in matches (such as verbally abusing team-mates) are judged by their peers. These peers can examine evidence such as chat logs and game scores, then decided whether to "pardon" or "punish" the offending player (**??**).

A more covert attempt to manipulate users' behaviour can be found in certain implementations of human-computer interaction design. HCI can be leveraged to "trick" users into performing (or not performing) an action desirable to the designer. These so-called "malicious interfaces" (**?**) are often used to trick users into spending time or money that they otherwise would not (for example, advertising banners that suddenly cover page content). In 2008, YouTube temporarily added an "Audio Preview" button to its comment system that would read aloud what the user intended to post. This was placed in the previous place of the "post" button (which had been moved further to the right), such that a user was likely to unintentionally preview their comment before posting it (**?**).

2.2.3 Semantically-Interlinked Online Communities

The Semantically-Interlinked Online Communities project (SIOC) aims to enable the cross-platform, cross-service representation of data from the social web (**?**). SIOC allows for semantic representations of Sites, which hold Forums, which contain Posts, authored by the owner of a UserAcount. This structure is shown in Figure 2.7. SIOC is often used in conjunction with the Friend of a Friend (FOAF) ontology, to show how individuals map to their online personas.

While an extension to SIOC, for the purposes of capturing and representing argumentation, does exist (**?**), it is based on the Issue Based Information System (IBIS) principals of modelling an argument as an issue that needs to be solved, with users suggesting ideas, then providing arguments for or arguments against these ideas. While this approach is highly useful when dealing with arguments centred around deliberation, and to a lesser extend criticism or inquiry, they are not as suitable when modelling negotiations or eristic arguments.

[17] http://leagueoflegends.com
[18] http://sioc-project.org/ontology

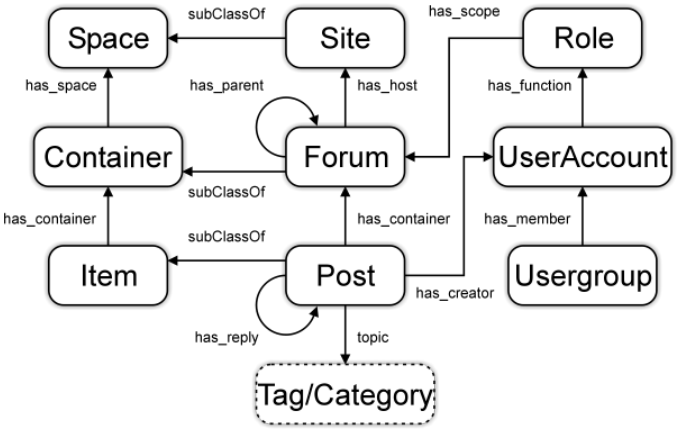


Figure 2.7: An overview of the core SIOC ontology[18]

Chapter 3

Investigating Current Modelling Frameworks

3.1 Purpose and Approach

To determine how capable current tools and frameworks are for capturing social argumentation, and the nuances between dialectic and eristic argumentation, a preliminary investigation was conducted. This aimed, firstly, to show how these tools and frameworks can be combined in a way that makes them fit for this particular purpose and, secondly, to determine the key strengths and weaknesses of this combination in relation to modelling social argumentation.

The AIF was determined to be the closest fit for purpose ontology for modelling argumentation on the social web, due to the goals of capturing practical, language-based argumentation, with the additional benefit of being readily extensible. Alongside the SIOC, the key elements of these ontologies have been combined to explicitly capture the social component of argumentation on the social web, while also modelling the formalised argument structure. This is achieved by linking the concept of a SIOC Post with that of an AIF Locution, treating a social web thread as a separate dialogue, or argument$_2$ and each post as an atomic unit within the dialogue (containing zero or more individual arguments$_1$). In the majority of cases, a single locution will translate to a single self-contained argument$_1$. However, a single post can contain a number of arguments$_1$ – each with a number of premises and a single conclusion. In this situation a single L-node will link to multiple YA-nodes, as shown in Figure 3.1. In rare cases (often caused by constraints imposed on the length of a post by the service, such as the 140 character limit on Twitter), a user will spread the premises of a single argument across multiple posts to construct their argument$_1$. Figure 3.2 shows how, in such a situation, multiple L-nodes will link to a single YA-node. If two users post identical statements, they still contribute two distinct locutions. However, they will both be linked to the same I-node(s), and therefore the same argument$_1$. In this situation, multiple YA-nodes may point to the same I-node, such as in Figure 3.3.

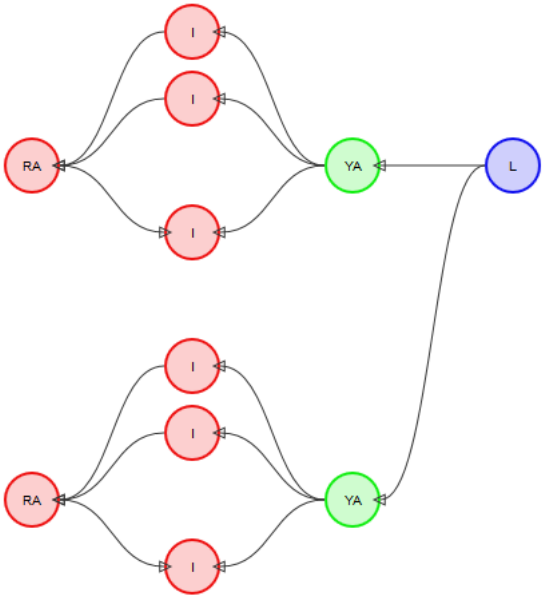


Figure 3.1: Visualisation of one post making two distinct arguments[1]

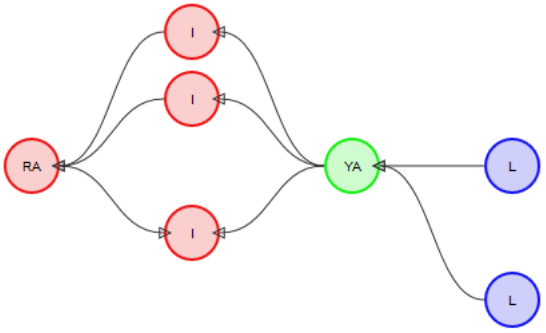


Figure 3.2: Visualisation of two posts, used to construct a single argument[1]

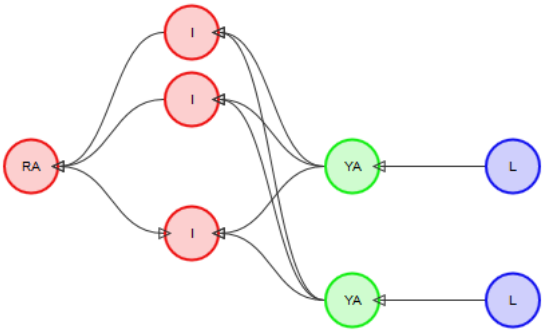


Figure 3.3: Visualisation of two posts, repeating the same argument[1]

3.2 Methodology

3.2.1 Data Collection

A single topic of argumentation was chosen to be examined for three case studies, each representing a different social ecosystem. To ensure the stimulation of debate, the selected topic needed to be controversial, have a large number of respondents and have been active for a long enough period of time to generate a rich and complete content. The October 2013 United States government shutdown caused by Congress's failure to agree on a budget, and the following condemnation this received from the presidency, was a suitable match for these requirements.

This topic was then tracked across three of the social media categories identified by **?**: Twitter, a microblogging service that allows users to publish messages of up to one-hundred and forty characters; Facebook, a social network, that allows users to create a network of "friends" and share text or images; and YouTube, a content creation site where users can create and upload videos, or playlists of videos.

The source of the posts themselves again needed to be both publicly available and have a large number of followers to ensure a maximally stimulated debate. As an authoritative public figure at the heart of the crisis, content from or relating to Barack Obama's social media profiles was chosen, and three posts that were broadly similar in content were selected for study. The first post, initially posted on 8th October 2013 from the White House's YouTube channel[1], is a 14m 40s video recording of Obama delivering a statement to press from the West Wing of the White House, condemning the shutdown. The post taken from Obama's official Twitter account[2] (which is managed by a third party, Organizing for Action), dated 15th October 2013, reads: *"This is unacceptable. Tell Tea Party Republicans to stop holding our economy hostage: http://OFA.BO/qNmA3Y"*. The included hyperlink leads to an Organising for Action page, which encourages users to to voice their displeasure at the shutdown by allowing them to automatically generate and send tweets. The post taken from Obama's official Facebook account[3] (also managed by Organizing for Action), also dated 15th October 2013, reads: *"Tea Party Republicans in the House of Representatives forced a government shutdown, and now they're threatening an economic shutdown. This has gone on for too long. Tell them to #EndThisNow: http://OFA.BO/ACC7qB"*.

The discussions surrounding these posts were acquired by collecting comments replying to each initial post, and those replying to subsequent posts in the discussion (taking into account only direct replies, rather than mentions within the text of the post), with the use of the public Twitter, Facebook and Youtube APIs respectively. This data was translated to an RDF triple-store using SIOC to record the data specific to the social media platform, such as which User created which

[1] https://www.youtube.com/watch?v=7LwoudGfug0
[2] https://twitter.com/BarackObama/status/390288744235823104
[3] https://www.facebook.com/photo.php?fbid=10151874920756749

Table 3.1: Metrics of total dataset collected from YouTube, Twitter and Facebook

Metric	YouTube	Twitter	Facebook
Total number of posts	2719	137	9494
Total number of users	1255	33	6224
Average posts per user	2.17	4.15	1.53
Average words per post	26.74	15.91	40.12
Average characters per post	150.13	97.63	241.14
Time between first and last posts	101d 16h 19m 12s	0d 13h 40m 48s	90d 19h 55m 12s
Average time between posts	53m 52s	3m 2s	13m 47s

Post and which Thread stores which Posts. This was used in conjunction with the DCTerms ontology, which held supplementary data such as timestamps.

3.2.2 Data Sampling and Annotation

Because of the volume of the data produced over the course of the tracked event and the time-intensive nature of manually annotating the data, it was necessary to sample the data to a more manageable size before annotation could take place. To prevent information being lost when the dataset was scaled down, it was important to ensure that the sampled graph maintained properties (such as diameter and average path length) similar to those of the raw data. To maintain these characteristics, "forest fire" sampling (**??**) was used to create a sub-graph that preserved the overall structure of the parent. The algorithm for forest fire sampling is as follows:

1. Choose a "forward burning probability" p – in this instance a value of 0.7 was chosen based on the recommendation by **?** for scaling down a larger graph
2. Choose a random starting node
3. Add this node to the sample graph. Select x nodes at random from all nodes linked to the chosen node, where x is a random number geometrically distributed with mean $\frac{p}{1-p}$. If the selected node has fewer than x linked nodes, select all available nodes, and return to step 2.
4. With each selected node, recursively repeat step 3 until the desired sample size has been reached.

Thirty posts from within the following discussion (i.e. not including the original posts) were selected using this method. This data was then manually annotated with the formal argument$_1$ information. Specifically, from each L-node, both explicit and implicit I-nodes were extracted and related together using the most appropriate S-nodes.

Table 3.2: Metrics of discussions sampled from YouTube, Twitter and Facebook

Metric	**YouTube**	**Twitter**	**Facebook**	**Total**
Total number of posts	30	30	30	90
Total number of users	23	12	30	65
Average posts per user	1.30	2.50	1.00	1.38
Average words per post	26.77	16.33	42.10	33.18
Average characters per post	147.90	101.20	259.67	201.70
Time between first and last posts	4d 0h 54m 56s	0d 5h 13m 33s	3d 12h 13m 18s	n/a
Average time between posts	3h 20m 31s	0h 10m 49s	2h 54m 15s	0h 17m 10s

Table 3.3: Aspects of raw data from social media APIs capable of being modelled using the AIF or SIOC ontologies

Features present in social media APIs	**Represented in:**	
	AIF	**SIOC**
Locution (explicit content)	✓	✓
Illocution (premises/conclusions)	✓	
Argumentation structure (attacks/support)	✓	
Author	✓	✓
Avatar		✓
Replies	✓	✓
Creation Date	✓	✓
Reputation (e.g. “Likes”)		
Location		
User “Type” (i.e. individual/business/etc.)		
Sentiment (implicit content)		

3.3 Results and Analysis

An overview of the raw data collected from each platform is shown in Table 3.1 and the sampled data in Table 3.2. In total, the discussion generated by the Twitter post has slightly over one-hundred and thirty replies – in contrast, the YouTube comments total nearly three thousand posts, and the Facebook discussion has well over nine-thousand. Each platform sees the vast majority of posts contributed soon after the initial post. However, each has a “long tail” of responses that gradually decrease in frequency as time goes on. The discussion on Twitter seems particularly ephemeral, with participants only contributing for a short time before moving onto other topics; while the Facebook and YouTube posts appear more “permanent”, with users finding and contributing to them months later.

In addition, when collecting this data it became apparent there was information that had no appropriate representation in either ontology, such as reputation systems (for example, the “Likes” used by Facebook), the sentiment of the post (for example, sarcasm, humour, abuse) or information about the type of user making the remark (whether they are an individual, a celebrity, a corporation, etc.); these omissions are shown in Table 3.3. These features could have substantial bearing on the perception of the argument$_2$. Consider the example of reputation systems: a

Table 3.4: Summary of AIF nodes found in annotated discussions collected from YouTube, Twitter and Facebook

Metric	YouTube	Twitter	Facebook	Total
L-nodes	30	30	30	90
TA-nodes	0	20	0	20
YA-nodes	31	30	41	102
I-nodes	88	116	110	314
S-nodes	13	30	26	69
L- to I-node ratio	15:44	8:29	3:11	45:157

retort stating *"You're an idiot"* may be perceived very differently by the audience if it has no up-votes, one up-vote or one hundred thousand up-votes. Alternatively, consider a user making the argument$_1$ that *"I really love using this product"*: whether the statement is made by an individual, or the company selling the product would likely influence the validity and value of the statement.

Table 3.4 shows the statistics collected after annotating the data with premises and conclusions, represented as AIF nodes. Given this data it can be seen that Twitter is the only sample that contains Transition-nodes; that is, replies to other posts within the thread. While this may appear to suggest that the platform is used more fore debate than the others, it is possible this is down to deficiencies in the APIs of the other platforms, which often do not accurately highlight replies. It can also be observed that the debates on Twitter and Facebook have a much higher information content than that of YouTube. The resulting structures are visualised in Figure 3.4, which shows a side-by-side comparison of the three different samples.

On the surface, the sample of posts taken from Twitter and Facebook appear to have similar information content. However, upon manual inspection, it can be seen that this average is actually heavily skewed by one particular Facebook post that is thirteen paragraphs long and contains a total of twenty six information nodes. The argument in question is reproduced on a number of different websites, and is likely reused in full as a boilerplate "cut and paste" rebuttal by many users when engaging in an argument on that topic.

To highlight the overall information disparity take, for example, the tweet *"@BarackObama Stop expanding government, spying on Americans and driving up the deficit."*. This is an enthymeme – the literally derived I-node acts as a conclusion, while the premises (that Obama is expanding government, spying on Americans and driving up the deficit and that to do so is a bad thing) are left implicit. In turn, contrast with the posts *"first"*, *"wow obama"* and *"lolollll i love this"* which contain very little information, either explicit or implicit. In addition, not all posts with a large amount of literal content have a comparatively large amount of information. For example, posts such as *"Give DIRETIDE Give DIRETIDE Give DIRETIDE..."* (repeated upwards of fifty times in a single post) show a desire to derail the discussion by flooding it with completely irrelevant information ("Diretide" refers to a cancelled seasonal event in the popular

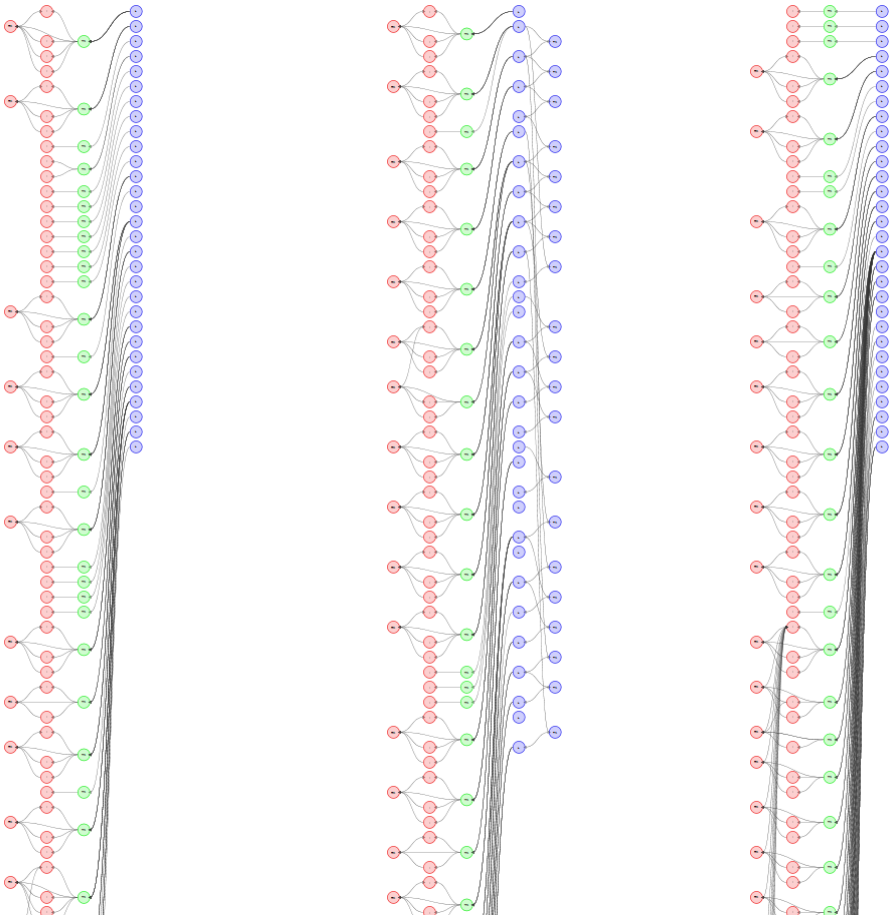


Figure 3.4: A side-by-side comparison of the emergent structures of discussions taken from YouTube (left), Twitter (centre) and Facebook (right)

online game *Defence of the Ancients 2*; the cancellation sparking uproar from the fanbase which led to a number of social media platforms being flooded with this message).

In addition, there are other posts that have deeper contextual meaning that would first appear. Consider, for example, "RedScareBot"[4]: this is an automated Twitter account that, using the avatar of Joseph McCarthy (an American politician famous for making claims at the height of the Cold War that their were numerous Soviet agents in the US government), replies to any tweet that includes phrases such as "communism" or "commie" with quips such as *"Commie Chameleon"*, *"Oh noes, Socialism"* or *"Rise of the USSA"*. While this may seem nonsensical or a non-sequitur without context, *with* context it can be viewed by the audience as a derisive or satirical retort to a knee-jerk insult, despite being posted by a machine.

There are of course limitations on the conclusions that can be drawn from a relatively small dataset when working with proverbial "big data". As such, these findings cannot be used to justify broad claims that state that *all* arguments on a particular example of social media are structured in this way. These examples instead serve to demonstrate the important fact that different types of structures *can* evolve, and provide some examples of the argumentative and

[4]https://twitter.com/RedScareBot

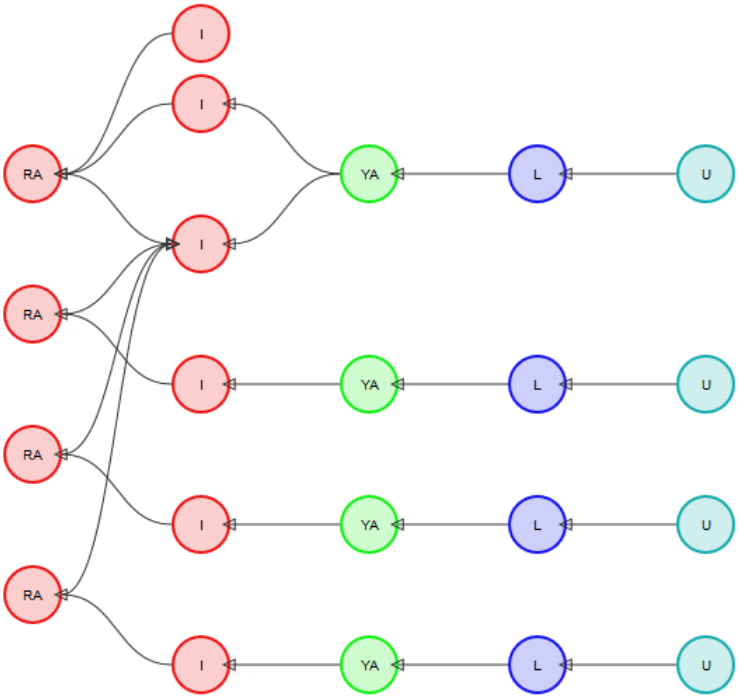


Figure 3.5: Proposal for representing reputation systems by modelling up- and down-votes as individual Locutions

rhetorical tactics people use when arguing over social media and how the conjunction of the AIF and SIOC projects (as well as any extensions made to these) can be used in attempts to map them.

3.4 Proposals

These results formed the basis for the work presented in the Workshop on Computational Models of Natural Argument, in which a number of suggestions on how these ontologies could be adapted to model the socio-rhetorical aspects of argumentation were proposed (**?**).

These proposals included suggestions for modelling social web specific features, such as the use of reputation systems. Reputation systems make up a key aspect of non-verbal argumentation on the social web, allowing users to show agreement or disagreement to a position, sometimes anonymously, without the need to articulate their own position. Thought must be given as to how to accurately represent this in a formal model. Figure 3.5 shows one such approach; namely, modelling each vote as a separate Locution, linking to an I-node that either (logically) supports or attacks the voted-on post. Alternatively, Figure 3.6 shows an approach which aggregates this information into a single Reputation node. This has the advantage of keeping to social information distinct from the logical graph structure, but the disadvantage of omitting how much each UserAccount contributed to the reputation.

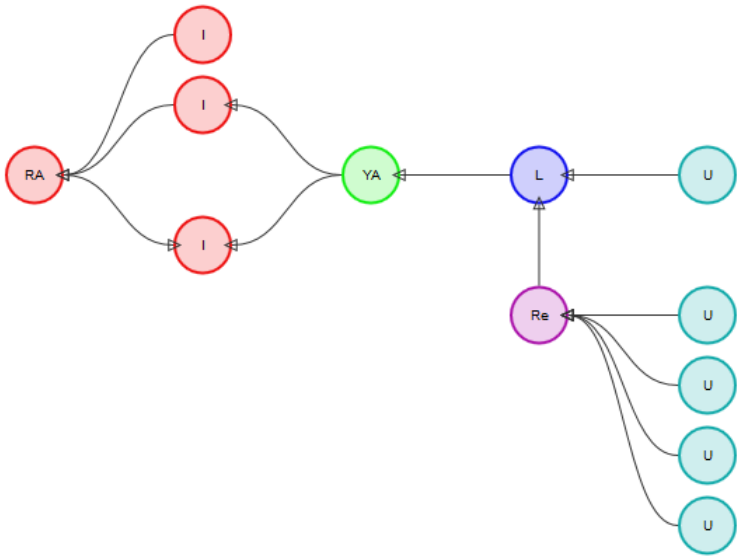


Figure 3.6: Proposal for representing reputation systems with the introduction of a Reputation node

Another focus of the discussions was that of abusive rhetorical attacks. Figure 3.7 shows the simplest approach, similar to the current way the AIF models the use of *ad hominem* attacks, by linking the attack to the opponent's argument$_1$ with a CA-node. However, this is insufficient for the majority of abusive attacks; while *ad hominem* tactics attack an opponent's argument$_1$ by claiming they are not qualified, or otherwise unfit, to make such an argument$_1$, abuse often does not attack their position at all, but seeks to undermine them emotionally in front of their peers. This mapping can be modelled by linking the content of the locution to the targeted user's account as shown in Figure 3.8. However, a UserAccount can be involved in any number of topics, and be attacked for any number of reasons. Furthermore, a person can choose to present themselves as a dramatically different person (having different credentials, skills, opinions or even race, religion or gender) when they are on the web as opposed to off. They may even choose to represent themselves differently between individual threads and discussions. To this end, another type of node is needed to represent the abstract notion of the "persona" a user presents. This is illustrated by Figure 3.9. Introducing the idea of personas allows each UserAccount to present a different view of themselves (that can be supported or attacked accordingly) when engaging in multiple discussions or topics.

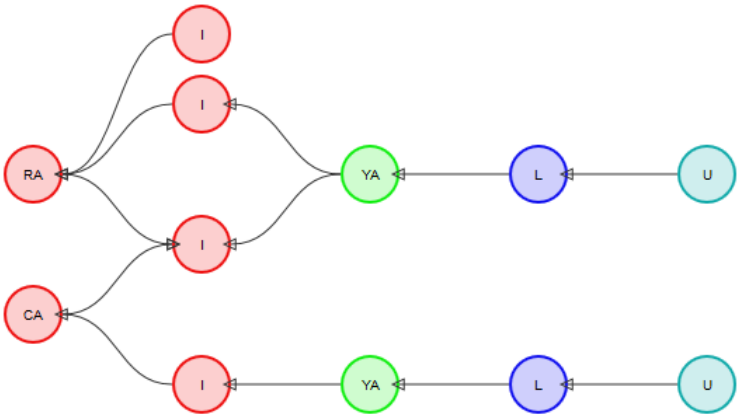


Figure 3.7: Proposal for representing abusive attacks as solely within the argument$_1$ structure

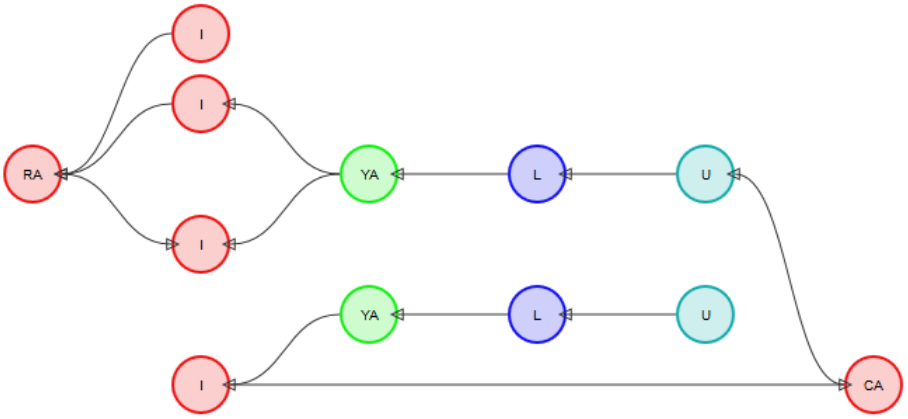


Figure 3.8: Proposal for representing abusive attacks as connected with the social aspect of the argument$_2$, attacking the author directly

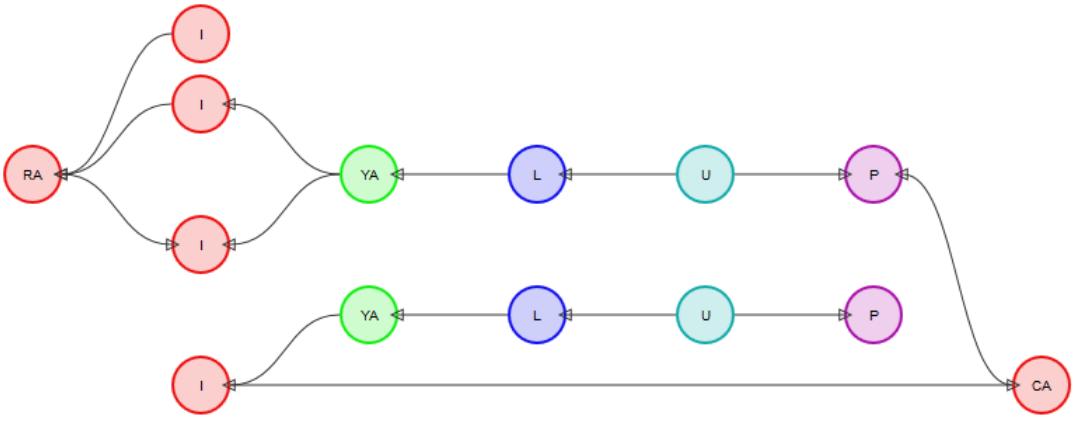


Figure 3.9: Proposal for representing abusive attacks, extending that shown in Figure 3.8 with the addition of Persona nodes

Chapter 4

Rhetorical Extensions for Eristic Argument

In the preliminary investigation, the capability of existing frameworks and their use in capturing and modelling argumentation and social communities was examined and evaluated (**?**). It became apparent that the AIF, while a powerful tool for modelling (dialectic) argument, lacked the ability to capture the eristic aspects of social argumentation. While some logical fallacies, such as the *ad hominem attack* can be suitably modelled within the AIF, the rhetorical force of "simple" abuse is difficult to capture.

However, there is reason to suggest that while such abuse (for example) may not be valuable to the argument_2 itself, that does not mean it is not valuable to model such outbursts. A heckler in a debate, for example, may not have any well-reasoned argument_1 to hand and resort to throwing vulgarities, but by simply disrupting the proceedings they are voicing their dissent at the positions offered. This is reason enough not to discard the contribution; however, it can also act to catalyse further argumentation on the subject between the main participants. Likewise, a participant in a debate may, instead of putting forth their own argument or attacking their opponent's, make some sort of joke to endear themselves to the audience. While the AIF can model the locution, the rhetorical force behind it goes uncaptured.

In addition, there are other socio-rhetorical tactics that are often employed on social media. These include spamming (posting large volumes of a repetitive nature) to drown out other posters, deliberate deviation from the topic at hand, bringing up non-sequiturs in an attempt to derail the argument and "meta-argumentation" – criticising the way in which an opponent argues, but not the argument itself (e.g. if a user claims another is breaking the rules of the forum, or of not arguing in good faith). There are also the non-textual features of social media to consider; that is, the feature of posts other than their content. For example, the number of "Likes" or "Favourites" a post has demonstrates popular (or audience) support for this opinion or position.

4.1 The Argumentation on the Social Web Ontology

Based on the observations made in Section 3.3 and the proposals discussed in Section 3.4, the principal features from the AIF and SIOC ontologies are combined alongside means to model rhetorical tactics in the Argumentation on the Social Web Ontology (ASWO).

The principal focus here is the inclusion of rhetorical support and attack. While these features are only one aspect of rhetorical argument, they feature heavily in eristic dialogue (particularly rhetorical attacks), showcase both the positive and negative aspects of rhetorical argument and are important due to the impact they can have within discussions on the social web and the culture surrounding it (**?**).

Rhetorical support is often relatively benign. It can be used to show solidarity with other members of the dialogue, to incorporate oneself into a social group, or to encourage . Consider the extracts *"bro fist bump"*, a short declaration of support for another user, and *"I commend you for admitting that debt & deficits are important...If only more [people] felt the way you do"*, which disagrees with the overall stance presented by their opponent, but commends them for conceding some common ground, in attempt to further dialectic argument.

Conversely, rhetorical attacks are often extremely hostile. They differ from logical attacks by attacking the person behind the argument rather than the argument itself (this is not to be confused with an *ad hominem* argument which attacks a person's argument by calling their character or credentials into question – these are logical, even though they are fallacious). Rhetorical attacks often contain extremely vulgar language: consider the extracts *"Some of you fucks need a good ass whipping"* and *"Fuck off cunt"*. The purpose of these statements can be interpreted in a number of ways, from intentionally silencing dissenting voices with threats, to showing the audience how impassioned and emotive the rhetor is on the subject, to cathartically blowing off steam.

The notion of rhetorical support and attack is modelled by introducing three new types of nodes to the ontology. Firstly, as described in Section 3.4, it is not enough to use a UserAccount to represent a person during an argument$_2$. Instead, the Persona node represents a user's character and authority on a given subject, or rather the character and authority they present themselves to have online, and is bound to one or more UserAccounts. Two more node types are introduced to differentiate these types of rhetorical interaction from logical supports or attacks. PersonalConflict (PC-nodes) nodes link from a YA-node to a Persona node to denote this type of personal support and, likewise, PersonalSupport (PS-nodes) nodes follow the same structure to denote support of a person's intentions and character. These are broad, umbrella definitions that serve to catalogue all discovered variants of personal support and attack; they could of course be refined to differentiate between different subcategories of support, such as simple agreement, or self-support (through humour, etc.), or attack such as threats, insults, or provocations. A simple example of this is shown in Figure 4.1, which shows the same argument visualised in Figure 2.6 (with the addition of User and Persona nodes) being rhetorically attacked by another user.

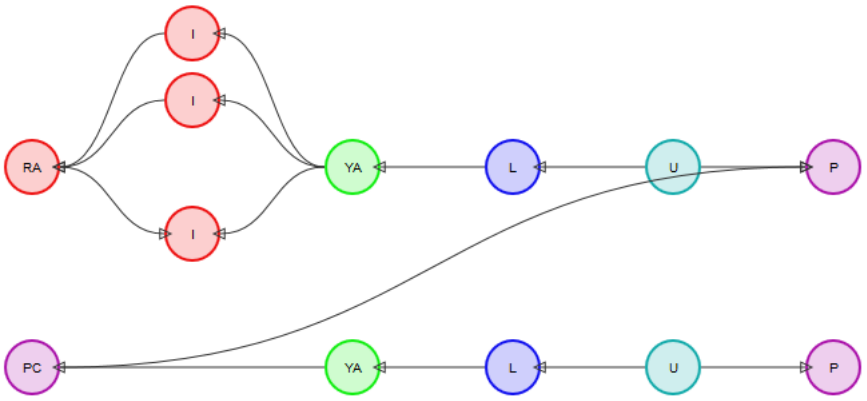


Figure 4.1: Visualisation showing a PersonalConflict node, an example of the augmentations made to the ASWO

The ASWO also models reputation systems (i.e. Favourites, Likes and up- and down-votes); however, for simplicity, they are not modelled as nodes in the graph structure, but are included as literal values attached to the relevant Locution. In addition, because the SIOC ontology also accounts for replies to and from a post, the use of AIF TA-nodes has also been refined in relation to the social web. They now no longer need to be used whenever an L-node directly responds to another; instead they can be used solely to refer to "transitions" in the argument. These transitions are used when a Locution contributes to the argument$_2$ without providing any information, but instead helps move the discussion to the "next stage", usually by asking questions or prompting further debate. Note that these transitions don't necessarily move the discussion forwards, but can also be used to take the argument$_2$ around in circles.

4.2 Investigations

The ASWO, and the augmentations made to the AIF and SIOC ontologies, were trialled in an investigation to study the application of logical versus rhetorical techniques in eristic dialogue on the social web. As before, this investigation focused on three different areas of the social web, but used a much larger sample size than previously: in total, two hundred and seventy posts were collected and annotated. These were used to analyse the proportion of rhetorical contributions throughout the argument$_2$, analyse the relation between logical and rhetorical arguments$_1$ used, and compare the features of the annotation structure with the content of each post.

4.2.1 Methodology

4.2.1.1 Data Collection

During the course of this work, the Google YouTube API v2.0 was deprecated before the API v3.0 fully supported the retrieval of explicit replies to comments. Due to the importance of the

Table 4.1: Metrics of discussions sampled from Twitter, Facebook and Reddit

Metric	Twitter	Facebook	Reddit	Total
Posts	90	90	90	270
Direct replies	77	0	67	144
Number of users	26	85	43	154
Average posts per user	3.5	1.1	2.1	1.8
Average words per post	15.83	41.36	42.34	33.18
Average characters per post	96.51	265.27	243.31	201.70
Time between first and last posts	0d 6h 53m 40s	3d 4h 51m 27s	3d 0h 50m 12s	n/a
Average time between posts	04m 39s	51m 49s	49m 06s	35m 11s

ability to capture replies, the decision was made to use an alternative medium in this case study. To this end, YouTube was replaced with the social news and networking site Reddit. Reddit has a variety of topic-specific boards or "subreddits" that allow users to post to a collaborative pool of information; posts can then be up-voted or down-voted to show interest and/or accuracy.

Obama's official account on Reddit was inactive over the period of the shutdown; however, another user (unaffiliated in any official capacity with Obama) posted a link to Obama's official website (managed by Organizing for Action) to Reddit's politics subreddit[1] on 15th October 2013 (the same date as the official posts to Twitter and Facebook). The post reads *"Tea Party Republicans in the House of Representatives have already shut down the government because they couldn't derail Obamacare. Now they're threatening to cause an economic shutdown"*. This thread was used alongside the previously acquired threads from Twitter and Facebook described in Section 3.2.1. Each UserAccount involved in the three threads was automatically designated a single Persona, as only one topic was monitored. This could be expanded if the same UserAccount took part in multiple threads on multiple topics, for example.

As with the preliminary work, forest fire sampling of the graphs was undertaken to provide a representative sample of the arguments that was feasible to annotate manually. For this investigation a larger sample size of ninety posts was used from within each discussion. Table 4.1 shows an overview of the sample structures and some key characteristics of each thread.

4.2.1.2 Annotation

With the changes to the ontologies in use (such as the decision regarding TA-nodes discussed in Section 4.1), and a larger amount of data needing to be annotated, the annotation method itself needed to be properly formalised to solidify reproducibility and minimise subjectiveness. Posts are annotated according to the scheme below.

Each post is considered to contain zero or more separate arguments$_1$. A YA-node is created for each argument$_1$ made in a single post, and links the L-node to each I-node in the argument$_1$. Repeated information does not create a new I-node; instead the YA-node links to the I-node

[1] http://reddit.com/r/politics/1oij25

already present. All participants are assumed to have some implicit knowledge about the world in general and the topic at hand. This is to avoid the inclusions of trivial I-nodes that state information such as *"Barack Obama is president of the United States"*, or even *"Barack Obama is a human being"*. Any information explicitly contained in a post that is deemed to be not in this set and relevant to the discussion at hand was included as an I-node. Information that meets one (or more) of the following criteria is not considered relevant:

- Off topic: posts that do not relate to the topic being discussed are not considered relevant. Example: *"Ataturk did revolution ! building moderate muslim network is oxymoron which has been destroy secular , democratic, rule of law in Turkey."*
- Conversational: similar to off-topic posts, those that are conversational in nature are not annotated as information-containing. Example: *"I thank you, have a good night!"*
- Meta-argumentation: while argumentation about how to argue "properly" is an interesting construct in itself, and an important aspect of rhetorical and eristic argumentation, but was out of scope for this particular study. Example: *"Down voting = disagree Upvoting = agree" "The rules say explicitly not to do that....."*

A TA-node is created to link two Locutions whenever a transition is present in the argument_2 – a step that contributes to the overall structure without providing any information (new or repeated). This is most often in the form of an interrogative (for example, asking for further information or evidence for claims). Support and attack between different I-nodes is denoted as described above: logical support through the use of RA-nodes, attack through the use of CA-nodes and preference with PA-nodes, while rhetorical support and attack utilises the new PS- and PC-nodes.

Some nodes in the graph may not be complete as a result of the nature of sampling the graph. For example, it may be possible to detect that a user attacks another user's persona, but not exactly which user they are attacking. Table 4.2 shows an overview of the number of AIF and ASWO nodes added during the annotation process.

4.2.2 Results and Analysis

4.2.2.1 Argumentation Tactics Over Time

Firstly, the way in which the argumentation structure changes and grows over time is presented, in both a logical and rhetorical capacity, by graphing how the number of logical support and attack nodes (i.e. RA- and CA-nodes) and rhetorical support and attack nodes (i.e. PS- and PC-nodes) changes with each post contributed to the argument_2. Logical contributions are displayed above the x-axis, and rhetorical contributions below. It must be emphasised that values below

Table 4.2: Summary of AIF and ASWO nodes found in annotated discussions collected from Twitter, Facebook and Reddit

Metric	Twitter	Facebook	Reddit	Total
L-nodes	90	90	90	270
TA-nodes	52	9	15	76
YA-nodes	58	74	70	202
I-nodes	56	98	86	240
RA-nodes	13	20	24	57
CA-nodes	18	1	34	53
PA-nodes	4	4	2	10
PS-nodes	2	2	3	7
PC-nodes	26	6	12	44
L- to I-node Ratio	45:28	45:49	45:43	9:8

the x-axis of each graph should *not* be considered as inherently negative, hostile or anti-social; they simply differentiate between the two types of content.

Figures 4.3 and 4.4 show that use of rhetorical tactics in the Facebook and Reddit case studies rise slowly compared to the use of logical tactics. However, Figure 5.1 shows that in the Twitter case study, the rhetorical contributions rise in parallel to the logical contributions.

In both samples from Twitter and Reddit, the distribution of logical supports and attacks also remain approximately equal. Due to the tendency of RA-nodes to be used for logical support within an argument$_1$, and the tendency of CA-nodes to be used between arguments$_1$, this highlights a greater engagement between participants within these debates than the Facebook sample, which has only one CA-node and comparatively much fewer instances of logical or rhetorical contribution overall. In all three examples however, rhetorical conflict far outweighs rhetorical support.

Overall, it appears that there is no sudden shift in tactics from arguing logically to adopting a rhetorical approach – rhetorical argument forms an underlying and consistent strategy throughout the argument$_2$.

4.2.2.2 Argumentation Tactics per User

In addition, proportion of logical versus rhetorical contributions made by each user is examined. These graphs show the contributions made by each user (ordered by total contributions overall). Once more, it must bed stressed that values below the x-axis should not considered anti-social solely due to their rhetorical nature.

Figures 4.5 and 4.7 show that users in the Twitter and Reddit samples made more individual contributions to the argumentation structure than those in the Facebook sample, shown in Figure 4.6. This, along with the data in Table 4.1, also supports the suggestion that there is more engagement in these communities than in the Facebook sample.

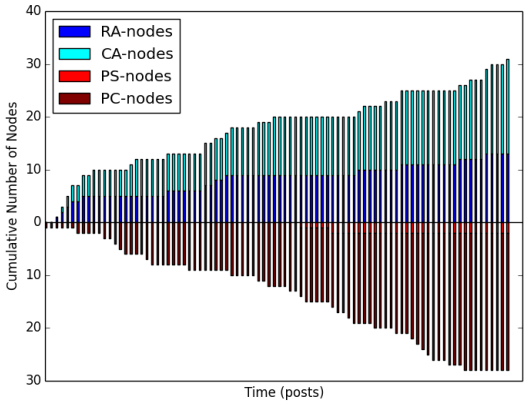


Figure 4.2: Cumulative use of logical and rhetoric tactics over time on Twitter

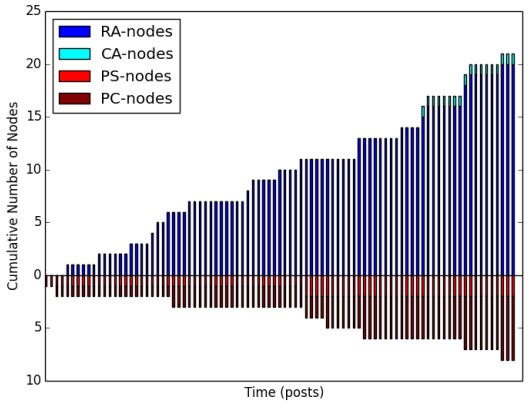


Figure 4.3: Cumulative use of logical and rhetoric tactics over time on Facebook

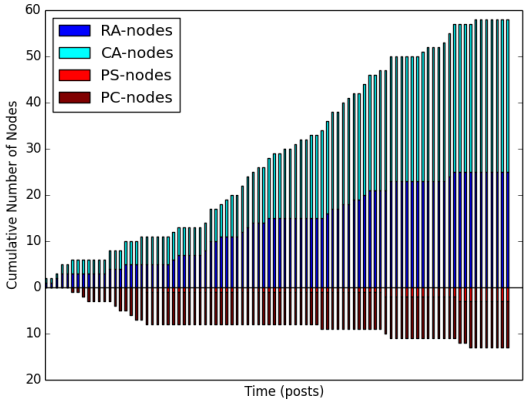


Figure 4.4: Cumulative use of logical and rhetoric tactics over time on Reddit

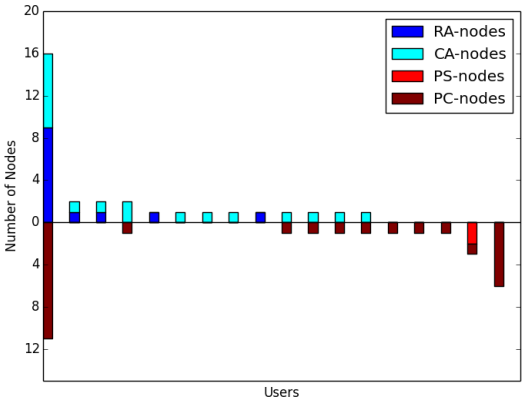


Figure 4.5: Logical and rhetorical contributions per sampled user on Twitter

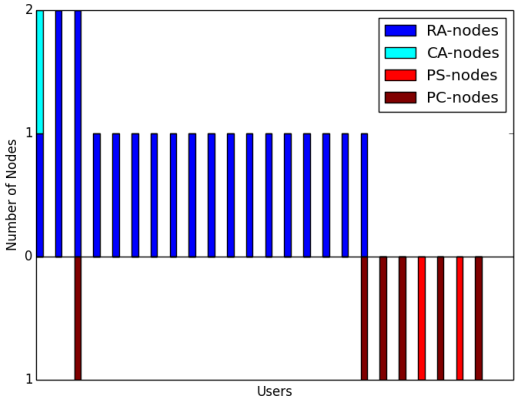


Figure 4.6: Logical and rhetorical contributions per sampled user on Facebook

All samples also display a tendency for rhetorical contributions to be distributed across the scale, with (weak) grouping towards either end. This implies that the users most likely to employ rhetorical techniques are those that contribute the most posts to the discussion overall, and those that make no logical contributions at all.

4.2.2.3 Correlation Between Argumentation Structure and Post Features

Correlations were drawn between the structure of the annotated argument graph, including elements such as the number of logical or rhetorical supports or conflicts and replies to and from each post, and features of the post content and structure, such as post length, number of expletives, percentage of spelling errors and again, replies to and from the post. Replies in particular were viewed from both sides: that is, to analyse whether certain types of posts were more likely to be made in reply, or whether posts that were made in reply tended to contribute similar argumentation structures.

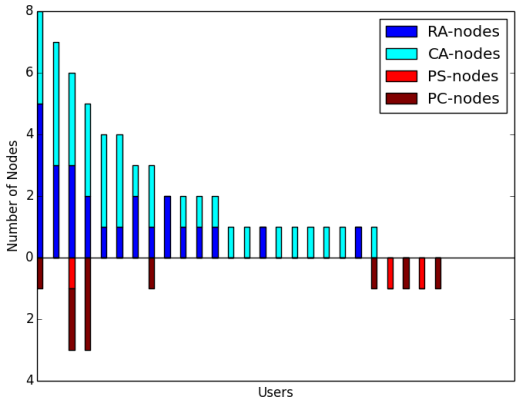


Figure 4.7: Logical and rhetorical contributions per sampled user on Reddit

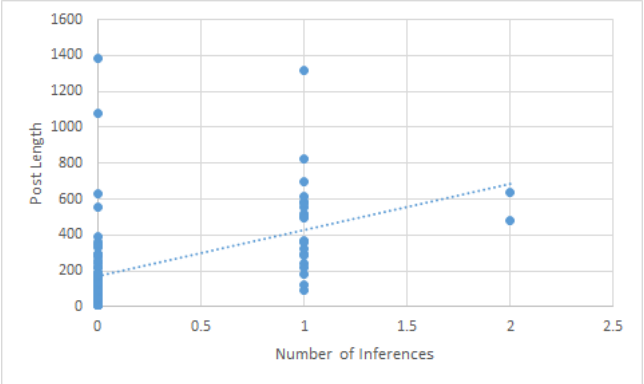


Figure 4.8: Post length correlated against number of logical inferences, on Reddit

Due to the largely discrete (and often binary) nature of the features and values studied (the majority of posts, for example, are likely to contain either zero or one logical or rhetorical conflict) the correlations are relatively weak, as show in Figure 4.8. However, some notable correlations are presented in Table 4.3. These show potential early indicators of the structure and value of an argument. For example, as might be expected, longer posts are more likely to have greater contributions to the discussion. Posts that use a large number of expletives are likewise more likely to contain a rhetorical attack. When examining all three case studies together, posts made in reply correlated with posts that were replied to, implying that when one or more users engage in a discussion, they are more likely to be engaged with in return.

Table 4.3: Notable correlations between structural argumentation annotations and post features

Case Study	**Argumentation Structure**	**Post Feature**	**Pearson's Correlation (3 d.p.)**	**p-value (3 d.p.)**
Twitter	Personal attacks	Number of replies to this post	0.325	0.002
Twitter	Personal attacks	Percentage of spelling errors	0.301	0.004
Twitter	Personal attacks	Expletives	0.462	0.000
Facebook	Original premises	Reputation ("Likes")	0.332	0.001
Facebook	Original conclusions	Reputation ("Likes")	0.329	0.002
Facebook	Logical inferences	Reputation ("Likes")	0.343	0.001
Facebook	Logical conflicts	Emoticons	0.500	0.000
Facebook	Logical conflicts	Expletives	0.397	0.000
Reddit	Original premises	Post length	0.335	0.001
Reddit	Original conclusions	Post length	0.333	0.001
Reddit	Logical inferences	Post length	0.476	0.000
Reddit	Logical conflicts	Number of posts replied to	0.435	0.000
Overall	Number of replies to this post	Number of posts replied to	0.417	0.000

Chapter 5

Conclusions and Future Work

5.1 Findings

The work described in this report covers an examination of the capability of current argumentation models, in particular the application of a combination of the AIF and SIOC ontologies to the social web, and the extension of these models to capture social and rhetorical information. Case-studies were carried out on three different areas of the social web to determine the strengths and weaknesses of modelling social, eristic argument on the web. This preliminary work indicated that existing techniques for modelling argumentation were insufficient to capture the structure and dynamic of argumentation taking place on the social web, which led to the publication of a paper in *the 14th workshop on Computational Models of Natural Argument*, detailing these omissions and proposing a set of augmentations to capture additional socio-rhetorical tactics (**?**). These extensions were implemented and trialled as part of an investigation re-examining the previous case-studies to determine the prevalence of rhetorical tactics in argumentation within areas of the social web and look for correlations that can be drawn between the use of these tactics and the machine-readable characteristic of the post such as length or readability. The results of this will published in the upcoming *ACM Conference on Hypertext and Social Media* (**?**). These investigations reveal the following findings.

Firstly, and most importantly, rhetorical tactics are shown to be present throughout the argumentation in the case studies, even when only accounting for a small subset of rhetorical argumentation. Clearly, failure to accurately model these social argumentation strategies is detrimental to the goal of studying how discussions evolve on the social web. Secondly, in the three use cases, rhetorical tactics are most often used by either those contributing the most to the discussion overall, or by those who do not contribute logically at all. Whether this effect is related to a participant's engagement is unknown. However, this raises the possibility that there is a tipping-point in a dialectic logical debate where participants feel the need to expand their use of tactics; alternatively, these users simply interleave both types of tactics throughout their arguments$_2$.

Finally, while the features of the argumentation structure above are challenging to detect automatically and expensive to manually annotate, the markers present in the social media sphere are relatively trivial to detect, and some correlations between the two can be observed.

The primary limitation of this work is the necessity to manually annotate all the data. This is time consuming and subjective, but as yet there is no way to circumvent this process and automatically extract premises and conclusions. A further constraint is that only English-language sites are examined. There are, of course, many other social media services that cater to audiences of different languages, such as *Renren*[1] for China or *VKontakte*[2] for eastern Europe. However, this separation is mitigated by the fact that different languages (and different cultures) have their own rhetorical structures and argumentation schemes (**?**, p. 21). As a result, attempting to analyse multiple sites with different primary languages concurrently would distort any patterns that might emerge in the argument structure of the users.

5.2 Future Work

Based on the investigations that have been carried out, and the findings described in Section 5.1, there are three particular avenues of future work that could be approached, using this extended model of social argumentation at their core.

Firstly, as is the focus of many researchers in this field, attention can be given to the use of artificial intelligence and argumentation, whether by reasoning over a model of argument in an attempt to determine the most valid argument and subsequent course of action (**?**) or by using the model to influence the techniques and strategies of intelligent agents involved in dialogue games (**?**). However, the fact that the eristic features of the model are unlikely to be practical (or appropriate) for the use of reasoning, or governing inter-agent negotiations is likely what has caused them to be currently excluded from the majority of formal models. Disregarding this, the weakness of this approach is that the model cannot, at this stage, be automatically constructed, but must be created through a time and labour intensive process of manual annotation. Therefore, using the model as a basis of reasoning over argumentation in general is ultimately flawed. Any gains that were achieved in this area would be rendered moot by the cost of creating a model for every argumentation to be reasoned over, and rendered impractical on a web-scale.

With this in mind, the second avenue would be to generate this model from the arguments$_2$ themselves, by means of natural language processing (**?**), the use of social machines (**?**) or some combination thereof. This would go some way towards solving a large outstanding issue in the field (**?**, p. 31-32). While working towards a means of automatically generating the model has potential, it is likely that the social and eristic nature of the arguments to be modelled is the very thing that hinders this approach. Web-based culture and language is made up of many disparate groups, and continues to rapidly and constantly evolve, which renders current natural

[1] http://renren.com/
[2] http://vk.com/

language processing impractical in the short term and ineffective in the long term, without the use of domain-specific normalisation techniques that are expensive or inaccurate (**??**). While the findings in Section 4.2.2 point towards a means of broadly classifying a post as containing different types of logical or rhetorical elements, with reasonable probability, the overall structure may be difficult to model automatically. Clearly, at this stage, human input cannot be wholly eliminated. However, with the use of crowd-sourcing or social machines, the large effort cost of annotating arguments$_2$ could be distributed across participants to a manageable level.

Finally, emphasis could be placed on the social aspect of argument. Because argumentation is a social process conducted by people, it is important to recognise the fact that individuals may perceive the same argument$_2$ in many different ways due to cultural beliefs (**?**), pre-existing cognitive biases (**?**), as well as features surrounding the content of the argument$_1$ such as avatars (**?**). The advantage of this approach is that it uses the existing model as a platform for experimentally evaluating how the use and prevalence of different argumentation tactics affect users' perceptions of an argument$_2$, and the way in which they engage with the thread (and one another) as a result. By using the model as a tool for analysing individual case studies, the requirements for creating and annotating the necessary argumentation structures are greatly constrained, while allowing the findings to be used in further work in the research area. This contribution to the field can then be used to assist further work in a number of other areas, such as another metric for use with adaptive recommendation techniques to match people based on preferred argumentation strategies (**?**), or the development of argumentation frameworks that integrate with the social web (**?**).

5.2.1 Hypothesis

This direction of future work forms the following research hypothesis:

"A model of eristic argumentation on the social web should include both logical and rhetorical tactics, as the inclusion of rhetorical techniques affects the way in which users perceive and engage with the argument"

This can be resolved into two distinct research questions:

1. *Do rhetorical techniques affect the way in which users perceive and engage with the argument?*
2. *Which rhetorical techniques should be included in a model of eristic argumentation on the social web?*

To effectively answer the former, it is important to define what is meant by the terms perception and engagement. Perception can be thought of as the way in which users understand the tone, persuasiveness, entertainment value or information content of an argument$_2$ (**?**), whereas

engagement is how they respond to or participate in the argument$_2$ itself. This not limited to replying to a post: users of social media can engage in multiple ways, including replying, sharing or voting (**?**).

The latter of the questions is closely tied to the former. While an exhaustive list of all categorisations of rhetorical tactics would be ideal, in practice it is infeasible. Therefore, the tactics that have the most impact on the perception and engagement of the arguments$_2$ should be included first to provide the most value, which can then be build on by further extensions.

5.2.2 Work Package 1a: Formalise Rhetorical Additions to Model

Description: It has been shown in Section 4.2.2 that rhetorical tactics, even when only considering a highly constrained subset, are prevalent throughout three case studies. Therefore, there is a clear need for expanding the capabilities of modelling such tactics, the relationships between them and how these integrate with the AIF and SIOC. To begin with, important rhetorical features would need to be determined and rigorously defined, based on existing literature. An example of additions that could be made include the notion of meta-argumentation described in Section 4.2.1.2 or of arguments$_1$ that actively pursues a particular call to action (rather than simply putting forward or advocating a position). These can then be incorporated as part of the ASWO.

Outcome: A formal ontology defining a broader subset of rhetorical tactics in social argumentation, and the relationships between them and the AIF and SIOC.

Estimated Time: 1.5 months

5.2.3 Work Package 1b: Multi-Contributor Classification

Description: Work Package 1a addresses the need to further model rhetorical tactics used on social media. However, for the purposes of preparing an experiment investigating the result of these tactics on user perception, there is no need to exhaustively model all aspects of this form of argumentation (such as individual information nodes, or the relations between them). Instead, a simpler classification format can be developed, that focuses on categorising a broad range of tactics from a number of sources. This will dramatically reduce the time cost of the annotation process, allowing for a larger number of posts to be included in the analysis.

Table 5.1 shows an example of these potential classifications. By using non-exclusive categories, arguments$_1$ can be annotated as employing multiple tactics (the only exception to this is the "other" category which, by definition, is exclusive). These tactics are grouped into fine-grained

categories (such as logical support, transitions, meta-argumentation, etc.) and coarser-grained categories (i.e. logical tactics, rhetorical tactics and other).

To achieve this, annotators will be recruited and given basic training in differentiating between tactics. A pool of ≈ 200 threads from three different social media platforms on the same topic will be chosen, from which three random thread will be chosen per site from the interquartile range of the multivariate probability mass function of features of the social web, such as thread length, number of contributors, intra-thread replies, etc., for the purpose of removing outliers.

From these nine threads, each annotator will be required to annotate posts based on the categories determined in Work Package 1a (with overlap between annotators, to allow for consensus checking). This method allows for a larger breadth of samples, greater amounts of data to be marked up for future experiments and, by incorporating inter-rater reliability checks, removes potential subjectivity bias compared to a single annotator. Expert validation will be required to verify a random sample of posts to determine the annotations are accurate, and mediate in instances of broad disagreement.

These annotations can then be used for quantitative analysis of different areas of the social web: for example, whether certain argumentation tactics are more common on certain platforms, or whether they are affected by features such as number of participants, number of posts, etc. In addition, the annotation process itself can be analysed, with a view to determine the ease with which such a process could be crowd-sourced (or even gamified) in future: for example, types of arguments that generated the highest and lowest consensus between annotators.

Outcome: A dataset annotated with a broader sub-set of rhetorical tactics used in nine different argumentative discussions and an analysis of the uses of granular rhetorical tactics across different spheres of the social web.

Estimated Time: 4 months

5.2.4 Work Package 2a: Interpretation and Engagement Pilot

Description: To determine an appropriate bounding on the length of experiment and participant overload, a short pilot study will be conducted. This will aim to asses how the number of posts per thread affects the required time for participants to complete the study and the quality and quantity of responses.

Outcome: Appropriate weightings for participant load during the main experiment described in Work Package 2b

Table 5.1: Example classifications of argumentation posts

Information	This post contains (purportedly) factual information	Logical tactics
(example)	*"Here's a List of 313+ Employers Who Have Cut Hours Because of Obamacare..."*	
Logical Support	This post supports another post or point of view by providing supplementary evidence, attempting to invoke the authority of the author, or another logical tactic	
(example)	*lol, right ? They don't get that if everyone has access to affordable healthcare then everyone pays their fair share*	
Logical Attack	This post attacks another post or point of view by providing contrary evidence, attempting to undermine the authority of the author, or another logical tactic	
(example)	*"No one "negotiates" over laws that have already passed" "Really? Then why isn't the Volstead Act still the law of the land?"*	
Transitionary	This post attempts to move the argument forwards by asking questions or prompting further debate	
(example)	*"If you know the numbers, then please tell me how many Dems lost their seat the last two rounds?"*	
Personal Support	This post expresses support for another user (rather than their argument)	Rhetorical tactics
(example)	*"I commend you for admitting that debt & deficits are important...If only more [people] felt the way you do"*	
Personal Attack	This post attacks, abuses or threatens another user (rather than their argument)	
(example)	*Fuck off cunt*	
Calls to action	Posts that advocate a particular course of action	
(example)	*"Kill them now, impeach them now. The american people dont need masters"*	
Meta-argumentation	Posts that argue about the argument itself – whether commenting on the rules of the medium or proposing a way participants should argue "properly"	
(example)	*"Down voting = disagree Upvoting = agree" "The rules say explicitly not to do that....."*	
Conversational	Posts that do not put forward, support or attack a particular view, but make small talk or converse with participants and/or the audience	Other
(example)	*"...I think I am all politically talked out for the night lol, I need to finish some work"*	
Off topic	Posts that do not relate to the topic being discussed	
(example)	*"Ataturk did revolution ! building moderate muslim network is oxymoron which has been destroy secular , democratic, rule of law in Turkey..."*	
Other	The only exclusive category, posts which match none of the above criteria	
(example)	*"[This post has been deleted]"*	

Table 5.2: Proposed potential participant groupings

Participant Group	$R+O$	$L+O$	$R+L+O$
1	A1	B2	C3
2	C3	A1	B2
3	B2	C3	A1
4	A2	B3	C1
5	C1	A2	B3
6	B3	C1	A2
7	A3	B1	C2
8	C2	A3	B1
9	B1	C2	A3

Estimated Time: 2 months

5.2.5 Work Package 2b: Interpretation and Engagement Study

Description: To determine the effect of rhetorical techniques on the perception of eristic argumentation on the social web, a within-participant experiment will be conducted in which voluntary participants are shown the argumentation threads annotated in Work Package 1b.

Each participant will be shown three different argumentation threads, each of which originates from a different social media platform. Each thread will be "pruned" according to the coarse-grained groups from Work Package 1b so that each users sees one thread containing only rhetorical tactics, one thread containing only logical tactics and one thread containing both rhetorical and logical tactics. Posts that are annotated as containing multiple tactics will be included on a non-exclusive basis (i.e. if a post is marked as containing both logical and rhetorical tactics, it could be displayed in any of the three combinations of tactics). The groups containing rhetorical content will also display social features such as reputation systems. These may need to be normalised across each social biome to prevent participants inferring the likely source platform. The annotations in Work Package 1b cover three different biomes (A, B and C) with three different threads from each (1, 2 and 3), which leads to the proposed potential participant grouping show in Table 5.2.

Datapoints per experimental factor (D) can be calculated from the number of threads shown to each participant (T), the number of participants (N), total tactic combinations (C) and the number of different social media biomes used (B) using the formula $D = \frac{T \times N}{C \times B}$. Given that the experiment is within-participants, each participant should be shown an equal number of threads and combinations of tactics ($T = C$). This constrains the number (and hence, granularity) of categories that can be examined through this experiment, but ensures that any variance between participants should be controlled for. Therefore, given that three social media platforms will be annotated, for an adequate number of datapoints (> 30), the number of participants required is $N > 90$.

The presentation of the arguments$_2$ themselves will be in a uniform format, to avoid leading participants to make judgements based on the (perceived) culture of the original platform. Usernames will be semi-anonymised; real names will be removed, as will artefacts revealing the source site (such as the "@" prefix used on Twitter), but "screen names" (such as *DemsAbroad* or *Tea4gunsSC*) can give an insight to a user's views and motivations (**?**, p. 379) and while it is conceivable that a participant may have interacted with the user before it is sufficiently unlikely in practice to warrant their inclusion. Participants will need to be regular users of the social web. Given the particular topic of discussion in the dataset, care must be taken to ensure that biases are identified during selection and accounted for during analysis of results. This can also be mitigated through use of a pre-test questionnaire to capture demographic data, topic interest and account for any biases – due to the topic at hand, this may also require asking participants what they consider their political affiliations.

The majority of questions in the questionnaire will ask participants to rate their agreement with a series of statements on a Likert scale. To determine how participants' perception of the argument$_2$ changes, statements will be based on the work of **?**, which examines perception of news media by asking participants to rate news stories a series of adjectives including accurate, biased, comprehensive, factual, informative, persuasive, sensationalistic and well-written. The precise adjectives to be used in the survey will need to be resolved to match the platform being examined, but may include statements such as:

- *Overall, I found the debate polite*
- *Overall, I found the debate informative*
- *Overall, I found the debate entertaining*

These can be interleaved with qualitative questions of the form *Please expand on the justification for your choices.*

To determine how participants' engagement may be altered, the Likert statements will take into account the work of **?**, in which they discuss the different types of engagement within social media: consumption, curation, creation and collaboration. These are reflected in the statements chosen:

- *I would like to see more posts by these users*
- *I would consider responding to this debate by replying with a comment of my own*
- *I would consider responding to this debate by voting on these posts*
- *I would consider sharing this debate with my friends*

Such questions could be further supplemented with questions of the form *Which user(s) did you find most informative? (Select up to three)*, *Which user(s) did you find least polite? (Select up to three)* or *Which user(s) did you feel had the most powerful argument? (Select up to three)*. This allows, to some degree, the examination of how an individual's posting style can impact the debate, and might also highlight any biases towards certain users and/or points of view.

The experiment itself will be run for a period of three months, which should be adequate time to accumulate the necessary participants, with sufficient additional time beforehand to prepare, and afterwards to analyse the results. Analysis will compare the responses of participants who have seen the same thread, but different combinations of tactics used, to determine how their viewpoints differ. Comparative evaluation will also show how each user reacts to each tactic-grouping. This will then feedback into the formalised model developed in Work Package 1a, and be written up as a journal article.

Outcome: An analysis of the experiment, and a journal paper detailing the process and results.

Estimated Time: 5.5 months

5.2.6 Work Package 3: Write-up of Thesis

Description: Having completed these experiments and the analysis of the results, a thesis will be written to describe the findings, determine the effect of rhetorical features on eristic argument and resolve the hypothesis.

Outcome: Printed and bound thesis

Estimated Time: 6 months

5.2.7 Gantt Chart

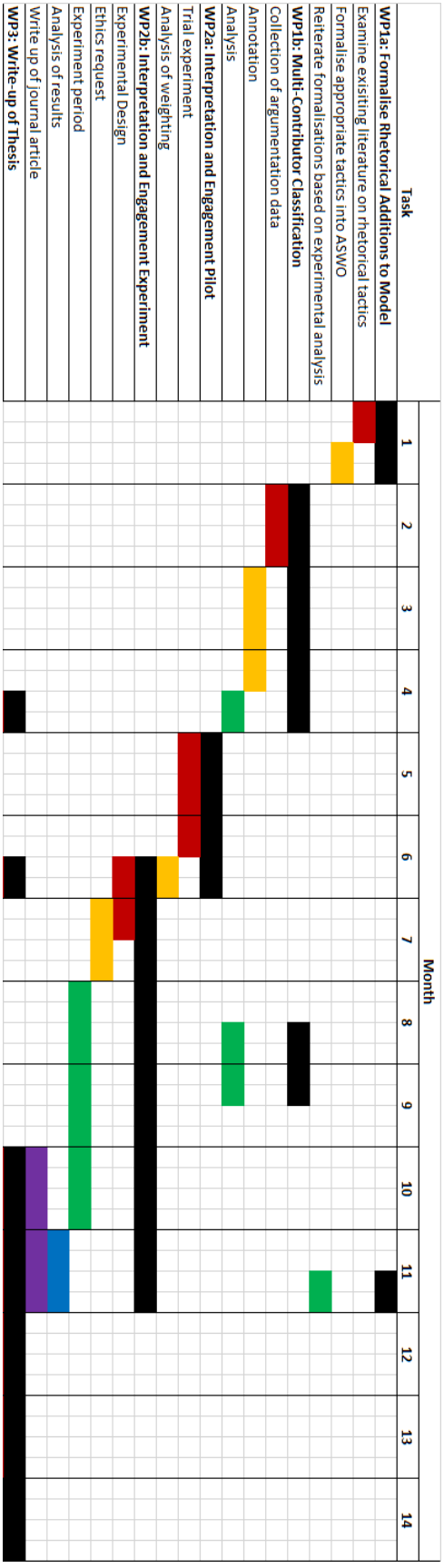


Figure 5.1: Gantt chart detailing the next three Work Packages

5.3 Conclusions

Argumentation, like the social web itself, is a diverse construct that is challenging to model but has huge potential if correctly harnessed. Rhetoric and logic are both important aspects of online social argumentation; to accurately model how arguments occur and evolve across social media it is important to take into account all the techniques and tactics that are employed. While it is difficult to determine the value of a contribution, to define all logical contributions (and only logical contributions) as valuable is a naive approach. Being able to accurately record all aspects of argumentation on social media is the first step towards being able to accurately analyse informal argument on an enormous scale. The work presented in this report provides a novel framework for modelling a subset of rhetorical argumentation, ideal for use in modelling social argumentation, and demonstrates some of the structures that may be observed when applied to three case studies. Bringing rhetorical and logical models of argumentation together with the computational modelling of social media argumentation has the potential to be a powerful tool in both our understanding of social media use and social argumentation. This raises the prospects for the development of new tools that could help communities manage argumentation, and counter diverse problems, from echo-chambers and groupthink to trolling and anti-social behaviour.

References

Alagoz, E. (2013). Social argumentation in online synchronous communication. *International Journal of Computer-Supported Collaborative Learning*, 8(4):399–426.

Anderson, A., Huttenlocher, D., Kleinberg, J., and Leskovec, J. (2012). Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 850–858. ACM.

Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. *American Journal of Political Science*, 56(2):271–285.

Aurisicchio, M. and Bracewell, R. (2013). Capturing an integrated design information space with a diagram-based approach. *Journal of Engineering Design*, 24(6):397–428.

Bench-Capon, T. J. (2003). Try to see it my way: Modelling persuasion in legal discourse. *Artificial Intelligence and Law*, 11(4):271—-287.

Bench-Capon, T. J. and Dunne, P. (2002). Value based argumentation frameworks. In *Research Report ULCS-02-001*, pages 444–453. Department of Computer Science, The University of Liverpool.

Bench-Capon, T. J. M. and Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial intelligence*, 171(10):619–641.

Berland, L. K. and Forte, A. (2010). When students speak, who listens?: constructing audience in classroom argumentation. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1*, pages 428–435. International Society of the Learning Sciences.

Bertea, S. (2004). Certainty, reasonableness and argumentation in law. *Argumentation*, 18(4):465–478.

Bex, F., Lawrence, J., Snaith, M., and Reed, C. (2013). Implementing the Argument Web. *Communications of the ACM*, 56(10):66–73.

Blount, T., Millard, D. E., and Weal, M. J. (2014). Towards Modelling Dialectic and Eristic Argumentation on the Social Web. In *14th workshop on Computational Models of Natural Argument*.

Blount, T., Millard, D. E., and Weal, M. J. (2015). An Investigation into the Use of Logical and Rhetorical Tactics within Eristic Argumentation on the Social Web. In *ACM Conference on Hypertext and Social Media*.

Braet, A. C. (1992). Ethos, pathos and logos in Aristotle's Rhetoric: A re-examination. *Argumentation*, 6(3):309–315.

Breslin, J. G., Decker, S., Harth, A., and Bojars, U. (2006). SIOC: an approach to connect web-based communities. *International Journal of Web Based Communities*, 2(2):133–142.

Budzynska, K. and Reed, C. (2012). The Structure of Ad Hominem Dialogues. In *Proceedings of Computational Models of Argument*, pages 410–421.

Caminada, M. and Amgoud, L. (2007). On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5):286–310.

Chalamish, M., Gabbay, D., and Schild, U. (2011). Intelligent Evaluation of Evidence Using Wigmore Diagrams. In *Proceedings of the 13th International Conference on Artificial Intelligence and Law*, ICAIL '11, pages 61–65, New York, NY, USA. ACM.

Chalamish, M., Hazoom, M., and Schild, U. (2013). Semi-automatic Creation of Wigmore Diagrams. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law*, ICAIL '13, pages 181–185, New York, NY, USA. ACM.

Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., South, M., Vreeswijk, G., and Willmott, S. (2006). Towards an argument interchange format. *Knowledge Engineering Review*, 21(4):293–316.

Conklin, J. and Begeman, M. L. (1987). gIBIS: A hypertext tool for team design deliberation. In *Proceedings of the ACM conference on Hypertext*, pages 247–251.

Conti, G. and Sobiesk, E. (2010). Malicious interface design: exploiting the user. In *Proceedings of the 19th international conference on World wide web*, pages 271–280.

Corbett, E. P. and Connors, R. J. (1999). Classical rhetoric for the modern student.

Cornetto, K. M. and Nowak, K. L. (2006). Utilizing usernames for sex categorization in computer-mediated communication: Examining perceptions and accuracy. *CyberPsychology & Behavior*, 9(4):377–387.

Davis, J. P. (2002). The experience of 'bad' behavior in online social spaces: A survey of online users. *Social Computing Group, Microsoft Research*.

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

Eisenstein, J. (2013). What to do about bad language on the internet. In *Human Language Technologies*, pages 359–369. North American Chapter of the Association for Computational Linguistics.

Gilbert, E., Bergstrom, T., and Karahalios, K. (2009). Blogs are echo chambers: Blogs are echo chambers. In *42nd Hawaii International Conference on System Sciences, HICSS'09*, pages 1–10. IEEE.

Goodwin, J. and Fisher, A. (2000). Wigmore's Chart Method. *Informal Logic*, 20(3):223–243.

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., and Uziel, E. (2010). Social media recommendation based on people and tags. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 194–201. ACM.

Hahn, U., Oaksford, M., and Corner, A. (2005). Circular arguments, begging the question and the formalization of argument strength. In *Proceedings of AMKLC'05, International Symposium on Adaptive Models of Knowledge, Language and Cognition*, pages 34–40.

Han, B. and Baldwin, T. (2011). Lexical normalisation of short text messages: Makn sens a #twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 368–378. Association for Computational Linguistics.

Hanna, R., Rohm, A., and Crittenden, V. L. (2011). Were all connected: The power of the social media ecosystem. *Business horizons*, 54(3):265–273.

Hendler, J. and Berners-Lee, T. (2010). From the semantic web to social machines: A research challenge for ai on the world wide web. *Artificial Intelligence*, 174(2):156–161.

Herring, S., Job-Sluder, K., Scheckler, R., and Barab, S. (2002). Searching for Safety Online: Managing "Trolling" in a Feminist Forum. *The Information Society*, 18(5):371–384.

Hodson, H. (2013). Moderate your language. *New Scientist*, 218(2912):18.

Hooi, R. and Cho, H. (2013). Deception in avatar-mediated virtual environment. *Computers in Human Behavior*, 29(1):276 – 284.

Jane, E. A. (2014). "Your a Ugly, Whorish, Slut" Understanding E-bile. *Feminist Media Studies*, 14(4):531–546.

Jiménez-Aleixandre, M. P. and Erduran, S. (2007). Argumentation in science education: An overview. In *Argumentation in science education*, pages 3–27. Springer.

Jørgensen, C. (1998). Public Debate – An Act of Hostility? *Argumentation*, 12(4):431–443.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1):59–68.

Karunatillake, N. C., Jennings, N. R., Rahwan, I., and McBurney, P. (2008). Formal Semantics of ABN Framework. Technical report, University of Southampton.

Kennedy, G. A. (1991). *Aristotle on rhetoric: a theory of civic discourses*. Oxford University Press.

Kerferd, G. B. (1981). *The Sophistic Movement*. Cambridge University Press.

Klein, M. (2010). Using metrics to enable large-scale deliberation. In *Proceedings of Collective Intelligence In Organizations: A Workshop of the ACM Group*, pages 103–233.

Konijn, E. A., Utz, S., Tanis, M., and Barnes, S. B. (2008). How technology affects human interaction. In Barnes, S. B., editor, *Mediated Interpersonal Communication*, chapter 1, pages 3–13. Routledge New York.

Kou, Y. and Nardi, B. (2013). Regulating anti-social behavior on the Internet: The example of League of Legends. In *Proceedings of iConference*. iSchools.

Krause, B., Schmitz, C., Hotho, A., and Stumme, G. (2008). The anti-social tagger: detecting spam in social bookmarking systems. In *Proceedings of the 4th international workshop on Adversarial information retrieval on the web*, pages 61–68. ACM.

Kunz, W. and Rittel, H. W. (1970). *Issues as elements of information systems*, volume 131. Institute of Urban and Regional Development, University of California Berkeley, California.

Lange, C., Bojars, U., Groza, T., Breslin, J. G., and Handschuh, S. (2008). Expressing Argumentative Discussions in Social Media Sites. In *Proceedings of the Workshop on Social Data on the Web*, Karlsruhe, Germany.

Lee, E.-J. and Shin, S. Y. (2014). When do consumers buy online product reviews? effects of review quality, product type, and reviewer's photo. *Computers in Human Behavior*, 31:356–366.

Leskovec, J. and Faloutsos, C. (2006). Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 631–636. ACM.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM.

Markova, S. and Petkovska-Mirčevska, T. (2013). Social media and supply chain. *Amfiteatru Economic*, 15(33):89–102.

Moor, A. d. and Aakhus, M. (2006). Argumentation support: From technologies to tools. *Communications of the ACM*, 49(3):93–98.

Munroe, R. (2008). Youtube Audio Preview. URI: http://blog.xkcd.com/2008/10/08/youtube-audio-preview/ (Accessed: 18/07/2014).

Newman, S. and Marshall, C. (1992). Pushing Toulmin too far: Learning from an argument representation scheme. In *Xerox PARC Tech Rpt SSL-92-45*.

O'Keefe, D. J. (1992). *Readings in argumentation*, volume 11, chapter 5, pages 79–91. Walter de Gruyter.

Palau, R. M. and Moens, M.-F. (2009). Argumentation mining: the detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, pages 98–107. ACM.

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Peroni, S. and Shotton, D. (2012). FaBiO and CiTO: ontologies for describing bibliographic resources and citations. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17:33–43.

Plato (380BC). *Book V*. The Republic. Basic Books. (Bloom, A.D. Trans. 1991).

Pólos, L. and Hannan, M. T. (2002). Reasoning with partial knowledge. *Sociological methodology*, 32(1):133–181.

Rahwan, I., Zablith, F., and Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial intelligence*, 171(10):897–921.

Reed, C., Wells, S., Devereux, J., and Rowe, G. (2008). AIF+: Dialogue in the Argument Interchange Format. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 172:311.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Communications of the ACM*, 43(12):45–48.

Rittel, H. W. and Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, 4(2):155–169.

Rowe, M., Angeletou, S., and Alani, H. (2011). Predicting discussions on the social semantic web. In *The Semanic Web: Research and Applications*, pages 405–420. Springer.

Schneider, J., Davis, B., and Wyner, A. (2012). Dimensions of argumentation in social media. *Lecture Notes in Computer Science*, 7603:21–25.

Schneider, J., Groza, T., and Passant, A. (2013). A review of argumentation for the Social Semantic Web. *Semantic Web*, 4(2):159–218.

Schneider, J., Villata, S., and Cabrio, E. (2014). Why did they post that argument? Communicative Intentions of Web 2.0 Arguments. In *Arguing on the Web 2.0*, Amsterdam. SINTELNET, European Network for Social Intelligence.

Schopenhauer, A. (2009). *The Art of Always Being Right*. Gibson Square.

Sherchan, W., Nepal, S., and Paris, C. (2013). A survey of trust in social networks. *ACM Computing Surveys (CSUR)*, 45(4):47.

Smith, T. (2009). The social media revolution. *International journal of market research*, 51(4):559–561.

Sood, S. O., Churchill, E. F., and Antin, J. (2012). Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, 63(2):270–285.

Spitzberg, B. H. and Hoobler, G. (2002). Cyberstalking and the technologies of interpersonal terrorism. *New Media & Society*, 4(1):71–92.

Suler, J. R. and Phillips, W. L. (1998). The bad boys of cyberspace: Deviant behavior in a multimedia chat community. *CyberPsychology & Behavior*, 1(3):275–294.

Sundar, S. S. (2000). Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads. *Journalism & Mass Communication Quarterly*, 77(3):480–499.

Suzuki, S. (2011). Perceptions of the qualities of written arguments by japanese students. *Written Communication*, 28(4):380–402.

Tindale, C. W. (2007). *Fallacies and argument appraisal*. Cambridge University Press.

Torroni, P., Prandini, M., Ramilli, M., Leite, J., and Martins, J. (2010). Arguments against the troll (position paper). In *1st Workshop on Argumentation in Artificial Intelligence and Philosophy*.

Toulmin, S. E. (1958). *The Uses of Argument*. University Press, Cambridge.

Van Eemeren, F. H. and Grootendorst, R. (2004). *A systematic theory of argumentation: The pragma-dialectical approach*, volume 14. Cambridge University Press.

Verheij, B. (2005). Evaluating arguments based on Toulmins scheme. *Argumentation*, 19(3):347–371.

Wigmore, J. H. (1913). *The Principles of Judicial Proof As Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials*. Little, Brown and Co., Boston, MA, 1st edition.

Willard, N. E. (2007). *Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress*. Research Press.