Machine Learning and Sentiment Analysis Approaches for the Analysis of Parliamentary Debates

Thesis submitted in accordance with the requirements of the
University of Liverpool for the degree of
Doctor in Philosophy

by

Zaher Ibrahim Saleh Salah

Faculty of Science
Department of Computer Science

May 2014

"It is not best that we should all think alike; it is a difference of opinion that makes horse races."

Mark Twain

Abstract

In this thesis the author seeks to establish the most appropriate mechanism for conducting sentiment analysis with respect to political debates; firstly so as to predict their outcome and secondly to support a mechanism to provide for the visualisation of such debates in the context of further analysis. To this end two alternative approaches are considered, a classification-based approach and a lexicon-based approach. In the context of the second approach both generic and domain specific sentiment lexicons are considered. Two techniques to generating domain-specific sentiment lexicons are also proposed: (i) direct generation and (ii) adaptation. The first was founded on the idea of generating a dedicated lexicon directly from labelled source data. The second approach was founded on the idea of using an existing general purpose lexicon and adapting this so that it becomes a specialised lexicon with respect to some domain. The operation of both the generic and domain specific sentiment lexicons are compared with the classification-based approach. The comparison between the potential sentiment mining approaches was conducted by predicting the attitude of individual debaters (speakers) in political debates (using a corpus of labelled political speeches extracted from political debate transcripts taken from the proceedings of the UK House of Commons). The reported comparison indicates that the attitude of speakers can be effectively predicted using sentiment mining.

The author then goes on to propose a framework, the Debate Graph Extraction (DGE) framework, for extracting debate graphs from transcripts of political debates. The idea is to represent the structure of a debate as a graph with speakers as nodes and "exchanges" as links. Links between nodes were established according to the exchanges between the speeches. Nodes were labelled according to the "attitude" (sentiment) of the speakers, "positive" or "negative", using one of the three proposed sentiment mining approaches. The attitude of the speakers was then used to label the graph links as being either "supporting" or "opposing". If both speakers had the same attitude (both "positive" or both "negative") the link was labelled as being "supporting"; otherwise the link was labelled as being "opposing". The resulting graphs capture the abstract representation of a debate where two opposing factions exchange arguments on related content.

Finally, the author moves to discuss mechanisms whereby debate graphs can be structurally analysed using network mathematics and community detection techniques. To this end the debate graphs were conceptualised as networks in order to conduct appropriate network analysis. The significance was that the network mathematics and community detection processes can draw conclusions about the general properties of debates in parliamentary practice through the exploration of the embedded patterns of connectivity and reactivity between the exchanging nodes (speakers).

Keywords: Sentiment Analysis, Machine Learning, Debate Visualisation, Debate Analysis & Information Retrieval.

Contents

List of Figures

List of Tables

List of Algorithms

Dedication

To the memory of my beloved Mother and Father.

Acknowledgement

Chapter 1

Introduction

"I love argument, I love debate. I don't expect anyone just to sit there and agree with me, that's not their job."

Margaret Thatcher

This chapter provides an introduction to the research work described in this thesis. The introduction commences in Section 1.1 with a brief overview of the background to the research. Section 1.2 then describes the motivations and potential benefits of the research work and Section 1.3 presents the research question and the associated research issues that the work is directed at. Section 1.4 describes how the research issues are addressed (the research methodology) while Section 1.5 describes the contributions of the work presented in this thesis. Section 1.6 describes how the rest of this thesis is organised and Section 1.7 presents the published work to date resulting from the content of the thesis. This chapter is concluded with a brief summary in Section 1.8.

1.1 Overview

Political analysis, whether this occurs in the form of "official" media (newspapers, television reports and so on) or "unofficial" media (such as blogs and social network sites), is an everyday part of our lives. Consequently the study of political debates is a popular area of sociological and cultural research. For example in [Welch, 1985] a study was undertaken to determine weather US congress women are more liberal than congress men by conducting a study of voting patterns. In [Porter et al., 2005] network analysis techniques were use to determine how the committees and sub-committees of

the US House of Representatives were interconnected. The study of political debates is also of interest in terms of how such debates operate, see for example the work of [Rissland, 1999] or [Thomas et al., 2006].

1.1.1 Political sentiment mining

One way to perform political analysis is using sentiment mining. In general, sentiment mining is concerned with various techniques to extract positive and negative feelings, opinions, attitudes or emotions, typically embedded within some form of text, concerning some object of interest [Liu, 2012, Asmi and Ishaya, 2012]. This object may be a product, a person, some legislation, a movie, or some kind of happening or topic. Sentiment mining is thus directed at the automatic retrieval and categorisation of subjective information embedded in various types of textual data as opposed to objective or factual information. Identifying the subjective information within a text is a challenging process on account of the fuzzy border between subjectivity and objectivity.

Sentiment mining is typically applied to a "document corpus" comprising either structured or unstructured free text. There are a variety of techniques that can be used for this purpose. One commonly used approach is the classification-based approach where a pre-labelled "training" corpus is used to build a classifier that can then be applied to previously unseen texts [Kim and Hovy, 2004, Pang and Lee, 2008] so as to extract the sentiment expressed within these texts. For example in [Dang et al., 2010] the approach was used in the context of product reviews and in [Kennedy and Inkpen, 2006] in the context of film reviews. Classifier-based sentiment mining techniques have been shown to perform well; an additional benefit is that classifier generation processes tend to be language independent. However, a disadvantage is that a pre-labelled training set (prior knowledge) is required; the resource needed to build such a training set is often prohibitive. A solution is the lexicon-based approach where sentiment lexicons are used to estimate the sentiment value/score and polarity (attitude) expressed within documents in a corpus by first identifying subjective words (words that convey feelings or judgement) and then "looking up" the identified words in a sentiment lexicon to obtain sentiment values (intensities) and polarities (positive or negative) for each word. These values and polarities can then be used to predict the overall polarity (attitude) for each document in the corpus [Esuli and Sebastiani, 2006, Denecke, 2009b, Montejo-Raez et al., 2012, Ohana and Tierney, 2009].

The most commonly used sentiment lexicon is the SentiWordNet 3.0 general purpose lexicon[1] which has the key advantage, over other such lexicons, that it covers a larger number of words. The problem with such general purpose lexicons is that they tend to

---

[1]sentiwordnet.isti.cnr.it.

not operate well with respect to specific domain corpora, because of the use of special purpose words (reserved words) and/or domain specific style and language that may be a feature of specialised domains such as the political debate domain. For example, given a specific domain, certain words and phrases may be used in a different context than their more generally accepted usage, in which case the words and phrases may reflect different sentiments than those that would be normally expected. Hence there is a view that general purpose lexicons are not well suited to sentiment mining in specialised domains. A solution is to use domain specific lexicons, however these tend not to be readily available and thus have to be generated. There are two techniques to generating such domain specific lexicons: (i) direct generation and (ii) adaptive generation. The first, as the name suggests, is founded on the idea of generating the desired domain-specific lexicon directly using the biased occurrence of words in a given pre-labelled training corpus (thus obviating the claimed advantage of lexicon-based sentiment mining approaches over classification-based approaches that a training set is not required). The second technique is founded on the idea of using an existing general purpose lexicon and adapting this so that it becomes a domain specific lexicon, again using pre-labelled training data.

In the context of political sentiment mining there are thus three main potential approaches that can be adopted: (i) classification-based, (ii) generic lexicon-based and (iii) domain specific lexicon-based (using either a direct or an adaptive lexicon generation technique). An obvious application of sentiment mining in the political context is with respect to the prediction of election outcomes by applying sentiment mining techniques to social media data (see for example [Metaxas et al., 2011]). Alternatively in [Tsytsarau and Palpanas, 2011] a sentiment mining-based mechanism is described for tracking how the sentiment expressed by members of the public or politicians evolves with respect to some piece of legislation or political topic over time.

1.1.2 UK House of Commons debates

To act as a focus for the work described in this thesis the political debates conducted in UK House of Commons are considered. Both houses in the UK parliament, the House of Commons and the House of Lords, reach their decisions by debating and then voting with either an *Aye* or a *No* vote at the end of each debate. The advantages offered are: (i) the proceedings of these debates are published on-line and (ii) the outcomes of these debates are known. The effectiveness of sentiment analysis approaches presented in this thesis can thus be evaluated by considering the known final outcomes of these debates.

1.2 Motivation

Thus from the foregoing, this thesis is concerned with the application of sentiment analysis to political debates. Little work has been conducted with respect to political sentiment mining focusing on parliamentary debates. The motivation for the work described in this thesis is thus a desire to be able to effectively predict the "attitude" of individual debaters within a political debate by deploying sentiment analysis techniques. The idea espoused in this thesis is that the extracted information about the attitude of the debaters, in addition to information about the exchanges made between them, can be used to create debate graphs that will in turn allow for the graphical summarisation and visualisation of the high-level structure of such debates. The expectation is that such debate graphs will provide an efficacious visualisation of the high-level structure of the debate such as, critically, who talks about similar issues (and to what extent), and who opposes whom (and how strongly). Once such debate graphs have been generated, the graphs may be analysed using network analysis techniques (see the examples on combining sentiment analysis and networks analysis presented in: [Bermingham et al., 2009, Gloor et al., 2009, Rabelo et al., 2012, Wang et al., 2013, Shams et al., 2012, Deitrick and Hu, 2013, Deng et al., 2013, Miller et al., 2011, Tan et al., 2011]) to identify their structural properties and highlight some of their features such as: how debaters are likely to vote; how parties interact in the debate; which debaters are more influential. The work described in this thesis is also likely to have practical benefits in contexts such as political campaign and debate management. To the best knowledge of the author there has been no previous work that has attempted to describe and analyse House of Commons debates in this manner.

1.3 Research Objectives

The broad objective of the research described in this thesis was thus to investigate the use of sentiment mining techniques for political debate analysis. More specifically the research work presented here is directed at three objectives:

Objective1: The application of sentiment mining techniques to predict the *attitude* of individual debaters, whether they are *for* or *against* a motion.

Objective2: The extraction of debate graphs describing and overviewing political debates from political verbatim transcripts.

Objective3: The analysis of the embedded graph structures, featured in debate graphs, with respect to how the individual participants interact.

Given the above three Objectives, two research Questions (RQs), that encompass a number of supplementary research questions, were identified as follows:

RQ1: *Is it possible to effectively predict the* attitude *of individual debaters, whether they are for or against a motion within the context of political debates?*
More specifically:

- How to use sentiment mining approaches to analyse political debates?
- What are the most appropriate sentiment mining approaches to predict the attitude of individual debaters?

RQ2: *Is it possible to represent and analyse debates as graphs using tools from the field of network analysis?*
More specifically:

- How best to extract graph structures from debate records?
- Which metrics and algorithms from network analysis to use to highlight structural features of debates?

1.4 Research Methodology

To address the research issues identified in the previous section a three phase programme of work was adopted:

- **The first phase** comprised an investigation of the most appropriate sentiment mining approaches issue associated with the first research question (**RQ1**) and how this might be resolved by a comparison between the three main identified sentiment mining approaches: (i) classification-based, (ii) generic lexicon-based and (iii) domain specific lexicon-based (two techniques, direct and adaptive). The comparison was conducted in terms of attitude prediction accuracy, the accuracy with which the approaches could be used to predict the known attitude of individual debaters in terms of how they eventually voted. Note that to evaluate the proposed attitude prediction processes, the predicted attitudes were compared with the known *attitudes* of the speakers defined according to whether, at the end of each individual debate, they voted *Aye* or *No*. Votes are often held to passing or rejecting a new piece of legislation proposed by the government or by an MP, Lord or even a member of the public or a private group or simply registering the opinion of the MPs on a subject. Because of party discipline the debates are typically not aiming at persuading other MPs to change their point of

5

view but rather to justifying why they voted with an *Aye* or a *No* vote at the end of the debate and reflecting their constituents' concerns and interests about the running events. In doing so it was assumed that the speakers' attitudes during their speeches reflect how the MP was going to vote. It was thus also assumed that speakers never change their minds during a debate.

- **The second phase** comprised consideration of the nature of a mechanism to extract graph structures from textual debates so as to partly address the research issues associated with the second research question (**RQ2**). The high level idea was to represent debates (visualise their structure) using a graph structure where the nodes represent speakers (debaters) and the links significant interactions (according to either (i) semantic similarity, (ii) interruptions made or (iii) combination of both semantic similarity and interruptions made) between debaters. Nodes and links were then labelled according to the attitude. Nodes were labelled with the attitude of the speaker, either "positive" or "negative" according to whether they are for or against the motion of the debate. Once the attitude of the debaters (nodes) is known the links may be labelled as follows. If two nodes connected by a link both have the same attitude label (both positive or both negative) then the link is labelled as being "supporting". If both nodes have different attitude labels (one is positive and the other is negative) the link is labelled as being "opposing".

- **The final phase** comprised an investigation into the application of network analysis techniques with respect to the second research question (**RQ2**). This phase focused on the identification of structural features of debates in terms of established network analysis metrics and community detection algorithms; where communities are identified by clustering debaters into groups (for example according to party affiliation, opinion or influences).

1.5   Research Contributions

The main contributions of the research work described in this thesis can be summarised as follows:

- A set of benchmark datasets extracted from proceedings of the UK House of Commons debates using information retrieval techniques to extract the required elements and attributes from the XML document archives.

- A domain specific list of parliamentary stop-words to support the preprocessing of such data.

- A framework for using machine learning classifiers in the context of political sentiment mining to classify the attitude (for or against a motion) of individual speakers in a political debate.

- A framework for using generic sentiment lexicons in the context of political sentiment mining to predict the attitude (for or against a motion) of individual speakers in a political debate.

- A framework for using domain specific sentiment lexicons in the context of political sentiment mining to predict the attitude (for or against a motion) of individual speakers in a political debate.

- A mechanism to determine the sentiment scores and polarities for terms in a pre-labelled corpus with regard to the biassed occurrences of these terms in this corpus.

- Two domain specific (political) sentiment lexicons, PoLex and PoliSentiWordNet, generated by applying the techniques described in this thesis to UK House of Commons benchmark data.

- A comparison of the performance, in terms of attitude prediction, of the three identified sentiment mining approaches.

- A Debate Graph Extraction (DGE) framework designed to extract debate graphs embedded within debate transcriptions.

- The conceptualisation of the extracted debate graphs as networks and an indication of how such networks might be used to analyse the structural properties of a debate graph.

1.6 Thesis Structure

The rest of this thesis is structured into nine chapters as follows:

Chapter 2   Presents a literature review of previous work relevant to the research work presented in this thesis. Background is presented concerning the three identified approaches to political sentiment mining commencing with the classification-based approach and then going on to consider the lexicon-based approaches (generic and domain specific). The chapter is completed with a review of recent work on sentiment analysis in the political domain.

Chapter 3 Introduces The UK House of Commons political debates corpus used for evaluation purposes with respect to the work described in this thesis and discusses the format and the characteristics of such parliamentary debates in addition to presenting the extraction process. Samples of selected debates are presented to illustrate the points raised. Statistics concerning a number of extracted "benchmark" datasets are presented. The automated data extraction and preparation processes are also described in detail.

Chapter 4 Considers political sentiment mining in terms of machine learning classification. In this chapter attitude classification, using off-the-shelf machine learning classifiers, in the context of mining the UK House of Commons political debates data is presented. The input to the generated classifier is the set of concatenated speeches that make up a single debate, the output is a set of attitude labels one per concatenated speech. More formally the input is a set of n concatenated speeches $S = \{s_1, s_2, \ldots, s_n\}$, and the output is a set of attitude class labels $C = \{c_1, c_2, \ldots, c_n\}$ taken from the set $\{positive, negative\}$ such that there is a one-to-one correspondence between the elements in S and C. The process encompasses two stages: (i) preprocessing and (ii) attitude prediction. Each of these stages is described in detail in the chapter.

Chapter 5 Presents political sentiment mining using generic sentiment lexicons. In this chapter attitude prediction, using the off-the-shelf generic SentiWordNet 3.0 sentiment lexicon, in the context of mining the UK House of Commons political debates data is considered. Given a new text which needs to be classified as expressing either a "positive" or a "negative" attitude, the subjective words in the text act as sentiment indicators. The first stage in the process comprises performing part-of-speech tagging so as to assign a part-of-speech tag to each word in the input text. The second stage is text preprocessing. Once the data has been pre-processed the attitude prediction (mining) phase can be commenced. To this end, the sentiment lexicon is used to look-up words firstly to identify the subjective words (as opposed to objective words) and secondly to determine the degree of sentiment and polarity (positive or negative) associated with the identified subjective words. The idea is to combine the subjective word-level sentiment values to give a whole document sentiment value. Each of these two stages is described in more detail in the chapter.

Chapter 6 Presents political sentiment mining using domain specific sentiment lexicons. This chapter considers attitude prediction using domain spe-

cific lexicons which operate in a similar manner to when using generic lexicons, with the exception that dedicated lexicons are used, as described in Chapter 5. The challenge is how best to generate the required specialist lexicons. Two approaches can be identified: (i) direct generation and (ii) adaptive generation. In both cases the input is a set of n binary labelled parliamentary speeches (conceptualised as documents) $D = \{d_1, d_2, \ldots, d_n\}$. The labels are drawn from the set $\{positive, negative\}$. The output in both cases is a lexicon where each term is encoded in the form of a set of tuples $\langle t_i, post_i, s_i \rangle$, where t_i is the term, $post_i$ is the part-of-speech tag associated with term t_i and s_i is the associated sentiment score. Both domain-specific lexicon generation approaches comprise four steps: (i) part-of-speech tagging (to identify the POS tags), (ii) document preprocessing, (iii) sentiment score (s_i) and polarity calculation and (iv) lexicon generation. Each of these steps is described in more detail in the chapter. With respect to evaluation, these two techniques were used to create two political-domain sentiment lexicon from the UK House of Commons political debates data: (i) PoLex produced using direct generation and (ii) PoliSentiWordNet produced using adaptive generation. The domain specific lexicons generation approaches is fully described in the chapter in addition to a description of how these lexicons may be used in the context of political sentiment mining.

Chapter 7 Compares experimentally the three approaches to political sentiment mining considered in Chapters 4, 5 and 6 by considering debater attitude prediction effectiveness. With respect to the machine learning approach, six classifiers were considered: Naïve Bayes, Support Vector Machine SMO, J48 decision trees learner, JRip rule-based classifier, IBk nearest neighbour classifier and ZeroR (the last as a baseline classifier). The generic lexicon used was SentiWordNet 3.0 and the domain specific lexicons used were PoLex and PoliSentiWordNet. The comparison was conducted using a corpus of the House of Commons political debates collection comprising 2,068 concatenated speeches (generated as described in Chapter 3). Recall that the classifiers were used to assign predefined attitude class labels ($\{positive, negative\}$) to each record, while the lexicons were used to assign sentiment scores to each record which were then used to determine the attitude label ($\{positive, negative\}$). Because the attitude of individual speakers with respect to each debate was known from the way that the speakers eventually voted, the predicted attitude could be compared with the known attitude. The metrics used for the comparison were precision, recall, the F-measure and average accuracy.

The F-measure (the harmonic mean of precision and recall) combines the precision and recall values and is thus a good overall measure. The obtained results are recorded and discussed in detail in the chapter. The outcome addresses the first research objective (**Objective1**).

Chapter 8 Introduces the design and the implementation of the proposed Debate Graph Extraction (DGE) framework. This chapter describes how the proposed DGE framework can be used for extracting embedded graph structures from transcripts of debates and generating the corresponding debate graphs to allow for graphical visualisation of the high-level structure of such debates. The idea is to represent the structure of a debate as a graph with speakers as nodes and exchanges (according to either (i) semantic similarity, (ii) interruptions made or (iii) combination of both semantic similarity and interruptions made) between debaters as links. Nodes are labelled with speaker attitude ("positive" or "negative"), and links are labelled as being "supporting" if both nodes (connected by a link) have the same attitude labels (both positive or both negative) or "opposing" if both nodes (connected by a link) have different attitude labels (one is positive and the other is negative). In total three different types of debate graph are generated using the proposed DGE framework: (i) semantic similarity debate graph, (ii) interruption debate graph and (iii) relevant interruption debate graph. The resulting graphs capture the abstract representation of a debate as two opposing factions exchange arguments on related content. The work described in this chapter is thus intended to address the second research objective (**Objective2**).

Chapter 9 Discusses mechanisms whereby debate graphs can be analysed using network metrics and community detection algorithms. This chapter describes the conceptualisation of debate graphs as networks in order to conduct appropriate network analysis. The significance is that the network metrics and community detection processes can lead to the prediction of debate outcomes through the exploration of the embedded patterns of connectivity and reactivity between the exchanging nodes (speakers). The work presented in this chapter was designed to address the third research objective (**Objective3**).

Chapter 10 Concludes the thesis by reviewing the contributions and main findings in terms of the identified research questions and issues. The chapter also revisits the research objectives and presents some ideas for future work.

1.7 Published Work

Some of the work described in this thesis has been published previously in a number of refereed publications as follows:

1. **Book Chapter**

    (a) Zaher Salah, Frans Coenen and Davide Grossi (2013). A Data Mining Approach to Extracting Debate Graphs. In Katie Atkinson, Henry Prakken and Adam Wyner (Eds.), From Knowledge Representation to Argumentation in AI, Law and Policy Making: A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday, London: College Publications, 2013, pp. 79-96. This paper described a framework for extracting debate graphs from debate transcriptions. The described framework used the machine learning-based approach to sentiment analysis, described in this thesis, in order to predict the attitude of speakers and then used this information to label the nodes and links in a debate graph. Similar work is presented in Chapters 4, 7 and 8.

2. **Technical Report**

    (b) Zaher Salah, Frans Coenen and Davide Grossi (2014). Political sentiment analysis: Predicting speaker attitude in the UK House of Commons. Technical Report ULCS-14-002, Department of Computer Science, University of Liverpool, UK, 2014. This technical report presented a comparison between the operation of the three different proposed mechanisms for conducting sentiment mining in the context of political debates with the objective of predicting their outcome. The content of this report provided the basis of work presented in Chapters 4, 5, 6 and 7.

3. **Conference Papers**

    (c) Zaher Salah, Frans Coenen and Davide Grossi (2013) Extracting debate graphs from parliamentary transcripts: A study directed at UK House of Commons debates. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law (ICAIL 2013), Rome, Italy, ACM Press, pp. 121-130. This paper introduced a framework, the Debate Graph Extraction (DGE) framework, for extracting debate graphs from transcripts of political debates. The described framework used the generic sentiment lexicons-based approach to sentiment analysis in order to predict the attitude of speakers and then used the attitude information to label the nodes

and links in a debate graph. The content of this paper is the precursor of work presented in Chapters 5, 7 and 8.

(d) Zaher Salah, Frans Coenen and Davide Grossi (2013) Generating domain-specific sentiment lexicons for opinion mining. In: Proceedings of the 9th International Conference on Advanced Data Mining and Applications (ADMA 2013), Hangzhou, China, 2013, Springer, Part I, LNAI 8346, pp. 13-24. In this paper, the two approaches to generating domain-specific sentiment lexicons were first proposed: (i) direct generation and (ii) adaptative generation. The work presented in this paper acted as the foundation for the work presented in Chapters 6 and 7.

(e) Zaher Salah, Frans Coenen and Davide Grossi (2014) Network Analysis of Parliamentary Debates: A Pilot Study on Two UK House of Commons Debates. To be presented at the First European Conference on Social Networks (EUSN), Barcelona, Spain. This paper described a pilot study on the conceptualisation of parliamentary debates as networks and their analysis by means of standard network analysis techniques. As a focus for the study two debates were chosen. For each debate two types of networks were built: (i) the interruption network and (ii) the relevant interruption network. The paper aimed to answer two research questions: (i) Do speeches by MPs normally respond to speeches of MPs with different party affiliation and/or different voting behavior? and (ii) Are standard community detection algorithms effective in singling out parties or sets of MPs with similar voting behavior? The content of this paper provided the basis of work presented in Chapter 9.

1.8   Summary

This introductory chapter has presented the context and the background of the research described later in this thesis. Details were given about the motivation for the research, the research questions and associated issues to be addressed, the adopted research methodology and contributions made. The following chapter presents a literature review, concerning the research work described in this thesis, aimed at providing the reader with the relevant background to the described work in much more detail than as presented in this introductory chapter.

Chapter 2

Previous Work

"Study the past if you would divine the future."

Confucius

2.1   Introduction

As noted in the previous chapter the work described in this thesis is concerned with the analysis of political debates. More specifically it is concerned with: (i) the application of sentiment analysis mechanisms and techniques to determine the "attitude" of debaters and (ii) with the visualisation and analysis of debates in the form of graphs. This chapter provides an overview of existing work that is of relevance with respect to the remainder of the thesis.

In the context of sentiment analysis most published research is directed at either classifier-based or lexicon-based techniques [Thelwall and Buckley, 2013] supporting the view expressed in this thesis. In classifier-based techniques labelled corpora (exhibiting prior knowledge) are used to learn some form of classifier using established approaches such as: Naïve Bayes, Support Vector Machines (SVM), Decision Trees or Neural Networks. In lexicon-based techniques a sentiment lexicon is used to retrieve word sentiment scores after which the sentiment attitude of a word, sentence or whole document is determined by summing, averaging or counting the individual sentiment scores. Classifier-based (machine learning) approaches have been shown to outperform lexicon-based approaches. However, the need for an appropriate training data set is often seen as a disadvantage. Learning a classifier is also a computationally expensive

process in terms of time and storage space. The research work described in this thesis thus considers both the classifier-based approach and the lexicon-based approach. In this previous work chapter, Section 2.2 considers previous work on classifier-based approaches and Section 2.3 previous work on lexicon-based approaches.

The author believes that visualisation is an important element of understanding the nature of political debates and the sentiments that may feature in such debates. Later in this thesis a mechanism for visualising debates is proposed using the concept of debate graphs. There is not a significant amount of reported work on the visualisation of political debates, however the work of [Kaptein et al., 2009] and [Marx, 2009] is notable. Thus Section 2.4 in this previous work chapter provides some background and commentary on relevant previous work in the context of debate structure visualisation. Although it is suggested in this thesis that the use of graphs to visualise the structure of debates is useful in its own right, such graphs also provide an opportunity for further sentiment analysis. Relevant background and selected previous work are thus also provided concerning the modelling and visualising of debates (Section 2.5).

There also exists some significant work on the analysis of political debates which is of relevance to the work described in this thesis. Early examples include [Welch, 1985] who investigated whether US congress women are more liberal than congress men by conducting a study of voting patterns and [Porter et al., 2005] who used network analysis techniques to determine how the committees and sub-committees of the US House of Representatives were interconnected. Section 2.6 thus reviews some of this work, concentrating on the application of sentiment analysis within the political domain. This chapter is completed with a brief summary in Section 2.7.

2.2 The classifier-based approach to sentiment extraction

In the classification-based (data mining or machine learning-based) approach to sentiment analysis/extraction a pre-labelled training corpora (exhibiting prior knowledge) is used to learn a "classifier" using some established supervised learning mechanism. The training data comprises a collection of ordered pairs $\langle s, c \rangle$ where s is an instance (observation) comprised of a set of attribute (feature) values and c is a known class label for the instance taken from a set of class labels C. Once the classifier has been generated it can be used to assign documents to the "fittest" class; essentially performing a mapping $s_i \rightarrow c_i$ where $c_i \in C$ (the set of known class labels). It has been argued that classification-based approaches in political sentiment mining tend to work well [Grijzenhout et al., 2010]. However, the need for appropriate training data is a limiting factor, and the learning process is highly dependent on the quality of the prior knowledge (historical data) available. Figure 2.1 shows the process of training (learn-

ing) and testing a classifier [Bird et al., 2009]. Once a classifier has been generated it can be applied to "unseen" data, provided that the unseen data is pre-processed in the same manner as that used with respect to the training data originally used to produce the classifier. Confidence in a generated classifier is typically gained by applying the classifier to pre-labeled test data.

Classifiers can be generated in a variety of different ways which in turn also dictates their usage. The following is a brief description of some of the most commonly used machine learning classifiers (which were used with respect to the work described in this thesis as reported in Chapter 4) for sentiment classification in sentiment mining.

- **Naïve Bayes**: The classifier uses training data to learn the conditional probability of each attribute given the class label and generates a probabilistic model of the features. This model is then used to predict the class of new instances using the highest posterior probability [Duda et al., 2001].

- **Support Vector Machine**: Results in a discriminative classifier-based on the concept of a separating hyperplane[1] (class boundary) placed between a set of objects having different class memberships [Theodoridis and Koutroumbas, 2008]. In other words, given a labelled training dataset, an optimal hyperplane (decision plane) is defined which can then be used to classify the new instances. Support Vector Machines (SVMs) have been shown to work well with respect to textual data [Joachims, 1998]. However, two notable disadvantage of SVMs are: (i) they are directed at binary classification problems and thus tend to be not suited to multi-class classification, and (ii) they are a black box technique in that it is unclear how a particular SVM, once generated, operates.

- **Decision Trees**: The algorithm learns a classifier from labelled training data by considering each data attributes in turn using some measure, such as information gain, to determine the discriminative power of each attribute. The splitting procedure stops if all instances in a subset belong to the same class. In this manner a "decision tree" is built where the internal nodes represent individual nodes. Leaf nodes (terminals) represent class labels [Duda et al., 2001]. Decision tree classifiers offer the advantage that they are easily understandable in that explanations as to why a certain classification is so can be easily generated.

- **Rule-based**: Classifiers built using rule-based approaches consist of a set of conditional " if ... then ... " style rules. A training dataset of labelled observations is used to extract the classification rules and to build the classifier. Classification

---

[1] A separating hyperplane is a decision boundary which can be used for classification. The best hyperplane is the one that represents the largest separation between the two classes.

rules are used in a given order during the prediction process so as to assign a class label to a new unlabelled observation (instance) [Duda et al., 2001]. Rule-based classifiers offer the advantage, as in the case of decision tree classifiers, that they are easily understandable by non-experts and that explanations can be easily generated.

- **Nearest neighbour classifier**: Nearest neighbors-based algorithms have been extensively used for classification purposes. The idea is to simply find a predefined constant number k of the most adjacent (closest in distance) training instances to a new instance and then use the labels from the k identified instances to predict the label for the new instance. This is typically done using a simple majority vote [Theodoridis and Koutroumbas, 2008]. The K Nearest Neighbour (KNN) form of classification is an instance-based, or non-generalizing, learning method in that a "general model" of the application domain is not built (as in the case of all the foregoing methods). KNN classification has been shown to be successful with respect to classification tasks with very irregular decision boundaries [Duda et al., 2001]. A disadvantage of KNN classification is the complexity of searching for the nearest neighbours, especially in the context of high dimensional feature spaces [Theodoridis and Koutroumbas, 2008].

- **Baseline classifier**: Baseline classifiers simply predict the most common class [Nasa and Suman, 2012]. The ZeroR algorithm [Witten et al., 1999] is an exemplar baseline classifier. Baseline classifier have little practical usage, however they are useful in experimental contexts to provide a "baseline" with which the operation of other (real) classifiers can be compared.
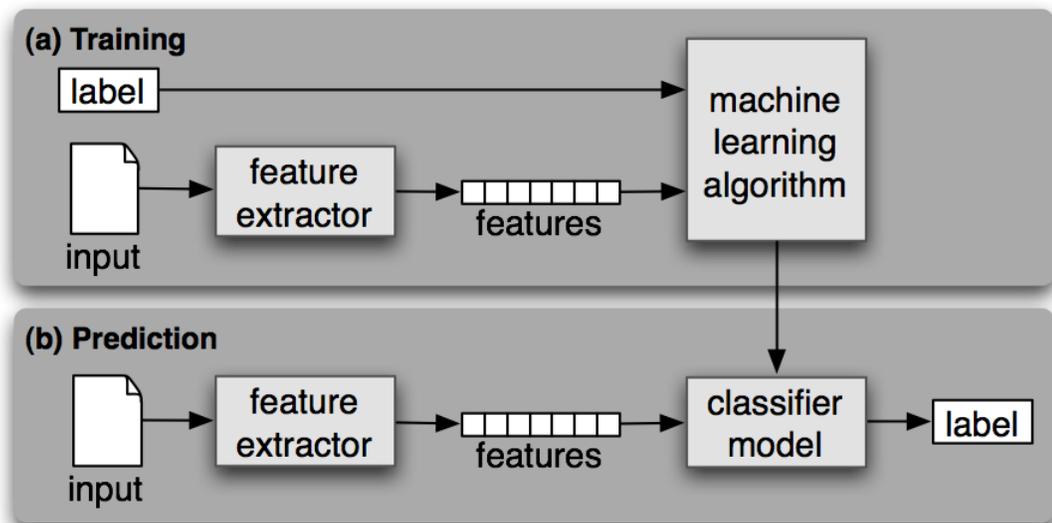
Figure 2.1: Training and testing a machine learning classifier.

2.3 The lexicon-based approach to sentiment extraction

Sentiment lexicons are lexical resources used to support sentiment extraction (and by extension analysis). More specifically they are used to assign a sentiment *value* (or *score*) and a *polarity* (or *orientation*) to a word. A sentiment value is a numeric value indicating some degree of subjectivity. The polarity (positive or negative) of a word is an indicator of whether the word expresses assent or dissent with respect to some object or concept. Consequently, document polarity can be judged by counting the number of positive and negative subjective words, summating their sentiment values and then calculating the difference. The result represents the attitude (positive or negative) of the document. Relatively small sized sentiment lexicons, which are built manually, can be extended by applying lexical induction techniques that exploit the semantic relationships between terms and their synonyms and antonyms, or by measuring term similarities in large corpora. Two types of sentiment lexicon can be used in the context of sentiment analysis: (i) generic (domain-independent) and (ii) dedicated (domain-specific) sentiment lexicons. More detail concerning these two types of lexicons is presented below in Sub-Sections 2.3.1 and 2.3.2 respectively.

2.3.1 Generic lexicon-based sentiment mining

The most commonly used generic (topic-independent) sentiment lexicon is the "off-the-shelf" SentiWordNet 3.0[2] sentiment lexicon [Baccianella et al., 2010], which is founded on WordNet 3.0[3]. WordNet is a large lexical repository of English words grouped into sets of cognitive synonyms called *synsets* expressing distinct concepts. Synsets are interlinked by means of conceptual-semantic and lexical relations. SentiWordNet 3.0 is an extension of SentiWordNet 2.0 (based on WordNet 2.0) which in turn is derived from SentiWordNet 1.0 (based on WordNet 1.0) [Esuli and Sebastiani, 2006]. SentiWordNet 3.0 associates to each synset s of WordNet a set of three scores: $Pos(s)$ ("positivity"), $Neg(s)$ ("negativity"), $Obj(s)$ ("neutrality" or "objectivity"). The range of each score is $[0,1]$ and for each synset s, $Pos(s) + Neg(s) + Obj(s) = 1$. Table 2.1 presents some statistics with respect to a number of popular sentiment lexicons, including SentiWordNet 3.0 [Ohana and Tierney, 2009]. From the table it can be seen that, out of the four lexicons listed, SentiWordNet 3.0 has the key advantage of covering the largest number of words.

---

[2]SentiWordNet 3.0 is accessible at `sentiwordnet.isti.cnr.it`.

[3]WordNet is accessible at `http://wordnet.princeton.edu/`.

Sentiment lexicon	Total number of sentiment bearing terms
SentiWordNet 3.0	**117659**
Subjectivity Clause Lexicon	7650
General Inquirer	4216
Grefenstette	2258

Table 2.1: Coverage of SentiWordNet 3.0 compared to other (manually built) sentiment lexicons [Ohana and Tierney, 2009].

2.3.2 Domain specific lexicon-based sentiment mining

As noted above, sentiment analysis using generic sentiment lexicons is a challenging process in the context of topic-dependent domains [Thelwall and Buckley, 2013]. In such cases it is desirable to use dedicated domain specific sentiment lexicons. However, the main issue with the usage of such dedicated lexicons is that they are frequently not readily available and thus have to be specially generated, a process that may be both resource intensive and error prone.

As noted in the introduction to this thesis, two approaches may be identified to generating specialised (dedicated) lexicons for domain specific sentiment analysis: (i) creating a new dedicated lexicon or (ii) adapting an existing generic lexicon. Both techniques use labelled corpora (training data) from a specific domain. An example of the first technique (creating a new dedicated lexicon) can be found in [Birla et al., 2011] where a semi-automated mechanism is proposed to extract domain-specific health and tourism words from noisy text so as to create a domain-specific lexicon. Examples of the second technique (adapting an existing general lexicon) can be found in [Demiroz et al., 2012] and [Choi and Cardie, 2009]. In [Demiroz et al., 2012] a simple algorithm was proposed to adapt a generic sentiment lexicon to a specific domain by investigating how the words from the generic lexicon are used in the specific domain context in order to assign new polarities to these words. In [Choi and Cardie, 2009] Integer Linear Programming[4] was used to adapt a generic sentiment lexicon into a domain-specific lexicon; the method combined the relationships between words and opinion expressions so as to identify the most probable polarity of lexical items (positive, negative, neutral or negator) for the given domain.

There is also reported work that combines the two techniques (adapting the sentiment scores of the terms in the base lexicon and additionally appending new domain words to extend the base lexicon). For example [Weichselbraun et al., 2011] created a

---

[4]Integer Linear programming is a mathematical optimisation of a linear objective function in which some or all of the variables are restricted to be integers.

domain-specific sentiment lexicon using crowd-sourcing for assigning sentiment scores to sentiment terms and then automatically extending an initially proposed base lexicon using a bootstrapping process to add new sentiment indicators and terms. The lexicon is then customised according to some specific domain. The evaluation conducted indicated that the created lexicon outperformed a generic sentiment lexicon (the General Inquirer Sentiment Lexicon[5]). Further reported work concerned with the "dual approach" to generating domain-specific lexicons can be found in [Qiu et al., 2009, Lau et al., 2011, Ringsquandl and Petković, 2012].

The sentiment score to be associated with each term in a lexicon can be calculated either by: (i) investigating the biased occurrence of the term with respect to a labelled (positive or negative) "training set"; (ii) utilising the semantic, contextual or statistical relationships between terms (words) in an input domain corpus; or (iii) learning a classifier (see above) to assign sentiment polarity to terms. In the context of the calculation of sentiment scores with respect to specific domains [Zhang and Peng, 2012] proposed a method to calculate the sentiment score of each word or phrase in different domains and use these scores to quantify sentiment intensity. [Thelwall and Buckley, 2013] proposed two approaches to improve the performance of polarity detection using lexical sentiment analysis with respect to social web applications, focusing on specific topics (such as sport or music). The two approaches were: (i) allowing the topic mood to determine the default polarity for false-neutral expressive text, and (ii) extending an existing generic sentiment lexicon by appending topic-specific words. The mood method slightly outperformed the lexical extension method. On the other hand it was found to be very sensitive to the "mood base" used, thus it was necessary to analyse the corpus first in order to choose an appropriate mood base relative to the corpus. Both methods require human intervention to either annotate a corpus (mood method) or to select terms (lexical extension).

2.4   Related work on visualising the debate structure

Debate visualisation is a central theme of the work described in this thesis. More specifically this thesis espouses the idea of visualising the structure of debates using the concept of debate graphs. This section presents existing work on the visualisation of debates (without sentiment analysis) and, in addition, the visualisation of the structure of arguments which has some overlap with the work described in this thesis. Selected previous work is also presented with respect to the identification and visualisation of the content of a speech (for example as a word-cloud or as a tabular visualisation) as

---

[5]The General Inquirer Sentiment Lexicon was built using the sentiment information contained in the General Inquirer which is a lexicon containing part-of-speech tagged words associated with syntactic, semantic, and sentiment information (see [Stone et al., 1966] and [Psathas, 1969] for more information).

well as relevant work using sentiment analysis to support debate visualisation.

There has been some published work on the visualisation of debates which has some bearing on the generation of debate graphs as conceived with respect to the research described in this thesis. In this context the work of [Kaptein et al., 2009] is of note. A mechanism was described in [Kaptein et al., 2009] for capturing debate structure using annotations of meeting notes and knowledge of "interruptions" with respect to the operation of the Dutch Parliament. Individual speeches and interruptions were summarised using word clouds. This structure was then visualised in a graph format where the nodes represented individuals and weighted arrows represented "interruptions". This graph format bares some similarity to the debate graphs proposed later in this thesis. Knowledge of "interruptions" was also used in the work of [Marx, 2009] to visualise parliamentary debate structure in terms of "debate time-lines". Each time-line represents a speaker and shows how he/she was interrupted and by whom.

Although there has been very little reported work on the generation and/or usage of debate graphs, there has been significant work on the visualisation of arguments which has some overlap with the work described in this thesis. The field of argumentation is concerned with the study of logical reasoning to arrive at conclusions [Rahwan and Simari, 2009]. Argumentation is an interdisciplinary field encompassing dialogue, negotiation, persuasion and, to an extent, debate. In the context of computer science the study of argumentation is broadly concerned with automating the argumentation process. A "support"/"oppose" classification, and its usefulness in argument visualisation, is argued for in [Birnbaum, 1982] who identified a number of frequent patterns of interaction in arguments. The author concluded that useful structural properties, abstracting specific propositions, are embedded in arguments; and that by inspecting frequent patterns of support and attack relations that involve several propositions an argument structure can be identified and visualised as a network of propositions connected by "support" (if a proposition is an evidence relation) or "attack" (if a proposition is an attack relation). The process described in [Birnbaum, 1982] relies on the annotations associated with the debate transcripts and does not employ the concept of sentiment analysis as advocated by the research work described in this thesis. A number of graph-based mechanisms have been proposed to support theories of argumentation such as Dung's argument framework [Dung, 1995]. The distinction between debate graphs (see Figure 2.3 which gives an example of a debate graph), as conceived of in this thesis, and argument graphs is that in the latter case each node is usually taken to represent one argument (of some inferential kind) and each link (which are directed) taken to represent that the argument at the source node has some bearing on the argument at the target node (for example an attacking or supporting relation). Besides the abstract theory of argumentation frameworks, research in argumentation has also produced a number of

systems that support argument visualisation. An early review of the visualisation of arguments is available in [Kirschner et al., 2003]. This book presented a comprehensive guide for the application of Computer Supported Argument Visualisation (CSAV) tools in education, science, public policy and business. These sense-making tools may be used effectively to support a better understanding of the multi-stakeholder and ill-structured problems by providing suitable representations, services and user interfaces to visualise different argumentation models in different contexts.

A comprehensive review and comparison between different argumentation visualisation systems can also be found in [Scheuer et al., 2010]. For instance, some systems like Digalo [Glassner and Schwarz, 2005] allow users to freely create their own arguments, while others like Araucaria [Reed and Rowe, 2004] prompt users to analyse arguments extracted from a transcript. On the other hand, some systems are collaborative like DebateGraph (`http://www.debategraph.org`) or single-user like Athena [Rolf and Magnusson, 2002], Carneades [Gordon, 2007] and Rationale [Gelder, 2007]. The Digalo tool may be used to map knowledge in an argumentative structure by defining a space of discussions called a "map". Users contribute to the discussion by adding shapes (for examples, rectangle, pentagon, or hexagon), with a short text as the title of the shape, to represent argumentative ontology like claims, evidence, or explanation. Users may also add links between the added shapes using supporting, opposing, or a neutral arrows. Araucaria is an argument mapping software tool to visually represent arguments as diagrams that can be used for analysis and stored in Argument Markup Language (AML), while Carneades is a software tool for summarising, evaluating and visualising the arguments of a debate as an argument graph. Debategraph is a web-based collaborative visualisation tool, may be used by communities of any size, to navigate and visualise individual debates and dialogue maps through different types of bubble, box, tree and outline views, while Athena argument mapping software is designed to support analysis and production of reasoning and argumentation by individual users such that hierarchical argument structures with premises and conclusions may be built, represented and evaluated. Figure 2.2 shows an example of on argument visualisation produced using the Rationale[6] software tool.

By comparing this argument visualisation with the debate graph format proposed later in this thesis and illustrated in Figure 2.3 it can be seen that: (i) a debate graph represents the reactive structure of a debate, while an argument graph represents the logical structure of an argument; (ii) nodes[7] in a debate graph are labelled according to speaker attitude, and edges (which are undirected) are identified according to the similarity between speeches and labelled according to whether the end nodes support or

---

[6]http://rationale.austhink.com/.

[7]Each node represents a debater.

oppose one another, while nodes in an argument graph represent propositions (the conclusion is shown at the top of the argument graph) and arrows (which are directed links) represent evidential support/oppose relationships; and (iii) debate graphs are generated automatically and the input to the debate graph generation (DGE) framework (see Chapter 8) are speakers' concatenated speeches comprising a whole debate, while argument graphs are typically generated manually. For example, Araucaria provides a simple point-and-click interface to allow the user to manually model and visualise an argument. The user reads a text containing an argument and then creates an argument graph by dragging lines (representing inferences) from one node (proposition) to another [Macagno et al., 2006]. Figure 2.6 shows a screen-shot of the Araucaria interface. Thus systems such as Araucaria and Carneades are directed at visualising the internal representation of the logical structure of arguments. Finally, in the context of debate analysis, as opposed to visualisation, there is also existing work directed at identifying and displaying keywords presented in a tabular format. In this case a visual analysis tool, like Termite[8], can be used to support the analysis of topics embedded within text corpora; for example by identifying latent topics based on co-occurring words. Figure 2.4 shows a tabular visualisation for co-occuring keywords produced using Termite [Chuang et al., 2012]. Keywords can also be identified and presented as word-clouds using text visualisation tools like Wordle[9]. Figure 2.5 shows a word cloud (created using Wordle) from the text of Barack Obama's speech at the Democratic National Convention in 2008 [Viegas et al., 2009].

A key idea presented in this thesis is that the use of graphs to represent the results of the application of sentiment analysis techniques provides a useful mechanism for enhancing the usability of these results. This idea is supported by the work of a number of authors including [Wanner et al., 2009], [Sobkowicz et al., 2012], [Tavares et al., 2012], [Wang et al., 2013] and [Miller et al., 2011]. The work of [Wanner et al., 2009] provides a good example of the utility of sentiment visualisation in the context of RSS news feeds. In [Wanner et al., 2009] sentiment analysis is combined with a visualization technique to reveal the emotional content of RSS news feed data over time. More specifically, selected positive and negative news items about the presidential candidates Obama and McCain, the vice president candidates Biden and Palin and the two major parties in the US presidential elections in 2008 were analysed. A case study demonstrated how the graphical presentation of the news postings might be a useful alternative for inferring meaningful conclusions from the postings without first having to read all the postings. [Sobkowicz et al., 2012] proposed an "opinion formation" framework for online opinion tracking and the simulation of changes in social reactions to specific topics and policy implementations. The framework comprised three

---

[8]Available on-line at: https://github.com/StanfordHCI/termite.
[9]Available on-line at: http://www.wordle.net/.
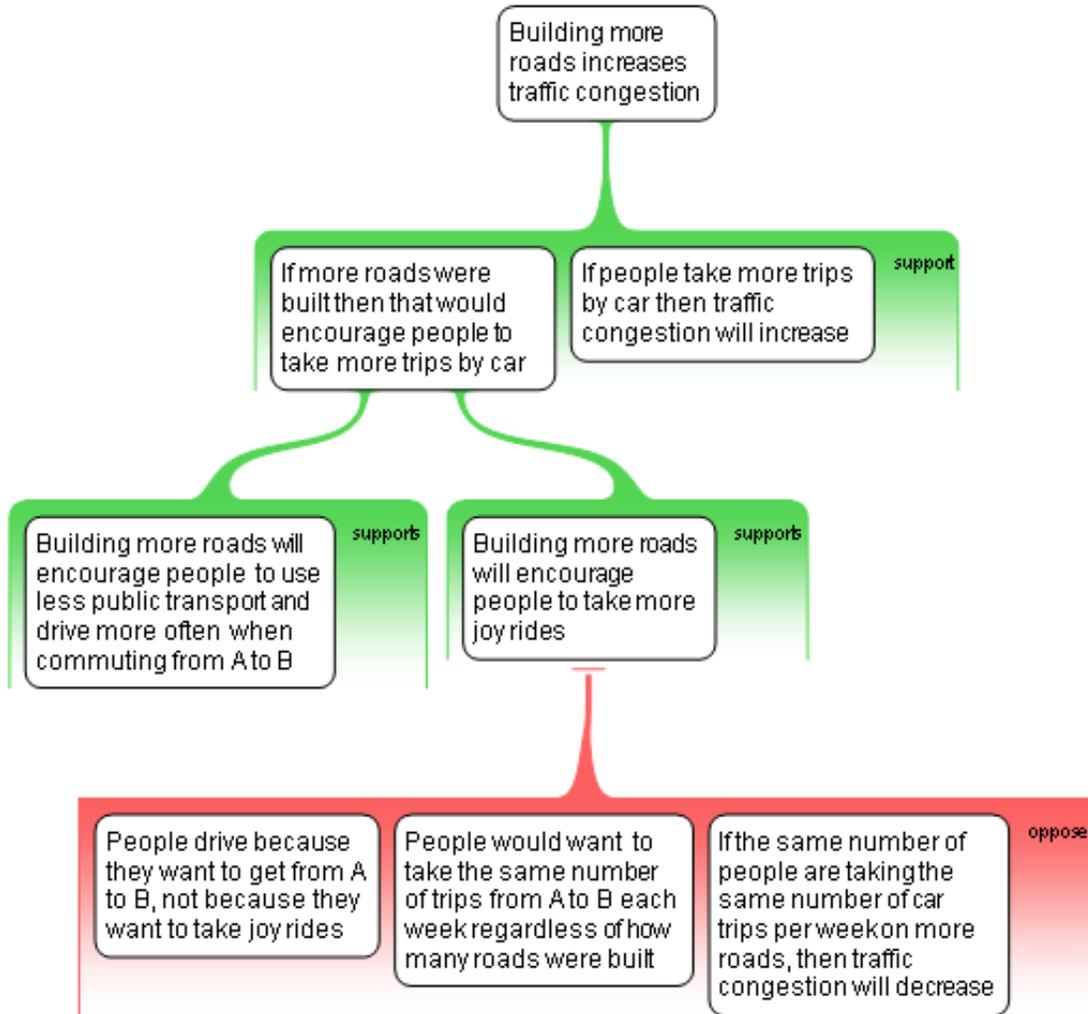
Figure 2.2: Argument structure visualisation produced using the Rationale software tool. Source: Wikimedia Commons.
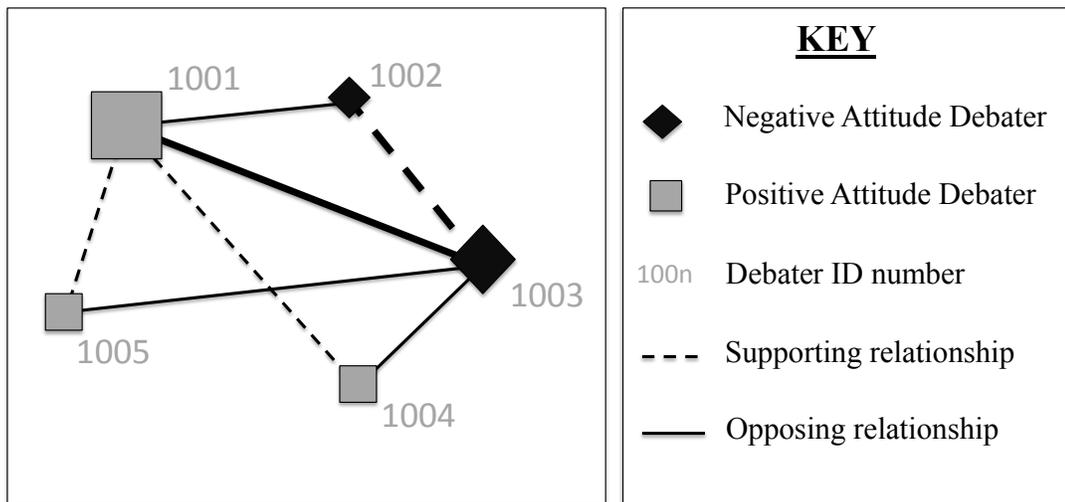
Figure 2.3: Simple debate graph of the form proposed in this thesis.
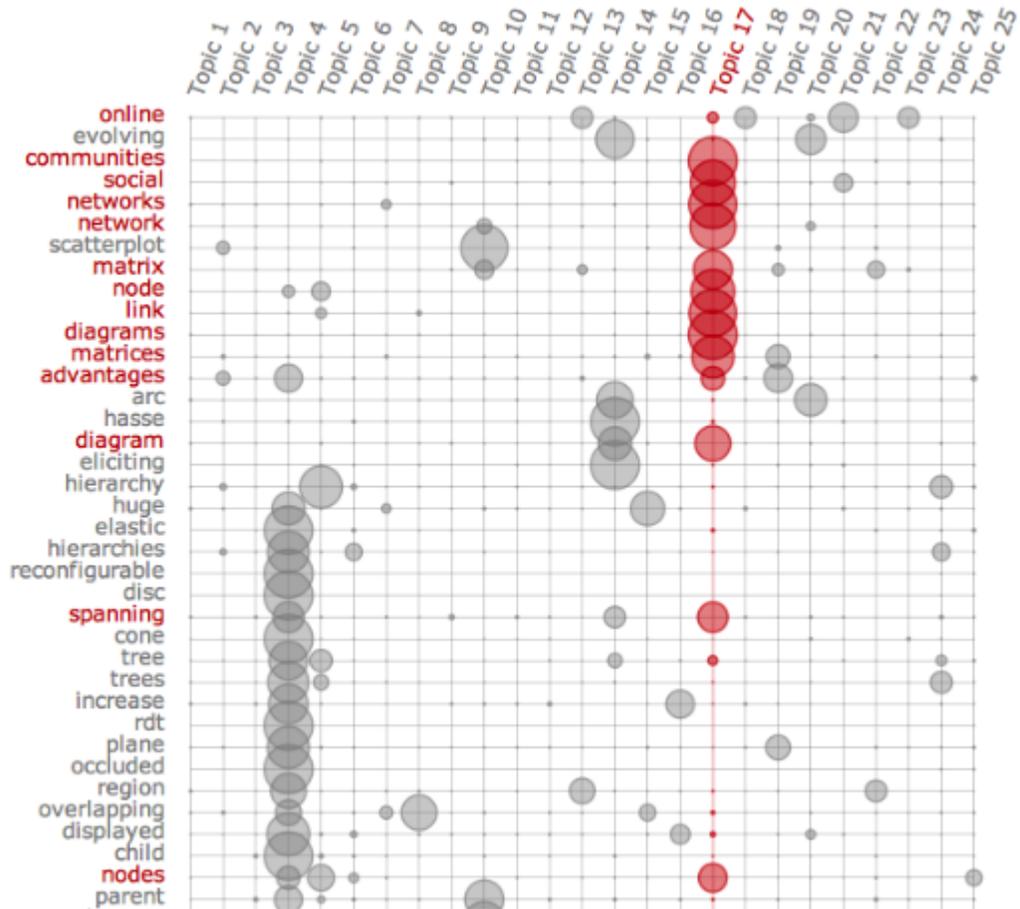
Figure 2.4: A tabular visualisation, produced using Termite, displaying the cooccurrences for a selected set of 30 terms.

Figure 2.5: A word-cloud from the text of Barack Obama's speech at the Democratic National Convention in 2008. Generated using Wordle.

Figure 2.6: Argument structure visualisation using the Araucaria software tool.

basic phases: (i) real-time opinion and topic detection based on content analysis, (ii) information flow modelling and simulation, and (iii) modelling of the opinion network. [Tavares et al., 2012] developed visualisation methods to support decision makers so as to allow them to follow political discussions. Their method commenced by extracting a set of relevant information about the most influential participants, their allies and "objector groups", and which messages or issues are the most popular, consensual and/or controversial. This extracted information was then presented in a graphical form. [Wang et al., 2013] proposed a visualisation system called *SentiView* which, unlike the other systems described above, was an interactive visualisation system directed at verifying and modelling the changes in public sentiment with respect to popular internet topics. Four sets of attributes were extracted as follows:

1. Number of forward blogs (i.e. new blogs following other blogs)

2. Number of commented blogs (i.e. blogs commenting on others blogs)

3. Total number of forward for the original blog

4. Total number of comments on the original blog

The system took into consideration changes to the above multiple attributes and complex relations among the attributes, such as numbers, location distribution, ages, and sentiment.

A text-based sentiment mining method and a model-driven prediction approach was included to analyse the public sentiments concerning popular topics. The relationships of interest among different participants (users debating some topics) were graphically presented using either a "sentiment helix" or a "sentiment relationship map". A Sentiment helix in this context is essentially an extension of line graphs that use line heights to represent temporally changing data (time-varying sentiments of participants). The ascent of a helix represents the overall tendency of public sentiment over time, while its width represents the number of participants, as shown in Figure 2.7. Figure 2.7 demonstrates the polarised opinions for "3Q WAR" which is a conflict between two popular Chinese IT companies. The sentiment relationship map highlights the participants who have a common interest in the forum (The Tianya community forum in China[10]) and facilitates the further analysis of the sentiments expressed concerning "hot topics". Figure 2.8 shows a sentiment relationship map where participants drawn as points and links as relationships. [Miller et al., 2011] explored the flow of sentiment information through a large collection of linked web pages. Pages and links between pages were represented graphically as a network where nodes represented web pages

---

[10]http://www.tianya.cn

and directed links represented the hyper-links between pages. The authors argued that the sentiment of a blog post is affected by its position in a cascade (an information propagation graph) as well as by the objectivity of its immediate parent. "Information cascades are phenomena in which individuals adopt a new action or idea due to influence by others". Sentiment intensity (objectivity/subjectivity) and polarity (positivity/negativity) were both determined by combining the sentiment scores of all words in a post using Harvard Inquirer and SentiWordNet.
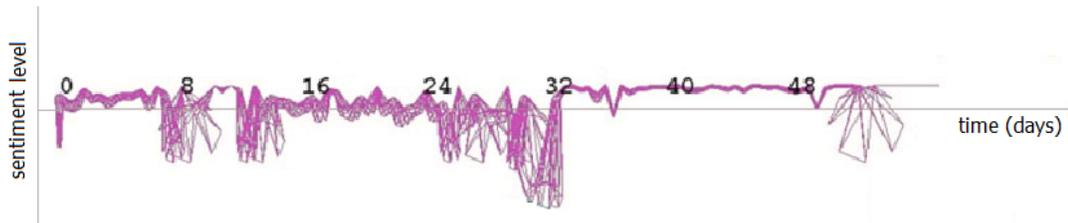
Figure 2.7: A sentiment helix illustrating the time distribution and evolution of sentiments for six popular Chinese IT companies debating about a conflict between other two IT companies. Source: [Wang et al., 2013].

2.5 Related work on graph networks analysis for political sentiment mining

As already noted the research work described in this thesis is directed at identifying and analysing the hidden and embedded structures contained in political debates (or more specifically text-based transcripts of debates) using sentiment analysis techniques. The intuition here is that knowledge of this structure may be used to extract useful information about how individual debate participants interact. This knowledge can be derived using network analysis techniques, particularly social network analysis techniques. Thus the work described in this thesis can also be viewed as being directed at a "bringing together" of sentiment and network analysis techniques to facilitate a better understanding of the structure of political debates. In this context there is some work that is of relevance, notably the work of [Shams et al., 2012], [Conover et al., 2011], [Rabelo et al., 2012], [Deitrick and Hu, 2013], [Younus et al., 2011] and [Tan et al., 2011] where network analysis techniques were used so as to improve sentiment analysis using machine learning approaches, while in the work of [Bermingham et al., 2009] and [Somasundaran et al., 2009] network analysis techniques were used to improve sentiment analysis using lexicon-based approaches.
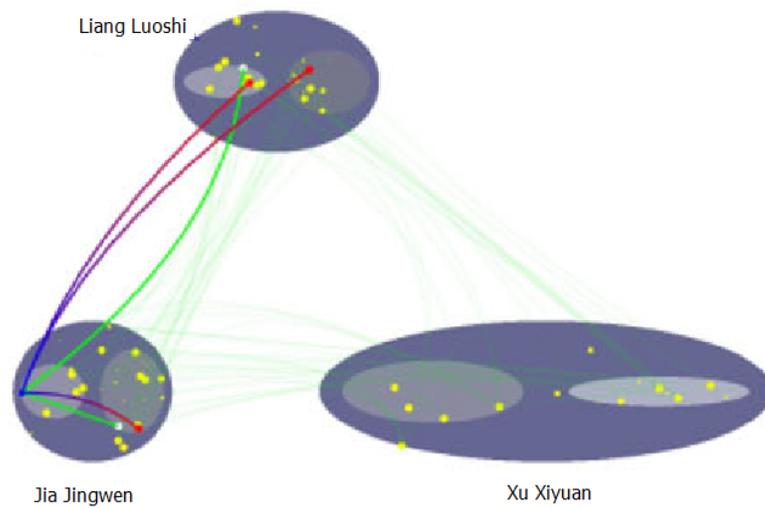
Figure 2.8: A sentiment relationship map for the hot topic: "Many actresses dating the rich and powerful". The size of a point represents the total number of subjective words in the post and the position of the point indicates the posts sentiment polarity. The green lines are used to link the same participants in positive (in the left side) and negative (in the right side) inner ellipses while the lines start with red and end with blue are used to indicate the relationship that the authors with red colour pay much attention to the information of blue points. Source: [Wang et al., 2013].

In [Shams et al., 2012] a method for combining both *textual* sentiment analysis and *structural* social network analysis was proposed directed at modelling the way that users of eCommerce sites operate. More specifically the combination of sentiment analysis and social network analysis was employed to identify rules with which to classify (using an SVM classifier) customers which in turn indicated customers preferences with respect to clusters of products. Experimental results, using an Amazon dataset, demonstrated that the proposed method produced better user models than systems based only on textual processing of reviews. [Conover et al., 2011] generated several machine learning classifiers for predicting the political alignment (affiliation) of politically-active Twitter users based on the content (sentiment-based information) and the structure (network-based information) of their political communication networks in the run-up to the 2010 U.S. midterm elections. Experiments based on the combination of content and network analysis were conducted using a dataset of 1,000 manually-annotated individuals. From these experiments it was found that the proposed sentiment and network analysis-based technique outperformed the techniques based on users' tweets. The partisan structure of the retweet network (community structure) and metadata (hashtag features) were very effective in predicting the political alignment of the users. Latent Semantic Analysis[11] (LSA) was applied to the users' tweets to detect the hidden structures of the embedded topics. The authors argued that the topic detection did not improve the political alignment prediction performance.

Another graphical approach for identifying opinion orientation was proposed in [Rabelo et al., 2012] where social network users were represented as nodes and the relationships between users as links. A classification approach was used to predict the unknown opinion (political polarity) of users. The collective classification approach mined the link structure information (connections between users) in a social network using a machine learning approach to infer opinions of users who have not posted their opinion about the subject under analysis. Experiments on a corpus of Twitter users to determine the political polarity of the users demonstrated promising results in predicting the polarity of users who have not posted their opinion about the subject under consideration.

In [Deitrick and Hu, 2013] work was presented on an approach integrating two popular techniques for studying online social networks to enhance each other. The two techniques were: (i) community detection (examining the structure of the social networks) and (ii) sentiment analysis (examining the content of the social networks). More specifically, using sentiment classification to enhance community detection, and community partitions to permit more in-depth analysis of sentiment data where mod-

---

[11]LSA solves the word mismatch problem by transforming the original feature space into a new space with a smaller number of features (words) with minimum loss of information. For example see [Cribbin, 2011].

ularity[12] values were increased with respect to community partitions detected in the networks. Furthermore, data collected during the community detection process enabled more granular, community-level, sentiment analysis on a specific topic referenced by users. The dataset used for experimental purposes contained the friend and follower networks of 62,346 Twitter users and 2,061,789 of their tweets, collected over a period of 32 days extracted from the publicly available Sanders Sentiment Corpus and four Microsoft-related social networks downloaded directly from the Twitter API. Community detection was performed on the friend/follower networks and then enhanced with three types of additional features: (i) replies, mentions, and retweets; (ii) hashtags and (iii) sentiment classifications. The sentiment classifications were derived using two Naïve Bayes classifiers. The social relationships between users represented as links in a social network representation were also utilised by [Younus et al., 2011] who proposed a method for sentiment analysis of tweets corresponding to the major political events during the Tunisian uprising in 2010. The proposed method considered the social features of large-scale opinions (over large corpus) expressed on social media platforms with the objective of increasing the accuracy and subjectivity of classification (using Naïve Bayes classifier) based on tweeting habits of users. [Tan et al., 2011] argued that information about social relationships (link information from a social network) can be used to improve user-level sentiment analysis. More specifically, users' opinions might be predicted more effectively by combining both users' utterances and connectedness, assuming that related users tend to share similar opinions. The authors' experiments using a Twitter dataset demonstrated that the proposed classification approach, using graphical models combined with information concerning both users' connectedness (link features) and users' utterances (text features), produced a better sentiment classification (concerning users' opinions, positive or negative, with respect to a given political topic) than when classifying on text features only.

[Bermingham et al., 2009] conducted sentiment, lexical and social network analysis on online social network content in the context of jihadi radicalisation, focusing on dedicated jihadist websites and forums. The authors collected a large dataset from a YouTube group that had been identified as potentially having a radicalising agenda. The data was analysed by conducting topic identification and then sentiment analysis was used to predict the sentiment polarity (positive or negative) towards the identified topics, focusing on the differences between male and female members of the group in terms of the nature of the discussion and interactions between them. A lexicon-based polarity scoring method was used to assign positivity and negativity scores to YouTube profiles and comments. Another example of combining sentiment and network analysis techniques can be found in [Somasundaran et al., 2009] who constructed discourse-level opinion graphs to which network analysis was applied so as to support lexicon-

---

[12]Newman's modularity metric measures how "modular" a network is.

based sentiment analysis methods. Both lexical information and link information were combined in order to improve the polarity classification process.

2.6 Related work on sentiment analysis in the political domain

In sentiment analysis most published research (see for example [Aleebrahim et al., 2011, Asmi and Ishaya, 2012, Denecke, 2008, 2009a, Martineau and Finin, 2009, Montejo-Raez et al., 2012, Ohana and Tierney, 2009, Paltoglou and Thelwall, 2012, Prabowo and Thelwall, 2009, Wilson et al., 2005]) is focused on what might be referred to as "traditional" types of subjective textual data found in blogs, social networks or specialised websites, for example reviews of movies, news articles, commercial products or services. The literature with respect to these traditional approaches is extensive, thus this section will be limited to focussing on approaches that are directly related to work on political sentiment analysis (the topic of interest with respect to the work described in this thesis).

In [Grijzenhout et al., 2010] two sentiment mining techniques were considered, based on two different models to automatically identify the subjectivity and orientation of text segments[13], to retrieve political attitudes or viewpoints from Dutch parliamentary publications. The outcomes were then compared with a manually compiled and annotated "gold standard". The first of the two techniques used machine learning classifiers (Naïve Bayes, Support Vector Machine SMO, BK1 nearest neighbour and ZeroR), while the second was a dictionary-based (lexicon-based) technique that used a subjectivity lexicon. Despite the fact that the machine learning approach outperformed the lexicon-based approach the results indicated that both sentiment mining techniques were applicable for investigating subjectivity and sentiment polarity in Dutch political semi-structured transcripts. [Rissland, 1999] manually surveyed and discussed different types of arguments made in the short (nearly one-minute) speeches given during the last hour of the debate (hearing) held within the United States House of Representatives in December 1998 on the articles of impeachment of President Clinton. The author demonstrated that these short speeches featured a reduced structure compared to the longer speeches more usually made in the House of Representatives. In [Thomas et al., 2006] work was presented on determining, using the transcripts of U.S. Congressional floor debates, the degree of agreement between opinions expressed by speakers' speeches supporting or opposing proposed legislation. By utilising information about the inter-document relationships between speeches (in particular, whether

---

[13]A "segment" in this context is a paragraph. The authors concluded that the paragraph level was most suitable to their task as it contains enough context information.

two speeches belonged to the same speaker, or whether they shared similar "content") it was demonstrated that this improved a "support" versus "oppose" classification over the classification of speeches in isolation. In a slightly different manner [Hirst et al., 2010] generated a classifier to distinguishing support from opposition without first detecting agreement or disagreement between individual speakers. [Hirst et al., 2010] argued that machine learning classifiers, that use words as features when classifying political texts (legislative speeches from both the English and French Canadian House of Commons Debates) according to ideology, are sensitive not to the expressions of ideology but rather to the expressions (language) of attack and defence of opposition and government. More specifically, the language of attack and defence, of government and opposition, seems to dominate and confound any sensitivity to ideology. One of the characteristics of Canadian politics is that Canadian parties have strong party discipline and agreement between speakers may be reliably predicted from their shared party affiliation. In contrast, "the weak party discipline of the U.S. and the separation of the Congress from the Executive branch motivates greater attention to ideological substance in debates than does the Canadian (Westminster-style) system in which an explicit governing party, including the head of government and all cabinet ministers, is represented as such in the legislature" [Hirst et al., 2010]. [Laver et al., 2003] estimated the policy positions (attitudes) of key political actors (political parties) on given societal issues from political texts (the ideological content of party manifestos and parliamentary speeches) using language-blind word scoring technique.

Most of the work related to political sentiment mining is directed at political text (from newspapers, forums, tweets, blogs) rather than parliamentary debates or political discussions as considered with respect to the work described in this thesis. In this work different sentiment analysis techniques were adopted, but mostly focussed on machine learning and lexicon-based approaches. The following is a summarisation of the most significant work conducted with respect to the sentiment analysis of political domain free text derived from sources other than structured parliamentary debates.

In [Delmonte et al., 2013] the authors extracted 750 articles (500,000 words) from three Italian newspapers and processed them so as to investigate the political stance of each newspaper with respect to a "deep political crisis situation". Two sentiment analysis approaches were considered: (i) a lexicon-based approach using off-the-shelf dictionaries with the addition of creating a lexicon, out of frequency lists, containing domain related concepts and associated keywords and (ii) a feature-based approach for pragmatic analysis (in linguistics and semiotics pragmatics is the study of the ways in which context contributes to meaning) which computes the comparative differences between the three newspapers in the use of three linguistic variables for each time period. [Stylios et al., 2010] introduced a method for extracting citizen opinions about

governmental decisions by using a machine learning approach (Support Vector Machine, K-Nearest Neighbor and Naïve Bayes classifiers) for classifying opinion phrases (652 distinct opinion phrases extracted from 124 comments, downloaded from a Greek forum focusing on policy issues, representing a set of real citizen opinions) in terms of their sentiment orientation. A machine learning approach was also used by [Durant and Smith, 2007] to identify the political sentiment of web log posts. In [Durant and Smith, 2007] the authors have investigated the utility of Naïve Bayes and Support Vector Machines (with and without feature selection) on a novel collection of datasets created from political web log posts to predict the political leanings of individual posting on a particular topic such as the Iraq War. [Singh et al., 2010a,b] proposed a framework using clustering and sentiment mining to conduct sentiment analysis on blog posts about a recent socio-political issue. The blog posts were grouped into clusters representing viewpoints, issues and concerns about the constitutional developments in the new democratic republic of Nepal. The identified topics were similar to those found in commonly available political literature and socio-political writings in newspapers and magazines. SentiWordNet was used to compute positivity, negativity and objectivity values of different words. It was suggested that these values might represent the expectations and feelings of people about the topic of analysis.

Other techniques have also been combined and applied, in addition to sentiment mining techniques, for example the morpho-syntactical analysis used by [Neri et al., 2010] where 1000 news articles or posts in forum/blogs, concerning the scandal surrounding Italian Prime Minister Silvio Berlusconi's sexual activity, have been morpho-syntactically analysed, labelled according to their semantic role and clustered in order to: (i) find meaningful similarities, (ii) highlight possible hidden relationships and (iii) evaluate their sentiment polarity. Other reported work has been directed at the development of a sentiment knowledge-base such as OpinioNetIt which was created by [Awadallah et al., 2012]. OpinioNetIt is a structured knowledge-base of opinions which has been used for political sentiment analysis applications such as the generation of "heat maps" showing political bias, identifying "flip-flopping" politicians, and identifying dissenters. The proposed sentiment knowledge-base contains information about people, topics and opinions as a sets of triples. Each triple $\langle O, P, T \rangle$ represents the opinion O of a person P on a specific topic T. The knowledge-base generation process comprises four steps: (i) identifying sources, (ii) input of seed phrases to acquire opinion "snippets", (iii) identifying opinion triples from opinion snippets and (iv) structuring the topic hierarchy. The difference between OpinioNetIt and SentiWordNet, or other sentiment lexicons, is that OpinioNetIt contains a limited amount of fixed sentiment information concerning specific persons with respect to some topic, and thus operates at a person-topic level. In contrast, sentiment lexicons operates at a lower level (the word level).

2.7 Summary

This chapter has presented a literature review and essential background for the research work described in this thesis focusing on: sentiment mining approaches, text visualisation and social network analysis. The review considered both the classification-based approach to sentiment analysis as well as the lexicon-based approaches (generic and domain specific). This chapter also reviewed some related work on sentiment analysis in the political domain and has described some related work on visualising debate structures and analysing such structures. The next chapter introduces the datasets used for evaluation purposes with respect to the work presented in the remainder of this thesis.

Chapter 3

The UK House of Commons Political Debates Corpus

"There is no such thing as public opinion. There is only published opinion."

Winston Churchill

This chapter presents an overview of the UK House of Commons parliamentary debate datasets which were used to act as a focus for the work described in this thesis in the context of the evaluation of the techniques proposed. More specifically the datasets were used for:

- Evaluation purposes with respect to attitude classification in the context of political sentiment mining using machine learning classification as described in Chapter 4.

- Evaluation purposes with respect to attitude prediction in the context of political sentiment mining using generic sentiment lexicons as described in Chapter 5.

- Evaluation purposes with respect to attitude prediction in the context of political sentiment mining using the proposed domain specific sentiment lexicons as described in Chapter 6.

- The comparison and contrast purposes with respect to the three approaches to political sentiment mining considered in Chapters 4, 5 and 6 by considering debater attitude prediction effectiveness as presented in Chapter 7.

- The extraction of debate graphs embedded within debate transcriptions, using the proposed Debate Graph Extraction (DGE) framework, so as to support the graphical visualisation of the high-level structure of political debates as described in Chapter 8.

- Evaluation purposes with respect to the determination of sentiment scores and polarities within a pre-labelled corpus, using the proposed ΔTF-IDF$'$ weighting scheme, as described in Chapter 6.

- The generation of the two proposed domain specific (political) sentiment lexicons, PoLex and PoliSentiWordNet, generated by applying the two techniques described in Chapter 6 to political benchmark data.

This chapter commences with an overview of the UK Parliamentry system in Section 3.1 so that the reader can more precisely understand the problem domain. Section 3.2 then provides a brief description of the political party system in operation in the UK, while Section 3.3 describes the nature and characteristics of UK parliamentary debates. Section 3.4 reviews the UK House of Commons datasets used with respect to the research described in this thesis. Collectively these datasets were referred to as the UK House of Commons Debate (UKHCD) datasets. In total four UKHCD datasets were used: UKHCD-1, UKHCD-2, UKHCD-3 and UKHCD-4. Finally, this chapter is completed with a brief summary in Section 3.5.

3.1 The UK Parliamentry System

The UK Parliament[1] consists of two Houses: (i) the House of Commons ("The Commons") and (ii) the House of Lords ("The Lords"). Both houses have similar responsibilities which comprise: (i) making laws (legislation), (ii) checking the work of Government (scrutiny) and (iii) debating running important political issues. The House of Lords is independent from the elected House of Commons and (generally) any decision made by one House needs to be approved by the other House in a two-chamber system. More precisely:

The Commons: The House of Commons has the most authority and comprises an elected membership. The UK public elects (currently) some 650 Members of Parliament (MPs) to represent them in the House of Commons. Almost all MPs are members of political parties and the party with the largest number of MPs in the Commons forms the Government (although at time of writing no party had

---

[1]This chapter contains information taken from the UK Parliament official website available on-line at (http://www.parliament.uk/).

sufficient votes to gain overall control, hence the Government comprised a coalition). MPs debate running political issues, propose new laws and make decisions on financial Bills and proposed new taxes. Debates in the House of Commons are chaired by what is known as the "Speaker" or one of his/her deputies. These debates are recorded and publicly available. It is these House of Commons debates that are of significance with respect to the work described in this thesis and hence they are discussed in more detail later in this Chapter in Section 3.4.

The Lords: The House of Lords is the second chamber of the UK Parliament and has less authority than The Commons. Members of the House of Lords are appointed by the Queen on the advice of the Prime Minister or recommended by the House of Lords Appointments Commission. The main roles of the House of Lords are to consider proposed legislation and to provide for a check on Government activities. The Lords can propose amendments to legislation, but ultimately cannot block legislation proposed by the House of Commons. The House of Lords also conducts its business through a process of debate and these are also recorded. However these debates are not considered in the context of this thesis, which concentrates on Commons debates although there is no reason why the proposed techniques cannot be equally applied to House of Lords debates.

Thus, so as to provide a focus for the research work described in this thesis UK House of Commons debates were used. To explain how UK House of Commons parliamentary debates are conducted the role of the House of Commons should first be considered in more detail:

1. Monitoring the work of the Government (scrutiny): Parliament monitors the work of the Government by: (i) directing questions to Government ministers, (ii) debating and (iii) constituting investigative committees. In more detail:

   - Questions: Minsters may attend Parliament to answer questions orally, or in writing. The Prime Minister answers questions every Wednesday.
   - Debates: Debates in the Commons are conducted so as to make or amend laws.
   - Committees: Committees consist of smaller groups of MPs. Their role is to look at specific policy issues and/or legislation in detail. Different committees have different roles ranging from offering advice to producing reports suggesting modifications to legislation.

2. Debating and passing of laws (legislation): Parliament is responsible for debating and voting on new laws or modifications to existing laws that are proposed by

the Government or in some cases an MP, a member of the House of Lords or a member of the public. Making Laws is conducted as follows:

- Introducing Bills: A Bill is a proposal to introduce a new law or modifications to an existing law, which is debated by Parliament.
- Approval: To become Law the text of a Bill must be agreed by both Houses[2].
- Royal Assent: The reigning monarch has to approve all new Laws. The Royal Assent is a formality and is not withheld. Once a Bill is approved by Royal Assent it becomes a Law and the Government is responsible for implementing that Law.

3. Enabling the Government to raise taxes.

3.2   Political parties

There are three main political parties in the UK: Labour, Conservative and Liberal Democrat. Most MPs are members of one or other of these three parties. MPs that do not represent a political party are known as "Independents". Typically the party having the majority of MPs forms the Government and the next largest party represents the official Opposition whose role is to:

1. Support and participate in the creation of policy and laws through a process of "constructive criticism".

2. Opposing Government proposals they disagree with.

3. Propose their own policies and laws.

At time of writing (January 2014) the elected Speaker was John Bercow and the Deputy Speakers were Lindsay Hoyle, Dawn Primarolo and Eleanor Laing. Table 3.1 details the composition of the House of Commons in January 2014, based on the number of MPs in each party. It should be noted that in January 2014 the Government comprised a coalition between the Conservative and Liberal Democrat parties. Table 3.2 details the seats won by each party at the 2010 General Election.

---

[2]Although if the Commons passes a Bill in two successive years then that Bill can become a Law without the agreement of the Lords.

Party	Seats
Conservative	303
Labour	256
Liberal Democrat	56
Democratic Unionist	8
Scottish National	6
Independent	5
Sinn Fein	5
Plaid Cymru	3
Social Democratic & Labour Party	3
Alliance	1
Green	1
Respect	1
Speaker	1
Vacant	1
Total number of seats	**650**
Current working Government Majority	**76**

Table 3.1: The composition of the UK House of Commons in January 2014.

Party	Seats
Conservative	306
Labour	258
Liberal Democrat	57
Democratic Unionist	8
Scottish National	6
Sinn Fein	5
Plaid Cymru	3
Social Democratic & Labour Party	3
Alliance	1
Green	1
Independent	1
Speaker	1
Total number of seats	**650**

Table 3.2: The seats won at the 2010 General Election.

3.3 Parliamentary debates

In the UK governmental decisions are arrived at through a process of political debate culminating in a vote. The debates are conducted in The House of Commons (the "lower house") and The House of Lords (the "upper house"). Members of both Houses debate in a dynamic style of discussion according to strict sets of rules. There are many rules and customs that affect how Parliament runs. In the House of Commons MPs refer to each other by their constituency name or official title, not by their actual names. MPs call each other either "`the honourable Member for ...`", or if the MP is a member of the Privy Council "`the right honourable Member for ...`". During House of Commons debates, debaters (MPs) can interrupt each other's speech to support or oppose what they are saying about the introduced subject. Additionally, debaters are entitled to be able to respond to points made by others without overwhelming background noise. Speeches must be made in English, but quotation in another language has been allowed. Unparliamentary language is not allowed. In the context of sentiment analysis, as will be demonstrated later in this thesis, the analysis of Commons debates is hindered by the nature of the "overly polite" parliamentary language used by debaters. For example "`I thank my right honourable friend for his intervention.`", "`As I hope the honourable gentleman is aware, ...`" and "`The honourable Lady makes a very important point.`" . The overly polite nature of the political debates, and the eccentric jargon sometimes used, means that the prediction of negative sentiments is a challenging process, as will become apparent later in this thesis in Chapter 7. Debates are directed at the establishment of new legislation, or modifications to existing legislation, according to Government policy. Debates end with a vote (a "division") where MPs vote to support (by voting "Aye") or oppose (by voting "No") the motion. The motions, usually beginning with the word "That", must be drafted in the affirmative, not in the style of a speech and without any objectionable, argumentative or irregular wording. Note that, in a debate the motion is first put forth by someone defending it and the debate develops then by responses (opposing the motion). So lexical methods make sense in as much we can assume that a debate is typically initiated by speakers with positive attitudes towards the motion (however it is phrased). Other than the passing (or rejecting) of legislation debates also provide MPs with an opportunity to reflect their constituents' concerns and interests. All debates are published in a verbatim record (called "Hansard") which is available both online and in print. Hansard also includes the outcome of votes, written ministerial statements and written answers to parliamentary questions.

3.4 The UK House of Commons political debates datasets

From the above MPs in the House of Commons reach their decision by debating and then voting with either an "Aye" or a "No" vote at the end of each debate. Proceedings of the Commons Chamber are published on-line in XML format three hours after they take place (at `TheyWorkForYou.com`). Figure 3.1 shows an extract from a debate transcript taken from the "Enterprise and Regulatory Reform Bill, Clause 4 - The UK Green Investment Bank: financial assistance, Tuesday 26 June 2012"[3] debate. The highlighted text indicates MPs who voted "Aye" at the end of the debate while the unhighlighted text indicates MPs who voted "No". Figure 3.2 shows the vote (division) results at the end of the debate while Figure 3.3 shows the XML mark-up for the same fragment of text in Figure 3.1. The advantage offered by this collection is that the outcome of the debates, i.e. the results of the divisions, are known and thus this collection can be used to evaluate the veracity of the outcomes of the application of sentiment analysis techniques such as those considered in this thesis.

QDAMiner4[4] was used to extract the desired textual information from the XML debate records. In this manner, the speeches associated with different UK House of Commons debates were obtained and the desired textual information extracted. For each debate the speeches associated with the same MP were concatenated together. Concatenated speeches by MPs who did not vote were ignored; as were speeches that contained fifty words or less, as it was conjectured that no valuable sentiment attitude could be associated with such short speeches. Four parliamentary debate datasets were constructed, collectively referred to as the UK House of Commons Debate (UKHCD) datasets, and labelled: UKHCD-1, UKHCD-2, UKHCD-3 and UKHCD-4. The reason for the existence of four different datasets is mostly historical. The UKHCD-1 and UKHCD-2 datasets were used early on in the programme of work. More specifically they were used in the evaluations published by the author in [Salah et al., 2013a] and [Salah et al., 2013b] to determine the effectiveness of the Debate Graph Extraction (DGE) framework described in Chapter 8 later in this thesis. The UKHCD-3 dataset was used in the context of the work on the generation of domain specific lexicons which is presented in Chapter 5 of this thesis. The UKHCD-4 dataset was used with respect to work using machine learning classification, generic sentiment lexicons and domain specific sentiment lexicons as described in Chapters 4, 5, and 6 respectively. The distinction between UKHCD-3 and UKHCD-4, which were both used with respect to the work presented in Chapter 5 is that UKHCD-3 was used, as a training dataset, to generate the two proposed domain-specific lexicons; while UKHCD-4 was used, as

---

[3]Hansard source citation: from Enterprise and Regulatory Reform Bill Deb, 26 June 2012, c230 to Enterprise and Regulatory Reform Bill Deb, 26 June 2012, c231.

[4]http://provalisresearch.com

Mark Prisk (Minister of State (Business and Enterprise), Business, Innovation and Skills; Hertford and Stortford, Conservative)

The only chance is that his singing might have been more harmonious than the economic analysis we were given. I did not notice at any point a mention of the enormous - indeed record- debt that we inherited. To be lectured by a party that left the worst Government debt in my lifetime on the prospects of one month-

Iain Wright (Hartlepool, Labour)

That is a long time.

Mark Prisk (Minister of State (Business and Enterprise), Business, Innovation and Skills; Hertford and Stortford, Conservative)

50 years is a long time. When I listened to that, I thought, It is all very well to say that we should be borrowing more and doing this, but it is a shame. It is a particular shame because there is an important issue here that people outside this room are concerned about: how the financial powers will work. It is a shame that there was a pitiful attempt to pretend that there were no borrowing issues, and that tomorrow we could simply borrow because it the money was available. It is a real shame, because there is an important issue at the heart of this.

John Cryer (Leyton and Wanstead, Labour)

Is the **Minister** aware that at the time of the last election, both the deficit and unemployment were falling? They are now both rising. The **Office for Budget Responsibility**, the body set up by the Government, predicts that the deficit will be £180 billion larger at the end of this Parliament than was predicted at the time of the last election.

Mark Prisk (Minister of State (Business and Enterprise), Business, Innovation and Skills; Hertford and Stortford, Conservative)

With respect to the hon. Gentleman, the other thing that we did not hear from the Labour party was mention of the eurozone. According to Labour Members, the only reason businesses are lacking in confidence is entirely to do with the UK's economic policies: there is nothing going on across the channel, it is all calm, they are enjoying their summer holidays and everything is entirely relaxed. When I deal with businesses on a weekly basis, seeking to encourage them to invest in green projects and elsewhere, they constantly refer to the international financial climate, particularly the eurozone, as the reason for hesitating over investing. I had hoped we would have a balanced debate on this issue, but let us address the **amendment** before us, because that is what matters.

On that basis, it will not come as a surprise to the hon. Gentleman that I intend to resist this **amendment** for two main reasons. First, the Government's approach to the bank's future borrowing is the right one. Secondly, legislation is not the right mechanism to govern the bank's borrowing. There are important issues which those wanting to look at the commitment of financial support for this institution are looking to hear about. Before I address these arguments in turn, let me restate that the coalition Government are committed to the **UK Green Investment Bank** growing into a successful, enduring green financial institution.

Figure 3.1: Fragment of a UK House of Commons debate as published on the They-WorkForYou.com www site.

Figure 3.2: The vote (division) results for the debate containing the fragment shown in Figure 3.1.

"unseen" test dataset, to evaluate the effectiveness of the lexicons in terms of attitude prediction. Some statistics concerning the four datasets are presented in Table 3.3. From the table it can be seen that the UKHCD-3 dataset was the largest. The most recent dataset, UKHCD-4, was used with respect to the majority of the work presented in this thesis.

	Number of Debates	Number of *Aye* speeches	Number of *No* speeches	Total Number of speeches
UKHCD-1	21	911	827	1,738
UKHCD-2	100	4,581	4,892	9,473
UKHCD-3	1101	147,559	180,566	328,125
UKHCD-4	29	1,119	949	2,068
Totals	1251	154,170	187,234	341,404

Table 3.3: UKHCD datasets Statistics.

The UKHCD-4 dataset thus merits some further discussion here. From Table 3.3 the UKHCD-4 dataset comprised 2,068 concatenated speeches (1,119 speeches made by speakers who voted Aye and 949 speeches made by speakers who voted No) associated with 29 different UK House of Commons debates held between August 2012 and March 2013 and 553 distinct Members of Parliament (MPs) belonging to 10 distinct political parties. Figure 3.4 shows the number of concatenated speeches made within each of the 29 debates. The blue part of each column[5] represents the portion of speeches made by speakers who voted Aye at the end of the debate, while the red part represents the portion of speeches made by speakers who voted No. Note that the number of concatenated speeches featured in a debate equates to the number of MPs taking part (MPs that made at least 50 words length speech during the debate and eventually voted

---

[5]A column represents a debate

```
<speech id="uk.org.publicwhip/standing/standing2012-06-19_ENTERPRISE_05-0_2012-06-26a.5.44"
speakerid="uk.org.publicwhip/member/40313" speakername="Mark Prisk" time="17:30"
url="http://www.publications.parliament.uk/pa/cm201213/cmpublic/enterprise/120626/pm/120626s01.h
tm#12062715000226" colnum="230">
<p>The only chance is that his singing might have been more harmonious than the economic analysis
we were given. I did not notice at any point a mention of the enormous—indeed record—debt that we
inherited. To be lectured by a party that left the worst Government debt in my lifetime on the
prospects of one month—</p>
</speech>
<speech id="uk.org.publicwhip/standing/standing2012-06-19_ENTERPRISE_05-0_2012-06-26a.5.45"
speakerid="uk.org.publicwhip/member/40302" speakername="Iain Wright" time="17:30"
url="http://www.publications.parliament.uk/pa/cm201213/cmpublic/enterprise/120626/pm/120626s01.h
tm#12062715000227" colnum="230">
<p>That is a long time.</p>
</speech>
<speech id="uk.org.publicwhip/standing/standing2012-06-19_ENTERPRISE_05-0_2012-06-26a.5.46"
speakerid="uk.org.publicwhip/member/40313" speakername="Mark Prisk" time="17:30"
url="http://www.publications.parliament.uk/pa/cm201213/cmpublic/enterprise/120626/pm/120626s01.h
tm#12062715000228" colnum="230">
<p>50 years is a long time. When I listened to that, I thought, "It is all very well to say that we should
be borrowing more and doing this, but it is a shame." It is a particular shame because there is an
important issue here that people outside this room are concerned about: how the financial powers will
work. It is a shame that there was a pitiful attempt to pretend that there were no borrowing issues, and
that tomorrow we could simply borrow because it the money was available. It is a real shame, because
there is an important issue at the heart of this.</p>
</speech>
<speech id="uk.org.publicwhip/standing/standing2012-06-19_ENTERPRISE_05-0_2012-06-26a.5.47"
speakerid="uk.org.publicwhip/member/40366" speakername="John Cryer" time="17:30"
url="http://www.publications.parliament.uk/pa/cm201213/cmpublic/enterprise/120626/pm/120626s01.h
tm#12062715000229" colnum="231">
<p>Is the Minister aware that at the time of the last election, both the deficit and unemployment were
falling? They are now both rising. The Office for Budget Responsibility, the body set up by the
Government, predicts that the deficit will be £180 billion larger at the end of this Parliament than was
predicted at the time of the last election.</p>
</speech>
<speech id="uk.org.publicwhip/standing/standing2012-06-19_ENTERPRISE_05-0_2012-06-26a.5.48"
speakerid="uk.org.publicwhip/member/40313" speakername="Mark Prisk" time="17:30"
url="http://www.publications.parliament.uk/pa/cm201213/cmpublic/enterprise/120626/pm/120626s01.h
tm#12062715000230" colnum="231">
<p>With respect to the hon. Gentleman, the other thing that we did not hear from the Labour party
was mention of the eurozone. According to Labour Members, the only reason businesses are lacking
in confidence is entirely to do with the UK's economic policies: there is nothing going on across the
channel, it is all calm, they are enjoying their summer holidays and everything is entirely relaxed.
When I deal with businesses on a weekly basis, seeking to encourage them to invest in green projects
and elsewhere, they constantly refer to the international financial climate, particularly the eurozone,
as the reason for hesitating over investing. I had hoped we would have a balanced debate on this
issue, but let us address the amendment before us, because that is what matters. On that basis, it will
not come as a surprise to the hon. Gentleman that I intend to resist this amendment for two main
reasons. First, the Government's approach to the bank's future borrowing is the right one. Secondly,
legislation is not the right mechanism to govern the bank's borrowing. There are important issues
which those wanting to look at the commitment of financial support for this institution are looking to
hear about. Before I address these arguments in turn, let me restate that the coalition Government
are committed to the UK Green Investment Bank growing into a successful, enduring green financial
institution.</p>
```

Figure 3.3: The XML mark-up for the debate fragment presented in Figure 3.1.

Selected attribute
Name: Debate-ID
Missing: 0 (0%)   Distinct: 29   Type: Nominal
                                                                                      Unique: 0 (0%)

No.	Label	Count
1	2012-07-16d	106
2	2012-09-05b	88
3	2012-09-10b	51
4	2012-09-11b	81
5	2012-09-12a	73
6	2012-10-15c	107
7	2012-10-16a	82
8	2012-10-18b	23
9	2012-10-24c	79
10	2012-10-29a	46
11	2012-10-31c	83
12	2012-11-05b	51
13	2012-11-06b	31
14	2012-11-07a	65
15	2012-12-18b	21
16	2012-12-19a	94
17	2013-01-08b	106

■ The portion of speeches made by speakers who voted Aye at the end of the debate.
■ The portion of speeches made by speakers who voted Aye at the end of the debate.
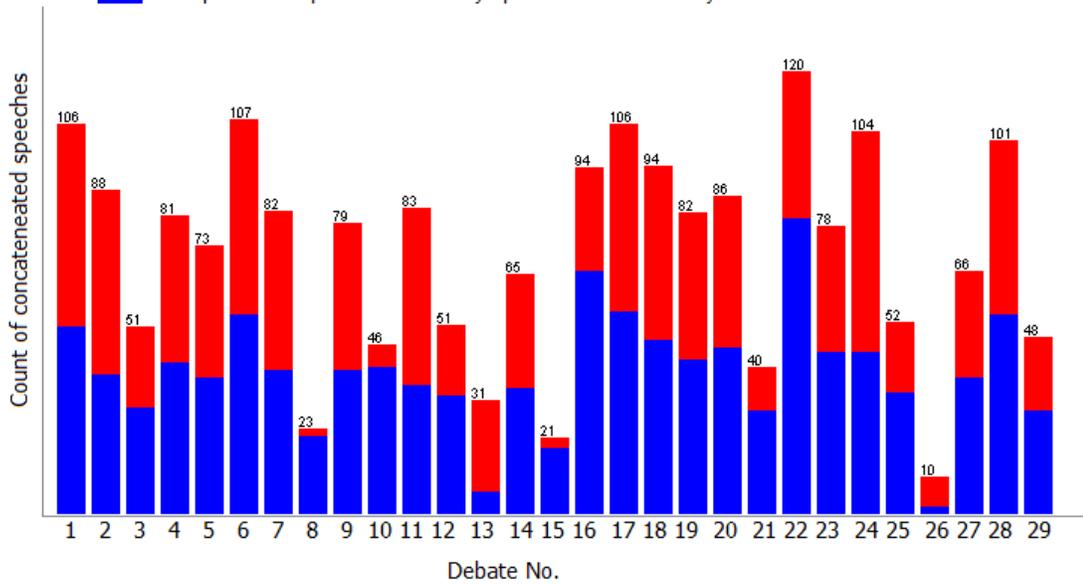
Figure 3.4: UKHCD-4 dataset: The distribution of speeches over the 29 debates.

| Name: Party | | | Type: Nominal |
| Missing: 0 (0%) | | Distinct: 10 | Unique: 1 (0%) |

No.	Label	Count
1	Lab	908
2	LDem	186
3	Con	875
4	PC	11
5	SDLP	14
6	DUP	28
7	Green	9
8	SNP	34
9	Alliance	2
10	Ind	1

■ The portion of speeches made by speakers who voted Aye at the end of the debate.
■ The portion of speeches made by speakers who voted Aye at the end of the debate.
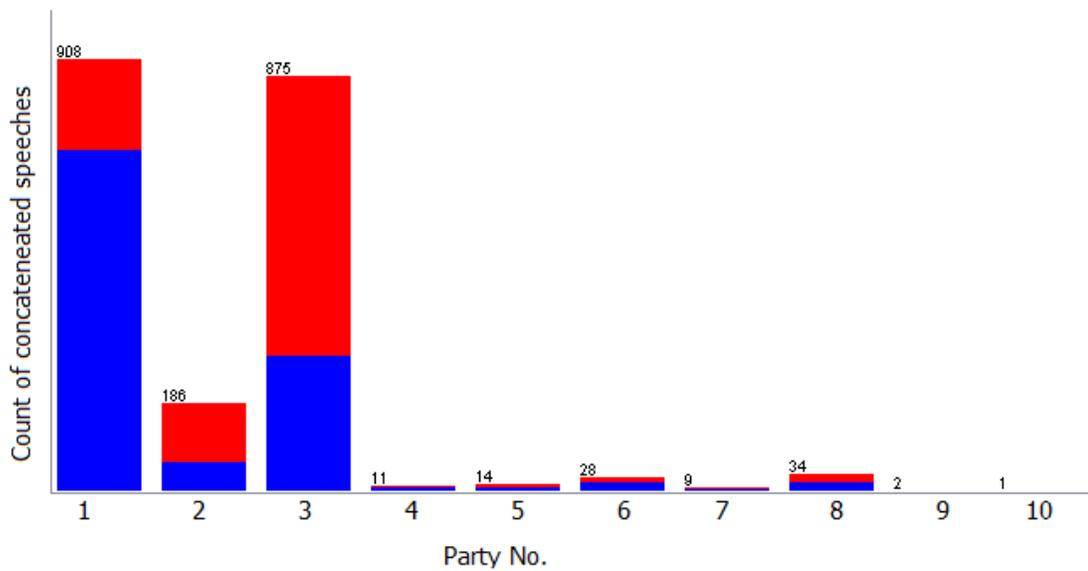
Figure 3.5: UKHCD-4 dataset: The distribution of speeches over the 10 participant parties.

at the end of the debate). The distribution of speeches over the 10 participant parties is shown in Figure 3.5 where the blue part of each column[6] represents the portion of speeches made by members of the party who voted Aye at the end of any debate, while the red part represents the portion of speeches made by members of the party who voted No. From Figure 3.5 the dominance of the three major parties (Labour, Conservative and Liberal Democrat) can be clearly seen.

More statistics concerning UKHCD-4 dataset are presented in Table 3.4. From the table, the average number of words in a concatenated speech is 975 and in a whole debate is 63,379. The average number of Aye and No speeches (39 and 33 respectively) is reasonably balanced. The speeches comprised a total of 1,837,996 words or unigrams (17,893 unique words after stemming and stop-word removal or 31,259 unique words after only stop-word removal). Figures 3.4 and 3.5 were produced using Weka[7] visualiser.

3.5   Summary

This chapter has presented an overview for the benchmark datasets that were used with respect to the research described in this thesis. A brief overview was also presented concerning the UK Parliament, the nature of the parliamentary debates and the UK political party system. This chapter has also provided an overview and some statistics concerning the extracted parliamentary speeches that formed four different versions of the dataset with which the work presented in this thesis was evaluated. The next three chapters present the three sentiment mining approaches of interest in the context of mining the UK House of Commons political debates, starting with the classification-based approach in Chapter 4.

---

[6]A column represents a political party

[7]Weka is available at: http://www.cs.waikato.ac.nz/ml/weka/

Debate	Aye Sp.	No Sp.	Total Sp.	Min length	Max length	Total words	Avg length
D1	51	55	106	50	4881	72376	682.792
D2	38	50	88	51	4974	82506	937.568
D3	29	22	51	51	4822	40766	799.333
D4	41	40	81	50	4843	65523	808.926
D5	37	36	73	51	4989	65214	893.342
D6	54	53	107	55	4804	50146	468.654
D7	39	43	82	50	4880	60474	737.488
D8	21	2	23	62	2878	22510	978.696
D9	39	40	79	50	4917	81095	1026.519
D10	40	6	46	51	4680	47207	1026.239
D11	35	48	83	54	4928	74375	896.084
D12	32	19	51	54	4951	75525	1480.882
D13	6	25	31	63	4846	39662	1279.419
D14	34	31	65	55	4896	77766	1196.400
D15	18	3	21	66	4872	34900	1661.905
D16	66	28	94	53	4804	59589	633.926
D17	55	51	106	51	4849	74703	704.745
D18	47	47	94	51	4915	72770	774.149
D19	42	40	82	50	4926	98845	1205.427
D20	45	41	86	50	5041	70457	819.267
D21	28	12	40	50	4999	37866	946.650
D22	80	40	120	51	4890	88518	737.650
D23	44	34	78	51	4817	74880	960.000
D24	44	60	104	54	4986	113701	1093.279
D25	33	19	52	50	4710	60114	1156.038
D26	2	8	10	86	4929	19107	1910.700
D27	37	29	66	51	4967	63392	960.485
D28	54	47	101	50	4961	81490	806.832
D29	28	20	48	58	4863	32519	677.479
MIN	2	2	10	50	2878	19107	468.654
MAX	80	60	120	86	5041	113701	1910.700
AVG	38.586	32.724	71.310	54.103	4821.310	63379.172	974.513
Total	1119	949	2068	1569	139818	1837996	28260.876

Table 3.4: UKHCD-4 dataset: Statistical overview.

Chapter 4

Political Sentiment Mining Using Classification

"There is little difference in people, but that little difference makes a big difference.
The little difference is attitude.
The big difference is whether it is positive or negative."

William Clement Stone

This chapter presents the first of the three sentiment mining approaches proposed in this thesis, namely the classification-based approach. The general idea was to use machine learning classifiers trained (learned) using an appropriately labelled training dataset and evaluated using test data (as previously described in Section 2.2). The generated classifiers were then used to predict the attitude of individual speakers participating in an "unseen" debate. An overview of this process is presented in Figure 4.1. In the context of the UK House of Commons debates used for evaluation purposes the input is a set of concatenated speeches that make up a single debate and the output is a set of attitude labels one per concatenated speech. More formally the input is a set of n concatenated speeches $S = \{s_1, s_2, \ldots, s_n\}$ and the output is a set of attitude class labels $C = \{c_1, c_2, \ldots, c_n\}$ taken from the set $\{positive, negative\}$ such that there is a one-to-one correspondence between the elements in S and C. The process encompasses two phases: (i) preprocessing (Phase 1) and (ii) attitude prediction (Phase 2). Each of these phases is described in more detail in the following two Sections 4.1 and 4.2. Some experimental results are recorded and discussed in Section 4.3, while the chapter is concluded with a summary presented in Section 4.4.
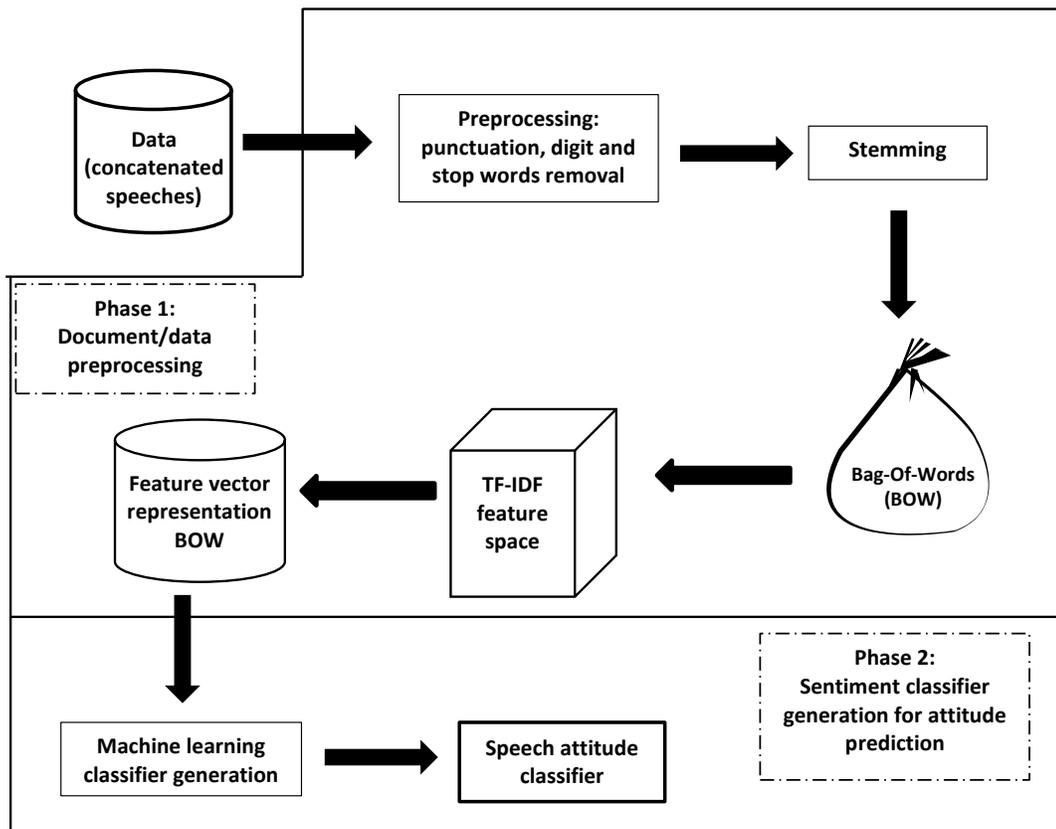
Figure 4.1: The classification-based approach to sentiment mining (classifier generation).

4.1 Preprocessing

This section presents an overview and discussion of the nature of the pre-processing required for effective sentiment classification. The preprocessing is described in terms of the House of Commons debate transcripts, however similar preprocessing would be required with respect to alternative forms of debate transcript. In the context of the required preprocessing each concatenated speech, associated with a speaker (MP), is conceptualised as a document. First all upper-case alphabetic characters were converted to lower-case letters followed by numeric digit removal. The latter was conducted on the grounds that no sentiment (subjective bias) information could be extracted from numeric data when considered in isolation. This was followed by a tokenisation process where by the content of each "document" was broken up into sets of primitive components called *tokens* (individual words). These tokens were identified using white space characters and/or punctuations (. , ; : ' " ( ) ? !). The resulting tokens were then indexed to form an initial Bag-Of-Words ($BOW = \{t_1, t_2, \ldots, t_{|BOW|}\}$). The next step was to reduce the size of the BOW by removing "stop words". Stop words in this respect were identified as words which are not expected to convey any significant meaning in the context of sentiment analysis, for example words such as "the", "a", "and", "is" and so on) [Chim and Deng, 2008, Hariharan and Srinivasan, 2008, Poomagal and Hamsapriya, 2011]. After the completion of stop word removal, each document was represented by some subset of the BOW. Given a specific domain there will also be additional words, other than stop words, that occur frequently. In the case of the House of Commons parliamentary debates words like: "hon.", "house", "minister", "government", "gentleman", "friend" and "member" are all very frequently occurring words. For similar reasons as for stop word removal these domain specific words were also removed. This was done by appending them to the stop-words list. The names of all the members of parliament, political parties and constituencies were also added to the words in default Weka's stop-word list[1]. Figure 4.2 shows a word-cloud (comprising the top 150 most frequent words) created using Wordle[2] for the UKHCD-4 dataset (one of the datasets used in the work described in this thesis) after initial stop words removal. From the Figure, it is plain to see that most of the very frequent words in the dataset were domain specific (parliamentary) words.

The size of the produced BOW was then further reduced by applying stemming. Stemming is concerned with the process of deriving the "stem" of a given word by removing the added affixes so that "inflated" words that belong to the same stem (root) will be "counted together" [Hariharan and Srinivasan, 2008]. For example "compute", "computes", "computer", "computed", "computation" and "computing" will all be

---

[1]Appendix A presents this bespoke stop-words list.
[2]available on-line at: http://www.wordle.net/

Figure 4.2: Word-cloud visualisation (comprising the top 150 most frequently occurring words after initial stop words removal) for the UKHCD-4 dataset. Generated using Wordle.

reduced to the common stem "compute"and thus less computations are required. Many mechanisms have been proposed to perform stemming, in the context of the work described in this thesis Porters Snowball Stemmer [Porter, 1997] was used.

On completion of the preprocessing and stemming stages the resulting BOW defines a feature space from which sets of feature vectors can be generated. The feature vector elements hold term weightings. The most widely used mechanism for generating term weightings, and that adopted with respect to the work described in this chapter, is the Term Frequency-Inverse Document Frequency (TF-IDF) weighting scheme which aims to "balance out the effect of very rare and very frequent" terms in a vocabulary [Kuhn et al., 2007]. TF-IDF also tends to reflect the significance of each term by combining local and global term frequency [Li et al., 2009]. TF-IDF is typically defined as follows:

$$w_{ij} = TFIDF(i,j) = tf(i,j) . \left(log \frac{N}{df(j)} \right) \tag{4.1}$$

where: (i) $tf(i,j)$ is the frequency of term j in document d_i (local weight for the term), (ii) N is the total number of documents in the corpus (concatenated speeches in the case of the House of Commons debates), and (iii) $df(j)$ is the number of documents (speeches) containing term j (global weight for the term).

Table 4.1 shows the document frequency counts for a number of example terms taken from UK House of Commons parliamentary debate collection (UKHCD-1). The table also shows the document count with respect to documents (speeches) where the MP in question voted "Aye" and where the MP voted "No". The final column gives the document frequency difference between the number of "Aye" and "No" counts. Inspection of this final column clearly indicates that some terms can be associated with an "Aye" vote, while other terms can be associated with a "No" vote. For example *cuts* is associated with an "Aye" vote while *european* is associated with a "No" vote (during the period when our political speeches were collected, August 2012 to March 2013, a right of centre political party was in government in the UK who had a tendency to favour tax *cuts* and oppose *european* integration).

Term	DF (Aye)	DF (No)	DF (Total)	Difference	Term	DF (Aye)	DF (No)	DF (Total)	Difference
people	406	338	744	68	timetable	23	23	46	0
cuts	87	38	125	49	taxpayer	11	29	40	-18
change	154	111	265	43	generous	10	28	38	-18
worse	52	17	69	35	fully	34	53	87	-19
simply	101	70	171	31	sustainable	11	33	44	-22
care	69	39	108	30	funding	41	64	105	-23
confidence	60	31	91	29	improve	40	63	103	-23
recession	42	13	55	29	assure	34	59	93	-25
women	64	36	100	28	inherited	9	38	47	-29
military	42	16	58	26	previous	101	131	232	-30
hope	136	120	256	16	raises	8	38	46	-30
existence	15	0	15	15	reduce	38	73	111	-35
wonderful	24	10	34	14	encourage	30	72	102	-42
deep	21	7	28	14	european	59	105	164	-46

Table 4.1: Document Frequency (DF) values (indicating biased occurrences) associated with selected terms occurring in UKHCD-1 with respect to MPs who voted "Aye" and "No".

Thus after the completion of the preprocessing phase the input collection of concatenated speeches were represented using a vector space model such that each speech was described by a feature vector. More formally a speech i is represented as a vector $V_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$ where w_{ij} is the TF-IDF value for term j in speech i. It should also be noted that each element in V_i corresponds to a term in the BOW. The list of terms associated with feature vector V_i were indicated using the notation $T_i = \{t_{i1}, t_{i2}, \ldots, t_{im}\}$. Thus a set of feature vectors $V = \{V_1, V_2, \ldots, V_z\}$ and a set of term lists $T = \{T_1, T_2, \ldots, T_z\}$, with a one-to-one correspondence between the two sets, were produced.

4.2   Classifier Generation

Once the input data was translated into the feature vector format, whereby the concatenated speeches for each speaker were defined by a subset of words contained in the BOW, classification could be applied to determine each speaker's "attitude" (positive or negative). To this end, a classifier was required. Classifier generation is a supervised machine learning mechanism (see Section 2.2) which, as noted previously, requires pre-labelled training data (something which we would only have with respect to historical data). In the work described in this thesis, the known vote associated with each speaker in the dataset was used as the label. Any number of different classifier generation techniques could have been adopted, however in the context of the comparison presented later in this chapter, a number of classifier generator techniques available within the Weka-3.6 workbench[3] were considered: (i) Naïve Bayes, (ii) Support Vector Machine SMO, (iii) J48 decision tree learner, (iv) JRip rule-based classifier, (v) IBk nearest neighbour classifier and (vi) ZeroR. A brief description of these machine learning classifiers was presented in Section 2.2.

Once an appropriate classifier has been generated it can be evaluated by applying it to pre-labelled test data and the labels produced compared with the known labels. Note that the training data used to generate a classifier and the data used to evaluate it need to be preprocessed in the same manner. Provided that the generated classifier was found to be sufficiently effective it can then be used with respect to unseen data. Algorithm 4.2.1 shows the process of attitude identification using a trained classifier. Note that the algorithm is designed to operate with respect to a collection of speeches to be labelled.

Algorithm 4.2.1 Attitude identification using trained classifier
---
1: INPUT: Set of Vectors $V = \{v_1, v_2, \ldots, v_z\}$, A classifier
2: OUTPUT: Set of Attitudes $C = \{c_1, c_2, \ldots, c_z\}$, *where $c_i \epsilon \{positive, negative\}$*
3: $C = \{\}$
4: **for all** $v_i \in V$ **do**
5:     $c_i = Classify(v_i)$ *into the fittest class*
6:     $C = C \bigcup c_i$
7: **end for**

4.3   Evaluation

In this section, a discussion concerning the evaluation of the classification-based approach is presented. The operation of the proposed classifier-based approach was evaluated by training the classifier on a proportion of the UKHCD-4 dataset, described in

---

[3]http://www.cs.waikato.ac.nz/ml/weka/downloading.html/

Chapter 3, and testing it on the remainder. More precisely Ten-fold Cross Validation (TCV) was adopted. Recall that the attitude of individual speakers, with respect to each debate, was known from the way that the speakers eventually voted[4]. In other words the assumption was made that speeches made during the course of a debate reflect how speakers would eventually vote, thus it was assumed that speakers never "change their mind" during a debate. The metrics used for the comparison were precision (the effectiveness of a system to correctly categorise records as being of a particular class), recall (the effectiveness of a system to distinguish between classes), the F-measure (the harmonic mean of precision and recall) and average accuracy (the ratio of correct classifications over all classifications). The F-measure combines the precision and recall values and is a good overall measure. The following equations (4.2a - 4.2g) show how the metrics are calculated:

$$Precision_{Aye} = \frac{TP}{TP + FP} = PositivePredictiveValue \tag{4.2a}$$

$$Precision_{No} = \frac{TN}{TN + FN} = NegativePredictiveValue \tag{4.2b}$$

$$Recall_{Aye} = \frac{TP}{TP + FN} = TruePositiveRate = Sensitivity \tag{4.2c}$$

$$Recall_{No} = \frac{TN}{TN + FP} = TrueNegativeRate = Specificity \tag{4.2d}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2e}$$

$$F - Measure_{Aye} = \frac{2 \times Precision_{Aye} \times Recall_{Aye}}{Precision_{Aye} + Recall_{Aye}} \tag{4.2f}$$

$$F - Measure_{No} = \frac{2 \times Precision_{No} \times Recall_{No}}{Precision_{No} + Recall_{No}} \tag{4.2g}$$

where TP, TN, FN and TN are the True Positive, True Negative, False Negative and True Negative counts as follows:

True Positive : Number of test records where the speaker votes "Aye" and sentiment miner predicts "Aye".

True Negative : Number of test records where the speaker votes "No" and sentiment miner predicts "No".

False Negative : Number of test records where the speaker votes "Aye" and sentiment miner predicts "No".

---

[4]In the few debates in the data set that were followed by multiple votes, it was assumed that the first vote better represented the speaker's attitude. The assumption appeared to be justified based on the direct inspection of these debates.

False Positive : Number of test records where the speaker votes "No" and sentiment miner predicts "Aye".

Three sets of experiments were conducted (the results of which are presented in the following three sub-sections). Each set of experiments had a different objective as follows:

- To test the operation of the proposed classification-based approach, in the context of classifier generation, using speeches on their own.

- To test the operation of the proposed classification-based approach, in the context of classifier generation, by augmenting the speech data with additional information, namely "party affiliation" and "debate ID".

- To test the operation of the proposed classification-based approach, in the context of classifier generation, by using only "party affiliation" and "debate ID" as the input data.

The reasoning associated with the above was to check whether individual speakers with the same political affiliation, and taking part within the same debate, will vote the same way or not. Sub-sections 4.3.2 and 4.3.3 discuss some observations concerning this issue in more detail.

4.3.1   Classification using speech data only

Regarding the classification-based approach using speech data only the results are presented in Table 4.2. The table shows the overall precision, recall and F-measure (with respect to both the "Aye" and the "No" classes), and the average accuracy, values obtained. From the table it can be observed that good results were obtained using the J48 classifier generator, which outperformed all the other classifiers including the SMO classifier. This was a surprising result as SVMs are usually considered to be best suited to text classification [Joachims, 1998] and were expected to outperform all the other classifiers. The reasons of J48 outperforming SMO are still unclear to us and point to a natural direction of future research. Reasonable results were also obtained using the JRip classifier. The worst recorded average F-measure (0.380) and worst recorded average precision (0.293) were obtained using the ZeroR classifier, while the worst recorded average recall (0.538) were obtained using the Naïve Bayes classifier. Inspection of Table 4.2 also indicates that there is no discernible difference with respect to the operation of the first five classifiers with respect to either the "Aye" or the "No" class (not the case when using lexicons as will become apparent later in this thesis).

Note also that the ZeroR classifier has only been included to provide a baseline classifier so as to establish a baseline accuracy. ZeroR is a simple rule-based classifier that only predicts the majority (most common) class. Table 4.3 shows the confusion matrix data used to calculate the metrics given in Table 4.2. With respect to Table 4.3 the True Positive (TP) counts with respect to each classifier are given in the top-left quadrant, the True Negative (TN) counts in the bottom-right quadrant; whilst the False Positive (FP) and the False Negative (FN) counts are given in the top-right and bottom-left quadrants respectively.

	Precision (P)			Recall (R)			F-Measure (F)			Accuracy (A)
Classifier	Aye	No	Avg.	Aye	No	Avg.	Aye	No	Avg.	Avg.
J48	0.628	0.601	0.615	0.719	0.497	0.618	0.671	0.544	0.613	61.751%
JRip	0.581	0.529	0.557	0.685	0.418	0.562	0.629	0.467	0.555	56.238%
SMO	0.602	0.531	0.569	0.604	0.529	0.570	0.603	0.530	0.570	56.963%
NB	0.576	0.493	0.538	0.531	0.538	0.534	0.552	0.515	0.535	53.433%
IBk	0.547	0.500	0.525	0.888	0.132	0.541	0.677	0.209	0.462	54.110%
ZeroR	0.541	0.000	0.293	1.000	0.000	0.541	0.702	0.000	0.380	54.110%
Min	0.541	0.000	0.293	0.531	0.000	0.534	0.552	0.000	0.380	53.433%
Max	0.628	0.601	0.615	1.000	0.538	0.618	0.702	0.544	0.613	61.751%
Average	0.579	0.442	0.516	0.738	0.352	0.561	0.639	0.378	0.519	56.101%
SD	0.033	0.220	0.114	0.176	0.230	0.031	0.056	0.223	0.084	3.089%

Table 4.2: Evaluation results obtained using the classification-based approach to sentiment mining using only speech data (values generated from confusion matrix data given in Table 4.3).

	Class label = Aye						Class label = No					
	J48	JRip	SMO	NB	Ibk	ZeroR	J48	JRip	SMO	NB	Ibk	ZeroR
Speaker votes Aye	805	766	676	594	994	1119	314	353	443	525	125	0
Speaker votes No	477	552	447	438	824	949	472	397	502	511	125	0

Table 4.3: Confusion matrix for results presented in Table 4.2.

4.3.2 Classification using speech data augmented with "party affiliation" and "debate ID" information

As already noted the classification-based approach allows for additional information to be included in the feature vector representation. Hence additional experiments that included "party affiliation" and "debate ID" information in the data representation were conducted. The intuition being that if a speaker with a particular political affiliation in debate x voted "Aye" then other people with the same political affiliation debating within debate x are also likely to vote "Aye". The results are presented in Table 4.4.

Comparing the results presented in Table 4.4 with the results presented previously in Table 4.2 it can be observed that significantly better results were obtained when adding "party affiliation" and "debate ID" than when using the speech information on its own (a best average accuracy of 85.397% compared to a best average accuracy of 61.751%). Overall best performance was again obtained using the J48 classifier. Good results were also obtained using the JRip classifier. The worst recorded average F-measure (0.451) was this time obtained using the IBk classifier, while the worst recorded average precision (0.542) and average recall (0.538) were again obtained using the Naïve Bayes classifier.

Classifier	Precision (P)			Recall (R)			F-Measure (F)			Accuracy (A)
	Aye	No	Avg.	Aye	No	Avg.	Aye	No	Avg.	Avg.
J48	0.865	0.841	0.854	0.865	0.841	0.854	0.865	0.841	0.854	85.397 %
JRip	0.892	0.710	0.809	0.688	0.902	0.786	0.777	0.795	0.785	78.627 %
SMO	0.707	0.655	0.683	0.709	0.653	0.683	0.708	0.654	0.683	68.327 %
NB	0.580	0.497	0.542	0.533	0.545	0.538	0.555	0.520	0.539	53.820 %
IBk	0.551	0.589	0.569	0.945	0.094	0.554	0.696	0.162	0.451	55.416 %
ZeroR	0.541	0.000	0.293	1.000	0.000	0.541	0.702	0.000	0.380	54.110 %
Min	0.541	0.000	0.293	0.533	0.000	0.538	0.555	0.000	0.380	53.820 %
Max	0.892	0.841	0.854	1.000	0.902	0.854	0.865	0.841	0.854	85.397 %
Average	0.689	0.549	0.625	0.790	0.506	0.659	0.717	0.495	0.615	65.949 %
SD	0.158	0.293	0.205	0.177	0.379	0.137	0.102	0.344	0.189	13.732 %

Table 4.4: Evaluation results obtained using the classification-based approach to sentiment mining applied to speeches augmented with "party affiliation" and "debate ID".

4.3.3 Classification using "party affiliation" and "debate ID" only

Given the results presented in Table 4.4 it was considered interesting to investigate whether a better indicator of attitude is simply "party affiliation" and "debate ID" on its own. The author thus constructed a data set comprising only two features, "party affiliation" and "debate ID", and conducted some further classification experiments with this aim in mind. The results are presented in Table 4.5. From the table it might be observed that best results were again obtained using the J48 classifier generator which outperformed all the other classifiers (average accuracy of 87.089% compared to a best average accuracy of 85.397% obtained using speeches, "party affiliation" and "debate ID"). Good results were also obtained using the JRip classifier, but there were also surprisingly good results obtained using IBk, NB and SMO. Whatever the case the results clearly show that the best indicator of attitude, given a particular debate (identified by a unique ID), is "party affiliation" and not the content of concatenated speeches made by individuals. In other words, as might be expected, speakers that belong to the same party are likely to vote in the same way. Thus if we wish to predict

the likely outcome of a debate while it is in progress we should determine attitude (using the techniques proposed in this thesis) with respect to groups of speeches belonging to speakers with the same political affiliation. We could do this either by further concatenating the speeches belonging to speakers with the same affiliation and analysing each as one very large "document" or alternatively by using some voting system to produce an aggregated attitude for groups of speakers with the same affiliation.

Classifier	Precision (P) Aye	No	Avg.	Recall (R) Aye	No	Avg.	F-Measure (F) Aye	No	Avg.	Accuracy (A) Avg.
J48	0.876	0.865	0.871	0.887	0.851	0.871	0.881	0.858	0.871	87.089 %
JRip	0.896	0.722	0.816	0.705	0.903	0.796	0.789	0.802	0.795	79.594 %
SMO	0.840	0.776	0.811	0.799	0.821	0.809	0.819	0.798	0.809	80.899 %
NB	0.781	0.707	0.747	0.735	0.757	0.745	0.757	0.731	0.745	74.468 %
IBk	0.851	0.840	0.846	0.868	0.821	0.846	0.859	0.830	0.846	84.623 %
ZeroR	0.541	0.000	0.293	1.000	0.000	0.541	0.702	0.000	0.380	54.110 %
Min	0.541	0.000	0.293	0.705	0.000	0.541	0.702	0.000	0.380	54.110 %
Max	0.896	0.865	0.871	1.000	0.903	0.871	0.881	0.858	0.871	87.089 %
Average	0.798	0.652	0.731	0.832	0.692	0.768	0.801	0.670	0.741	76.797 %
SD	0.132	0.325	0.218	0.109	0.342	0.119	0.066	0.331	0.182	11.933 %

Table 4.5: Evaluation results obtained using a classifier built using only "party affiliation" and "debate ID".

4.4 Summary

This chapter has presented the classification-based approach for political sentiment mining in the context of mining the UK House of Commons political debates. The two phases of the process (preprocessing and attitude prediction) were described in detail. The operation of the machine learning classifier approach was described and a full evaluation was presented. The next two chapters present the remaining two proposed sentiment mining approaches, again discussed in the context of mining the UK House of Commons political debates, starting with the generic sentiment lexicons-based approach in Chapter 5.

Chapter 5

Political Sentiment Mining Using Generic Sentiment Lexicons

"When you have too many people and you're trying to satisfy everybody's input, you usually end up with something so incredibly generic that it has no point of view."

Rob Zombie

As already noted earlier in this thesis, two approaches to sentiment mining using sentiment lexicon can be identified: (i) using off-the-shelf generic sentiment lexicons and (ii) using domain specific sentiment lexicons. This chapter is directed at the generic lexicon-based approach and the following chapter (Chapter 6) considers the domain specific approach.

Given a new text to be categorised as expressing either a "positive" or a "negative" sentiment, the subjective words in the text act as sentiment indicators. However, subjective word identification is a challenging process because of the complexity of natural language. One solution is to use sentiment lexicons, which can be used to look-up words to firstly identify subjective words (as opposed to objective words) and secondly to determine the degree of sentiment and polarity (positive or negative) associated with the identified subjective words. This information can then be used to make a judgement about the overall sentiment represented by a text. The main idea is to combine the subjective word-level sentiment values to give a whole document sentiment value. More precisely, the "primitive" word-level sentiment values can be combined (accumulated) to form higher level sentiment values with respect to different levels of "text inclusion" (sentence-level, paragraph-level, and so on) so that a judgement can be made about the polarity of these higher levels. [Grijzenhout et al., 2010] has presented a review of

previous work directed at different levels of sentiment identification at different levels of text inclusion. Thus sentiment mining can be performed at a number of different levels of text inclusion. In the context of the debate transcripts under consideration with respect to the work described in this thesis seven potential levels, similar to those argued by [Grijzenhout et al., 2010], can be identified as follows (listed from highest whole debate level to lowest word level):

- whole debate

- party concatenated speeches

- speaker's concatenated speeches within a single debate

- speaker's single speech

- a paragraph within a single speech

- a sentence within a paragraph

- a word within a sentence

Here the debate level is the most general. In this case the whole debate is considered and assigned a "positive" or "negative" attitude value that represents the overall polarity of the debate (the outcome). In the context of the nature of the debate analysis considered in this thesis this level is too generic in that it conceals the individual attitudes of single speakers (Members of Parliament). The second level, the party level, considers the complete set of speeches made by all speakers belonging to each individual political party by concatenating them together to form one text segment representing the party during the debate. The polarity of each set of speeches, representing participant parties, is then classified as having either a "positive" or "negative" attitude. The third level, the speaker level, considers the complete set of speeches presented by each individual speaker by concatenating them together to form one text segment representing the speaker during a debate. The polarity of each set of speeches is separately classified as having either a "positive" or "negative" attitude. The fourth level, the speech level, considers each of the speeches made by each individual speaker, during a debate, separately. In this case the polarity of each speech is separately classified according to whether it displays a "positive" or "negative" attitude. The paragraph level, as the name suggests considers the speeches in terms of paragraphs. The remaining two levels operate at a very low level of granularity. Consequently both the word and sentence level are considered not to be applicable with respect to the political debate application considered in this thesis because MP's viewpoints or attitudes cannot be effectively represented by a single word or sentence. The speaker level is considered

to be the most appropriate choice of level with respect to the parliament transcripts, assuming that speakers do not change their opinion during the debate, and thus this was adopted with respect to the work considered in this thesis chapter (and the next).
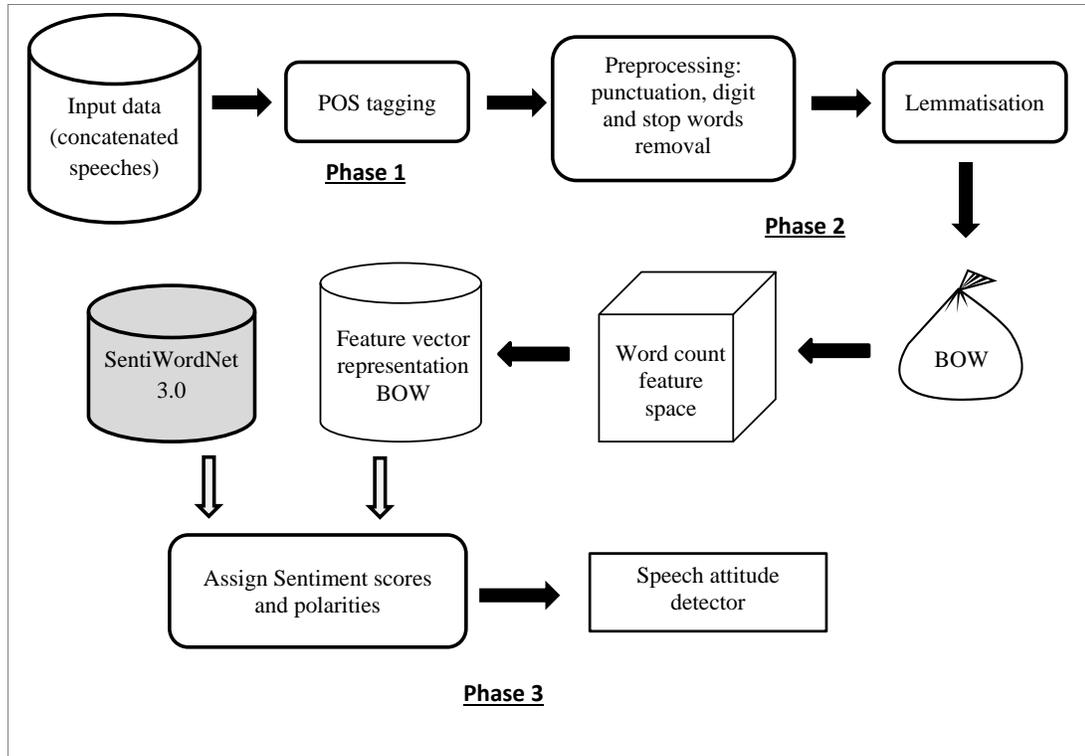
Figure 5.1: The generic sentiment lexicon-based approach to sentiment mining.

An overview of the "phases" that make up the overall lexicon base approach is presented in Figure 5.1. The most significant phases are: (i) part-of-speech tagging (POST), (ii) text preprocessing and (iii) attitude detection. Part-of-speech tagging was conducted so as to assign a part-of-speech tag to each word in the input text; this is described further in Section 5.1. The second phase was text preprocessing which is described in Section 5.2. Once the data has been preprocessed the attitude detection phase was commenced, this is described in Section 5.3. Some experimental results are presented and discussed in Section 5.4, while the chapter is concluded with a summary presented in Section 5.5.

5.1 Part-Of-Speech Tagging (POST)

Part-Of-Speech Tagging (POST) is a process whereby each word in a given text is assigned a POS tag according to its context in the sentence or a phrase in which it is used [Bellegarda, 2010]. There are different tag sets that can be used to assign a POS tag for a particular instance of a word. Tag sets may be very coarse (using a small tag

set of the form {N, V, Adj, Adv}[1]) or fine-grained like the Penn Treebank POS tag set which contains a set of 35 different part-of-speech tags [Petrov et al., 2011]. Table 5.1 shows an alphabetical list of standard Penn Treebank tags and the parts of speech corresponding to them. For example, the sentence:

```
Further to the previous unclear answer, is the Secretary of State
categorically ruling out revisiting the ''cat and trap'' system for the
aircraft carriers?
```

after assigning a tag for each part of speech will be as follows:

```
Further/RBR to/TO the/DT previous/JJ unclear/JJ answer/NN is/VBZ the/DT
Secretary/NP of/IN State/NP categorically/RB ruling/VBG out/RP
revisiting/VBG the/DT cat/NN and/CC trap/NN system/NN for/IN the/DT
aircraft/NN carriers/NNS
```

With respect to sentiment mining, POST is important because many related words (for example "suffice", "sufficiency", "sufficient" and "sufficiently"), which have different POS tags, will typically have different sentiment scores. POST is also significant with respect to sentiment mining because it can contribute to Word Sense Disambiguation (WSD), the process of dealing with the polysemy problem (different meanings for the same word) by discriminating the proper "semantic" sense of a word in a specific context or circumstance [Wilks and Stevenson, 1998]. At the end of the POST phase a list of terms $T = \{t_1, t_2, \ldots, t_m\}$ each associated with a POS tag $post_i$ will have been produced, thus a set of pairs of the form $\langle t_i, post_i \rangle$.

5.2   Preprocessing

Once the POST phase is complete the preprocessing phase can be commenced. As in the case of the classification-based approach presented in the previous chapter (in Section 4.1) the first steps in the preprocessing phase were tokenisation and stop word removal. Stemming was not used in the context of the two lexicon-based approaches proposed in this thesis, because (as noted above) words like "suffice", "sufficiency", "sufficient" and "sufficiently" will have different Part Of Speech (POS) tags and will consequently have different sentiment scores. When stemming is applied, these words will be reduced to a single word (stem) and thus share the same sentiment score therefore possibly losing the more appropriate individual sentiment values. Instead a lemmatisation approach

---

[1]N: noun, V: verb, Adj: adjective and Adv: adverb.

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NP	Proper noun singular
NPS	Proper noun plural
PDT	Predeterminer
POS	Possessive ending
PP	Personal pronoun
PP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	To
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle
VBP	Verb, noun-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP$	Possessive wh-pronoun
WRB	Wh-adverb

Table 5.1: Standard Penn Treebank POS tag set.

was adopted. Lemmatisation is different from stemming in that the aim is to reduce a given word to its "conventional standard form" instead of its root or stem form. For example all verbs would be converted to their infinitive form and all nouns to their singular form [Amine et al., 2010].

On completion of tokenisation, stop-word removal and lemmatisation a Bag-Of-Words (BOW) representation was again used, as in the case of the classification-based approach, however in this case each word in the BOW was linked to a POS tag. Thus each document (concatenated speech) was represented by some subset of the BOW which in turn was translated into a feature vector form. The feature vector elements hold term occurrence counts. More formally a speech i was represented as a vector $V_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$ where w_{ij} is the occurrence count of term j in speech i. It should also be noted that each element in V_i corresponds to a term in the BOW where it is stored with its POS tag. The list of terms associated with feature vector V_i is indicated using the notation $T_i = \{t_{i1}, t_{i2}, \ldots, t_{im}\}$. Thus a set of feature vectors $V = \{V_1, V_2, \ldots, V_z\}$ and a set of term lists $T = \{T_1, T_2, \ldots, T_z\}$, with a one-to-one correspondence between the two sets, were produced.

5.3 Attitude detection using generic sentiment lexicons

In this section a discussion concerning the effectiveness of the proposed generic sentiment lexicon approach with respect to sentiment mining is presented. More specifically, a test set was used to evaluate the application of sentiment analysis using the generic sentiment lexicon approach to detect the attitude (sentiment polarity) of speeches made by individual debaters. To this end a parliamentary debate corpus (UKHCD-4) was used. Further detail regarding this data set is presented in Chapter 3 in this thesis. The adopted generic sentiment lexicon was used to assign sentiment scores to the test data to determine the attitude of the debaters (speakers). Because this attitude was known from the way that the speakers eventually voted, the predicted attitude could be compared with the known attitude.

Given the produced feature vectors representations (see the end of Section 5.2), sentiment analysis was applied to the terms in each vector to determine the attitude reflected by the vector and consequently the document (concatenated speech) it represents. The "sentiment" score (value) associated with each term t_{ij} in feature vector T_i is obtained by "looking up" the term in a sentiment lexicon. As noted previously in Section 2.3 a sentiment score is a numeric value indicating some degree of subjectivity. The orientation of a word is an indicator of whether a word expresses assent or dissent with respect to some object or concept. Consequently document polarity can be judged by counting the number of positive and negative terms and calculating the difference.

The resulting polarity then describes the attitude reflected by the document.

With respect to the generic lexicon-based approach SentiWordNet 3.0 was used. SentiWordNet 3.0 assigns a positive and a negative numerical score (ranging from 0.0 to 1.0) to each synset that exists in WordNet 3.0 so as to generate polarity scores. The sentiment scores were assigned automatically for WordNet 3.0 synsets using an algorithm that includes a semi-supervised learning step in addition to a random-walk step for refining the scores. The synsets in SentiWordNet 3.0 were broken down into single terms in order to produce a list of terms which were then used to retrieve the corresponding score. Terms which originate from the same synset were taken to have the same sentiment score and polarity. However, if a term features in different synsets then: (i) if the different forms of the term in the different synsets have different grammatical tagging (POS tag), then the "word-sense distinction" was resolved simply by considering the different POS tags of the term (as suggested in [Wilks and Stevenson, 1998]) and thus it was split into distinguished terms; (ii) otherwise if the term has the same grammatical tagging in the different synsets, the highest sentiment score was selected.

More formally, from the above, the accumulative sentiment score st_i associated with a speech i was computed using:

$$st_i = \sum_{j=1}^{j=m} (\text{SenLex}(term_j) \times w_{ij}) \qquad (5.1)$$

where: (i) $term_j$ is a term in the the feature vector representing the current speech i; (ii) SenLex is a function that returns the sentiment score ($-1.0 \leq \text{SenLex}(term_j) \leq +1.0$) for each $term_j$, from the adopted sentiment lexicon (SentiWordNet 3.0), where the score is the summation of the term's positivity and negativity scores (positive and negative values); (iii) m is the number of terms in the given feature vector and (iv) w_{ij} is the occurrence count for term j in feature vector i. The occurrence count can be a true frequency count of the number of times $term_j$ appears in document i, or simply a binary value (1 or 0) to indicating the presence or absence of the term. In the upcoming sections we refer to these two techniques using the labels *TF* and *Binary* respectively. The summation (which accumulate all of the sentiment values) was used because of the nature of the political parliamentary speeches used in this thesis which are relatively long texts and typically contain large numbers of sentiment indicators. Thus the summation will ensure that the sentiment of the majority will dominate.

The attitude (class label) for each document (speaker) i is then determined according to st_i. To this end the class label set is {*positive, negative, objective, neutral*} where: (i) *positive* indicates a positive text (for the motion in the case of our political

debates), (ii) *negative* indicates a negative text (against the motion), (iii) *objective* indicates that no sentiment scores were found and (iv) *neutral* that the sentiment scores negate one another. In practice it was found that the last two class labels are rarely encountered. Algorithm 5.3.1 describes the attitude identification process. The algorithm loops through the input set of speeches, represented in terms of the sets S and T (see end of Section 5.2), a sentiment score for each speech is calculated from lines 7 to 23, the attitude from lines 24 to 38.

5.4 Results obtained using the generic lexicon-based approach

The generic lexicon-based approach was evaluated, in a similar manner to the classification based approach described in Chapter 4, by comparing the predicted class label (positive or negative) with the known manner in which the individual MPs voted with an "Aye" vote equating to a positive label, and a "No" vote equating to a negative label. The comparison was conducted using the UKHCD-4 House of Commons political debate corpus. Detail regarding this data set was presented previously in Chapter 3.

	SentiWordNet 3.0 TF		
	Aye	**No**	**Avg.**
Precision	0.554	0.495	0.524
Recall	0.766	0.271	0.518
F-Measure	0.643	0.350	0.497
Avg. Accuracy		53.894%	
	SentiWordNet 3.0 Binary		
	Aye	**No**	**Avg.**
Precision	0.546	0.502	0.524
Recall	0.909	0.109	0.509
F-Measure	0.682	0.179	0.430
Avg. Accuracy		54.167%	

Table 5.2: Evaluation results produced using the generic lexicon-based approach to sentiment mining (values generated from confusion matrix data given in Table 5.3).

The evaluation was conducted using the general purpose SentiWordNet 3.0 lexicon. The results are presented in Table 5.2 (the associated confusion matrix data is given in

| | SentiWordNet 3.0 | |
	TF	Binary
True Positive (TP)	857	1016
False Negative (FN)	262	103
True Negative (TN)	258	104
False Positive (FP)	691	845
Total	2068	2068

Table 5.3: Confusion matrix for the generic lexicon-based approach to sentiment mining.

Table 5.3). Note that a comparison between the use of the feature vector term frequency occurrence count approach and the use of the binary occurrence count approach (the columns labelled TF and Binary respectively) is included. Inspection of the results presented in Table 5.2 reveals the interesting observation that the generic lexicon-based approach exhibited a poorer performance when identifying "No" attitudes than when identifying "Aye" attitudes. An analysis of the data was undertaken (by direct inspection of the transcripts) to determine why the performance with respect to the "No" vote was relatively poorer than for the "Aye" vote and the reason for this, it is argued here, is due to the often overly polite parliamentary jargon used with respect to House of Commons political debates which means that positive sentiment is easier to identify than negative sentiment. The best generic lexicon-based average accuracy was achieved using binary feature vectors (54.167%).

5.5   Summary

This chapter has presented the generic sentiment lexicons-based approach for political sentiment mining in the context of mining the UK House of Commons political debates. The three phases of the process (part-of-speech tagging, preprocessing and attitude identification) were described in detail. The operation of using the generic sentiment lexicon-based approach, in the context of sentiment mining, was described and an evaluation of the approach was presented and discussed. Chapter 6 discusses the remaining proposed sentiment mining approach in the context of mining UK House of Commons debates; namely the domain specific sentiment lexicon-based approach.

Algorithm 5.3.1 Attitude identification using sentiment lexicon

1: INPUT: Sentiment Lexicon $SenLex$, set of sets of terms $T \subset BOW$, set of feature vectors S
2: OUTPUT: Set of Attitudes labels $A = \{a_1, a_2, \ldots, a_z\}$
3: $PosCount = 0$
4: $NegCount = 0$
5: $PosScore = 0$
6: $NegScore = 0$
7: **for all** $T_i \in T^1$ **do**
8:     **for all** $t_{ij} \in T_i$ **do**
9:         **if** $t_{ij} \in SenLex$ **then**
10:            $score_{ij} = SenLex(t_{ij}) \times w_{ij}$
11:         **else**
12:            $score_{w_{ij}} = 0$
13:         **end if**
14:         **if** $Score_{ij} > 0$ **then**
15:            $PosCount = PosCount + w_{ij}$
16:            $PosScore = PosScore + Score_{ij}$
17:         **else if** $Score_{w_{iu}} < 0$ **then**
18:            $NegCount = NegCount + w_{ij}$
19:            $NegScore = NegScore + Score_{ij}$
20:         **else**$[Score_{ij} = 0]$
21:            DO NOTHING
22:         **end if**
23:     **end for**
24:     **if** $PosCount = 0 \ \wedge \ NegCount = 0$ **then**
25:         $a_i = Objective$
26:     **else if** $PosScore > NegScore$ **then**
27:         $a_i = Positive$
28:     **else if** $NegScore > PosScore$ **then**
29:         $a_i = Negative$
30:     **else**$[PosScore = NegScore]$
31:         **if** $PosCount > NegCount$ **then**
32:            $a_i = Positive$
33:         **else if** $NegCount > PosCount$ **then**
34:            $a_i = Negative$
35:         **else**$[PosCount = NegCount]$
36:            $a_i = Neutral$
37:         **end if**
38:     **end if**
39: **end for**

Chapter 6

Political Sentiment Mining Using Domain Specific Sentiment Lexicons

"Each man is capable of doing one thing well. If he attempts several, he will fail to achieve distinction in any."

Plato

This chapter is directed at sentiment mining using the domain specific lexicons-based approach. Sentiment mining using domain specific lexicons operates in a similar manner to that using generic lexicons with the exception that dedicated lexicons are used. The challenge is obtaining the required specialist lexicons. As discussed in Sub-section 1.1.1 two approaches to generating domain-specific sentiment lexicons can be identified: (i) direct generation and (ii) adaptive generation. In this context the author has proposed and developed techniques for both direct and adaptive domain specific lexicon generation, and theses are presented in this chapter. In both cases the input is a set of n binary labelled parliamentary speeches (documents) $D = \{d_1, d_2, \ldots, d_n\}$. The labels are drawn from the set $\{positive, negative\}$. The output in both cases is a lexicon where each term is encoded in the form of a set of tuples $\langle t_i, post_i, s_i \rangle$, where t_i is a term that appears in the document collection D, $post_i$ is the part-of-speech tag associated with term t_i and s_i is the associated sentiment score. As shown in Figure 6.1 both domain-specific lexicon generation techniques comprise four phases as follows:

1. Part-of-speech tagging (to identify the POS tags)

2. Document preprocessing

3. Sentiment score (s_i) and polarity calculation

4. Lexicon generation

Each of these phases is described in more detail in the following four Sections 6.1, 6.2, 6.3 and 6.4 respectively. For evaluation purposes, these two techniques were used to create two political domain specific sentiment lexicons using the UK House of Commons political debate data (namely the UKHCD-3 dataset): (i) *PoLex* and (ii) *PoliSenti-WordNet*. The first was produced using the proposed direct generation technique, while the second was produced using the proposed adaptive generation technique (founded on SentiWordNet 3.0). PoLex comprised 170,703 terms and PoliSentiWordNet 258,353 terms. The generated lexicons were applied, in the same manner as SentiWordNet 3.0 was applied as described in Chapter 5, to assign sentiment scores to the UK House of Commons political debate test data (namely the UKHCD-4 dataset) so as to determine the attitude of the debaters (speakers). The experimental results that were obtained are presented and discussed in Section 6.6. This chapter is concluded with a summary presented in Section 6.7.
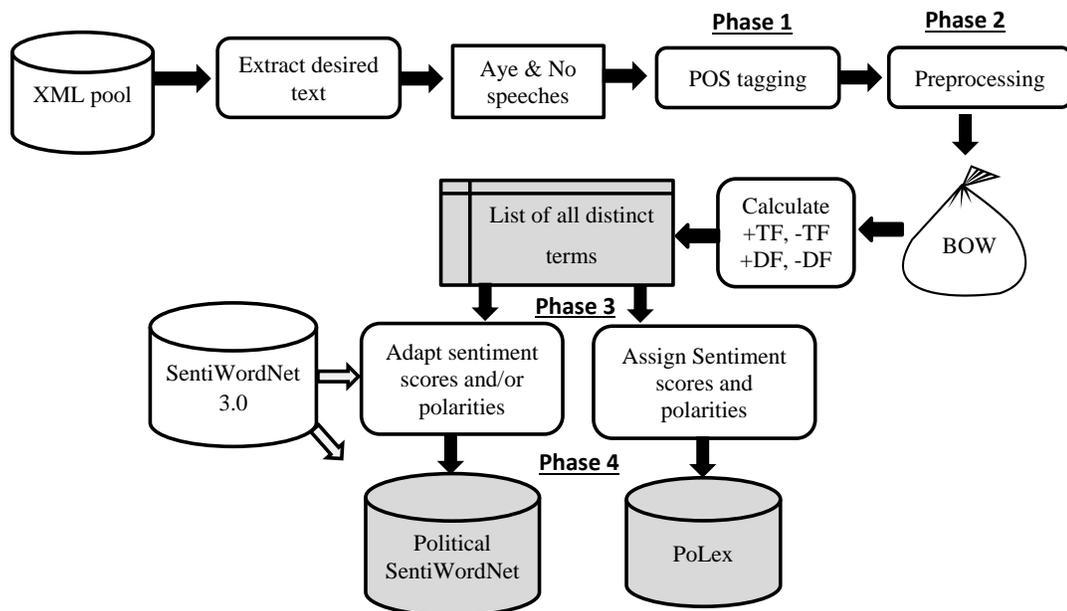
Figure 6.1: High-level schematic illustrating the domain-specific lexicon generation process.

6.1   Part-Of-Speech-Tagging (POST)

The first phase with respect to the two lexicon generation techniques was part-of-speech tagging (POST) so that each word in the input was assigned a particular part-of-speech (POS) tag to produce a list of terms $T = \{t_1, t_2, \ldots, t_m\}$, each associated with a POS tag $post_i$. The practice of POST was discussed previously in Section 5.1 in the previous chapter.

6.2   Preprocessing

The pre-processing phase for the domain specific lexicon generation process commenced with the conversion of all upper-case alphabetic characters to lower-case, this was then followed by punctuation mark and numeric digit removal. Next, given the produced list of terms T from the previous steps, a Bag-Of-Words (BOW) representations was created for all t_i in T (all the terms in the input document collection D). Each term t_i in the BOW is defined using a 6-tuple of the form $\langle t_i, post_i, tf_i^+, tf_i^-, df_i^+, df_i^- \rangle$, where: (i) t_i is the term of interest (term number i); (ii) $post_i$ is the associated POS tag as identified in the previous phase (Part-Of-Speech-Tagging), (iii) tf_i^+ is the associated term frequency (number of occasions that the term t_i appears in a text collection) with respect to texts that display a positive attitude ("Aye" labelled texts in the case of our political speeches), (iv) tf_i^- is the associated term frequency with respect to texts that display a negative attitude ("No" labelled texts in the case of our political speeches), (v) df_i^+ is the associated document frequency (number of texts in which t_i appears) with respect to texts that display a positive attitude ("Aye" labelled documents) and (vi) df_i^- is the associated document frequency with respect to texts that display a negative attitude ("No" labelled documents).

6.3   Sentiment score and polarity calculation

On completion of the pre-processing phase, sentiment scores (sentiment weightings) were calculated with respect to each term contained in the generated BOW so far. The TF-IDF weighting value W_{ij} for a term t_i in a text j is obtained using:

$$W_{ij} = TF - IDF = tf_{i,j} \cdot \left(log_2 \frac{n}{df_i} \right) \qquad (6.1)$$

where: (i) $tf_{i,j}$ is the frequency of term t_i in document (speech) j (thus the local weight for the term), (ii) n is the total number of documents in the corpus (concatenated

speeches in the debate), and (iii) df_i is the number of documents (speeches) containing term t_i (thus the global weight for the term). A disadvantage of TF-IDF, in the context of sentiment mining, is that it does not reflect a term's sentiment tendency (orientation) and thus for the purposes of generating lexicons to be used for sentiment analysis there was a need for either: (i) an alternative sentiment intensity weighting scheme or (ii) an alternative form of the TF-IDF scheme that takes into consideration the situation where a term t_i appears in both positive and negative documents. With respect to the latter (an alternative form of the TF-IDF scheme) the ΔTF-IDF provides "an intuitive general purpose technique to efficiently weight word scores" [Martineau and Finin, 2009]. Thus ΔTF-IDF considers the biased occurrence of terms with respect to individual classes (sentiment in our case). The ΔTF-IDF value W_{ij} for a term t_i in a text j is obtained as follows:

$$
\begin{aligned}
W_{i,j} &= \Delta TF - IDF = tf_{i,j} \cdot \left(log_2 \frac{N^+}{df_i^+} \right) - tf_{i,j} \cdot \left(log_2 \frac{N^-}{df_i^-} \right) \\
&= tf_{i,j} \cdot \left(log_2 \frac{N^+}{df_i^+} \frac{df_i^-}{N^-} \right)
\end{aligned}
\tag{6.2}
$$

where: (i) N^+ is the number of *positive* texts in the input document collection D ("Aye" labelled with respect to the proposed political sentiment mining application), (ii) N^- is the number of *negative* texts ("No" labelled), (iii) $tf_{i,j}$ is the term frequency for term t_i in text j, (iv) df_i^+ is the document frequency for term t_i with respect to positive texts in the input document collect D and (v) df_i^- is the document frequency for term t_i with respect to negative texts.

However, the ΔTF-IDF scheme is directed at sentiment classification of individual texts according to their ΔTF-IDF values [Martineau and Finin, 2009]. The work described in this chapter was focused on building domain-specific lexicons and thus a slightly adapted ΔTF-IDF weighting scheme was proposed so that term weightings were considered with respect to the entire document collection D and not per document. This scheme is referred to as ΔTF-IDF$'$. Thus the ΔTF-IDF$'$ value $W_{i,D}$ for a term t_i with respect to a document collection D was obtained as follows:

$$
W_{i,D} = \Delta TF - IDF' = tf_i^+ \cdot \left(log_2 \frac{N^+}{df_i^+} \right) - tf_i^- \cdot \left(log_2 \frac{N^-}{df_i^-} \right)
\tag{6.3}
$$

where: tf_i^+ is the term frequency with respect to positive texts and tf_i^- is the term frequency with respect to negative texts. The advantages offered by the proposed ΔTF-IDF$'$ scheme were that it could be used to assign sentiment scores to each term in a

text (speech) collection in such a manner that the term occurrences in both negative and positive texts can be taken into consideration.

Thus the ΔTF-IDF$'$ scheme was used to determine sentiment scores for each term. On completion of this phase, each term in the BOW comprised a 7-tuple of the form $\langle t_i, post_i, tf_i^+, tf_i^-, df_i^+, df_i^-, s_i \rangle$, where s_i is the sentiment score associated with term t_i.

6.4 Lexicon generation

As the name implies, using the direct generation technique, the desired domain-specific sentiment lexicon was directly generated from the labelled source data processed as described earlier in this chapter. Thus the generated BOW, which contains the terms and their associated POS tags, sentiment scores and polarities, was converted directly into a domain specific lexicon. The proposed $PoLex$ political domain specific lexicon was generated in this manner. The PoLex lexicon comprises 170,703 terms such that each term was encoded in the form of a tuple $\langle t_i, post_i, s_i \rangle$, thus each term in PoLex was combined with its: (i) associated part-of-speech tag, (ii) sentiment score (determined using ΔTF-IDF$'$ scheme) and (iii) polarity (the sign of the sentiment score).

In the case of adaptive generation the idea was to use an existing, domain independent, sentiment lexicon[1] Lex, and adapt this to produce a domain-specific lexicon Lex'. The process commenced by copying the content of Lex over to Lex'. The adaptation was then as follows. Given a term t_i that is both in T (the list of all distinct terms derived from the document collection) and Lex, if the two associated sentiment scores have different polarities (thus one negative and one positive) the polarity from the calculated sentiment score was adopted (but not the magnitude) with respect to Lex'. Note that the polarity of a term is the sign of its associated sentiment score. The magnitude was not also adopted to: (i) produce a different combination other than that one produced from the direct generation technique where both of polarity and magnitude were used and (ii) preserve the compatibility between the sentiment scores of all terms (those that copied from Lex and that originated from T) in Lex'. Terms included in T, but not in Lex, were simply appended to Lex' after rescaling the range of their sentiment scores' magnitude to be compatible with the other scores already exist in Lex'. The proposed $PoliSentiWordNet$ political domain specific lexicon was generated in the above manner. PoliSentiWordNet comprises 258,353 terms.

---

[1]SentiWordNet 3.0 was used in the case of the work described here.

6.5 Evaluation framework for domain specific sentiment lexicons

In this section a discussion concerning the effectiveness of the domain specific sentiment lexicons, generated using the proposed domain-specific lexicon generation techniques, with respect to sentiment mining is presented. The UKHCD-3 training set was used to generate two exemplar domain specific lexicons: (i) PoLex using the direct generation technique and (ii) PoliSentiWordNet using the adaptive technique. Once the domain specific sentiment lexicons were generated, as described in Section 6.4, these lexicons were ready for usage in the context of sentiment mining.

To evaluate the techniques the generated lexicons, PoLex and PoliSentiWordNet, were applied to the UKHCD-4 parliamentary debate corpus and the predicted sentiment compared to the "known" sentiment (the way that speakers actually voted). As in the case of the attitude detection using the generic sentiment lexicon-based approach (presented in Chapter 5) it was first necessary to prepare the test data. This is a two phase process: (i) Part-Of-Speech Tagging (POST) and (ii) pre-processing. During the first phase particular part-of-speech tags were assigned to each word in the input data as described in Section 5.1. The required preprocessing (phase two) was conducted in the same manner as applied to the original training data used to generate the lexicons as described in Section 5.2. Once the test data had been prepared attitude detection could be commenced in the same manner as described in Section 5.3 with respect to the usage of generic lexicons for sentiment analysis. In this manner, the generated domain specific sentiment lexicons were used to assign sentiment scores to the test data to determine the attitude of the debaters (speakers). Because this attitude was known from the way that the speakers eventually voted, the predicted attitude could be compared with the known attitude. The results obtained are presented in Section 6.6 below.

6.6 Evaluation results for domain specific lexicons

The sentiment analysis results produced using PoLex and PoliSentiWordNet lexicons, are presented in Table 6.1 (the associated confusion matrix data is given in Table 6.2). As in the case of the previous evaluations reported earlier in this thesis, experiments to determine the effectiveness of the domain-specific lexicon-based techniques were again conducted using feature vectors comprised of both term frequency occurrence counts and binary (yes-no) occurrence counts (the columns in Table 6.1 labelled TF and Binary respectively). Inspection of the results presented in Table 6.1 indicates that both PoLex and PoliSentiWordNet produced almost the same results. As in the case of the use of

generic lexicons closer inspection of the table reveals the interesting observation that both lexicons worked better with respect to predicting positive "Aye" attitudes than negative "No" attitudes. The reason for this, it is again argued, is due to the often overly polite parliamentary jargon used which means that positive sentiment is easier to identify than negative sentiment. The best domain specific lexicon-based average accuracy was achieved using PoLex and binary feature vectors (55.464%).

	PoLex TF			PoliSentiWordNet TF		
	Aye	No	Avg.	Aye	No	Avg.
Precision	0.554	0.503	0.528	0.554	0.500	0.527
Recall	0.798	0.241	0.520	0.777	0.262	0.520
F-Measure	0.654	0.326	0.490	0.647	0.344	0.496
Avg. Accuracy	54.255%			54.110%		

	PoLex Binary			PoliSentiWordNet Binary		
	Aye	No	Avg.	Aye	No	Avg.
Precision	0.560	0.534	0.547	0.556	0.522	0.539
Recall	0.831	0.229	0.530	0.831	0.217	0.524
F-Measure	0.669	0.320	0.495	0.666	0.307	0.486
Avg. Accuracy	55.464%			54.937%		

Table 6.1: Evaluation results produced using the domain specific lexicons-based approach to sentiment mining (values generated from confusion matrix data given in Table 6.2).

	PoLex		PoliSentiWordNet	
	TF	Binary	TF	Binary
TP	893	930	870	930
FN	226	189	249	189
TN	229	217	249	207
FP	720	732	700	742
Total	2068	2068	2068	2068

Table 6.2: Confusion matrix for the domain specific lexicons-based approach to sentiment mining.

6.7 Summary

This chapter has presented the domain specific sentiment lexicons-based approach for political sentiment mining in the context of applying sentiment mining to the UK House of Commons political debates. The four phases of the proposed domain-specific lexicon generation techniques (part-of-speech tagging, preprocessing, sentiment score and polarity calculation, and attitude identification) were described in detail. The usage of the domain specific sentiment lexicon-based approach in sentiment mining was described using two domain specific sentiment lexicons, PoLex and PoliSentiWordNet, generated using the two proposed lexicon generation techniques. A full evaluation was also presented and discussed.

This completes the individual discussions on the three proposed political sentiment mining approaches presented in this thesis. In the following chapter a comparison of the operation of the three approaches is presented.

Chapter 7

Global Comparison Between The Sentiment Mining Approaches

"The best way to show that a stick is crooked is not to argue about it or to spend time denouncing it, but to lay a straight stick alongside it"

Dwight Lyman Moody

This brief chapter reports on an overall (global) comprehensive comparison of the three sentiment mining approaches described in the foregoing chapters: (i) straightforward classification, (ii) using generic sentiment lexicons and (iii) using domain specific sentiment lexicons. Recall that with respect to straight forward classification, six machine learning classifiers were considered: (i) Naïve Bayes, (ii) Support Vector Machine SMO, (iii) J48 decision trees learner, (iv) JRip rule-based classifier, (v) IBk nearest neighbour classifier and (vi) ZeroR (the last as a baseline classifier). The generic lexicon used was the off-the-shelf SentiWordNet 3.0, while the domain specific lexicons used were PoLex and PoliSentiWordNet generated as described in Chapter 6. The comparison reported upon here is based on the individual experimental analysis conducted with respect to each of the proposed approaches as discussed in the foregoinig chapters (Chapters 4, 5 and 6). This chapter concludes the work directed at the first objective (**Objective1**) identified with respect to the work described in this thesis. Namely the application of sentiment mining techniques to predict the *attitude* of individual debaters, whether they are *for* or *against* a motion.

7.1 Comparison

As already noted, for evaluation purpose, standard data mining performance measures
were used: precision, recall and F-measure (see Section 4.3) with respect to both the
"Aye" and the "No" classes in addition to the average of both. The reader should note
that the accuracy with respect to the "Aye" class (true positive rate) is equal to recall
with repect to the "Aye" class (sensitivity) and the accuracy with respect to the "No"
class (true negative rate) is equal to recall with repect to the "No" class (specificity)[1].
Table 7.1 summarises the results discussed previously. The table shows the overall
precision, recall and F-measure values obtained (with respect to both the "Aye" and
the "No" classes in addition to the average of both), and the average accuracy values
obtained. The table also includes minimum, maximum, average and standard deviation
values of the results. From the table it can be observed that best average results were
recorded using the J48 classifier generator which outperformed all the other classifiers
and both the lexicon-based approaches (see Figures 7.1 and 7.2). Inspection of Table
7.1 also indicates that there is no significant difference with respect to the operation of
the first five classifiers with respect to either the "Aye" or the "No" class (not the case
when using lexicons as can be seen from later on in the table).

With respect to the lexicon-based approaches, the results produced using the general
purpose sentiment lexicon (SentiWordNet 3.0), and both domain specific sentiment lex-
icons (PoLex and PoliSentiWordNet), are presented in Table 7.1. Inspection of the re-
sults presented in Table 7.1 indicates that: (i) there is a small improvement with respect
to the average values obtained when using domain specific lexicons (compared to gen-
eral purpose lexicons), where the best domain specific lexicon-based average accuracy
was achieved using PoLex and binary feature vectors (55.464%) and the best generic
lexicon-based average accuracy was achieved using binary feature vectors (54.167%)
and (ii) both PoLex and PoliSentiWordNet produced almost the same results using
term frequency feature vectors. As noted previously all the lexicon-based techniques
worked significantly better with respect to predicting "Aye" (positive) attitudes than
"No" (negative) attitudes. This observation is especially the case with respect to the
recall and F-measure metrics. The reason for this, as also noted previously, is due to
the often overly polite parliamentary jargon used which means that positive sentiment
is easier to identify than negative sentiment. This was a surprising result as domain
specific lexicons are considered to be well suited to predicting "No" (negative) atti-
tudes and thus providing an interesting avenue for further research considering deeper
analysis. Comparing all the results presented in Figure 7.1, it can be seen that the
classification-based approach tends to produce a better prediction than the lexicon-

---

[1]See equations (4.2c) and (4.2d).

based approaches, a best classification-based average accuracy was achieved using J48 (61.751%) compared to a best lexicon-based average accuracy using the PoLex domain specific lexicon coupled with binary feature vectors (55.464%).

	Precision (P)			Recall (R)			F-Measure (F)			Accuracy (A)
	Aye	No	Avg.	Aye	No	Avg.	Aye	No	Avg.	Avg.
J48	0.628	0.601	0.615	0.719	0.497	0.618	0.671	0.544	0.613	61.751%
JRip	0.581	0.529	0.557	0.685	0.418	0.562	0.629	0.467	0.555	56.238%
SMO	0.602	0.531	0.569	0.604	0.529	0.570	0.603	0.530	0.570	56.963%
NB	0.576	0.493	0.538	0.531	0.538	0.534	0.552	0.515	0.535	53.433%
IBk	0.547	0.500	0.525	0.888	0.132	0.541	0.677	0.209	0.462	54.110%
ZeroR	0.541	0.000	0.293	1.000	0.000	0.541	0.702	0.000	0.380	54.110%
SentiWordNet 3.0 (TF)	0.554	0.495	0.524	0.766	0.271	0.518	0.643	0.350	0.497	53.894%
SentiWordNet 3.0 (Binary)	0.546	0.502	0.524	0.909	0.109	0.509	0.682	0.179	0.430	54.167%
PoLex (TF)	0.554	0.503	0.528	0.798	0.241	0.520	0.654	0.326	0.490	54.255%
PoLex (Binary)	0.560	0.534	0.547	0.831	0.229	0.530	0.669	0.320	0.495	55.464%
PoliSentiWordNet (TF)	0.554	0.500	0.527	0.777	0.262	0.520	0.647	0.344	0.496	54.110%
PoliSentiWordNet (Binary)	0.556	0.522	0.539	0.831	0.217	0.524	0.666	0.307	0.486	54.937%
Min	0.541	0.000	0.293	0.531	0.000	0.509	0.552	0.000	0.380	53.433%
Max	0.628	0.601	0.615	1.000	0.538	0.618	0.702	0.544	0.613	61.751%
Average	0.567	0.476	0.524	0.778	0.287	0.541	0.650	0.341	0.501	55.286%
SD	0.026	0.153	0.077	0.131	0.174	0.030	0.040	0.161	0.063	2.286%

Table 7.1: Comparison of evaluation results (attitude prediction performance) obtained, with respect to the three proposed approaches to sentiment mining, using only speech data.

7.2 Summary

This chapter has presented an overall comparison between the three approaches to sentiment mining considered in Chapters 4, 5 and 6 with respect to debater attitude prediction effectiveness in the context of the UK House of Commons political debates. The main findings from this evaluation were that the classification-based approach outperformed the lexicon-based approaches and that there is a distinction between their operation with respect to the "Aye" class and the "No" class (this is particularly pronounced with respect to the lexicon-based approaches). In the context of the first research objective (**Objective1**) identified with respect to this thesis we can therefore conclude that: (i) it is possible to effectively predict the attitude of individual debaters, whether they are for or against a motion within the context of political debates and (ii) the classification-based approach can be chosen as the most appropriate sentiment mining approach to predict the attitude of individual debaters. The next chapter introduces the design and the implementation of the proposed Debate Graph Extraction (DGE) framework.
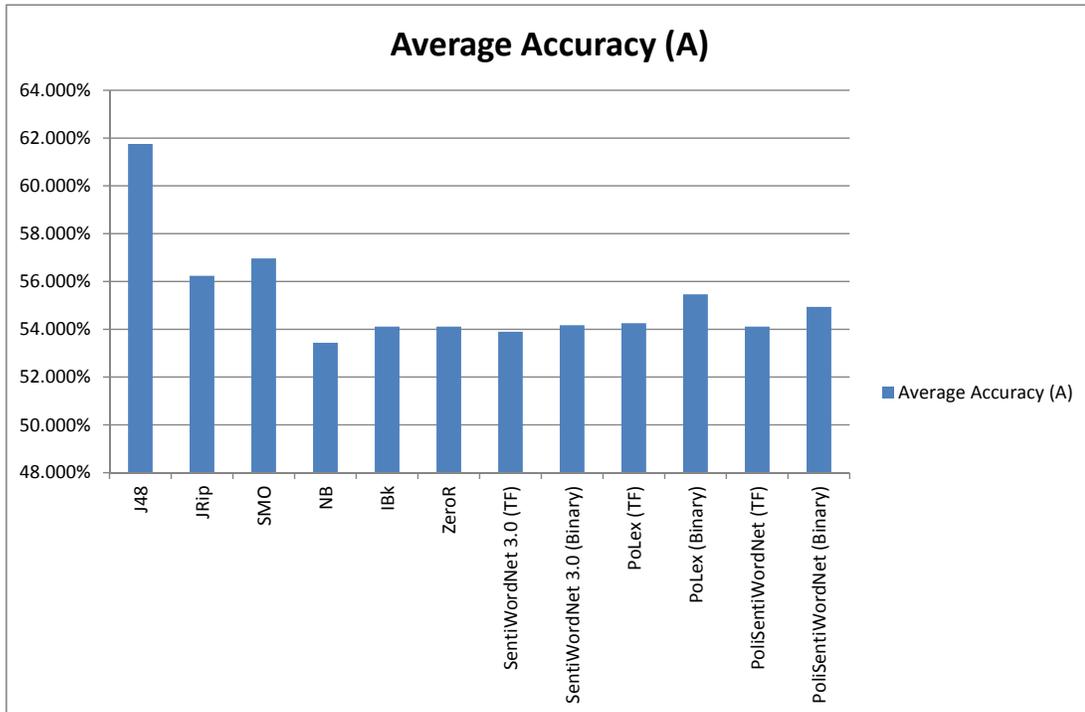
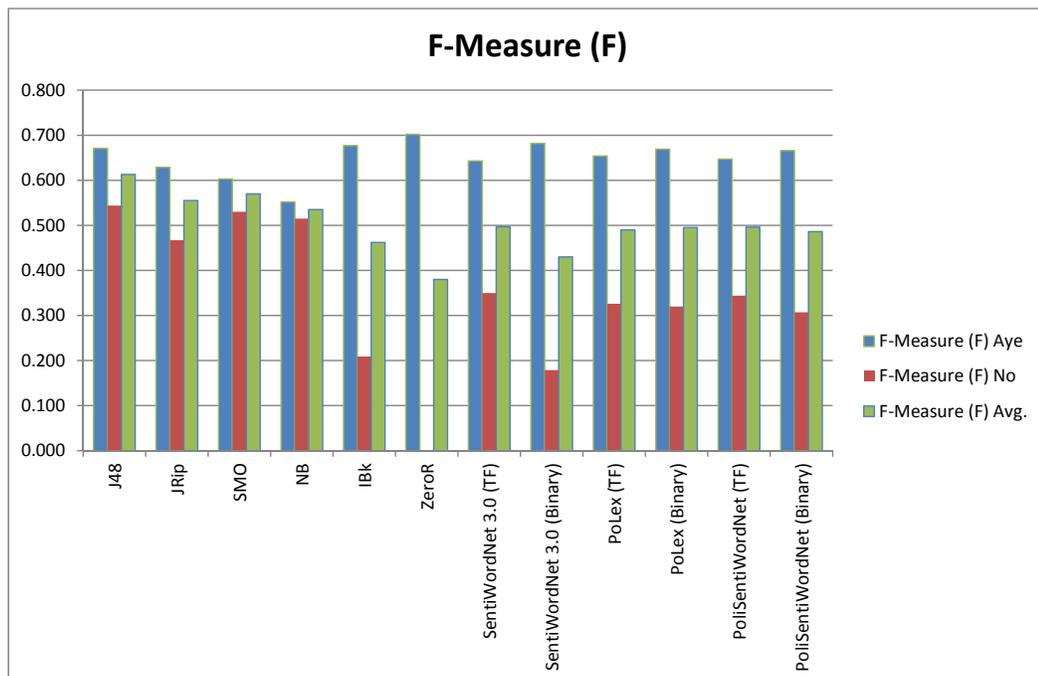Figure 7.1: Bar chart representation for the recorded Average Accuracy (A) values presented in Table 7.1.

Figure 7.2: Bar chart representation for the recorded F-Measure (F) values presented in Table 7.1.

Chapter 8

The Debate Graph Extraction (DGE) Framework

"One look is worth a thousand words."

Frederick R. Barnard

This chapter presents the design and the implementation of the proposed Debate Graph Extraction (DGE) framework and describes how the DGE framework may be used to: (i) extract embedded graph structures from transcriptions of debates and (ii) generate the corresponding debate graphs to allow for graphical visualisation of the high-level structure of such debates. The idea was to represent the structure of a debate as a graph with speakers (concatenated speeches) as nodes and significant interactions (according to either (i) semantic similarity, (ii) interruptions made or (iii) combination of both semantic similarity and interruptions made) between debaters as links. Nodes are labelled with speaker attitude ("positive" or "negative"), and links are labelled as being "supporting" if both nodes (connected by a link) have the same attitude labels (both positive or both negative) or "opposing" if both nodes (connected by a link) have different attitude labels (one is positive and the other is negative). In total three different types of debate graph are considered as follows:

1. Semantic similarity debate graphs where links between speakers (nodes) indicate similarity between their speeches.

2. Interruption debate graphs where links between speakers (nodes) indicate interruptions made between them.

3. Relevant interruption debate graphs where links between speakers (nodes) indicate interruptions with respect to a similar topic (semantic similarity between speeches).

The resulting graphs capture the abstract representation of a debate in terms of two opposing factions exchanging arguments on related content. The work presented in this chapter is directed at **Objective2** defined in Chapter 1. Namely the extraction of debate graphs describing and overviewing political debates from political verbatim transcripts.

As noted earlier in this thesis, two approaches for attitude detection can be identified: (i) sentiment lexicon-based (generic or domain specific lexicon) and (ii) classification based. Both approaches can be used for attitude detection, in order to label the nodes in the debate graph to be generated, and thus two variations of the Debate Graph Extraction (DGE) framework are suggested. Overviews of the two proposed DGE framework variations are presented in Figures 8.1 and 8.2. From the figures, it can be seen that in both versions the input for the DGE framework is a set of concatenated speeches associated with a single debate and the output is a graph representing the structure of the debate. More formally the input to the DGE framework is a set of n concatenated speeches $S = \{s_1, s_2, \ldots, s_n\}$. The output is a graph of the form $G(V, E, L_V, L_E, Sz_V, Th_E, f_{map})$ where: (i) V is a set of n vertices (one per concatenated speech) such that $V = \{v_1, v_2, \ldots, v_n\}$, (ii) E is a set of m edges such that $E = \{e_1, e_2, \ldots, e_m\}$, (iii) L_V is a set of two vertex labels (positive or negative), (iv) L_E is a set of two edge labels (supporting or opposing), (v) Sz_V is a set of vertex sizes such that the size of a vertex reflects the number of edges connected to it (vertex degree), (vi) Th_E is a set of edge thicknesses such that the thickness of an edge between any two vertices reflects the significant interactions between the two vertices and (vii) f_{map} is some mapping function that maps the vertex and edge labels on to vertices and edges. Note that the differences between the two variations (lexicon-based and classification-based) are: (i) the two different approaches used for attitude detection and node labelling and thus (ii) two different types of preprocessing needed to produce the appropriate feature vector representations as required. Whatever the case the DGE frameworks describes a general four phase process as follows:

1. Input text preprocessing (as already noted two different types of preprocessing are considered; these are described in Sub-sections 8.1.1 and 8.1.2)

2. Attitude detection and node labelling (again two approaches for attitude detection described in Sub-sections 8.2.2 and 8.2.1)

3. Edge identification and labelling (described in Section 8.3)

4. Debate graph generation (described in Section 8.4)

8.1 Preprocessing

The input to the DGE framework, as noted in Figures 8.2 and 8.1, is a set of speeches. In terms of text processing each speech can be conceptualised as a document, and in this context each document represents a speaker and contains all the speeches, with respect to a particular debate, of that speaker concatenated together. Given that we have two different approaches for attitude detection each approach required a different type of preprocessing: the nature of the required preprocessing in the context of the lexicon-based approach is presented in Sub-section 8.1.1, while that required in the context of the classification-based approach is presented in Sub-section 8.1.2.

8.1.1 Preprocessing for the sentiment lexicon-based approach for attitude detection and node labelling

The preprocessing phase with respect to the sentiment lexicon-based approach commenced with the conversion of all upper-case alphabetic characters to lower case followed by punctuation mark and numeric digit removal. The next steps were tokenisation, and stop word and domain specific word removal. The following step was to produce a Bag-Of-Words (BOW) representation containing all the remaining words in the document collection (speeches), $BOW = \{t_1, t_2, \ldots, t_{|BOW|}\}$. Each document was then represented by some subset of the BOW. In fact, as can be seen from Figure 8.1, two BOWs were created, $BOW1$ and $BOW2$. As will be seen, $BOW1$ was used for attitude detection (using the lexicon-based approach) and $BOW2$ was used for edge identification, each was generated in a slightly different manner. The generation of $BOW1$ includes a lemmatisation process while the generation of $BOW2$ includes a stemming process[1].

The two bags of words were then used to define two feature spaces from which two sets of feature vectors were generated. As will be seen, one of the two sets of feature vectors was used for attitude detection and thus node labelling, while the other was used in link identification and labelling. The distinction between the two sets of feature vectors, other than that one incorporated lemmatisation and the other stemming, was that the feature vector elements in the first case hold term frequency counts[2] while the elements in the second case hold TF-IDF term weightings[3].

---

[1]See Section 5.2 for more detail about the difference between stemming and lemmatisation and why lemmatisation is used for $BOW1$ and stemming used for $BOW2$.

[2]Term frequency is the number of occasions that the term t_i appears in a document collection.
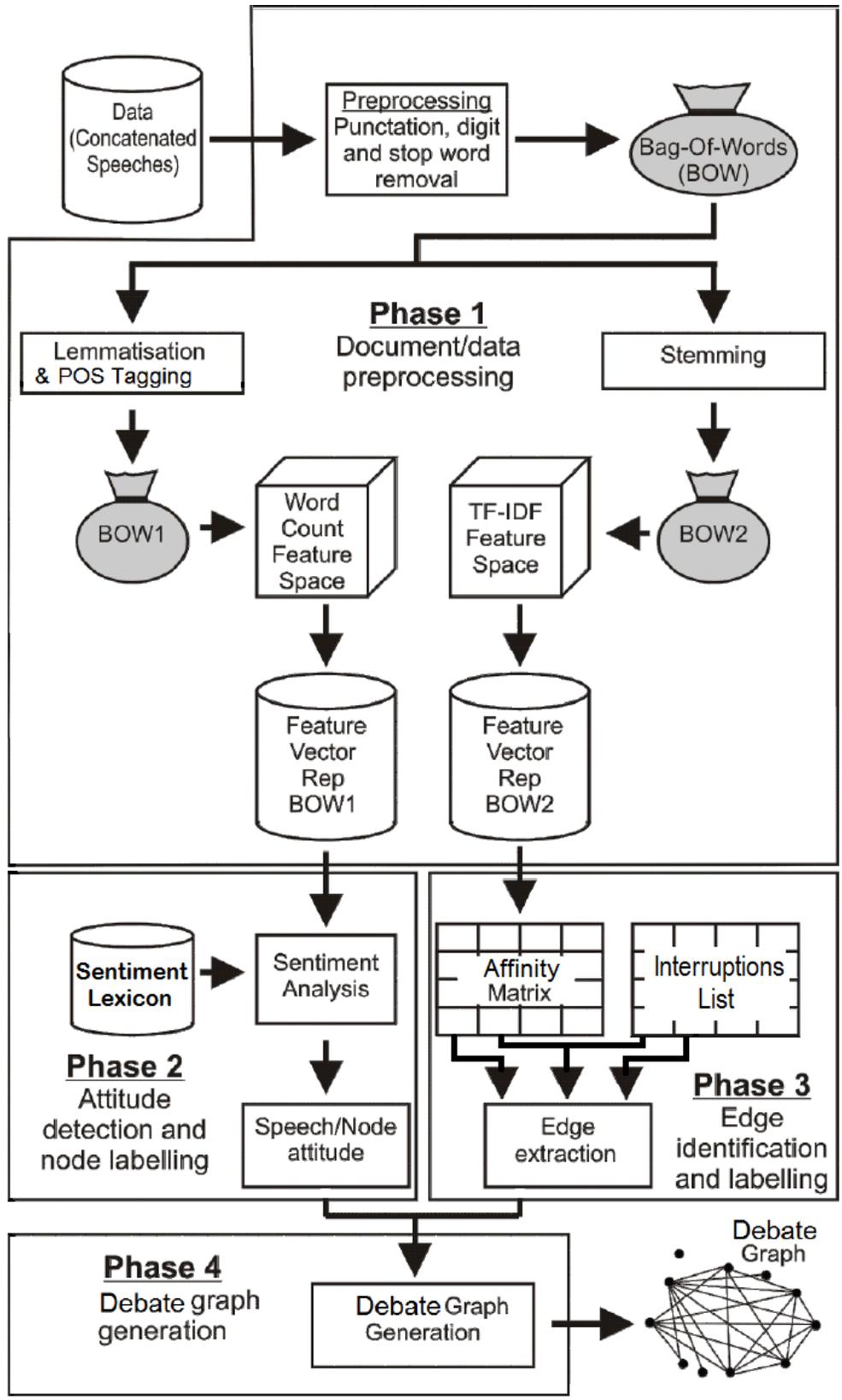
[3]See Equation (4.1).

Figure 8.1: The Debate Graph Extraction (DGE) framework using the sentiment lexicon-based approach for attitude identification.
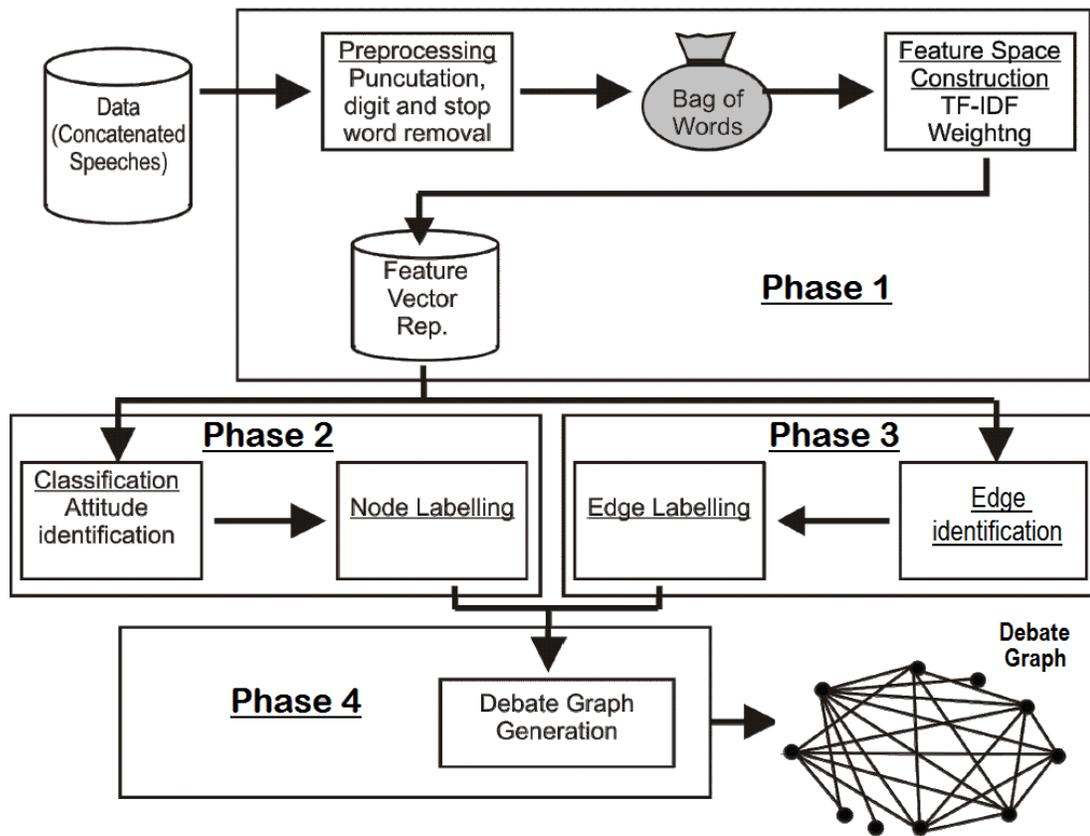
Figure 8.2: The Debate Graph Extraction (DGE) framework using the classification-based approach for attitude prediction.

On completion of the preprocessing phase the input collection of concatenated speeches (documents) were represented as a set of vectors such that each concatenated speech (document) was described by a feature vector. More formally a speech i was represented as a vector $S_i = \{w_{i1}, w_{i2}, \ldots, w_{iz}\}$ where, in the case of $BOW1$, w_{ij} is the occurrence count of term j in speech i, and in the cases of $BOW2$ w_{ij} is the TF-IDF value for term j in speech i. Note that each element in S_i corresponds to a term in either $BOW1$ or $BOW2$ as appropriate. The list of terms associated with feature vector S_i was indicated using the notation $T_i = \{t_{i1}, t_{i2}, \ldots, t_{iz}\}$. Thus we have a set of feature vectors $S = \{S_1, S_2, \ldots, S_z\}$ and a set of term lists $T = \{T_1, T_2, \ldots, T_z\}$ with a one-to-one correspondence between the two.

8.1.2 Preprocessing for classification-based approach for attitude detection and node labelling

The preprocessing phase, with respect to the classification-based approach, commenced in a similar manner to that described for the lexicon-based approach described above. Thus conversion of all upper-case alphabetic characters to lower case, and punctuation mark and numeric digit removal took place. As before, the next steps were tokenisation and then stop word and domain specific word removal. The following stage is to produce a Bag-Of-Words (BOW) representation containing all the remaining words in the document collection (speeches), $BOW = \{t_1, t_2, \ldots, t_{|BOW|}\}$. Each document will then be represented by some subset of the BOW. To reduce the number of individual words to be considered in the BOW stemming was applied. The resulting revised BOW was then used to define feature spaces from which sets of feature vectors were generated. The feature vector elements in this case hold TF-IDF term weightings.

As before, on completion of the preprocessing phase the input collection of speeches was represented using the vector space model such that each speech was described by a feature vector. More formally a speech i was represented as a vector $S_i = \{w_{i1}, w_{i2}, \ldots, w_{iz}\}$ where w_{ij} was the TF-IDF value for term j in speech i. It should also be noted that each element in S_i corresponds to a term in the BOW. The list of terms associated with feature vector S_i was indicated using the notation $T_i = \{t_{i1}, t_{i2}, \ldots, t_{iz}\}$. Thus we have a set of feature vectors $S = \{S_1, S_2, \ldots, S_z\}$ and a set of term lists $T = \{T_1, T_2, \ldots, T_z\}$ with a one-to-one correspondence between the two.

8.2 Attitude detection and node labelling

The nature of the required attitude detection and node labelling was dependent on whether a classification or a sentiment lexicon-based approach was adopted. The fol-

lowing two sub-sections consider each case.

8.2.1 Attitude detection and node labelling using the sentiment lexicon based approach

Once the data has been preprocessed the attitude detection and node labelling phase, using the sentiment lexicons-based approaches, commences in the same manner as described in Section 5.3 (using generic sentiment lexicon) or in Section 6.5 (using domain specific sentiment lexicons). The obtained attitude of each speaker is then used to label the associated node in the graph (recall that each speaker is represented by a node in the debate graph). Algorithm 5.3.1, presented in Chapter 5, described the attitude identification process and thus the node labelling process. Recall that four types of attitude might be identified: (i) positive (for the motion), (ii) negative (against the motion), (iii) objective (no sentiment scores found) or (iv) neutral attitude (sentiment scores add up to approximately zero). With respect to the evaluation that the author has carried out to date it was found that the last two class labels (objective and neutral) were rarely encountered. Whatever the case, objective or neutral attitude nodes will be excluded from the debate graph. The next Sub-section describes the classification-based approach for attitude detection and node labelling.

8.2.2 Attitude detection and node labelling using the classification-based approach

Once the input data has been translated into the desired *BOW* format, whereby each speaker (node in the debate graph) is defined by a subset of words contained in the *BOW*, text classification can be applied to determine each speaker's "attitude" (positive or negative) in the same manner as described in Section 4.2. The predicted attitude of each speaker is then used to label the associated node in the debate graph.

8.3 Link identification and labelling

As noted earlier in this chapter exchanges (Links) between speakers (nodes) can be identified using: (i) the semantic similarity between speakers' concatenated speeches, (ii) the interruptions (interventions) made by MPs during the debate (who interrupted whom) or (iii) the interruptions with respect to a similar topic made by MPs during the debate. Thus three types of debate graph can be generated by conducting the DGE frame coupled with one of the three proposed link identification methods: (i) semantic similarity debate graph, (ii) interruption debate graph and (iii) relevant interruption

debate graph. The following three sub-sections discuss the three methods to identify and establish links between speakers (nodes).

8.3.1 Link identification using semantic similarity

Using the concept of semantic similarity a link between a node pair was established when the speeches (conceptualised as documents) associated with two nodes (speakers) were deemed to be semantically similar. There are different techniques to compute the semantic similarity between documents including: (i) corpus-based similarity computation (using similarity metrics) and (ii) lexicon-based similarity computation (using semantic lexicon like WordNet which is a lexical database for the English language) [T.Sujatha et al., 2012]. For the work described in this thesis the corpus-based similarity computation, which utilise the context or proximity of words to compute semantic similarity, was adopted. More specifically, a text-based similarity measure was used instead of an ontology like WordNet which is difficult and costly because the length of the paths, that link the words in the hierarchies, are irregular and many related words are not in the same hierarchies [Varelas et al., 2005].

There are a number of measures that can be used to determine the similarity between two documents (represented as two feature vectors), such as: the Euclidean or Manhatten distance, the Jaccard measure or the cosine similarity measure [Madylova and Oguducu, 2009]. For the work described in this thesis the cosine similarity measure was adopted because it is simple and very efficient to evaluate sparse text vectors where only the non-zero dimensions need to be considered [Kalaivendhan and Sumathi, 2014]. Cosine similarity between a pair of documents i and j is typically defined as follows:

$$
\begin{aligned}
CosSim(d_i, d_j) &= \frac{d_i \cdot d_j}{|d_i| \times |d_j|} \\
&= \frac{\sum_{k=1}^{k=z} w_{ik} \times w_{jk}}{\sqrt{\sum_{k=1}^{k=z} w_{ik}^2 \times \sum_{k=1}^{k=z} w_{jk}^2}}
\end{aligned}
\tag{8.1}
$$

where (i) d_i and d_j are two z-dimensional feature vectors (representing a pair of documents i and j) over the set of terms (words) $T = \{t_1, t_2, \ldots, t_z\}$ and (ii) w_{ik} and w_{jk} represent the weight of term t_k in the documents i and j.

Cosine similarity is the normalised dot product between two document vectors. The obtained similarity values range between 0.0 and 1.0. A value of 1.0 indicates that the two documents under consideration are identical, and a value 0.0 means that the two documents are entirely unrelated. With respect to the DGE framework similarities between all document (node) pairs are determined by constructing an "affinity" matrix.

This matrix is then used to determine where links exist between nodes. A link between two nodes is deemed to exist if the similarity value is greater than the average of all pair-wise similarities.

8.3.2 Link identification using interruptions

An alternative option to using the semantic similarity between speakers' concatenated speeches to identify and establish the exchanges (links) between debaters is to use interruptions (interventions) made by the individual speakers during the debate (who interrupted whom), as for example considered in [Kaptein et al., 2009] (see Section 2.4). A link between a speaker (node) pair was established when one (or both) of the two speakers (MPs) represented by the end nodes has been interrupted, during a debate, by the other.

8.3.3 Link identification using relevant interruptions

The final link identification method considered was a combination of the semantic similarity and interruption-based approaches. A link between a speaker (node) pair was established when: (i) one (or both) of the two speakers (MPs) represented by the end nodes has been interrupted, during a debate, by the other and (ii) the interruption concerned a "sufficiently similar (relevant)" topic[4]. In other words interruption data was combined with semantic similarity. The term "relevant interruption" indicates that both speakers' speeches should be: (i) sensible (i.e. at least 50 words[5]) and (ii) exceeding some similarity threshold[6].

8.3.4 Link labelling

Links are labelled using the terms "support" and "oppose". The label "support" is applicable if both of the linked nodes have the same attitude, and the label "oppose" is used if they have different attitudes. The algorithm for determining graph links and their labels is presented in Algorithm 8.3.1. Note that the algorithm generates all three types of link (corresponding to the three different types of debate graph considered). The inputs are: (i) the set of feature vectors S^2 and (ii) the list of interruptions made. The output is a graph $XG(V, XE, L_V, L_{XE}, Sz_V, Th_{XE}, f_{map})$ where XE is a linked list in which each item comprises a tuple of the form $\langle start, label, end \rangle$, where $start$ and end are the start and end node identifiers. To indicate the start or end node, or the

---

[4]See Section 8.3 for more detail about determining the similarity between node pairs.

[5]This number was estimated empirically.

[6]This threshold was estimated empirically.

label, associated with a particular link xe_i the notation $xe_i.start$, $xe_i.end$ and $xe_i.label$ is used. In Algorithm 8.3.1 an affinity matrix is calculated in lines 2 to 5, this is then processed in lines 6 to 10 to establish the set of semantic similarity links (SE). The list of interruptions is then processed in lines 11 to 15 to establish the set of interruption links (IE), while the set of relevant interruption links (RE) is established in lines 16 to 23. The link labels are determined in lines 24 to 30 while the link thicknesses and node sizes are determined in lines 31 to 35.

8.4   Debate graph generation

The final phase of the DGE framework comprises debate graph generation. Graph generation is conducted using the outputs from Algorithm 5.3.1 (presented in Chapter 5) and Algorithm 8.3.1 (presented in this chapter), and is fairly straightforward. Although any suitable graph drawing package can be used to visualise the generated result, the author used the NetDraw[7] software tool for visualising social network data [Borgatti, 2002]. Recall that three types of debate graph can be generated: (i) sentiment similarity debate graph, (ii) interruption debate graph and (iii) relevant interruption debate graph. Note that the accuracy of these debate graphs is dependent on the nature of the speakers' attitude detection with respect to the node labelling. The following section shows an illustrative example with respect to each type of debate graph.

8.5   Illustrative example

The process supported by the DGE framework is illustrated in this section using one of the smaller debates contained in the UKHCD-1 dataset, the "Enterprise and Regulatory Reform Bill, Clause 4 - The UK Green Investment Bank: financial assistance" debate, a fragment of which was presented in Figure 3.1 for illustrative purpose. The following three sub-sections show the three types of debate graph generated by conducting the DGE framework coupled with the three methods.

8.5.1   Sentiment similarity debate graph

Applying the DGE framework to the chosen debate the semantic similarity graph presented in Figure 8.3 is generated. With reference to the figure each speaker's concatenated speeches is represented by a node labelled with a speaker-ID (the official MP ID numbers used in Hansard). A square (green) node indicates a positive attitude and a diamond (red) node a negative attitude. The size of a node reflects the number of links

---

[7] https://sites.google.com/site/netdrawsoftware/home

connected to it (node degree). The "thickness" of a link between any two speakers reflects the similarity between the two sets of concatenated speeches (documents) calculated in terms of cosine similarity. Supporting links are indicated using dashed (grey) lines while opposing links are indicated using solid (blue) lines[8]. Figure 8.4 shows an alternative visualisation for the semantic similarity debate graph presented in Figure 8.3. Figure 8.4 uses the name of the speaker to label the nodes, the size of the name indicates the sentiment intensity value (corresponding to the speaker's speech).
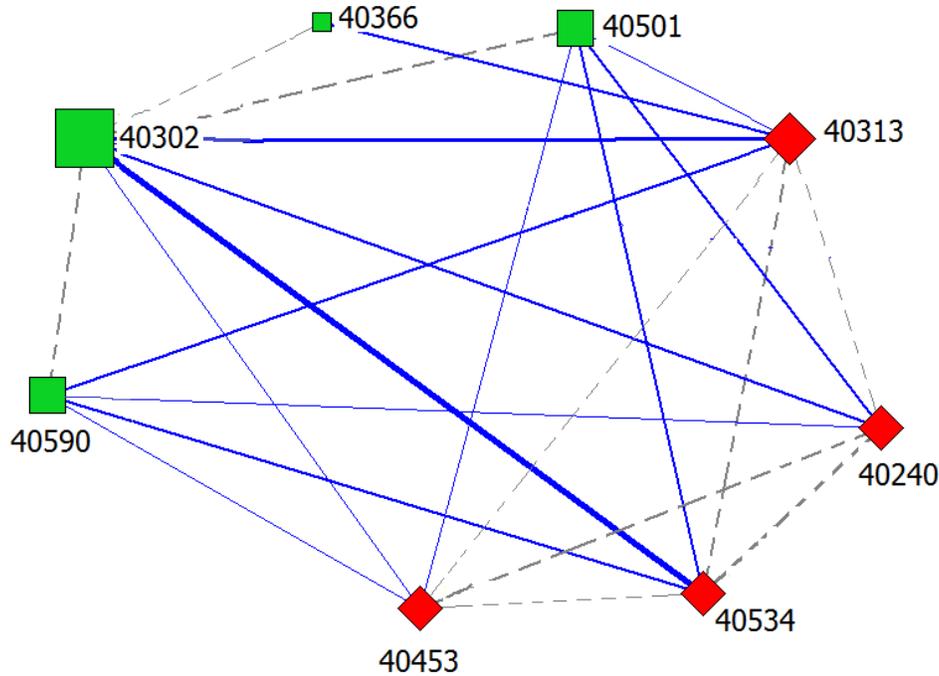

Figure 8.3: Semantic similarity debate graph generated from a UKHCD-1 debate using the DGE framework and semantic similarity data to indicate links.

8.5.2 Interruption graph

Applying the DGE framework to the chosen debate and then using the set of identified interruption links the interruption graph presented in Figure 8.5 is generated. The figure shows the same debate used to generate Figure 8.3 but using the interruptions made by individual speakers to establish links. Note that a square (green) node again indicates a positive attitude and a diamond (red) node a negative attitude. The size of a node reflects the number of links connected to it. The "thickness" of a link between any two speakers indicates the number of interruptions made. Supporting links are indicated using dashed (grey) lines while opposing links are indicated using solid (blue)

---

[8]As such debate graphs may be considered to be structurally similar to argument graphs *à la Dung* [Dung, 1995], except for two key differences: nodes represent speeches instead of arguments; edges are symmetric and of two types (supporting or opposing).
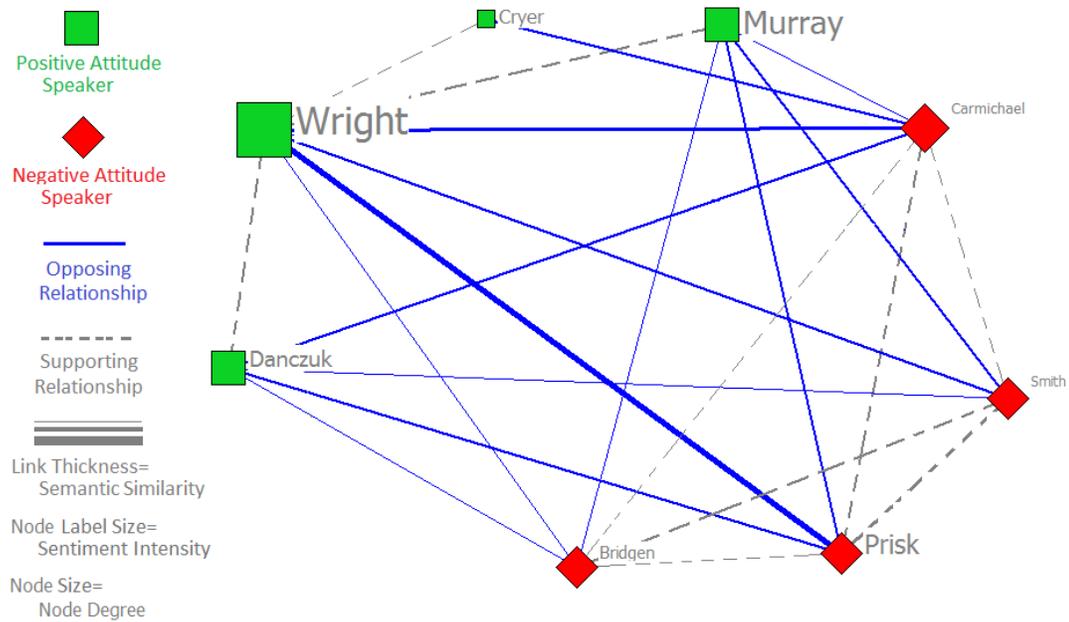
Figure 8.4: An alternative visualisation for the semantic similarity debate graph presented in Figure 8.3 using the name of the speaker.

lines.

8.5.3 Relevant interruption graph

Applying the DGE framework to the chosen debate and then using the set of identified relevant interruption links the relevant interruption graph presented in Figure 8.6 is generated. The figure shows the same debate used to generate Figures 8.3 and 8.5 but using the relevant interruptions made by individual speakers to establish links. As in the case of the two previous example graphs a square (green) node indicates a positive attitude and a diamond (red) node a negative attitude. The size of a node again reflects the number of links connected to it. The "thickness" of a link between any two speakers reflects the similarity between the two speakers' concatenated speeches. Supporting links are indicated using dashed (grey) lines while opposing links are indicated using solid (blue) lines.

8.6 Summary

This chapter has presented the Debate Graph Extraction (DGE) framework in the context of extracting embedded graph structures from transcripts of UK House of Commons political debates. Two variations of the DGE framework (each comprising a four phases process) were described, one founded on the use of sentiment lexicons to
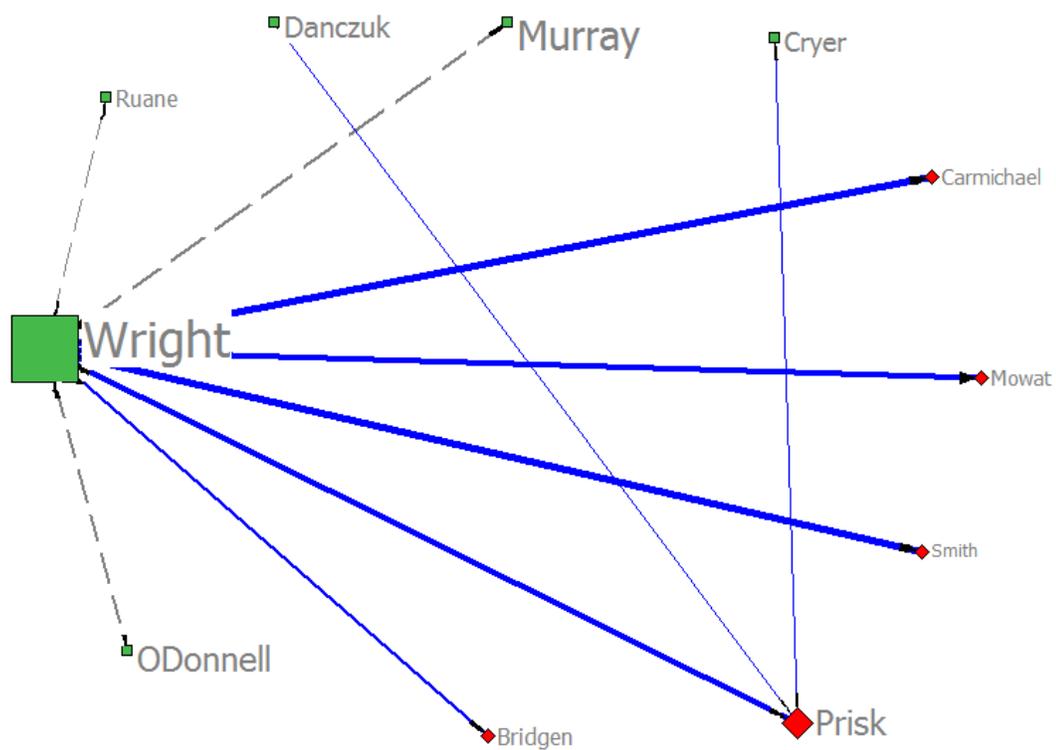
Figure 8.5: Interruption graph generated from a UKHCD-1 debate using the DGE framework and interruption data to indicate links. Link thickness denotes the number of interruptions made.
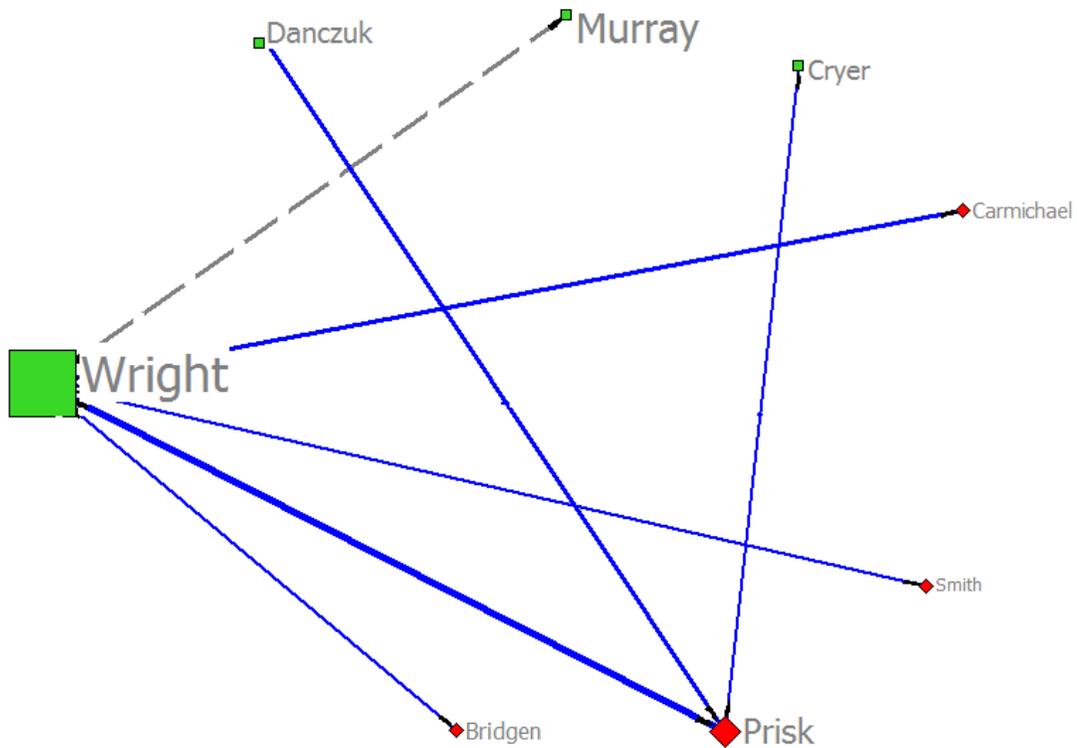
Figure 8.6: Relevant interruption graph generated from a UKHCD-1 debate using the DGE framework and relevant interruption data to indicate links.

determine speaker attitude and one founded on the concept of classification to determine speaker attitude. The operation of the proposed DGE framework was illustrated by applying the framework to a parliamentary debate and consequently generating the associated debate graph. The aim of the work presented in this chapter was to address the second research objective (**Objective2**) identified with respect to this thesis. Given the presented examples it is argued that a successful mechanism has been proposed to extract debate graphs from political debate data. The next chapter discusses mechanisms whereby a debate graph can be conceptualised as a network and thus analysed using appropriate network mathematics and community detection techniques. For this purpose only interruption and relevant interruption debate graphs are considered (although similar analysis could be applied to semantic similarity debate graphs).

Algorithm 8.3.1 Link Identification and Labelling

1: INPUT: Set of feature vectors S^2 and List of Interruptions LOI
2: Initialise $z \times z$ affinity matrix $Affinity$
3: **for all** document pairs $\langle s_i, s_{i'} \rangle \in S^2$, $i < i'$ **do**
4:      $Affinity_{i,i'} = CosineSimilarity(s_i, s_{i'})$
5: **end for**
6: **for all** $Affinity_{i,i'} \in Affinity$ **do**
7:      **if** $Affinity_{i,i'} > average\ similarity$ **then**
8:         add link se_i to SE
9:      **end if**
10: **end for**
11: **for all** document pairs $\langle s_i, s_{i'} \rangle \in S^2$, $i < i'$ **do**
12:      **if** $\langle s_i, s_{i'} \rangle \in LOI \vee \langle s_{i'}, s_i \rangle \in LOI$ **then**
13:         add link ie_i to IE
14:      **end if**
15: **end for**
16: **for all** document pairs $\langle s_i, s_{i'} \rangle \in S^2$, $i < i'$ **do**
17:      $Affinity_{i,i'} = CosineSimilarity(s_i, s_{i'})$
18:      **if** $Affinity_{i,i'} > average\ similarity$ **then**
19:         **if** $\langle s_i, s_{i'} \rangle \in LOI \vee \langle s_i, s_{i'} \rangle \in LOI$ **then**
20:            add link re_i to RE
21:         **end if**
22:      **end if**
23: **end for**
24: **for all** $xe_i \in XE$ **do** [where $x \in \{s, i, r\}$ and $X \in \{S, I, R\}$]
25:      **if** $xe_i.start == xe_i.end$ **then**
26:         $xe_i.label = Support$
27:      **else**$[xe_i.start \neq xe_i.end]$
28:         $xe_i.label = Oppose$
29:      **end if**
30: **end for**
31: **for all** $xe_i \in XE$ **do** [where $x \in \{s, i, r\}$ and $X \in \{S, I, R\}$]
32:      $Th_{xe_i} =$ the significant interactions between $xe_i.start$ and $xe_i.end$
33:      $Sz_{xe_i.start} = VertexDegree(xe_i.start)$
34:      $Sz_{xe_i.end} = VertexDegree(xe_i.end)$
35: **end for**
36: OUTPUT1: Semantic similarity graph $SG(V, SE, L_V, L_{SE}, Sz_V, Th_{SE}, f_{map})$
37: OUTPUT2: Interruption graph $IG(V, IE, L_V, L_{IE}, Sz_V, Th_{IE}, f_{map})$
38: OUTPUT3: Relevant interruption graph $RG(V, RE, L_V, L_{RE}, Sz_V, Th_{RE}, f_{map})$

Chapter 9

Debate Graph Analysis

"Birds of a feather flock together."

Lewis Carroll

Following on from the foregoing chapter this chapter illustrates how debate graphs, as generated using the proposed DGE framework, can be effectively utilised to support various forms of analysis. More specifically how debate graphs can be used to:

1. Investigate whether MPs consistently responded to MPs belonging to a different party and/or MPs that eventually voted in a different way.

2. Investigate whether "communities" in the debate network indicate party affiliation or a voting profile.

The first is specifically directed at interruption and relevant interruption debate graphs and is concerned with the nature of the interruptions made. The second is concerned with the identification of communities within debate graphs and the interpretations that can be accorded to the nature of these communities. Both forms of analysis are rooted in the theory of network analysis [Newman, 2010]. To act as a focus for the work two specific recent debates conducted in the UK House of Commons were considered together with two exemplar analysis question directed at both debates. More specifically in this chapter two types of debate graphs (networks) are considered: (i) interruption networks and (ii) relevant interruption networks (corresponding to interruption and relevant interruption debate graphs of the form described in the previous

97

chapter). The two debates considered are presented in Section 9.1, while the two exemplar questions are discussed in Section 9.2 and further detail of the specific debate graphs (networks) is given in Section 9.3.

The nature of the required network analysis in the context of the two types of analysis considered in this chapter, is presented in Section 9.4. To conduct the first type of analysis it is suggested that the concept of assortativity can be used [Newman, 2010]. This is discussed further in Sub-section 9.4.3. To conduct the second type of analysis network community detection algorithms can be applied as discussed in detail in Sub-section 9.4.4. The overall aim of the work described in this chapter is to address thesis **Objective3** defined previously in Chapter 1, namely the analysis of the embedded graph structures, featured in debate graphs, with respect to how the individual participants (MPs that made at least 50 words length speech during the debate and eventually voted at the end of the debate) interact.

9.1   The debates

As a focus for the work described in this chapter two specific UK House of Commons debates were chosen: the debate held on 18 March 2003 which led to the parliamentary approval for the invasion in Iraq[1], and the debate held on 29 August 2013 which led to the parliamentary refusal of a military intervention in Syria[2]. These debates were chosen because:

1. It was unclear at the outset of the debates whether they would have led to an approval (which happened in the case of the Iraq debate) or a defeat of the Government's motion (which happened in the case of the Syria debate). In other words the outcomes were not expected to be significantly influenced by party doctrine or affiliation, as in the case of many parliamentary debates in the UK House of Commons

2. They are substantial debates in that almost all MPs participated in them.

As in the case of the datasets used for evaluation purposes with respect to the work described earlier in this thesis the transcripts of these debates were readily available at the "They Work For You" www site (www.theyworkforyou.com) and information on the divisions (votes) on the motions raised during the debates can also be accessed

---

[1]For more information on this debate, including its political background, see: http://en.wikipedia.org/wiki/British_Parliamentary_approval_for_the_invasion_of_Iraq

[2]For more information on this debate see: http://www.parliament.uk/business/news/2013/august/commons-debate-on-syria/

online at `www.publicwhip.org.uk`. In the Iraq and Syria debates the MPs cast more than one vote (two divisions) at the end of each debate. The first division was on an amendment, which is a proposed textual changes to motions or bills, on the principal motion which was voted last. The way in which each MP voted in each division defines, what is termed here as, the debate *voting profile*. A voting profile is the combination of the two votes made during the two separate divisions at the end of each debate. For examples, "AyeNo" means the MP voted "Aye" in the first division and "No" in the second while "AbsAye" means the MP *abstained* in the first division and voted "Aye" in the second. Figure 9.1 shows some statistics concerning the voting profiles of MPs with respect to the two debates (six distinct profiles in the Iraq debate and seven in the Syria debate). Note that: (i) an MP has voted twice in the same division (voted with a "Both" vote) provided one vote is "Aye" and the other is "No" to cancel the effect of the vote as a signal of active abstention from the vote and (ii) some profiles are not listed (for example, "AyeAye") as there were no MPs who voted in that manner. The significance of the colour coding is that this is used later in this thesis to differentiate between voting profiles.
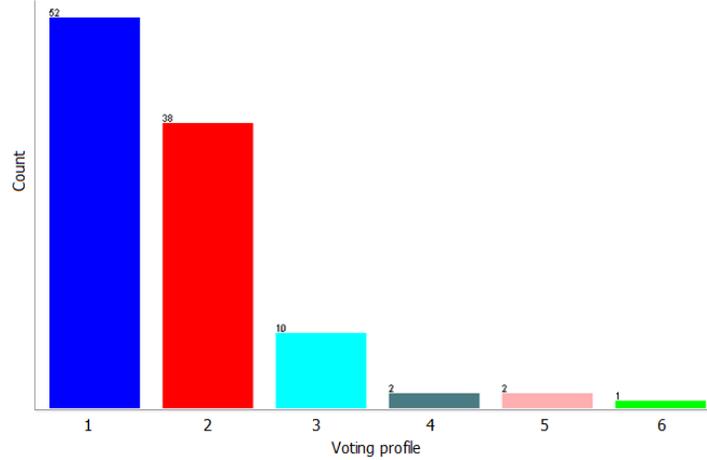
9.2 Exemplar questions

To illustrate the utility of debate graph in the context of the two kinds of analysis identified above the following exemplar questions were considered with respect to the selected parliamentary debate graphs:

Q1: Are speeches by MPs consistently responding to speeches by MPs belonging to a different party and/or speeches by MPs that eventually voted in a different way?

Q2: Do "communities" detected in the debate network manage to single out party affiliation or voting profile, at least roughly?

As already noted above, in the Iraq and Syria debates high level of defection from the party majority line can be observed. For example, in the Iraq debate the majority of Labour MPs (government party) voted against the government motion (against party discipline) and at the same time almost all Conservative MPs (opposition party) voted in favour of the government motion. This is demonstrated in Figure 9.2, which shows the distributions of voting profiles among parties with respect to the Iraq debate and the Syria debate. With reference to the figure each column (bar) represents a party and each coloured portion of a column represent the fraction of the MPs belonging to that party and having the same voting profile as indicated by the colour coding used in Figure 9.1. The absence of a joining between voting pattern and party affiliation is
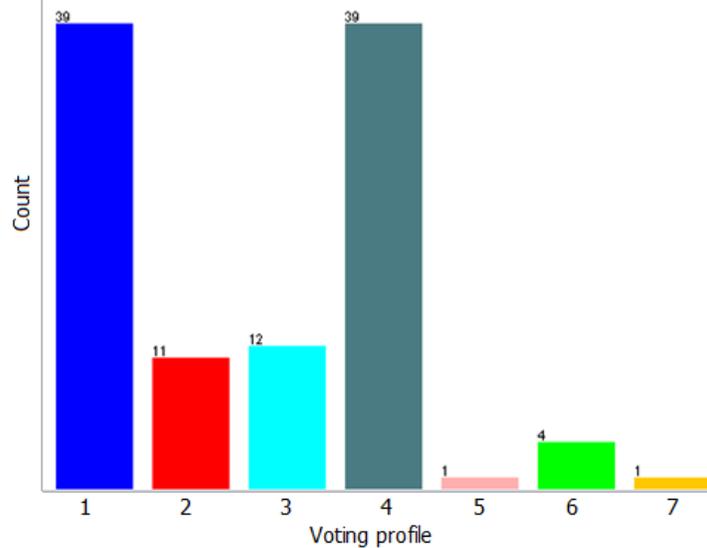
Figure 9.1: Statistics concerning the voting profiles of the MPs speaking during the Iraq debate (top) and the Syria debate (bottom); "Abs" = *abstain* and "Both" = "Aye and No". Screenshots from Weka Visualiser.

particularly distinct with respect to the Syria debate where the motion proposed by the governing coalition (Conservatives and Liberal Democrats) was defeated. The first of the above questions is therefore geared to demonstrating that the proposed mechanism supports the analysis of the interaction between speakers. More specifically checking whether an intuitive feature of the interaction in debates, namely that speeches tend to alternate with one another depending on the opinion they put forth, does indeed show up in the networks.

The second question was selected because it is more exploratory in nature. The question is designed to demonstrate that the proposed mechanism can be used to find joining between sets of nodes selected according to some structural properties of the network (for example communities detected using some community-detection algorithm, or exogenous node labels such as party affiliation and voting profile). This question also connects with one of the general topics of this thesis, namely "predicting" the attitude of individual debaters (vote), whether they are for or against a motion within the context of political debates[3].

9.3   The debate graphs (networks)

As noted in the foregoing chapter, generating a debate graph using the DGE framework can be conducted using: (i) the semantic similarity[4] between speakers' concatenated speeches, (ii) the interruptions (interventions) made by MPs during the debate (who interrupted whom) or (iii) a combination of both semantic similarity and interruptions made to indicate the exchanges (links) between debaters. Thus three types of debate graph (network) were identified: (i) Sentiment similarity network (ii) interruption network and (iii) relevant interruption network. In the context of the two chosen debates (presented in Section 9.1), the associated semantic similarity graphs (networks) were chaotic and almost fully connected[5] (too dense) mesh structures and thus this type of debate networks was not used in the research work described in this chapter. Only the associated interruption networks and relevant interruption networks were considered:

- **Interruption network**: Networks where the nodes represent all MPs that have actively participated in the debate, and the edges represent where one of the two MPs represented by the end nodes has been interrupted during a speech by the other. The reader should note that this type of networks is not necessarily

---

[3]See **RQ1** in Chapter 1.

[4]See Section 8.3 for more detail about determining the similarity between nodes.

[5]It is possible to choose a higher value to threshold the similarity values between the speakers' concatenated speeches to indicate links between the nodes representing the speakers and thus reduce the connectedness.

No.	Label	Count
1	Lab	51
2	Con	37
3	LDem	11
4	SNP	1
5	UUP	2
6	DUP	1
7	PC	2

No.	Label	Count
1	Con	46
2	Alliance	1
3	SDLP	2
4	LDem	10
5	Lab	39
6	Green	1
7	PC	1
8	DUP	3
9	SNP	2
10	Ind	1
11	Res	1

Figure 9.2: Distributions of voting profiles among parties with respect to the Iraq debate (top) and the Syria debate (bottom). Screenshots from Weka Visualiser.
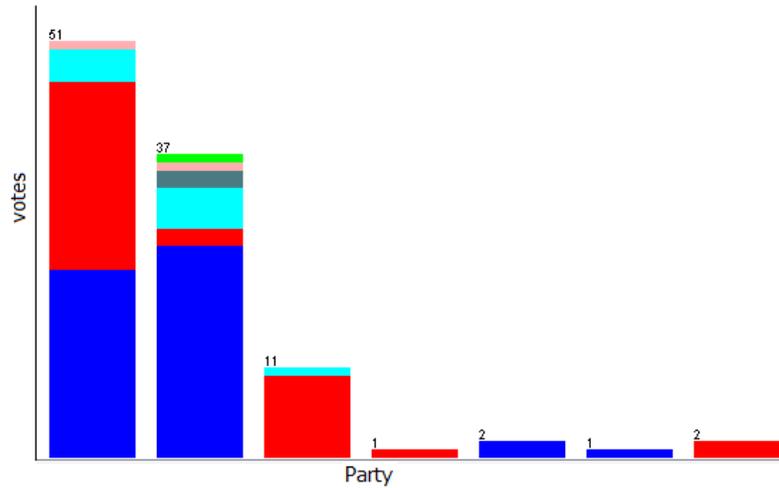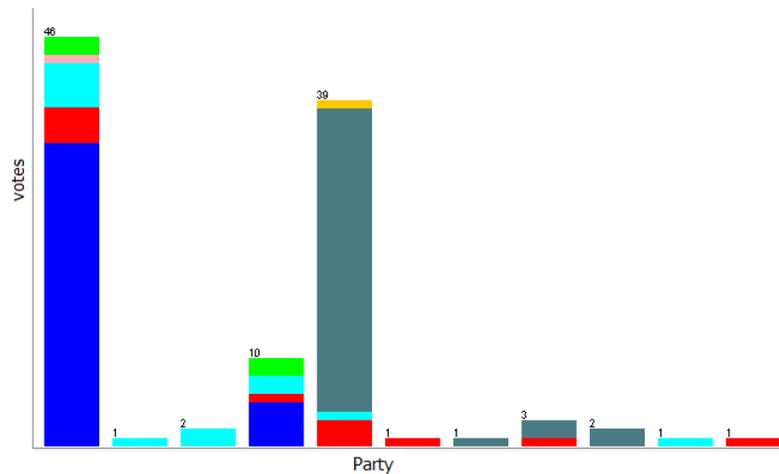
connected.

- **Relevant interruption network**: Networks where the nodes again represent all MPs that have actively participated in the debate, and the edges represent the situation where one of the two MPs has been interrupted during a speech by the other, and the interruption concerned a "sufficiently similar (relevant)" topic. The reader should note that the relevant interruption network is a sub-network of the interruption network. Again, this type of networks is not necessarily connected.

Figures 9.3 and 9.4 present the interruption and relevant interruption networks for the Iraq and Syria debates respectively. The network visualisations were produced using Wolfram Mathematica[6]. From the figures it can be seen that in both types of network the links are not labelled and nodes are labelled, in a slightly different manner than in Chapter 8, with: the name of the node (MP), the party affiliation of the MP and the MP's voting profile which records the voting behaviour of the MP in all the divisions following the debate represented by the network. See for instance Figure 9.1. Party affiliation and voting profile provides useful information, as will be seen below, to understand the patterns of interaction within the debate. The following two sub-sections consider the debates and the associated networks in more detail.

9.3.1 The approval of the invasion in Iraq debate networks

The interruption network for the Iraq debate consisted of 105 nodes, which represented all the MPs that made interruptions in the debate (except for the Speaker of the House), and 159 edges, which represent the interactions among the MPs (who interrupted whom at least once). Figure 9.5 (top) plots the degree distribution of the network. As might be expected highly connected nodes are fewer in number than poorly connected nodes.

The relevant interruption network for the Iraq debate consisted of 89 nodes, which represented all the MPs that made relevant interruptions concerning a sufficiently similar topic in the debate. Disconnected nodes and the nodes representing the Speakers of the House, who do not vote, were not included. The network contains 121 edges, which represent the interactions amongst MPs who "intentionally" interrupt, or are interrupted, at least once. Figure 9.5 (bottom) plots the degree distribution of the network. Again, as expected, highly connected nodes are fewer in number than poorly connected nodes. Note that the colour of the bars in the bar chart has been included for the purpose of clarity, no meaning should be attached to this colouring. A natural question from Network Analysis view is checking whether the degree distribution of a network follows a power-law distribution. In Figure 9.6 the histograms of the
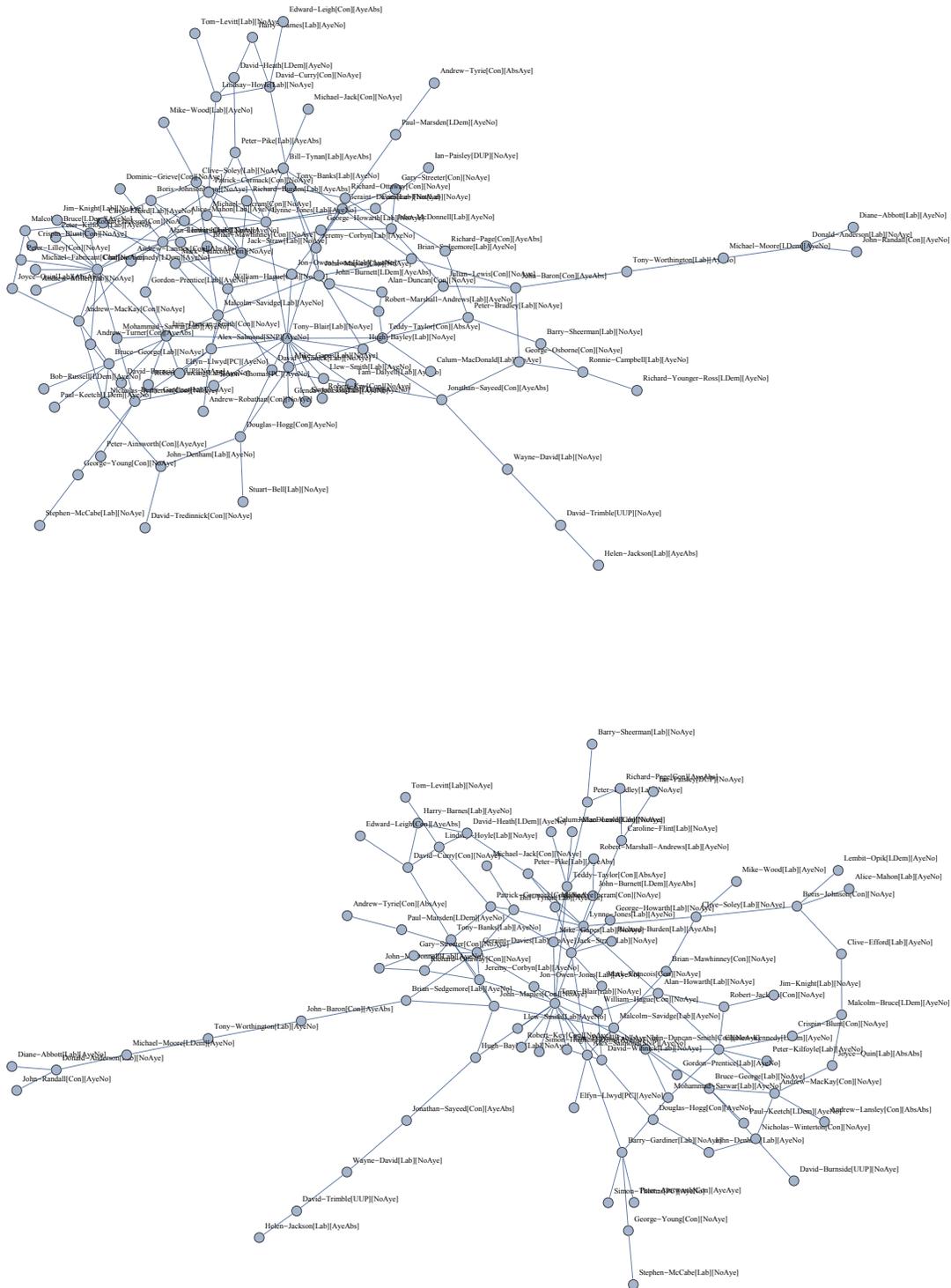
---

[6]http://www.wolfram.com/

Figure 9.3: Interruption network (top) and relevant interruption network (bottom) for the Iraq debate. The visualisations were produced using Wolfram Mathematica.
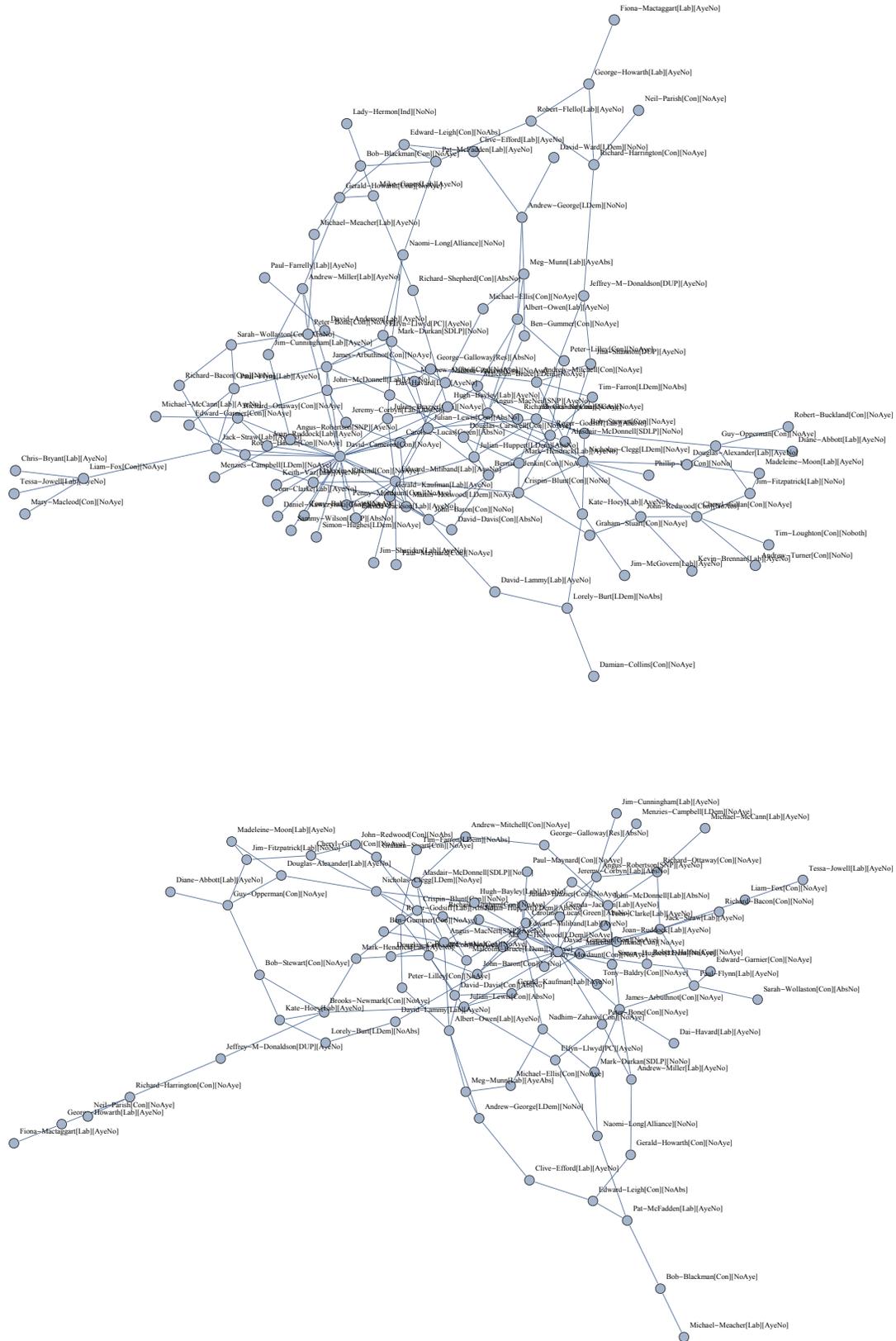
Figure 9.4: Interruption network (top) and relevant interruption network (bottom) for the Syria debate.

degree distributions for the Iraq debate interruption network and the relevant interruption network indicate that the degree distributions of both networks are right-skewed and roughly follow a power-law distribution. The degree distributions (green) are approximated with Zipfian distributions (magenta) with the same range. The Zipfian distribution is a discrete power law probability distribution. Networks with power-law degree distributions are so-called *scale-free* networks [Newman, 2010].

9.3.2 The military intervention in Syria debate networks

The interruption network for the Syria debate consisted of 107 nodes, which represented all of the MPs that participated in the debate (except for the Speaker of the House), and 167 edges, which represented the interactions among the MPs (who interrupted whom at least once). Figure 9.7 (top) plots the degree distribution for the network. As in the case of the Iraq debate highly connected nodes are fewer in number than poorly connected nodes.

The relevant interruption network for the Syria debate consists of 85 nodes, which represent the MPs that participated (made relevant interruptions concerning a sufficiently similar topic) in the debate. Disconnected nodes and the nodes representing the Speaker of the House, who does not vote, were removed. Disconnected sub-networks, if any, containing an insignificant proportion of the nodes (up to three nodes) were also removed. These, the author argues, represent side discussions about minor (marginal) topics. The network contains 129 edges, which represent the interactions among those MPs (thus pairs of MPs where an intentional interruption occurred). Figure 9.7 (bottom) plots the degree distribution of the network. Again highly connected nodes are fewer in number than poorly connected nodes. In Figure 9.8 the histograms of the degree distributions for the Syria debate interruption network and relevant interruption network indicate that the degree distributions of both networks are right-skewed and roughly follow a power-law distribution (i.e. *scale-free* networks). The degree distributions (green) are approximated with a Zipfian distributions (magenta) with the same range. Note that the colouring has been included for the purpose of clarity, no meaning should be attached to this colouring.

9.4 Analysis of debate graphs (networks)

This section presents the two types of network analysis considered in this chapter: (i) Assortativity is introduced in Sub-section 9.4.1 and discussed in detail, in the context of answering question **Q1**, in Sub-section 9.4.3 and (ii) Community structures detection is introduced in Sub-section 9.4.2 and discussed in detail, in the context of answering

Figure 9.5: Degree distributions for the Iraq debate interruption network (top) and relevant interruption network (bottom). Each bar indicates the number of nodes (vertical axis) with respect to a given degree (horizontal axis).

107

Figure 9.6: The power-law degree distributions for the Iraq debate interruption network (top) and the relevant interruption network (bottom). Each bar indicates the fraction of nodes (vertical axis) with respect to a given degree (horizontal axis). The degree distributions of both networks follow approximate power-law form indicated by the green curves, while the magenta curves indicate Zipfian distributions with the same range.
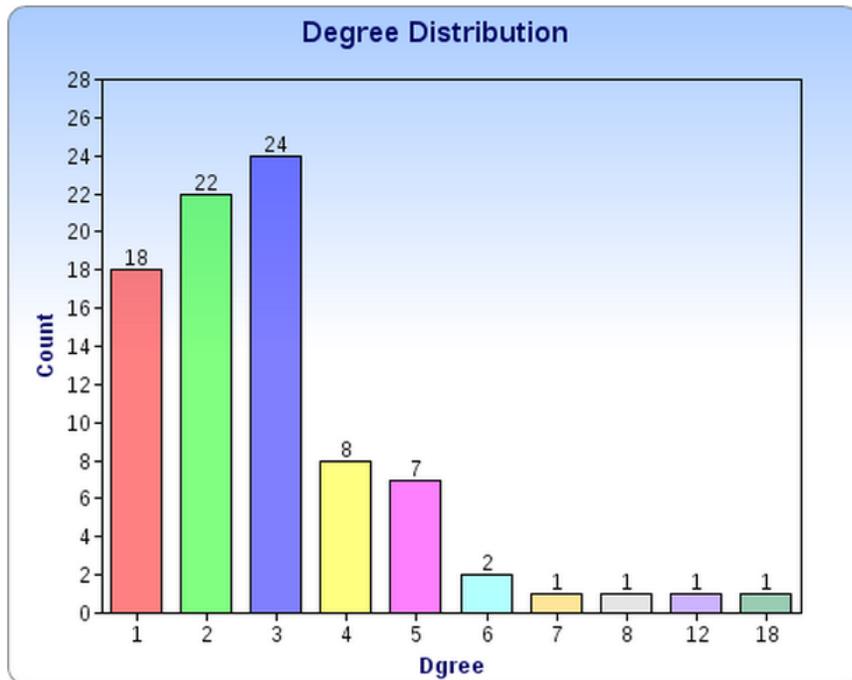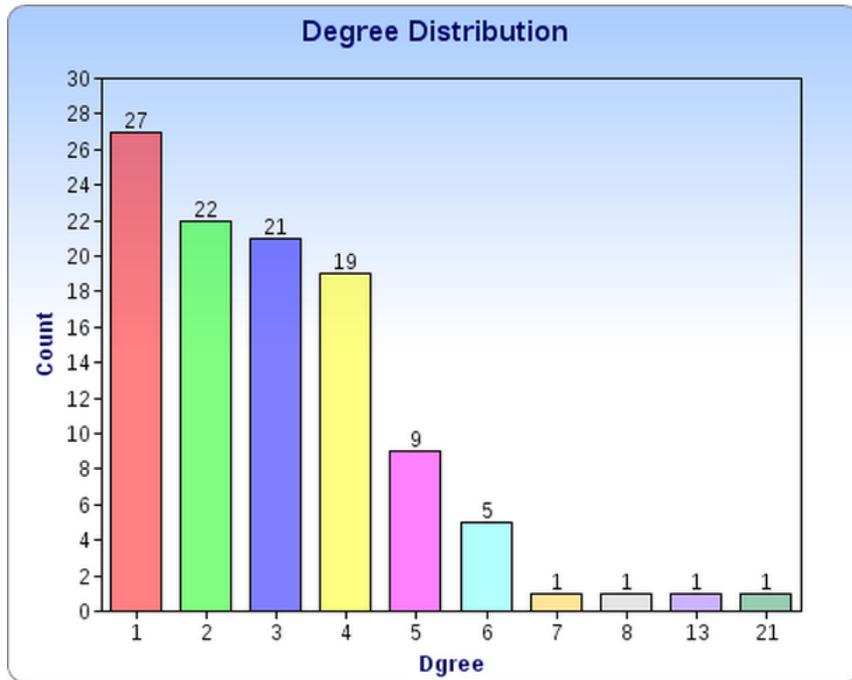
Figure 9.7: Degree distribution for the Syria debate interruption network (top) and relevant interruption network (bottom). Each bar indicates the number of nodes (vertical axis) with with respect to a given degree (horizontal axis).

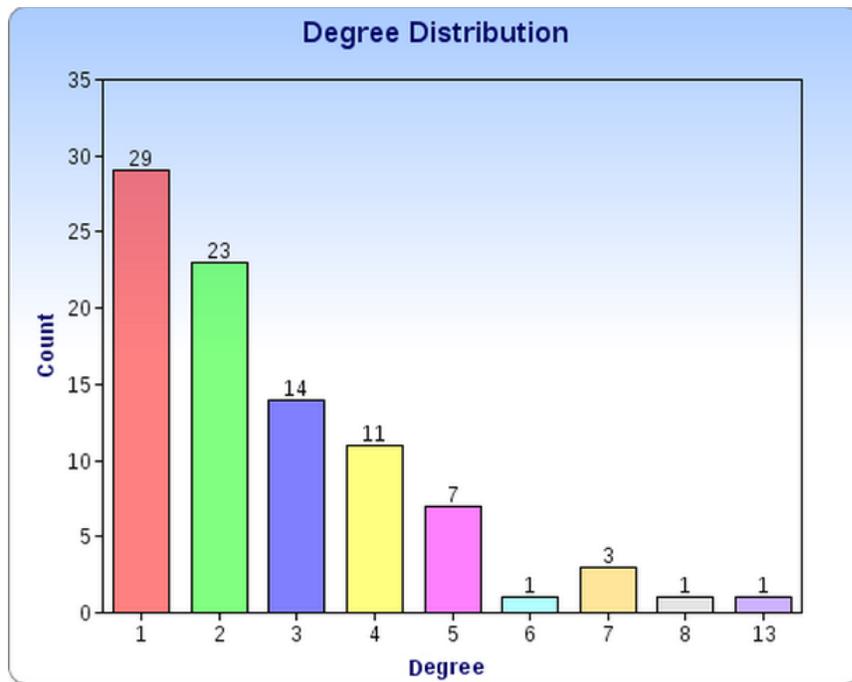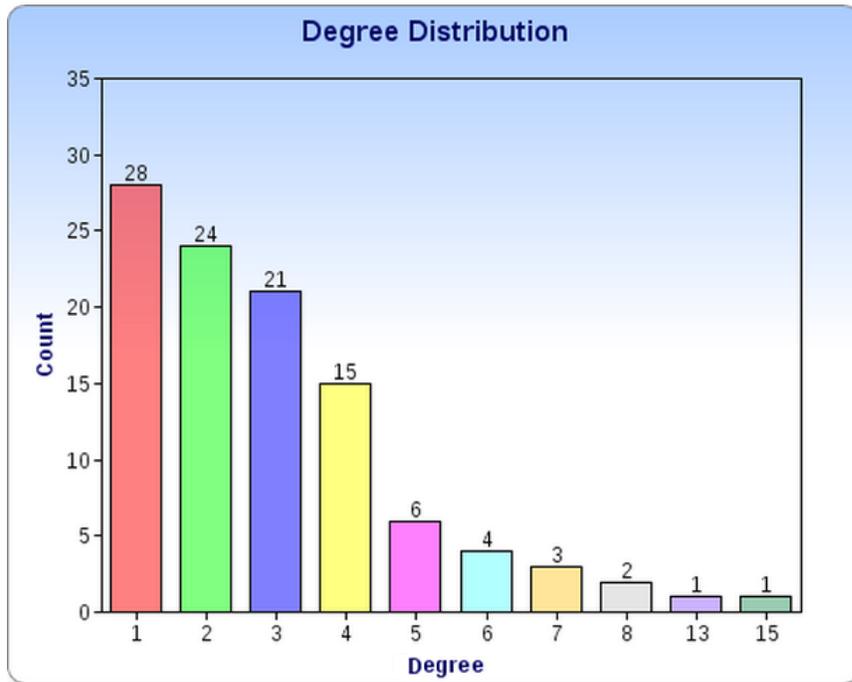Figure 9.8: The power-law degree distributions for the Syria debate interruption network (top) and relevant interruption network (bottom). Each bar indicates the fraction of nodes (vertical axis) with respect to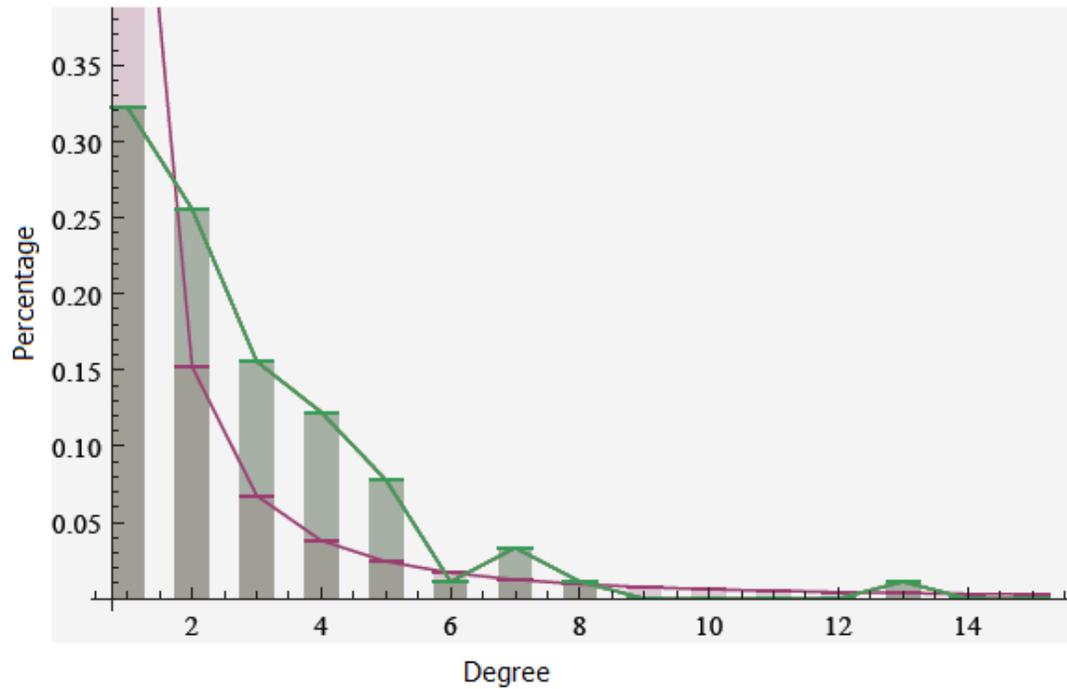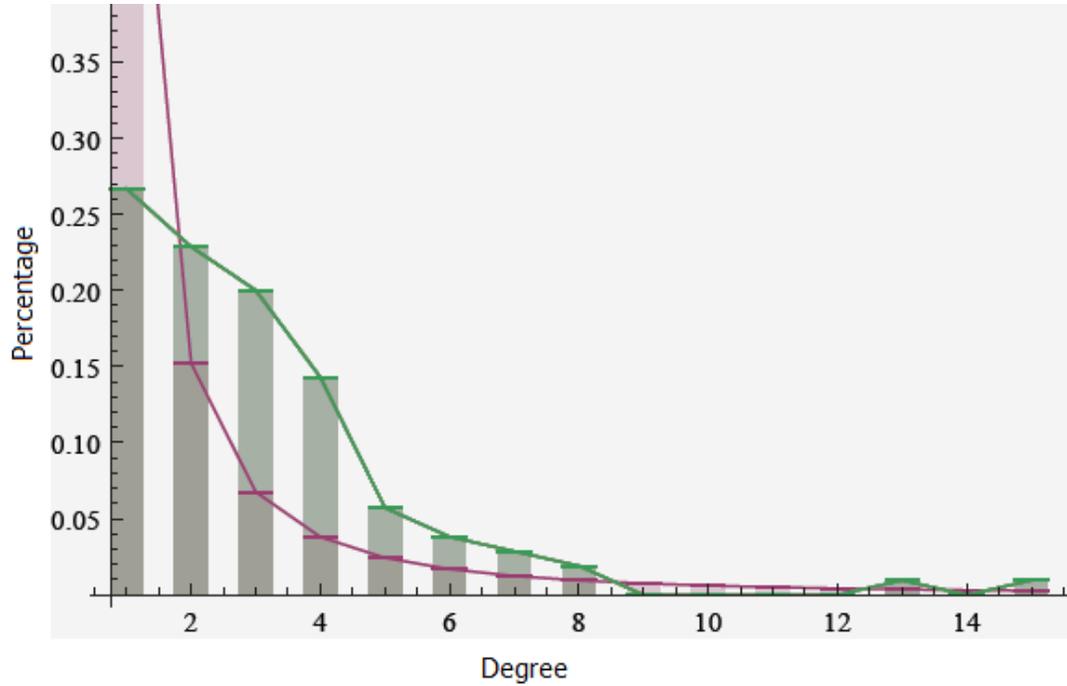 a given degree (horizontal axis). The degree distributions of both networks follow approximate power-law form indicated by the green curves, while the magenta curves indicate Zipfian distributions with the same range.
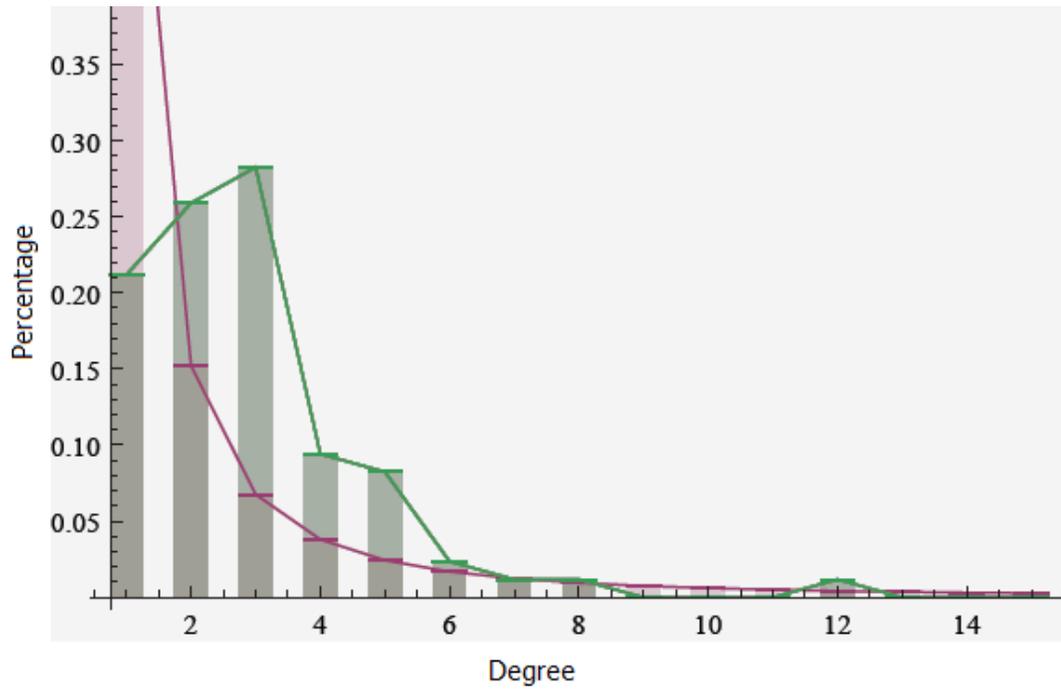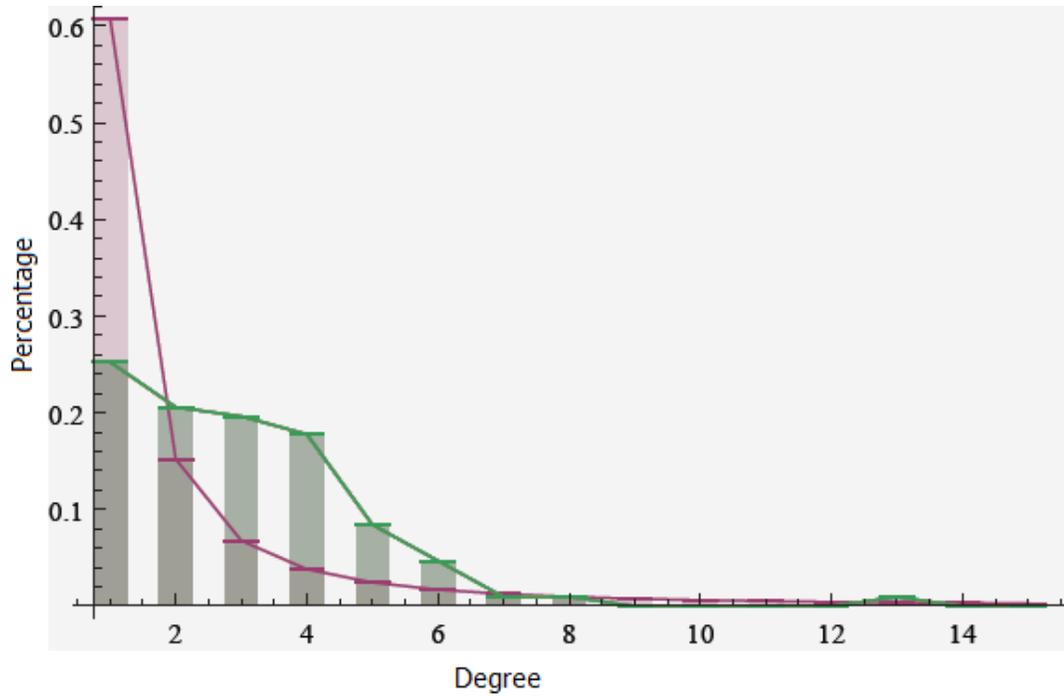
question **Q2**, in Sub-section 9.4.4.

9.4.1 Assortativity

In [Newman, 2010] it was shown that a network is said to be *assortative* (a significant proportion of the nodes in the network that are similar, with respect to some given attribute (label), tend to be connected to each other) if the network is positively correlated according to the value of an *assortativity coefficient*. On the other hand, a network is said to be *disassortative* (a significant proportion of the nodes in the network that are dissimilar, with respect to some given attribute (label), tend to be connected to each other) if the network is negatively correlated according to the value of an *assortativity coefficient*. The assortativity coefficient measures the level of *homophily* of the network based on some node labelling or values assigned to nodes. If the coefficient is high, this means that connected vertices tend to have the same labels or similar assigned values. [Newman, 2002] defined the assortativity coefficient as follows:

$$
r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2}(j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2}(j_i + k_i) \right]^2}
\tag{9.1}
$$

where: j_i and k_i are the degree of the two end nodes of the ith link, and M is the total number of links in the network.

9.4.2 Community structures

Community structures in a network are groups (sets) of nodes (potentially overlapping) and each set of nodes features a dense internal connectivity (within the same group) and sparser connections externally (between groups). Detecting communities within a network can be a computationally difficult task as the number of communities embedded within the network is typically unknown, and of unequal size and/or density [Newman, 2010]. Different methods and algorithms have been developed for community detection. Later in this chapter experiments are reported using five of the most commonly used algorithms for communities detection. The following is a brief description of these algorithms:

- **Modularity maximization**: Modularity is a measure of the quality of clusters calculated by dividing a network into communities (clusters). Networks with high modularity tend to present node groupings with dense connections internally and

sparser connections externally. The modularity maximization algorithm is an op-timisation method for detecting communities by searching all potential divisions of a network and selecting the one whereby the network has the highest possible modularity [Newman, 2010].

- **Hierarchical clustering**: The hierarchical clustering algorithm defines a mea-sure of topological similarity between node pairs (typically the Cosine similarity measure is used) based on the network structure. The most similar nodes are then joined together to form a hierarchy of clusters (communities). Nodes within a cluster have a similarity that exceeds some threshold. Two approaches for hi-erarchical clustering have been proposed: (i) bottom-up (agglomerative) where, at the start, each node represents a cluster and as the algorithm proceeds these clusters are merged; and (ii) top-down (divisive) where, at the start, all nodes are in a single cluster which is then recursively split. The results of hierarchical clustering can conveniently be presented in the form of a dendrogram. The hier-archical clustering process (using the agglomerative approach) can be described as follows [Newman, 2010]:

  1. Assign each node to a cluster of its own (to give a set of "singleton clusters" each containing a single node).

  2. Using some similarity measure determine the similarity between each cluster pair.

  3. Merge the pair of clusters that have the highest similarity into one cluster.

  4. Repeat from step 2 until no more clusters can be merged as determined by some threshold (or one single large cluster has been arrived at).

- **Spectral clustering**: Spectral clustering is a graph-based clustering algorithm that makes use of the eigenvectors of matrix representations of the network such that the clustering problem becomes a graph cut problem designed to isolate sets of nodes from each other [Newman, 2010].

- **Edge centrality**: This algorithm identifies "between-community" links (sparser external connections) in a network and then removes these links thus producing a set of isolated clusters (communities) that feature a dense internal connectivity. Between-community links are identified by measuring the "betweenness central-ity" value for each edge (link); this is a measure of the extent to which an edge lies on the "shortest" path between nodes. The process of the edge centrality al-gorithm for detecting communities can be described briefly as follows [Newman, 2010]:

  1. Calculate the betweenness scores of all edges in the network.

2. Remove the edge with the highest score.

3. Recalculate the betweenness scores.

4. Repeat from step 1 until no more edges can be removed as determined by some threshold (or there are no more edges left to be removed).

Again the results can be usefully presented in the form of a dendrogram.

- **k-Clique percolation**: This algorithm detects communities in a network for the purpose of analysing overlapping community structures. The clique percolation method builds up the communities from k-cliques. A k-clique is a complete (fully connected) sub-network of k nodes. A k-clique at k = 3 is equivalent to a triangle (tricomponent). The basic idea behind this algorithm is that a process of transitive closure of edges can be used to form a clique by filling up a part of a network with edges. The edges required to form a clique are more likely to come from the same component (where the process starts) where that internal edges (within the same community) are expected to have higher density than external edges (between communities). Thus cliques would be "trapped" inside the community they start from. A k-clique-community can be defined as a union of all k-cliques that can be reached from each other through a series of adjacent k-cliques (where adjacency means sharing k−1 nodes) [Palla et al., 2005].

9.4.3 Assortativity: Answering question Q1

As noted above, assortativity is a measure of the similarity of connected nodes with respect to some given attribute and it can be described in terms of an assortativity coefficient. The significance here is that this can be used to answer questions such as exemplar question **Q1**. More specifically **Q1** can be answered by using an assortativity coefficient described in terms of party affiliation and voting profile labels. Table 9.1 presents the assortativity coefficients values, with respect to party and voting profile, using the interruption and relevant interruption networks for both the Iraq and the Syria debates. The results show that all four networks are disassortative with respect to both party and voting profile. These results are discussed further in the following four sub-sections. Sub-sections 9.4.3.1, 9.4.3.2 and 9.4.3.3 present a qualitative inspection of the different values of disassortativity obtained, while Sub-section 9.4.3.4 conclude the results with a statistical test establishing that no significant difference in disassortativity, with respect to party affiliation and voting profile, can be observed across the different types of networks considered in the two chosen debates.

	Party Affiliation	Voting Profile
Iraq Debate		
interruption network	-0.205	-0.264
relevant interruption network	-0.262	-0.174
Syria Debate		
interruption network	-0.089	-0.103
relevant interruption network	-0.147	-0.146

Table 9.1: The assortativity coefficients, with respect to party affiliation and voting profile, for the interruption and relevant interruption networks for both the Iraq and the Syria debates.

9.4.3.1 Disassortativity with respect to party affiliation

With respect to Table 9.1 the relevant interruption network for the Iraq debate has a higher level of disassortativity with respect to party affiliation (-0.262) than the corresponding network for the Syria debate (-0.147), suggesting that in the former debate salient speeches occurred more markedly between members of different parties than in the latter. A similar relationship occurs between the Interruption Networks describing the two debates (-0.205 vs. -0.089). Overall the Syria debate contained many inter-party interruptions, while the Iraq debate appears to be much more "polarised" in term of party affiliation. This may be linked, the author conjectures, to the fact that in the Iraq debate the motion moved by the government (the majority party) was indeed accepted by the House of Commons, suggesting that, after all, the majority position was identifiable within the majority party.

It is interesting to note that the levels of disassortativity for both debates increases when moving from the interruption networks to the relevant interruption networks. That is, when we focus on relevant interruptions, it looks like the responses of MPs to each other's speeches may have alternated between MPs of different parties.

9.4.3.2 Disassortativity with respect to voting profile

With respect to Table 9.1 the interruption networks appear to be more disassortative with respect to the voting profiles of the MPs than party affiliation, and again the Iraq debate appears to consist of less speeches responding to speakers with the same voting profile (the coefficient is -0.264 in the interruption network) than in the Syria debate (-0.103). So, again, the Iraq debate appears to be more "polarised". Moving to the relevant interruption networks a similar pattern was identified, although the coefficient are now very close: -0.174 (Iraq) and -0.146 (Syria). So it still seems that the Iraq debate is more "polarised" with respect to voting profile. However, when salient exchanges

only are considered then the two debates become more closely disassortative.

Unlike in the case of the Iraq debate, here we notice that disassortativity decreases as we move from the interruption network to the relevant interruption network. Thus it is concluded that when we focus on relevant interruptions it looks like the responses of MPs to each other's speeches tend to serve to differentiate less well between voting profiles (the responses of MPs to each other's speeches may have alternated between MPs with the same voting profile).

9.4.3.3 Disassortativity in interruption vs. relevant interruption networks

It is also interesting to observe that in the Iraq debate, the relevant interruption networks are more disassortative with respect to party affiliation than with respect to voting profile (-0.262 and -0.174), while in the Syria debate they are very close (-0.147 and -0.146). That is, party affiliation seems a slightly worse predictor, in the Syria case, of differences in opinion (when focusing on relevant interruptions). It may be suggestive to take this as an indicator of the key difference between the two debates: in the Iraq case the motion of the majority party was accepted by the Chamber, while in the Syria case it was rejected.

A different trend can be observed with respect to interruption networks. Disassortativity increases in both debates when moving from party affiliation to voting profile: Iraq, from -0.205 to -0.262; Syria, from -0.089 to -0.103. This may be linked to the fact that, for both debates, it was unclear at the outset how a critical number of MPs would vote, so speeches may have alternated between MPs of the same party but with different voting intention.

9.4.3.4 Disassortativity significance testing

Table 9.1 summarises the assortativity coefficients values, with respect to party and voting profile, using the interruption and relevant interruption networks for both the Iraq and the Syria debates. With respect to the results presented in Table 9.1 the "t-test", also known as Welch's t-test, significance testing was applied and presented in this sub-section. The "t-test" was applied to checking whether there was a statistically significant difference in the disassortativity (negative assortativity coefficients) values with respect to party and voting profile, using the interruption and relevant interruption networks for both chosen debates. Figure 9.9 presents a box and whisker plot showing where the middle of the assortativity coefficient values lie, while Table 9.2 presents the results obtained by applying the "t-test" significance testing where: (i) Mean: is the central value of a discrete set of values and may be calculated as the sum of the

values divided by the number of values, (ii) Variance: is a measure of how far a set of values is dispersed around the mean and from each other, (iii) Observations: indicates the number of paired observations (four rows represented the four different networks) that were made on the two samples (two columns represented party affiliation and voting profile), (iv) Hypothesised Mean Difference: this is normally equal to zero for a hypothesis test of paired data such that the test will calculate the probability of obtaining the given results by chance assuming that there was no actual difference between the population means, (v) df: is the degree of freedom which refers to the number of independent observations in a set of data, (vi) t-Stat: the t-test produces a single value (t-Stat) which is calculated from the data as a ratio of the departure of an estimated parameter from its notional value and its standard error and (vii) P value is calculated from the t-Stat value and it is typically defined as the probability of obtaining a t-value that is at least as big as the one observed. α is the level of significance which is a value for which a P value less than or equal to $\alpha = 0.05$ is considered statistically significant [Fay and Gerow, 2005].

	Party Affiliation	Voting Profile
Mean	-0.176	-0.172
Variance	0.006	0.005
Observations	4	4
Hypothesised Mean Difference	0	
df	6	
t-Stat	-0.079	
P(T ≤ t)	0.470	

Table 9.2: The outcomes from applying the "t-test" significance testing (two-sample assuming unequal variances and $\alpha = 0.05$) with respect to party and voting profile, using the interruption and relevant interruption networks for both the Iraq and the Syria debates.

From Table 9.2, it can be noted that the calculated P value (0.470) is not less than or equal to the significance level $\alpha = 0.05$ and thus the null hypothesis (means are the same) cannot be rejected indicating that there is no statistically significant difference between the disassortativity values with respect to party and voting profile, using the interruption and relevant interruption networks for both chosen debates.

9.4.4 Community detection: Answering question Q2

In order to answer **Q2** it was first necessary to investigate and identify communities within the network data. A number of mechanisms for achieving this were identified earlier in Section 9.4.2 above: (i) modularity maximization, (ii) hierarchical cluster-
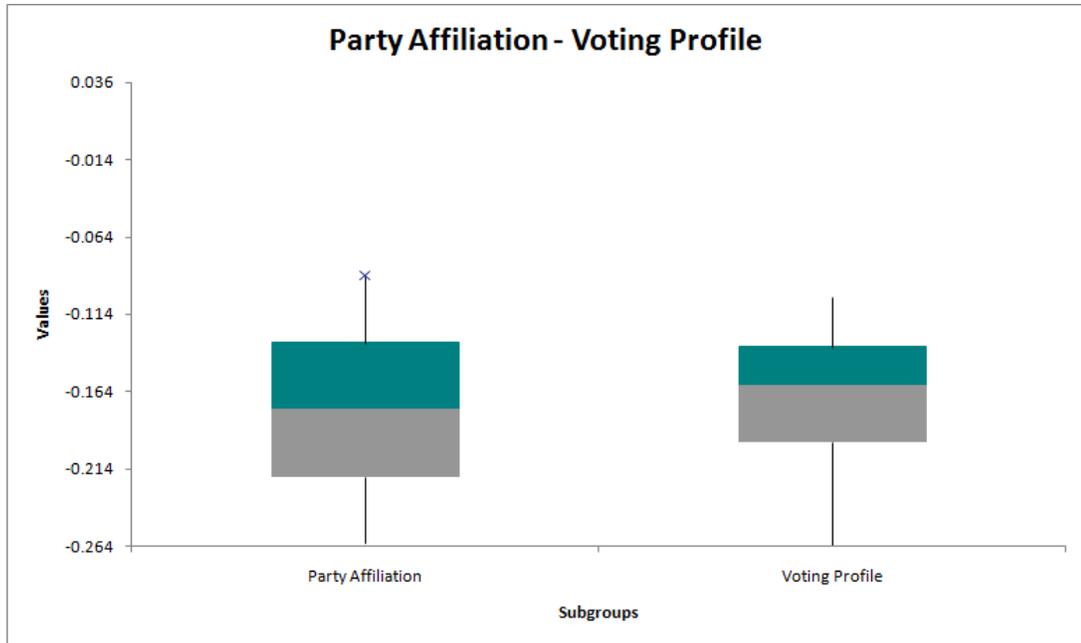
Figure 9.9: A box and whisker plot shows where the middle of the assortativity coefficients values, with respect to party and voting profile, lie.

ing, (iii) spectral clustering, (iv) edge centrality and (v) k-Clique percolation. The comparison was conducted by comparing the nature of communities identified using the community detection algorithms and the "known" communities defined in terms of either the known voting patterns or the known party affiliation. None of the detected communities presented a close match to the known (voting pattern) communities. More specifically it was found that, with respect to the chosen debate networks, none of the detected communities contained members of the same party or with the same voting profile.

Figures 9.10 and 9.11 present a comparison between the known communities (identified according to voting profile) and the detected communities (clusters predicted using the adopted community detection algorithms) for both the interruption and relevant interruption networks with respect to the Iraq debate; while Figures 9.12 and 9.13 present the same comparison with respect to the Syria debate. In each case the top left histogram presents the known communities identified from voting patterns. Note also that in each case the colour coding is defined in the top left histogram. From the figures two observations can be made: (i) all the community detection algorithms found more communities than the number of voting profiles and (ii) all detected communities contained comparable shares of MPs with the same voting profile.

On the other hand, Figures 9.14 and 9.15 present a comparison between the known communities (identified according to party affiliation) and the detected communities

117

Figure 9.10: Communities in the interruption network for the Iraq debate. The top left histogram presents the known communities, according to voting profile, while the rest of histograms present the predicted communities using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms. The colour coding is defined in the top left histogram.

118

Figure 9.11: Communities in the relevant interruption network for the Iraq debate. The top left histogram presents the known communities, according to voting profile, while the rest of histograms present the predicted communities using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms. The colour coding is defined in the top left histogram.

119

Figure 9.12: Communities in the interruption network for the Syria debate. The top left histogram presents the known communities, according to voting profile, while the rest of histograms present the predicted communities using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms. The colour coding is defined in the top left histogram.

120
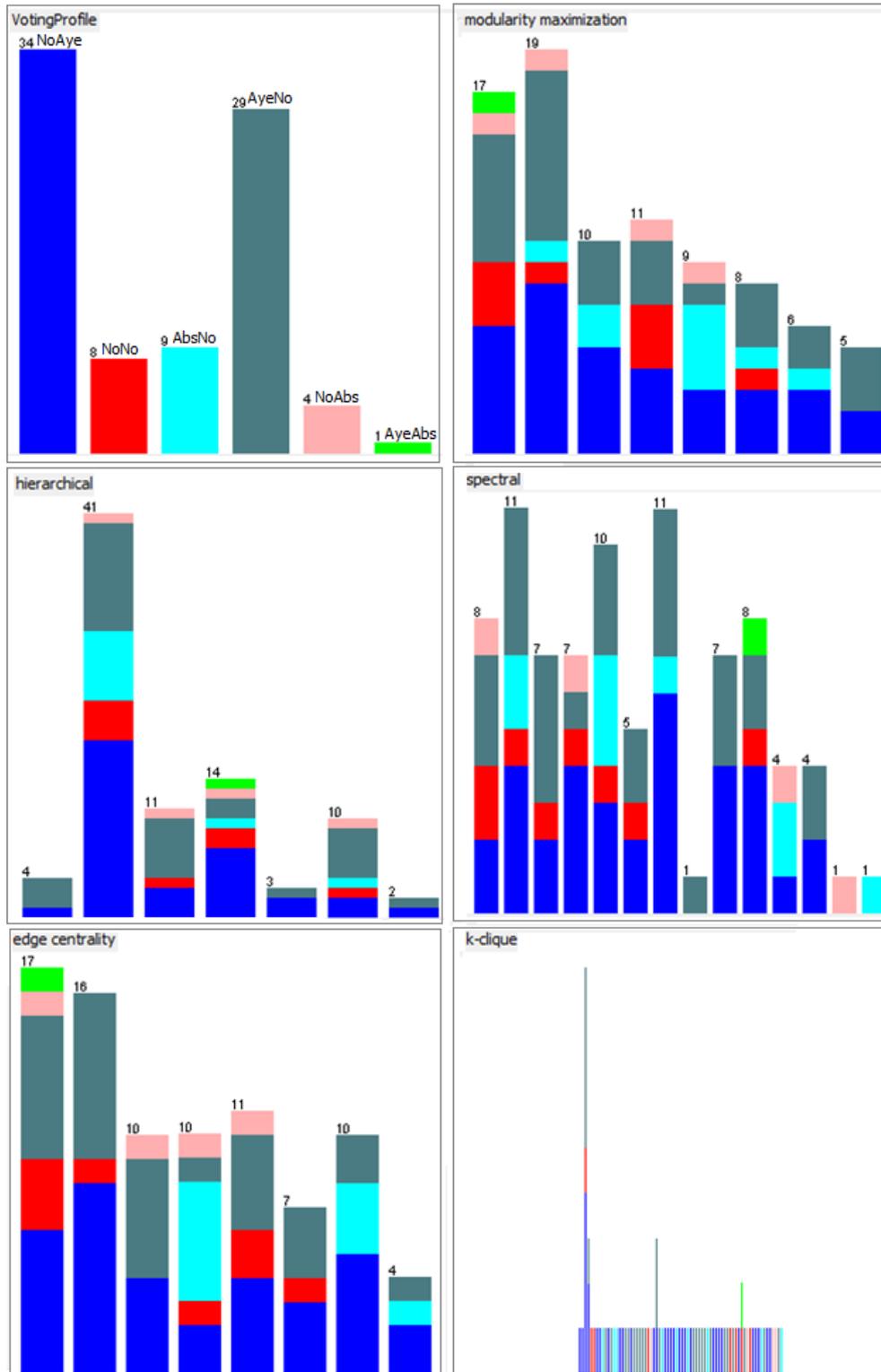
Figure 9.13: Communities in the relevant interruption network for the Syria debate. The top left histogram presents the known communities, according to voting profile, while the rest of histograms present the predicted communities using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms. The colour coding is defined in the top left histogram.
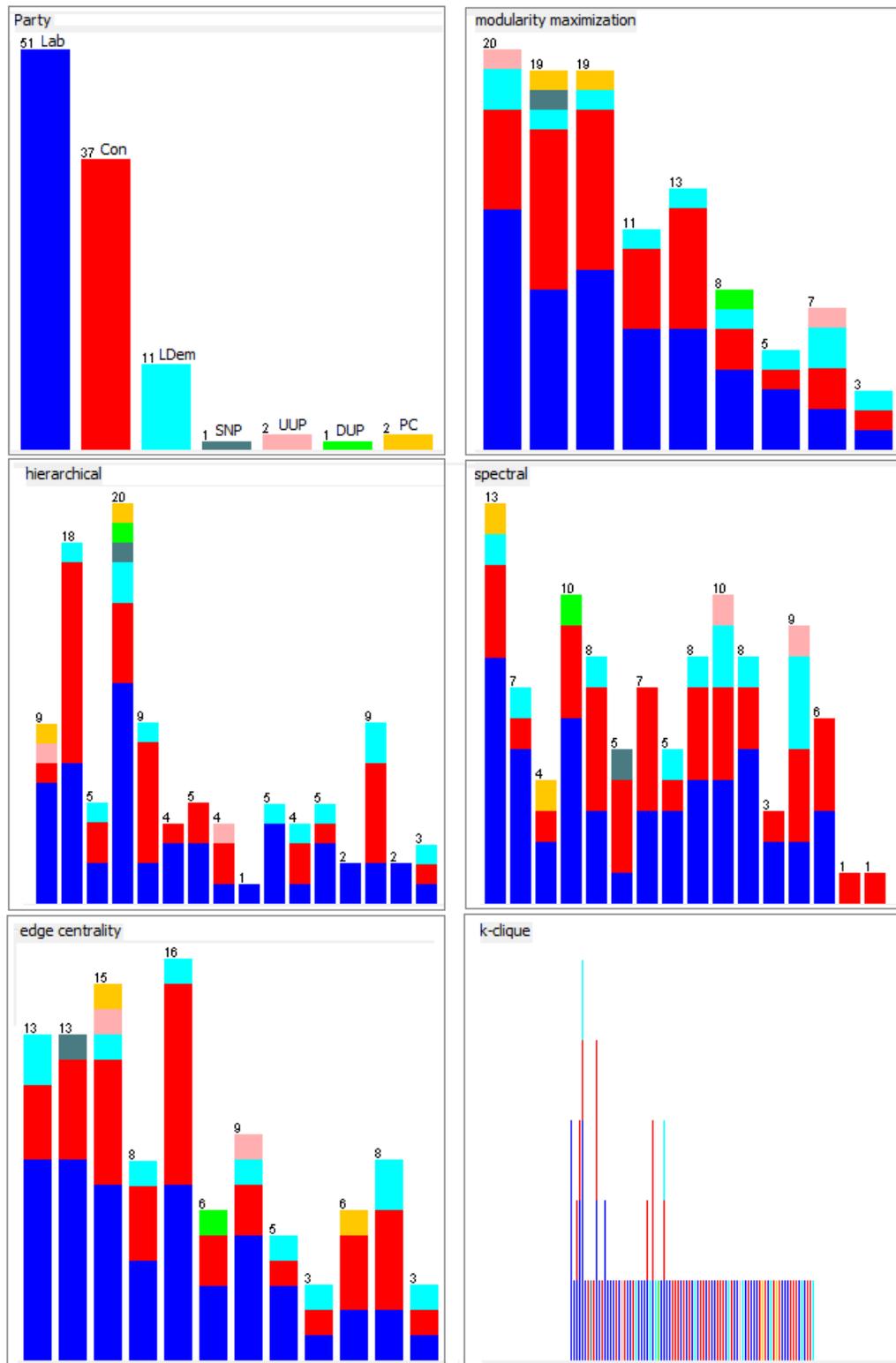
Figure 9.14: Communities in the interruption network for the Iraq debate. The top left histogram presents the known communities (according to party affiliation) while the rest of histograms present the detected communities (clusters predicted using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms). The colour coding is defined in the top left histogram.
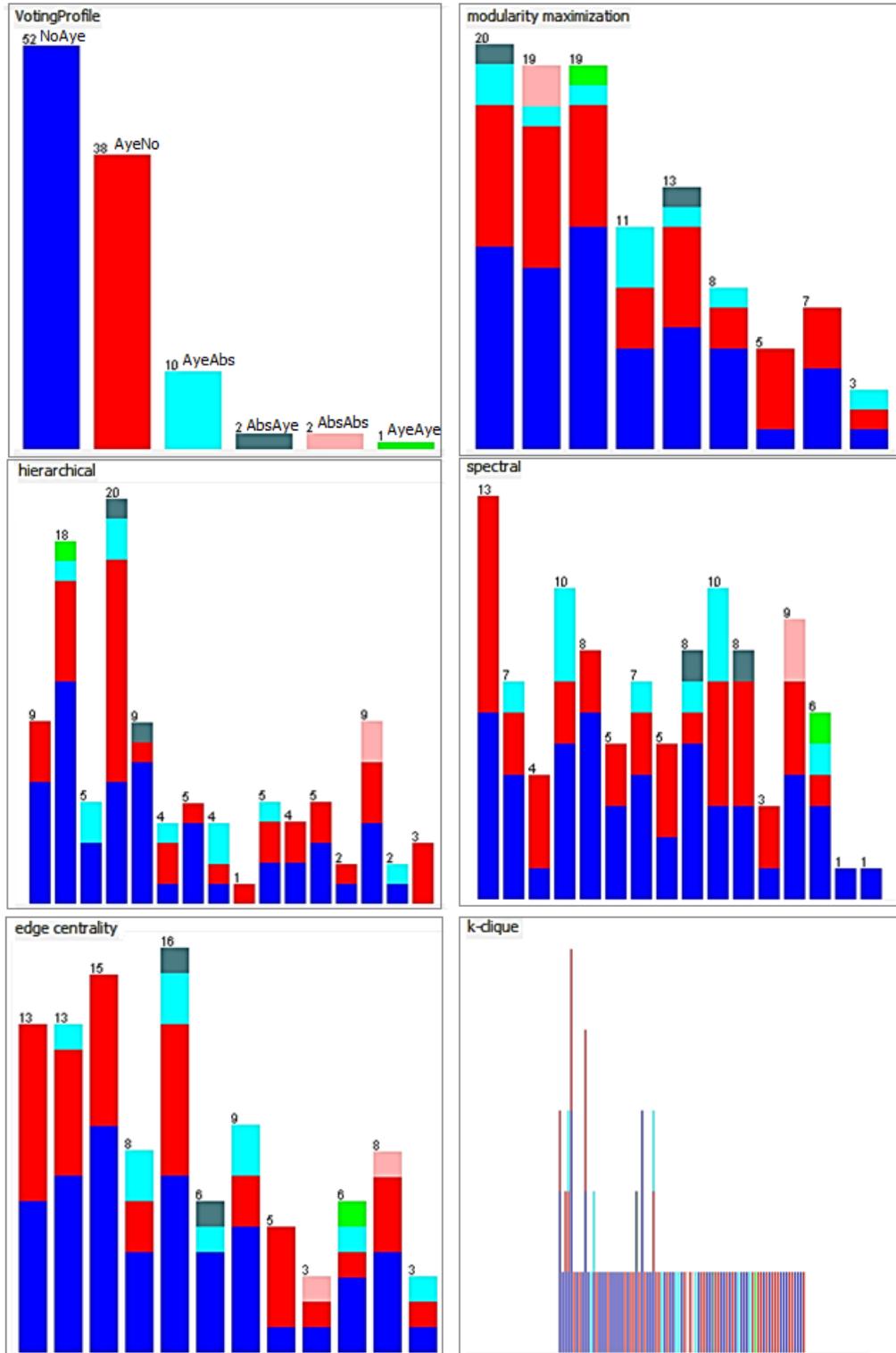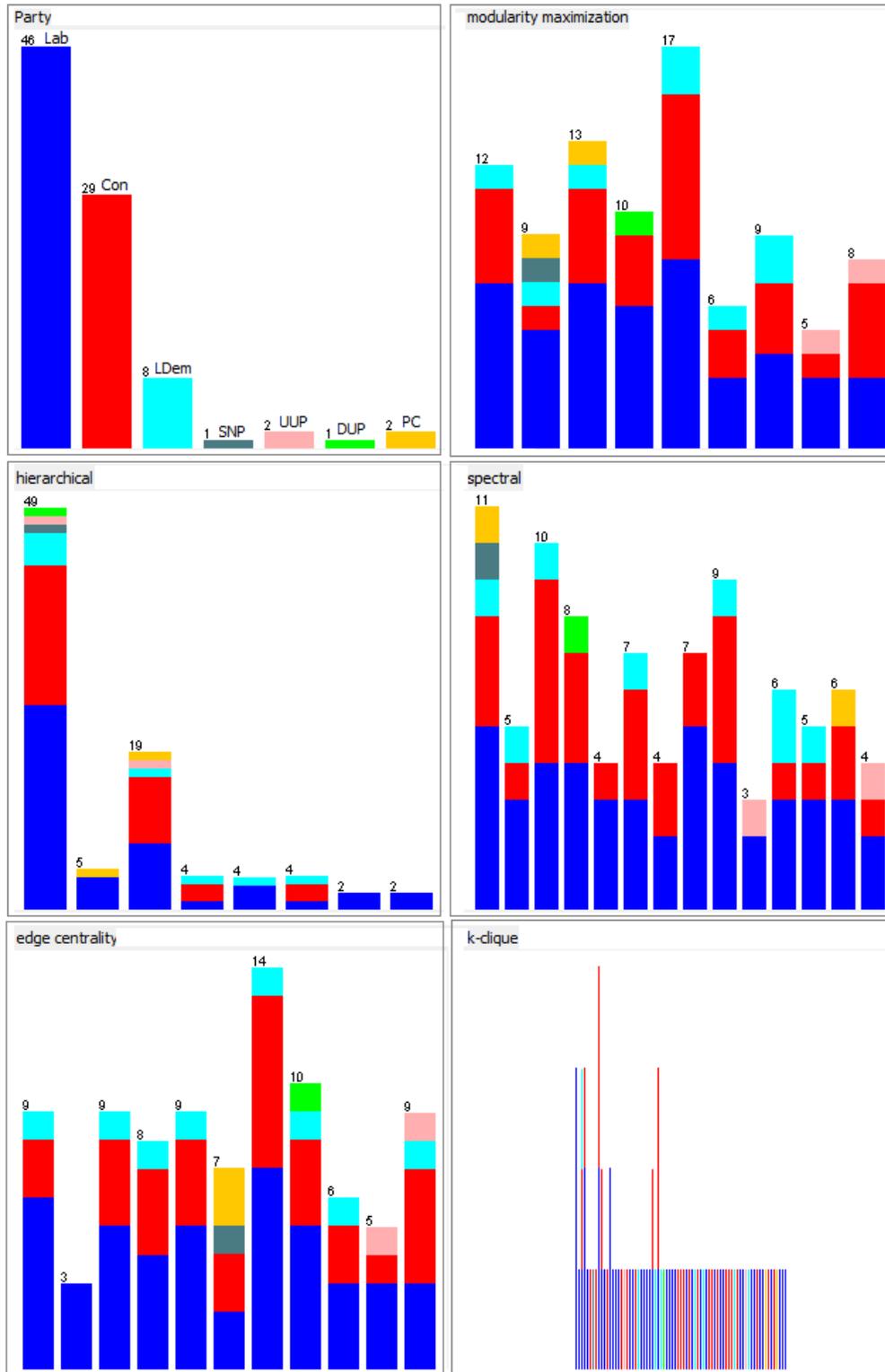
Figure 9.15: Communities in the relevant interruption network for the Iraq debate. The top left histogram presents the known communities (according to party affiliation) while the rest of histograms present the detected communities (clusters predicted using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms). The colour coding is defined in the top left histogram.
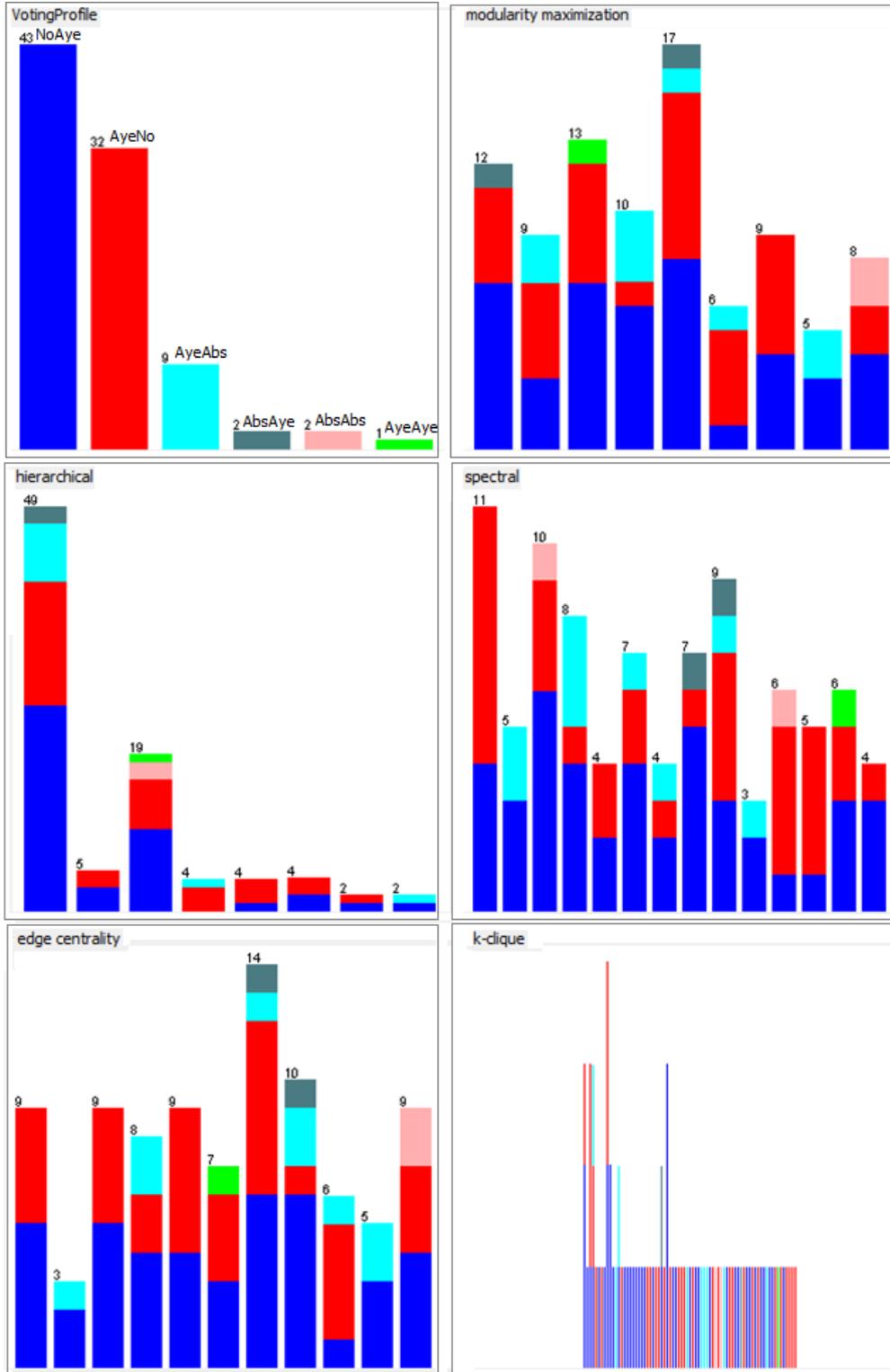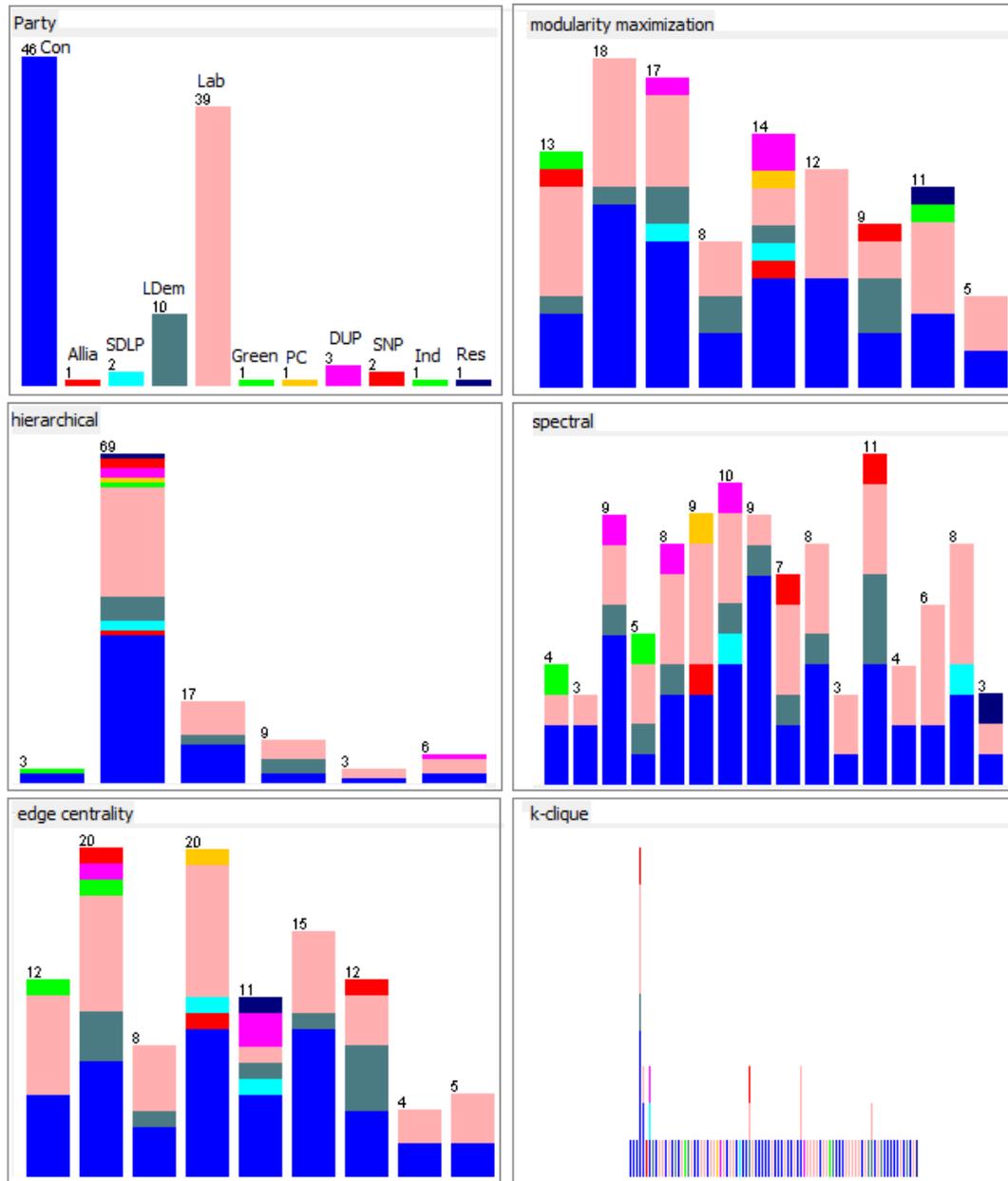
Figure 9.16: Communities in the interruption network for the Syria debate. The top left histogram presents the known communities (according to party affiliation) while the rest of histograms present the detected communities (clusters predicted using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms). The colour coding is defined in the top left histogram.
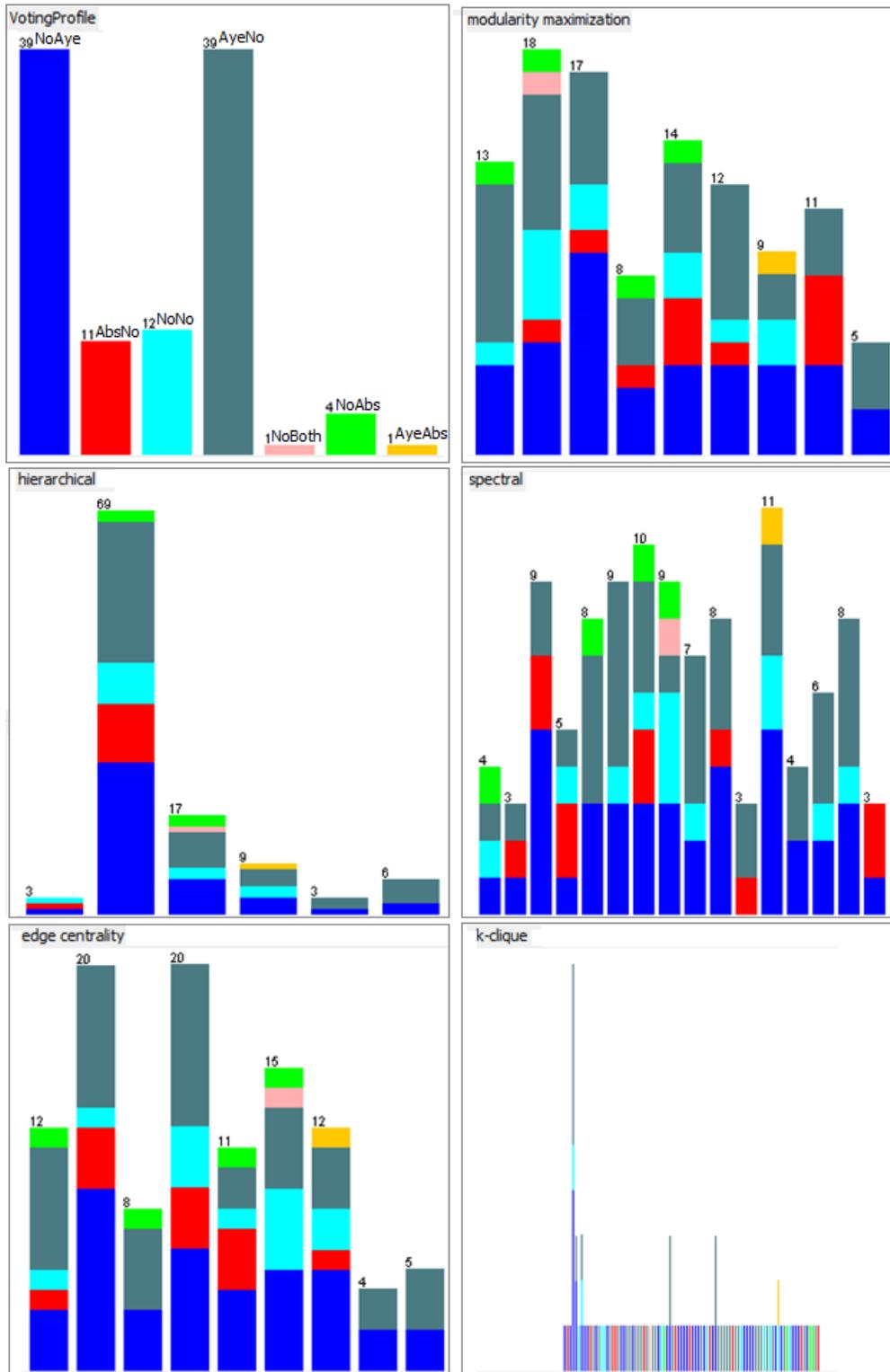
Figure 9.17: Communities in the relevant interruption network for the Syria debate. The top left histogram presents the known communities (according to party affiliation) while the rest of histograms present the detected communities (clusters predicted using modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation algorithms). The colour coding is defined in the top left histogram.
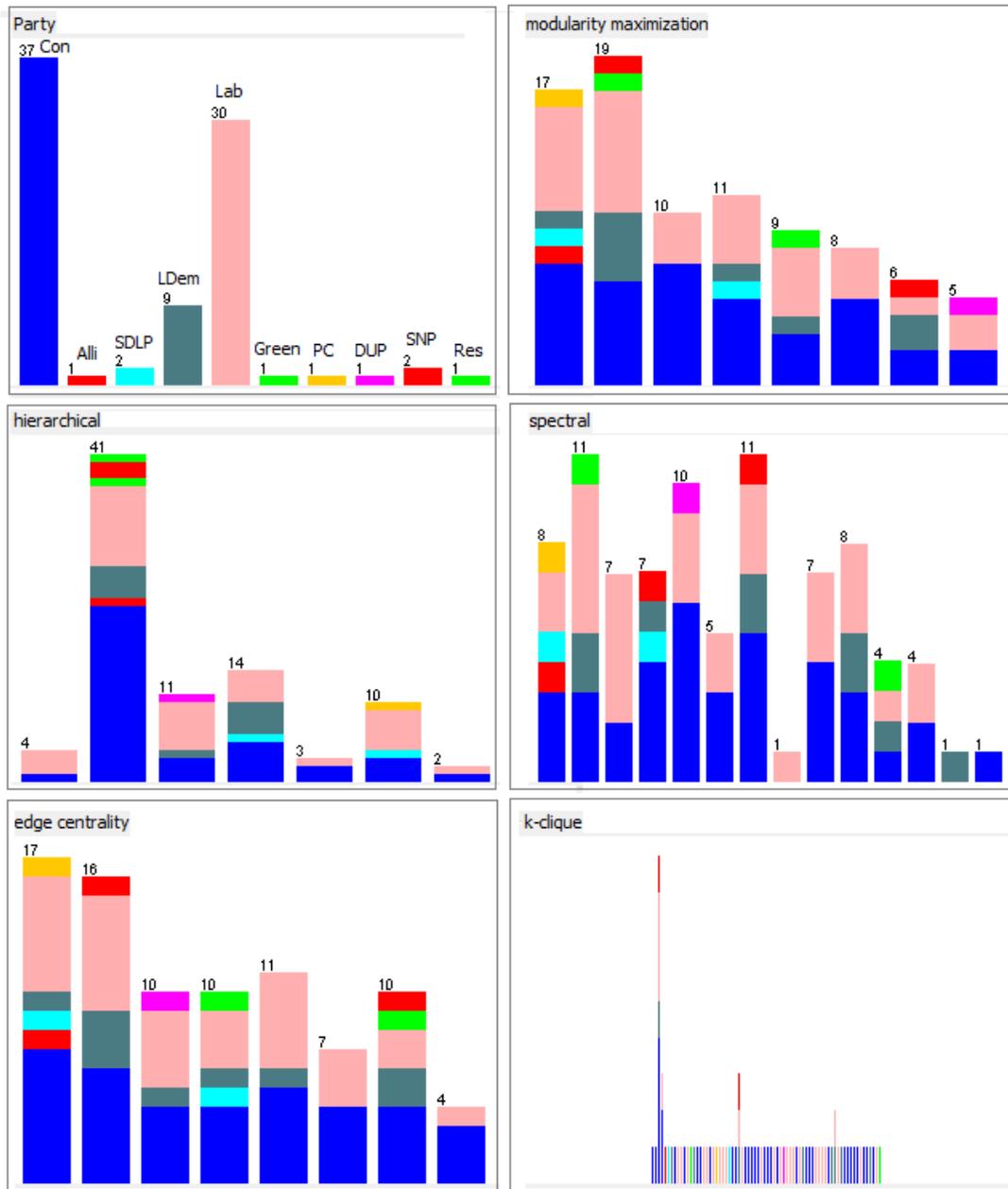
(clusters predicted using the adopted community detection algorithms) for both of the interruption and the relevant interruption networks with respect to the Iraq debate; while Figures 9.16 and 9.17 present the same comparison with respect to the Syria debate. In each case the top left histogram presents the known communities identified from party affiliation. Note also that in each case the colour coding is again defined in the top left histogram. From the figures two observations can be made: (i) all the community detection algorithms found either more or less communities than the number of party affiliations and (ii) all detected communities contained comparable shares of MPs with the same party affiliation. In other words it has not been possible to identify communities defined by voting pattern or party affiliation in the context of interruption or relevant interruption networks.

9.5   Summary

This chapter has presented a study on the conceptualisation of parliamentary debates as networks and their analysis by means of standard network analysis techniques through the exploration of the embedded patterns of connectivity and reactivity between the exchanging nodes (speakers). Two UK House of Commons political debates were chosen and for each debate two types of debate graphs (networks) were built: (i) interruption and (ii) relevant interruption networks. The process of the analysis of these debate graphs (conceptualised as networks) by means of standard network analysis techniques were described in detail and thus the third research objective (**Objective3**) identified with respect to this thesis was addressed.

From the foregoing it can be concluded that: (i) question (**Q1**) is answered affirmatively because all the networks exhibited high degrees of disassortativity with respect to party affiliation and voting behaviour, although with different values. This highlights a "discrepancy" between party affiliation and voting behaviour. MPs tend therefore to respond to speeches by members of other parties or by MPs with different voting inclinations and (ii) question (**Q2**) is answered negatively because none of the community detection algorithms considered (modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation) was able to detect communities of members of the same party or with the same voting behaviour. The following chapter concludes this thesis and presents a summary of the work, the main findings and some suggestions for future work.

Chapter 10

Conclusion

"This is not the end, this is not even the beginning of the end, this is just perhaps the end of the beginning."

Winston Churchill

This chapter concludes the research work described in this thesis by presenting an overall summary of the work, the main findings and some potential areas for future work. The summary is presented in Section 10.1, the main findings from the research are presented Section 10.2 and the research contributions are summarised in Section 10.3; while potential research future extensions are presented in Section 10.4.

10.1 Summary

In this thesis the author has proposed and compared the operation of three different approaches for conducting sentiment mining in the context of political debates with the objective of predicting their outcome (**Objective1**). The three different sentiment mining approaches considered were: (i) straightforward classification, (ii) using generic sentiment lexicons and (iii) using domain specific sentiment lexicons. The generic lexicon used was the off-the-shelf SentiWordNet 3.0 while in the context of the domain specific sentiment lexicons, two techniques to generating the desired domain specific sentiment lexicons were considered: (i) direct generation and (ii) adaptive generation. The two techniques were used to produce two political sentiment lexicons (using UK House of Commons debates for training purposes). The first (PoLex) was produced using the direct generation approach and the second (PoliSentiWordNet) using the adap-

tation approach. Recall that with respect to straightforward classification, six machine learning classifiers were considered: (i) Naïve Bayes, (ii) Support Vector Machine SMO, (iii) J48 decision trees learner, (iv) JRip rule-based classifier, (v) IBk nearest neighbour classifier and (vi) ZeroR (the last as a baseline classifier). An overall (global) comprehensive comparison of the use of the three sentiment mining approaches was conducted using the UKHCD-4 data set; a collection of 2086 concatenated speeches for 29 different debates extracted from the proceedings of the UK House of Commons.

To deploy the proposed sentiment mining approaches for the extraction of debate graphs that will in turn allow for the graphical visualisation of the high-level structure of such debates the author proposed the DGE framework for generating debate graphs from transcripts of debates (**Objective2**). The operation of the framework was illustrated using specified debates taken from the proceedings of the Commons Chamber.

The author also illustrated the utility of such debate graphs (interruption and relevant interruption graphs) by conceptualising them as networks and applying network analysis techniques (**Objective3**). More specifically the ideas of assortativity and communities were considered. Assortativity was applied so that the similarity between connected nodes (speakers) could be investigated in the context of party affiliation and voting profile. Community detection, using off-the-shelf community detection algorithms (modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation) was applied to assess whether there were identifiable communities with respect to party affiliation and voting behaviour. Experiments were conducted using the idea of interruption and relevant interruption networks for two chosen parliamentary debates.

10.2 Main Findings

The main aim of the research work described in this thesis, as initially stated in Chapter 1, was to investigating the use of sentiment mining techniques for the analysis of political debates. More specifically the research work detailed in this thesis was directed at the following three objectives (restated from Chapter 1):

Objective1: The application of sentiment mining techniques to predict the *attitude* of individual debaters, whether they are *for* or *against* a motion.

Objective2: The extraction of debate graphs describing and overviewing political debates from political verbatim transcripts.

Objective3: The analysis of the embedded graph structures, featured in debate graphs, with respect to how the individual participants interact.

To realise the research objectives, a number of research issues needed to be resolved expressed as a set of research questions, some of which encompassed supplementary research questions, reiterated here for completeness, as follows:

RQ1: *Is it possible to effectively predict the* attitude *of individual debaters, whether they are for or against a motion within the context of political debates?*
More specifically:

1. How to use sentiment mining approaches to analyse political debates?

2. What are the most appropriate sentiment mining approaches to predict the attitude of individual debaters?

RQ2: *Is it possible to represent and analyse debates as graphs using tools from network analysis?*
More specifically:

1. How best to extract graph structures from debate records?

2. Which metrics from network analysis to use to highlight structural features of debates?

The research questions and their associated supplementary research questions were considered throughout the thesis. The work described in the thesis addresses each research question as follows:

RQ1: *Is it possible to effectively predict the* attitude *of individual debaters, whether they are for or against a motion within the context of political debates?*

This research question was addressed as part of the evaluation of the proposed sentiment mining approaches (presented in Chapter 7): (i) classification-based, (ii) generic lexicon-based and (iii) domain specific lexicon-based as described in Chapters 4, 5 and 6 respectively. The comparison was conducted with respect to debater attitude prediction in the context of the UK House of Commons political debates and demonstrated that the attitude of individual debaters can be effectively predicted.

1. **How to use sentiment mining approaches to analyse political debates?** The main idea was to use either: (i) machine learning classifiers or

(ii) sentiment lexicons. The classifiers were trained (learned) using an appropriately labelled training dataset and evaluated using test data. The generated classifiers were then used to predict the attitude of individual speakers participating in an "unseen" debate. The sentiment lexicons (general purpose and domain specific) were used to look-up words to firstly identify their subjectivity and secondly to determine their degree of sentiment and polarity (positive or negative). It was then illustrated how this information can be used to make a judgement about the overall (accumulated) sentiment represented by a debater's speech so that a judgement can be made about the debater "attitude" (for or against a motion) in the context of a political debate.

2. **What are the most appropriate sentiment mining approach to predict the attitude of individual debaters?** The main findings from the overall experimental comparison between the three proposed approaches to sentiment mining considered in Chapters 4, 5 and 6 with respect to debater attitude prediction effectiveness in the context of the UK House of Commons political debates were that: (i) the classification-based approach outperformed the lexicon-based approaches and (ii) there is no discernible difference with respect to the operation of the classification-based approach with respect to the "Aye" class and the "No" class. With respect to the lexicon-based approaches, the results produced using the general purpose sentiment lexicon, and two domain specific sentiment lexicons, indicates that: (i) there is a small improvement with respect to the average values obtained when using domain specific lexicons (compared to general purpose lexicons), (ii) both domain specific lexicons produced similar results and (iii) the lexicon-based techniques worked significantly better with respect to predicting "Aye" (positive) attitudes than "No" (negative) attitudes (not the case when using the classification-based approach as noted). So the classification-based approach was chosen as the most appropriate sentiment mining approach to predict the attitude of individual debaters.

RQ2: *Is it possible to represent and analyse debates as graphs using tools from network analysis?*

A Debate Graph Extraction (DGE) framework was proposed and described with respect to how it might be used to extract embedded graph structures from transcriptions of debates. The idea was to represent the structure of a debate as a graph with speakers as nodes and significant interactions (either semantic similarity or interventions made) between debaters as links. Nodes and links were then labelled according to the detected attitude of the speakers. Nodes were labelled

with the attitude of the speaker, either "positive" or "negative" according to whether they are for or against the motion of the debate. Links were labelled as follows: If two nodes connected by a link both have the same attitude label (both positive or both negative) then the link is labelled as being "supporting". If both nodes have different attitude labels (one is positive and the other is negative) the link is labelled as being "opposing". The resulting graphs capture the abstract representation of a debate in terms of two opposing factions exchanging arguments on a related content. The operation of the proposed DGE framework was illustrated by applying the framework to UK House of Commons political debates and consequently generating the associated debate graphs. So it was possible to represent and analyse debates as graphs using tools from network analysis.

1. **How best to extract graph structures from debate records?** The Debate Graph Extraction (DGE) framework was proposed (in Chapter 8) to extract embedded graph structures from debate records, specifically transcripts of UK House of Commons political debates. Two variations of the DGE framework were described: (i) using a classification-based approach to determine speaker attitude and (ii) using sentiment lexicons to determine speaker attitude. Both variations of the DGE framework were found to produce effective debate graphs. The accuracy of these debate graphs is of course dependent on the nature of the speaker attitude detection. From earlier work it had been established that the classification-based approach produced the most effective sentiment analysis. Thus it is argued that version (i) of the DGE framework is the most appropriate. Furthermore, using the DGE framework, three types of debate graph can be generated: (i) semantic similarity debate graphs using the semantic similarity between speakers' concatenated speeches, (ii) interruption debate graphs where the interruptions (interventions) made by MPs during a debate (who interrupted whom) are used or (iii) relevant interruption debate graphs where a combination of both semantic similarity and interruption data is used. With respect to the three different types of graph it was found that semantic similarity graphs were the most informative; while interruption graphs provided for deeper debate analysis.

2. **Which are the most appropriate network analysis metrics and algorithms to use to highlight structural features of debates?** Once a debate graph has been extracted it can be conceptualised as a network in order to analyse this debate graph by means of standard network analysis techniques. With respect to structural analysis the assortativity coefficient was chosen as the most appropriate metric with which to calculate and measure the similarity of connected nodes (speakers) with respect to party

affiliation and voting profile of the interruption and relevant interruption networks. Furthermore, a number of off-the-shelf community detection algorithms have been run and assessed according to whether they were able to identify communities that reflected (to a satisfactory extent) either party affiliation or voting pattern within the interruption and relevant interruption networks. The main findings from the structural analysis, considered in Chapter 9, when applied to the two chosen networks were that: (i) the two networks exhibited high degrees of disassortativity, with different values, in the context of party affiliation and voting behaviour and this reflected a discrepancy between party affiliation and voting behaviour; and (ii) none of the main community detection algorithms considered (modularity maximization, hierarchical and spectral clustering, edge centrality and k-Clique percolation) was able to detect communities of speakers that accurately reflected party affiliation or voting profile.

10.3   Research Contributions

The main contributions of the research work considered in this thesis were presented in Chapter 1. For completeness these are reiterated here as follows:

- A set of benchmark datasets extracted from proceedings of the UK House of Commons debates using information retrieval techniques to extract the required elements and attributes from the XML document archives.

- A domain specific list of parliamentary stop-words to support the preprocessing of such data.

- A framework for using machine learning classifiers in the context of political sentiment mining to classify the attitude (for or against a motion) of individual speakers in a political debate.

- A framework for using generic sentiment lexicons in the context of political sentiment mining to predict the attitude (for or against a motion) of individual speakers in a political debate.

- A framework for using domain specific sentiment lexicons in the context of political sentiment mining to predict the attitude (for or against a motion) of individual speakers in a political debate.

- A mechanism to determine the sentiment scores and polarities for terms in a pre-labelled corpus with regard to the biassed occurrences of these terms in this corpus.

- Two domain specific (political) sentiment lexicons, PoLex and PoliSentiWordNet, generated by applying the techniques described in this thesis to UK House of Commons benchmark data.

- A comparison of the performance, in terms of attitude prediction, of the three identified sentiment mining approaches.

- A Debate Graph Extraction (DGE) framework designed to extract debate graphs embedded within debate transcriptions.

- The conceptualisation of the extracted debate graphs as networks and an indication of how such networks might be used to analyse the structural properties of a debate graph.

10.4   Research Future Extensions

Many promising directions for future research present themselves so as to extend the functionality and enhance the operation of the proposed approaches. Potential directions for future research can be summarised as follows:

With respect to sentiment mining:

- All approaches considered were less effective at predicting negative speaker attitudes, due to the often overly polite parliamentary jargon used, thus providing an interesting avenue for future research considering further linguistic analysis.

- Investigation of new mechanisms whereby the proposed sentiment mining techniques, especially the classification-based technique, can be used to better predict the attitude of individual speakers. For Example using feature selection techniques, like Chi-Square, Principle-Component or Information-Gain, to rate feature subsets and then select the feature subset that achieved the best performance. In this way the most relevant features may be selected and the misleading (erroneous) subsets ignored.

- In order to achieve a more accurate result for speaker attitude prediction in the future, a larger working corpora needs to be obtained for training purposes in the context of future experiments. The generation of such a corpora would provide a beneficial topic for future work.

- Combination of new features in the corpora (other than textual features or interruptions made in the current debate) like: (i) the recorded patterns of exchanges (interruptions or opposing/supporting interactions) between individual speakers (MPs) within previous debates and (ii) the recorded fixed vote (attitude) of individual speakers with respect to specific topics previously debated. It is suggested that these kinds of features can help predict/classify speaker votes by assuming that the speakers will exhibit the same behaviour in future debates.

- The addition of numbers and punctuation marks, which could actually be strong indicators of sentiment, in the debate analyses to achieve a more accurate result for speaker attitude prediction in the future.

- Application of the three sentiment mining approaches to other forms of debate (unlike the House of Commons debates considered in this thesis) and other forms of structured discussion, in different possible domains, where the result is not known (or not yet known). A study of this form would provide for a challenging research project.

- Work on the process of making a judgement about the polarity of speeches with respect to different levels of "text inclusion" (sentence-level, paragraph-level, and so on) and investigation of the most appropriate level of text inclusion with respect to speaker attitude prediction processes.

With respect to graph extraction:

- The integration of the DGE framework with word-clouds (on nodes and links) so as to provide for more comprehensible debate graphs.

- The generation of a substantial collection of debate graphs using the proposed DGE framework in order to conduct further, more elaborate and extensive, network analysis of community structures.

With respect to network analysis:

- The development of community detection algorithms better suited to identifying parties and voting profiles.

- In the longer term it would be interesting to focus on analysing large debate graph collections directly using graph mining techniques, rather than using simple tabular data mining techniques, so as to attempt to predict debate outcomes using the information embedded in the structure of such debate graphs. For example by identifying frequently occurring sub-graphs.

Overall the work on the use of sentiment mining techniques for the analysis of political debates, as presented in this thesis, has produced some interesting results and

provided a sound foundation for future work. Some of suitable non-political domains and applications that the work in the thesis could be applied to comprise auditing, logging and analysing public opinions about the products and the provided services in the context of marketing analysis, industry and commercial advertising.

Appendix A

Parliamentary Stop Words List

As noted in the main body of this thesis, stop words are words which are not expected to convey any significant meaning in the context of sentiment analysis, for example words such as "the", "a", "and", "is" and so on) [Chim and Deng, 2008, Hariharan and Srinivasan, 2008, Poomagal and Hamsapriya, 2011]. Given a specific domain there will also be additional words, other than stop words, that occur frequently. In the case of House of Commons parliamentary debates words like: "hon.", "house", "minister", "government", "gentleman", "friend" and "member" are all very frequently occurring words. For similar reasons as for stop word removal these domain specific words were also removed. This was done by appending them to a stop-words list. The names of all the members of parliament, political parties and constituencies were also added to the words in default Weka's stop-word list. For completeness this appendix presents the stop-words list adopted in this thesis for conducting stop word removal.

a, about, above, across, after, afterwards, again, against, all, almost, alone, along, already, also, although, always, am, among, amongst, amoungst, amount, an, and, another, any, anyhow, anyone, anything, anyway, anywhere, are, around, as, at, back, be, became, because, become, becomes, becoming, been, before, beforehand, behind, being, below, beside, besides, between, beyond, bill, both, bottom, but, by, call, can, cannot, cant, co, computer, con, could, couldnt, cry, de, describe, detail, do, done, down, due, during, each, eg, eight, either, eleven, else, elsewhere, empty, enough, etc, even, ever, every, everyone, everything, everywhere, except, few, fifteen, fifty, fill, find, fire, first, five, for, former, formerly, forty, found, four, from, front, full, further, get, give, go, had, has, hasnt, have, he, hence, her, here, hereafter, hereby, herein, hereupon, hers, herself, him, himself, his, how, however, hundred, i, ie, if, in, inc, indeed, interest, into, is, it, its, itself, keep, last, latter, latterly, least, less, ltd, made, many, may, me, meanwhile, might, mill, mine, more, moreover, most, mostly, move, much, must,

my, myself, name, namely, neither, never, nevertheless, next, nine, no, nobody, none, noone, nor, not, nothing, now, nowhere, of, off, often, on, once, one, only, onto, or, other, others, otherwise, our, ours, ourselves, out, over, own, part, per, perhaps, please, put, rather, re, same, see, seem, seemed, seeming, seems, serious, several, she, should, show, side, since, sincere, six, sixty, so, some, somehow, someone, something, sometime, sometimes, somewhere, still, such, system, take, ten, than, that, the, their, them, themselves, then, thence, there, thereafter, thereby, therefore, therein, thereupon, these, they, thick, thin, third, this, those, though, three, through, throughout, thru, thus, to, together, too, top, toward, towards, twelve, twenty, two, un, under, until, up, upon, us, very, via, was, we, well, were, what, whatever, when, whence, whenever, where, whereafter, whereas, whereby, wherein, whereupon, wherever, whether, which, while, whither, who, whoever, whole, whom, whose, why, will, with, within, without, would, yet, you, your, yours, yourself, yourselves, diane, debbie, nigel, adam, bob, peter, danny, douglas, heidi, rushanara, graham, david, stuart, james, jonathan, ian, richard, adrian, william, norman, steven, tony, harriett, edward, gordon, stephen, gregory, john, kevin, gavin, hugh, guto, margaret, anne, alan, henry, hilary, joe, paul, luciana, jake, clive, andrew, brian, roberta, nicola, hazel, tom, crispin, nicholas, karen, ben, angie, julian, steve, annette, lyn, nick, russell, jeremy, fiona, malcolm, chris, robert, aidan, andy, conor, simon, alistair, lorely, dan, liam, vincent, alun, menzies, ronnie, neil, martin, jenny, rehman, christopher, greg, katy, kenneth, geoffrey, ann, vernon, therese, damian, oliver, michael, rosie, yvette, mary, stella, tracey, jon, alex, jim, wayne, geraint, glyn, philip, gloria, caroline, frank, thomas, pat, jeffrey m, brian h, nadine, gemma, jackie, jack, iain, mark, angela, maria, julie, jane, louise, tobias, charlie, natascha, george, tim, lynne, don, yvonne, hywel, mike, lorraine, roger, barry, michelle, cheryl, sheila, zac, helen, kate, justine, lilian, dominic, nia, sam, duncan, fabian, matthew, harriet, rebecca, dai, charles, lady, meg, sharon, jimmy, kelvin, kris, stewart, gerald, lindsay, tristram, huw, glenda, margot, sin, cathy, sajid, bernard, diana, gareth, jo, kevan, marcus, susanelan, tessa, eric, daniel, barbara, liz, sadiq, kwasi, eleanor, pauline, andrea, jessica, phillip, charlotte, brandon, ivan, elfyn, naomi, angus, denis, khalid, shabana, seema, francis, theresa, kerry, jason, karl, gregg, siobhain, alasdair, alison, catherine, patrick, esther, austin, madeleine, penny, nicky, graeme, anne-marie, grahame, sheryll, lisa, pamela, brooks, sarah, jesse, chi, guy, sandra, albert, priti, owen, teresa, toby, claire, bridget, dawn, yasmin, jamie, jacob, rachel, emma, linda, laurence, amber, joan, laura, anas, lee, grant, alok, virendra, alec, keith, dennis, chloe, anna, rory, gary, mel, gisela, gerry, desmond, hugo, emily, justin, elizabeth, derek, chuka, shailesh, valerie, robin, dave, heather, eilidh, craig, jennifer, phil, rob, sammy, pete, shaun, nadhim, abbott, abrahams, adams, afriyie, ainsworth, aldous, alexander, ali, allen, amess, anderson, arbuthnot, ashworth, bacon, bailey, bain, baker, baldry, baldwin, balls, banks, barclay, barker, baron, barron, barwell, bayley, bebb, beckett, begg,

beith, bellingham, benn, benton, benyon, bercow, beresford, berger, berry, betts, bingham, binley, birtwistle, blackman, blackman-woods, blackwood, blears, blenkinsop, blomfield, blunkett, blunt, boles, bone, bottomley, bradley, bradshaw, brady, brake, bray, brazier, brennan, bridgen, brine, brokenshire, brooke, brown, browne, bruce, bryant, buck, buck-land, burden, burley, burnham, burns, burrowes, burstow, burt, byles, byrne, cable, cairns, cameron, campbell, carmichael, carswell, cash, caton, chapman, chishti, chope, clappison, clark, clarke, clegg, clifton-brown, clwyd, coaker, coffey, collins, colvile, con-narty, cooper, corbyn, cox, crabb, crausby, creagh, creasy, crockart, crouch, cruddas, cryer, cunningham, curran, dakin, danczuk, darling, davey, davidson, davies, davis, debois, depiero, denham, dinenage, djanogly, dobbin, dobson, docherty, dodds, doherty, donaldson, donohoe, doran, dorrell, dorries, dowd, doyle, doyle-price, drax, dromey, duddridge, dugher, duncan smith, dunne, durkan, eagle, edwards, efford, elliott, ellis, ellison, ellman, ellwood, elphicke, engel, esterson, eustice, evans, evennett, fabricant, fallon, farrelly, farron, featherstone, field, fitzpatrick, flello, flint, flynn, foster, fo-vargue, fox, francois, freeman, freer, fullbrook, fuller, gale, galloway, gapes, gardiner, garnier, gauke, gibb, gilbert, gildernew, gillan, gilmore, glass, glen, glindon, godsiff, goggins, goldsmith, goodman, goodwill, gove, gray, grayling, greatrex, green, greening, greenwood, grieve, griffith, griffiths, gummer, gwynne, gyimah, hague, hain, halfon, hames, hamilton, hammond, hancock, hands, hanson, harman, harper, harrington, har-ris, hart, harvey, haselhurst, havard, hayes, heald, healey, heath, heaton-harris, hem-ming, henderson, hendrick, hendry, hepburn, herbert, hermon, heyes, hillier, hilling, hinds, hoban, hodge, hodgson, hoey, hollingbery, hollobone, holloway, hood, hopkins, horwood, hosie, howarth, howell, hoyle, hughes, huhne, hunt, hunter, huppert, hurd, irranca-davies, jackson, jamieson, jarvis, javid, jenkin, johnson, jones, jowell, joyce, kaufman, kawczynski, keeley, kelly, kendall, kennedy, khan, kirby, knight, kwarteng, laing, lamb, lammy, lancaster, lansley, latham, lavery, laws, lazarowicz, leadsom, leech, lefroy, leigh, leslie, letwin, lewis, liddell-grainger, lidington, lilley, lloyd, llwyd, long, lopresti, lord, loughton, love, lucas, luff, lumley, macleod, macneil, macshane, mactag-gart, mahmood, main, malhotra, mann, marsden, maskey, maude, maynard, mccabe, mccann, mccarthy, mccartney, mcclymont, mccrea, mcdonagh, mcdonnell, mcfadden, mcgovern, mcguinness, mcguire, mcintosh, mckechin, mckenzie, mckinnell, mclough-lin, mcpartland, mcvey, meacher, meale, mearns, mercer, metcalfe, miliband, miller, mills, milton, mitchell, moon, moore, mordaunt, morden, morgan, morrice, morris, mosley, mowat, mudie, mulholland, mundell, munn, munt, murphy, murray, murrison, nandy, nash, neill, newmark, newton, nokes, nuttall, o'brien, o'donnell, offord, olleren-shaw, onwurah, opperman, osborne, ottaway, paice, paisley jnr, parish, patel, pater-son, pawsey, pearce, penning, penrose, percy, perkins, perry, phillips, phillipson, pick-les, pincher, poulter, pound, primarolo, prisk, pritchard, pugh, qureshi, raab, randall, raynsford, reckless, redwood, reed, rees-mogg, reevell, reeves, reid, reynolds, rifkind, ri-

ordan, ritchie, robathan, robertson, robinson, rogerson, rosindell, rotheram, roy, ruane, rudd, ruddock, ruffley, rutley, sanders, sandys, sarwar, scott, seabeck, selous, shannon, shapps, sharma, sheerman, shelbrooke, shepherd, sheridan, shuker, simmonds, simpson, skidmore, skinner, slaughter, smith, soames, soubry, spellar, spelman, spencer, stanley, stephenson, stevenson, straw, streeter, stride, stringer, stunell, sturdy, sutcliffe, swales, swayne, swinson, swire, syms, tami, tapsell, teather, thornberry, thurso, timms, timpson, tomlinson, tredinnick, trickett, truss, turner, twigg, tyrie, umunna, uppal, vaizey, vara, vaz, vickers, villiers, walker, wallace, walley, walter, ward, watkinson, watson, watts, weatherley, webb, weir, wharton, wheeler, white, whiteford, whitehead, whittaker, whittingdale, wiggin, willetts, williams, williamson, willott, wilson, winnick, winterton, wishart, wollaston, wood, woodcock, woodward, wright, yeo, young, zahawi, labour, conservative, liberal, democrat, dup, sinn, social, plaid, respect, independent, scottish, alliance, fein, democratic, national, cymru, party, constituency, hackney, oldham, selby, windsor, coventry, waveney, inverness,, paisley, lewisham, bethnal, nottingham, southend, blaydon, pudsey, north, leicester, dudley, south, west, glasgow, lewes, wycombe, banbury, morley, ochil, bexhill, basildon, rother, croydon, york, aberconwy, derby, aberdeen, berwick-upon-tweed, leeds, bootle, newbury, buckingham, mole, liverpool,, rossendale, sheffield, high, northampton, burnley, harrow, city, oxford, salford, middlesbrough, sheffield,, reigate, grantham, wellingborough, worthing, staffordshire, exeter, altrincham, carshalton, ealing, canterbury, cardiff, winchester, old, mid, kirkcaldy, newcastle, dumfries, taunton, congleton, rhondda, westminster, birmingham,, cannock, bournemouth, chelmsford, enfield,, sutton, solihull, twickenham, vale, witney, tynemouth, east, blyth, orkney, stroud, clacton, stone, gower, darlington, gillingham, christchurch, hertsmere, tunbridge, rushcliffe, coatbridge,, cynon, gedling, stockport, suffolk, folkestone, plymouth,, linlithgow, normanton,, islington, torridge, preseli, bolton, wakefield, walthamstow, edinburgh, chatham, dagenham, leyton, stockton, workington, scunthorpe, rochdale, kingston, caerphilly, monmouth, swansea, montgomeryshire, shipley, haltemprice, enfield, ashfield, southampton,, gosport, huntingdon, heywood, holborn, dunfermline, belfast, lagan, central, charnwood, thurrock, rochford, barnsley, rutland, chingford, ludlow, foyle, wallasey, garston, carmarthen, eltham, sunderland, battersea, dover, sefton, camborne, islwyn, weaver, ribble, bexleyheath, lichfield, sevenoaks, newcastle-under-lyme, westmorland, hornsey, birkenhead, cities, poplar, stoke-on-trent, newport, bath, makerfield, aberavon, rayleigh, finchley, bedford, bradford, ilford, brent, harborough, wyre, st, bognor, fermanagh, chesham, salisbury, wythenshawe, richmond, bishop, scarborough, surrey, gloucester, maidstone, epsom, rutherglen, ashford, stretford, putney, beaconsfield, llanelli, burton, ipswich, denton, neath, harlow, chippenham, midlothian, runnymede, wimbledon, portsmouth, chelsea, delyn, camberwell, forest, watford, castle, saffron, merthyr, wentworth, somerton, daventry, sittingbourne, preston, wealden, jarrow, arundel, ashton-under-lyne,

fareham, barking, washington, vauxhall, meon, kettering, gravesham, lanark, luton, keighley, cheltenham, dundee, knowsley, aldershot, henley, chorley, bermondsey, eastleigh, cheadle, cambridge, ruislip,, ogmore, hampstead, peterborough, stourbridge, kilmarnock, bromsgrove, harwich, dartford, orpington, harrogate, hyndburn, warrington, nuneaton, dulwich, falkirk, manchester,, shrewsbury, worsley, ross,, tooting, brighton,, spelthorne, epping, tottenham, wansbeck, yeovil, erewash, bracknell, stafford, gainsborough, bristol, great, bury, new, bridgwater, aylesbury, hitchin, eastbourne, manchester, dwyfor, filton, woking, edmonton, wrexham, redditch, brentford, na, rotherham, slough, feltham, bassetlaw, blackpool, horsham, maidenhead, colne, lincoln, cumbernauld,, mitcham, wolverhampton, wirral, stirling, thirsk, inverclyde, derbyshire, stevenage, mansfield, gateshead, fylde, newark, doncaster, ellesmere, basingstoke, guildford, bridgend, berwickshire,, loughborough, livingston, morecambe, easington, halesowen, dumfriesshire,, wells, newry, torfaen, wigan, airdrie, bromley, braintree, truro, romsey, hereford, eddisbury, hendon, hexham, tatton, ayr,, ynys, tiverton, witham, rugby, erith, hemel, weston-super-mare, brigg, chesterfield, devizes, sleaford, houghton, brentwood, tamworth, hertford, southport, esher, uxbridge, greenwich, rochester, wokingham, copeland, dewsbury, argyll, stalybridge, kensington, halifax, moray, faversham, tewkesbury, romford, motherwell, glenrothes, hastings, lewisham,, colchester, macclesfield, torbay, strangford, welwyn, reading, ealing,, huddersfield, elmet, aldridge-brownhills, boston, upper, broadland, kingswood, bolsover, hammersmith, penistone, norwich, crawley, skipton, blaenau, pontypridd, broxtowe, warley, meriden, sherwood, tonbridge, pendle, carlisle, beckenham, penrith, blackburn, blackley, beverley, redcar, poole, alyn, louth, caithness,, crewe, bosworth, hemsworth, isle, halton, chichester, streatham, wantage, walsall, cleethorpes, chipping, broxbourne, worcester, hornchurch, hove, thornbury, warwick, banff, calder, maldon, havant, arfon, ceredigion, brecon, sedgefield, perth, totnes, batley, barrow, telford, hartlepool, kenilworth, stratford-on-avon, whiston, hampshire, helens, perthshire, spen, furness, southam, antrim, barnet, dorset, bromwich, yate, buchan, test, valley, leamington, herefordshire, radnorshire, upminster, ainsty, strathspey, badenoch, renfrewshire, stoke, bow, saddleworth, newington, worcestershire, outwood, cambridgeshire, battle, billericay, nairn,, norfolk, darwen, sutherland, ham, swindon, wight, hull, nantwich, deeside, horncastle, finsbury, easter, ross, gwent, ripon, aberdeenshire, sussex, stocksbridge, kincardine, malling, broughton, devon, keynes, border, edgbaston, grove, outer, dunbartonshire, holderness, wavertree, peak, brightside, moorlands, bexley, deane, northfield, chase, southgate, warwickshire, londonderry, ayrshire, chryston, hallam, cotswolds, coastal, lancashire, pontefract, pembrokeshire, itchen, tyrone, bedfordshire, erdington, riverside, somerset, thanet, ives, regis, austell, tyneside, hall, park, auckland, wiltshire, (yorks), point, walden, tydfil, holland, yardley, northwood, durham, gorton, skye, kemptown, yorkshire, northamptonshire, withington, yarmouth, meirionnydd, pavilion, h-eileanan, ladywood, albans, selly,

kilbride,, kilsyth, ulster, dales, shields, port, coldfield, grimsby, roxburgh, abbot, clydesdale, heeley, lothian, carrick, mn, shropshire, hempstead, wrekin, leicestershire, cornwall, walton, deptford, moor, hatfield, southall, bann, eccles, stamford, sale, wallington, cowdenbeath, tyne, cheam, hill, glamorgan, fife, shetland, rainham, hythe, aylesford, wanstead, surbiton, howden, middleton, melton, woodford, halewood, redruth, crayford, lonsdale, london, limehouse, wickford, golders, hertfordshire, amersham, whitby, ewell, urmston, reddish, weybridge, fulham, peckham, dean, dearne, frome, sheppey, kilburn, loudoun, knaresborough, atcham, harpenden, isleworth, barr, heston, oak, strathaven, harlington, malton, lunesdale, rowley, chester, armagh, shotts, chislehurst, falmouth, southampton, fleetwood, honiton, thamesmead, goole, ongar, stortford, woolwich, strood, bute, hyde, wishaw, rye, edmunds, view, rothwell, skegness, cleveland, abingdon, hillsborough, acton, sidcup, arran, bellshill, devonport, castleford, pancras, penge, dinefwr, littlehampton, newquay, weald, rhymney, downs, shoreditch, southwark, pinner, essex, norwood, lochaber, leith, shoreham, iar, cleveleys, kirkintilloch, royton, penarth, neston, selkirk, tweeddale, cumnock, hykeham, ruislip, kent, lesmahagow, hessle, deepings, constituencies, gentleman, gentlemen, hon, amendment, minister, government, friend, committee, member, members.

Bibliography

Neda Aleebrahim, Mohammad Fathian, and Mohammad Gholamian. Sentiment classification of online product reviews using product features. In *Proc. 3rd International Conference on Data Mining and Intelligent Information Technology Applications (ICMiA 2011)*, pages 242–245, 2011.

Abdelmalek Amine, Zakaria Elberrichi, and Michel Simonet. Evaluation of text clustering methods using wordnet. *The International Arab Journal of Information Technology*, 7(4):349–357, 2010.

Amna Asmi and Tanko Ishaya. A framework for automated corpus generation for semantic sentiment analysis. In *Proc. World Congress on Engineering (WCE 2012)*, pages 436–444, 2012.

Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. Opinionetit: A structured and faceted knowledge-base of opinions. In Jilles Vreeken, Charles Ling, Mohammed Javeed Zaki, Arno Siebes, Jeffrey Xu Yu, Bart Goethals, Geoffrey I. Webb, and Xindong Wu, editors, *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 878–881. IEEE Computer Society, 2012. ISBN 978-1-4673-5164-5. URL http://dblp.uni-trier.de/db/conf/icdm/icdmw2012.html#AwadallahRW12.

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), 2010.

Jerome R. Bellegarda. Part-of-speech tagging by latent analogy. *Selected Topics in Signal Processing, IEEE Journal of*, 4(6):985–993, 2010. ISSN 1932-4553. doi: 10.1109/JSTSP.2010.2075970.

Adam Bermingham, Maura Conway, Lisa McInerney, Neil O'Hare, and Alan F. Smeaton. Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. In *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, ASONAM '09, pages 231–

236, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3689-7. doi: 10.1109/ASONAM.2009.31. URL http://dx.doi.org/10.1109/ASONAM.2009.31.

Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python.* O'Reilly Media, Inc., 1st edition, 2009. ISBN 0596516495, 9780596516499.

Vineet Kumar Birla, Rahul Gautam, and V.N. Shukla. Retrieval and creation of domain specific lexicon from noisy text data. In *In Proceedings of ASCNT-2011, CDAC, Noida, India, pp*, 2011.

Larry Birnbaum. Argument molecules: A functional representation of argument structure. In *Proceedings of the Third National Conference on Artificial Intelligence (AAAI-82)*, pages 63–65, 1982.

Stephen Borgatti. *NetDraw Software for Network Visualization.* Lexington, 2002.

Hung Chim and Xiaotie Deng. Efficient phrase-based document similarity for clustering. *IEEE Transactions on Knowledge and Data Engineerin*, 20(9):1217–1229, 2008.

Yejin Choi and Claire Cardie. Adapting a polarity lexicon using integer linear programming for domainspecific sentiment classification. In *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 590–598, 2009.

Jason Chuang, Christopher D. Manning, and Jeffrey Heer. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 74–77, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1287-5. doi: 10.1145/2254556.2254572. URL http://doi.acm.org/10.1145/2254556.2254572.

Michael Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of twitter users. In *Proceedings of 3rd IEEE Conference on Social Computing (SocialCom)*, 2011.

Timothy Cribbin. Discovering latent topical structure by second-order similarity analysis. *Journal of the American Society for Information Science and Technology*, 62(6):1188–1207, 2011. URL http://dblp.uni-trier.de/db/journals/jasis/jasis62.html#Cribbin11.

Yan Dang, Yulei Zhang, and HsinChun Chen. A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, 25(4):46–53, July 2010. ISSN 1541-1672. doi: 10.1109/MIS.2009.105. URL http://dx.doi.org/10.1109/MIS.2009.105.

William Deitrick and Wei Hu. Mutually enhancing community detection and sentiment analysis on twitter networks. *Journal of Data Analysis and Information Processing*, 1(3):19–29, 2013. doi: 10.4236/jdaip.2013.13004.

Rodolfo Delmonte, Rocco Tripodi, and Daniela Gîfu. Opinion and factivity analysis of italian political discourse. In *IIR*, pages 88–99, 2013.

Gulsen Demiroz, Berrin Yanikoglu, Dilek Tapucu, and Yucel Saygin. Learning domain-specific polarity lexicons. In *Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on*, pages 674–679, 2012. doi: 10.1109/ICDMW.2012.120.

Kerstin Denecke. Using sentiwordnet for multilingual sentiment analysis. In *Proc 24th IEEE International Conference on Data Engineering Workshop (ICDEW 2008)*, pages 507–512, 2008.

Kerstin Denecke. Are sentiwordnet scores suited for multi-domain sentiment classification? In *Proc. 4th International Conference on Digital Information Management (ICDIM 2009)*, pages 33–38, 2009a.

Kerstin Denecke. Are sentiwordnet scores suited for multi-domain sentiment classification? In *Digital Information Management, 2009. ICDIM 2009. Fourth International Conference on*, pages 1–6, 2009b. doi: 10.1109/ICDIM.2009.5356764.

Hongbo Deng, Jiawei Han, Heng Ji, Hao Li, Yue Lu, and Hongning Wang. Exploring and inferring user-user pseudo-friendship for sentiment analysis with heterogeneous networks. In *SDM*, pages 378–386. SIAM, 2013. ISBN 978-1-61197-262-7. URL http://dblp.uni-trier.de/db/conf/sdm/sdm2013.html#DengHJLLW13.

Richard Duda, Peter Hart, and David Stork. *Pattern Classification*. Wiley, 2nd edition, 2001. ISBN 978-0-471-05669-0.

Phan Minh Dung. On the acceptability of arguments and its fundamental role in non monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

Kathleen T. Durant and Michael D. Smith. Predicting the political sentiment of web log posts using supervised machine learning techniques coupled with feature selection. In Olfa Nasraoui, Myra Spiliopoulou, Jaideep Srivastava, Bamshad Mobasher, and Brij Masand, editors, *Advances in Web Mining and Web Usage Analysis*, volume 4811 of *Lecture Notes in Computer Science*, pages 187–206. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-77484-6. doi: 10.1007/978-3-540-77485-3_11. URL http://dx.doi.org/10.1007/978-3-540-77485-3_11.

Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings from the International Conference on Language Resources and Evaluation (LREC)*, 2006.

David S. Fay and Ken Gerow. *A biologist's guide to statistical thinking and analysis*. WormBook, 2005.

Tim Van Gelder. The rationale for rationale. *Law, Probability and Risk*, 6(1-4):23–42, 2007.

Amnon Glassner and Baruch B. Schwarz. The role of floor control and of ontology in argumentative activities with discussion-based tools. In *Proceedings of th 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: The Next 10 Years!*, CSCL '05, pages 170–179. International Society of the Learning Sciences, 2005. ISBN 0-8058-5782-6. URL http://dl.acm.org/citation.cfm?id=1149293.1149315.

Peter A. Gloor, Jonas Krauss, Stefan Nann, Kai Fischbach, and Detlef Schoder. Web science 2.0: Identifying trends through semantic social network analysis. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 215–222, 2009. doi: 10.1109/CSE.2009.186.

Thomas Gordon. Visualizing carneades argument graphs. *International Journal of AI Tools*, 6:109–117, 2007.

Steven Grijzenhout, Valentin Jijkoun, and Maarten Marx. Opinion mining in dutch hansards. In *Proceedings Workshop From Text to Political Positions (t2pp 2010)*, April 2010.

Shanmugasundaram Hariharan and Rengaramanujam Srinivasan. A comparison of similarity measures for text documents. *Journal of Information Knowledge Management*, 7(1):1–8, 2008.

Graeme Hirst, Yaroslav Riabinin, and Jory Graham. Party status as a confound in the automatic classification of political speech by ideology. In *Proceedings of 10th International Conference on Statistical Analysis of Textual Data / 10es Journes internationales d'Analyse statistique des Donnes Textuelles (JADT 2010)*, pages 731–742, 2010.

Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *10th European Conference on Machine Learning (ECML'98)*, pages 137–142, 1998.

K Kalaivendhan and P Sumathi. An efficient clustering method to find similarity between the documents. In *Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)*, pages 2532–2535, 2014.

Rianne Kaptein, Maarten Marx, and Jaap Kamps. Who said what to whom? capturing the structure of debates. In *Proc. 32nd international ACM SIGIR conference on Research and development in information retrieval (SIGIR' 09)*, pages 831–832, 2009.

Alistair Kennedy and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125, 2006. doi: 10. 1111/j.1467-8640.2006.00277.x. URL http://dx.doi.org/10.1111/j.1467-8640.2006.00277.x.

Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. doi: 10. 3115/1220355.1220555. URL http://dx.doi.org/10.3115/1220355.1220555.

Paul Arthur Kirschner, Simon Buckingham-Shum, and Chad S. Carr, editors. *Visualizing Argumentation : Software Tools for Collaborative and Educational Sense-Making*. Springer, January 2003. ISBN 1-85233-664-1. 240 pages, EUR 85,55.

Adrian Kuhn, Stephane Ducasse, and Tudor Gibra. Semantic clustering: Identifying topics in source code. *Information and Software Technology*, 49(3):230–243, 2007.

Raymond Lau, Wenping Zhang, Peter Bruza, and Kam-Fai Wong. Learning domain-specific sentiment lexicons for predicting product sales. In *e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on*, pages 131–138, 2011. doi: 10.1109/ICEBE.2011.55.

Michael Laver, Kenneth Benoit, and Trinity College. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331, 2003.

Hong-Mei Li, Chen-Xia Sun, and Ke-Jian Wan. Clustering web search results using conceptual grouping. In *Proc. 8th International Conference on Machine Learning and Cybernetics*, pages 12–15, 2009.

Bing Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

Fabrizio Macagno, Glenn Rowe, Chris Reed, and Douglas Walton. Araucaria as a tool for diagramming arguments in teaching and studying philosophy (2005). *Teaching Philosophy*, 29(2):111–124, 2006. URL http://ssrn.com/abstract=1751659.

Ainura Madylova and Sule-Gunduz Oguducu. Comparison of similarity measures for clustering turkish documents. *Intelligent Data Analysis*, pages 815–832, 2009.

Justin Martineau and Tim Finin. Delta tfidf: An improved feature space for sentiment analysis. In *Proc 3rd International ICWSM Conference*, pages 258–261, 2009.

Maarten Marx. Advanced information access to parliamentary debates. *J. Digit. Inf.*, 10(6), 2009.

Panagiotis T. Metaxas, Eni Mustafaraj, and Daniel Gayo-Avello. How (not) to predict elections. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 165–171, 2011. doi: 10.1109/PASSAT/SocialCom.2011.98.

Mahalia Miller, Conal Sathi, Daniel Wiesenthal, Jure Leskovec, and Christopher Potts. Sentiment flow through hyperlink networks. In Lada A. Adamic, Ricardo A. Baeza-Yates, and Scott Counts, editors, *ICWSM*. The AAAI Press, 2011. URL http://dblp.uni-trier.de/db/conf/icwsm/icwsm2011.html#MillerSWLP11.

Arturo Montejo-Raez, Eugenio Martínez-Cámara, Mara-Teresa Martin-Valdivia, and Luis-Alfonso Ureña-López. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In *Proc 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 3–10, 2012.

Chitra Nasa and Suman. Evaluation of different classification techniques for web data. *International Journal of Computer Applications*, 52(9):34–40, August 2012. Published by Foundation of Computer Science, New York, USA.

Federico Neri, Paolo Geraci, and Furio Camillo. Monitoring the web sentiment, the italian prime minister's case. In Nasrullah Memon and Reda Alhajj, editors, *Advances in Social Networks Analysis and Mining (ASONAM), 2010 International Conference on*, pages 432–434. IEEE Computer Society, 2010. ISBN 978-0-7695-4138-9. URL http://dblp.uni-trier.de/db/conf/asunam/asonam2010.html#NeriGC10.

M. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, Oct 2002. doi: 10.1103/PhysRevLett.89.208701. URL http://link.aps.org/doi/10.1103/PhysRevLett.89.208701.

M. Newman. *Networks. An Introduction.* Oxford University Press, 2010.

Bruno Ohana and Brendan Tierney. Sentiment classification of reviews using sentiwordnet. In *Proceedings of the 9th IT & T Conference.* Dublin Institute of Technology, 2009.

Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435: 814 – 818, 6/2005 2005. ISSN 0028-0836. doi: 10.1038/nature03607.

Georgios Paltoglou and Mike Thelwall. Twitter, myspace, digg: Unsupervised sentiment analysis in social media. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(4):1–19, 2012.

Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January 2008. ISSN 1554-0669. doi: 10.1561/1500000011. URL http://dx.doi.org/10.1561/1500000011.

Slav Petrov, Dipanjan Das, and Ryan T. McDonald. A universal part-of-speech tagset. *CoRR*, abs/1104.2086, 2011.

S. Poomagal and T. Hamsapriya. K-means for search results clustering using url and tag contents. In *Proc. International Conference on Process Automation, Control and Computing (PACC)*, pages 1–7, 2011.

M. F. Porter. Readings in information retrieval. chapter An Algorithm for Suffix Stripping, pages 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1997. ISBN 1-55860-454-5. URL http://dl.acm.org/citation.cfm?id=275537.275705.

Mason A. Porter, Peter J. Mucha, M. E. J. Newman, and Casey M. Warmbrand. A network analysis of committees in the u.s. house of representatives. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7057–7062, 2005. doi: 10.1073/pnas.0500191102. URL http://www.pnas.org/content/102/20/7057.abstract.

Rudy Prabowo and Mike Thelwall. Sentiment analysis: A combined approach. *Journal of Informatics*, 3(2):143–157, 2009.

George Psathas. The general inquirer: Useful or not? *Computers and the Humanities*, 3(3):163–174, 1969. ISSN 0010-4817. doi: 10.1007/BF02401609. URL http://dx.doi.org/10.1007/BF02401609.

Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international jont conference on Artifical intelligence*, IJCAI'09, pages 1199–1204, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc. URL http://dl.acm.org/citation.cfm?id=1661445.1661637.

Juliano Rabelo, Ricardo B. C. Prudencio, and Flavia Barros. Collective classification for sentiment analysis in social networks. In *Proceedings of the 2012 IEEE 24th*

International Conference on Tools with Artificial Intelligence - Volume 01, ICTAI '12, pages 958–963, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-0-7695-4915-6. doi: 10.1109/ICTAI.2012.135. URL http://dx.doi.org/10.1109/ICTAI.2012.135.

Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009. ISBN 0387981969, 9780387981963.

Chris Reed and Glenn Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools*, 14:961–980, 2004.

Martin Ringsquandl and Dušan Petković. Expanding opinion lexicon with domain specific opinion words using semi-supervised approach. In *In Proceedings of BRACIS WTI, Curitiba, Brasil, October 2012*, 2012.

Edwina Rissland. I yield one minute...: an analysis of the final speeches from the house impeachment hearings. In *Proceedings of the 7th international conference on Artificial intelligence and law*, ICAIL '99, pages 25–35, New York, NY, USA, 1999. ACM. ISBN 1-58113-165-8. doi: 10.1145/323706.323712. URL http://doi.acm.org/10.1145/323706.323712.

Bertil Rolf and Charlotte Magnusson. Developing the art of argumentation- a software approach. In *in Proceedings of the 5th International Conference on Argumentation. International Society for the Study of Argumentation (ISSA-2002)*, pages 919–926, 2002.

Zaher Salah, Frans Coenen, and Davide Grossi. Extracting debate graphs from parliamentary transcripts: A study directed at UK House of Commons debates. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law (ICAIL 2013)*, pages 121–130, Rome, Italy, 2013a.

Zaher Salah, Frans Coenen, and Davide Grossi. A data mining approach to extracting debate graphs. In Katie Atkinson, Henry Prakken, and Adam Wyner, editors, *From Knowledge Representation to Argumentation in AI, Law and Policy Making. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday*, pages 79–96. College Publications, 2013b. ISBN 978-1-84890-133-9.

Oliver Scheuer, Frank Loll, Niels Pinkwart, and Bruce M.McLaren. Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-Supported Collaborative Learning*, 5(1):43–102, January 2010. ISSN 1556-1607, 1556-1615. doi: 10.1007/s11412-009-9080-x. URL http://www.springerlink.com/content/j3p581601n3x1200/fulltext.html.

Mohammadreza Shams, Mohammadtaghi Saffar, Azadeh Shakery, and Heshaam Faili. Applying sentiment and social network analysis in user modeling. In *Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part I*, CICLing'12, pages 526–539, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-28603-2. doi: 10.1007/978-3-642-28604-9_43. URL http://dx.doi.org/10.1007/978-3-642-28604-9_43.

Vivek Kumar Singh, Rakesh Adhikari, and Debanjan Mahata. A clustering and opinion mining approach to socio-political analysis of the blogosphere. In *Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on*, pages 1–4, 2010a. doi: 10.1109/ICCIC.2010.5705807.

Vivek Kumar Singh, Debanjan Mahata, and Rakesh Adhikari. Mining the blogosphere from a socio-political perspective. In *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*, pages 365–370, 2010b. doi: 10.1109/CISIM.2010.5643634.

Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4):470 – 479, 2012. ISSN 0740-624X. doi: http://dx.doi.org/10.1016/j.giq.2012.06.005. URL http://www.sciencedirect.com/science/article/pii/S0740624X12000901.

Swapna Somasundaran, Galileo Namata, Lise Getoor, and Janyce Wiebe. Opinion graphs for polarity and discourse classification. In *Proceedings of the 2009 Workshop on Graph-based Methods for Natural Language Processing*, TextGraphs-4, pages 66–74, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. ISBN 978-1-932432-54-1. URL http://dl.acm.org/citation.cfm?id=1708124.1708138.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press, Cambridge, MA, 1966.

George Stylios, Dimitris Christodoulakis, Jeries Besharat, Maria-Alexandra Vonitsanou, Ioanis Kotrotsos, Athanasia Koumpouri, and Sofia Stamou. Public opinion mining for governmental decisions. *Electronic Journal of e-Government*, 8(2):202–213, 2010.

Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD

'11, pages 1397–1405, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0813-7. doi: 10.1145/2020408.2020614. URL http://doi.acm.org/10.1145/2020408.2020614.

Rafael Lage Tavares, Mariano Gomes Pimentel, and Renata Mendes de Araujo. Information visualization in political discussions. In *SBSC*, pages 75–83. IEEE Computer Society, 2012. ISBN 978-1-4673-4696-2. URL http://dblp.uni-trier.de/db/conf/sbsc/sbsc2012.html#TavaresPA12.

Mike Thelwall and Kevan Buckley. Topic-based sentiment analysis for the social web: The role of mood and issue-related words. *Journal of the American Society for Information Science and Technology*, 64(8):1608–1617, 2013. ISSN 1532-2890. doi: 10.1002/asi.22872. URL http://dx.doi.org/10.1002/asi.22872.

Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, 2008. ISBN 978-1-59749-272-0.

Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Empirical Methods in Natural Language Processing (EMNLP'06)*, pages 327–335. Association for Computational Linguistics, 2006.

T.Sujatha, Ramesh Naidu G, and P.Suresh B. Measuring semantic similarity between words using web pages. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(3):213–218, 2012.

Mikalai Tsytsarau and Themis Palpanas. Towards a framework for detecting and managing opinion contradictions. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 1219–1222, 2011. doi: 10.1109/ICDMW.2011.167.

Giannis Varelas, Epimenidis Voutsakis, Euripides G. M. Petrakis, Evangelos E. Milios, and Paraskevi Raftopoulou. Semantic similarity methods in wordnet and their application to information retrieval on the web. In *In: 7 th ACM Intern. Workshop on Web Information and Data Management (WIDM 2005)*, pages 10–16. ACM Press, 2005.

Fernanda B. Viegas, Martin Wattenberg, and Jonathan Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1137–1144, November 2009. ISSN 1077-2626. doi: 10.1109/TVCG.2009.171. URL http://dx.doi.org/10.1109/TVCG.2009.171.

Changbo Wang, Zhao Xiao, Yuhua Liu, Yanru Xu, Aoying Zhou, and Kang Zhang. Sentiview: Sentiment analysis and visualization for internet popular topics. *Human-*

Machine Systems, IEEE Transactions on, 43(6):620–630, 2013. ISSN 2168-2291. doi: 10.1109/THMS.2013.2285047.

Franz Wanner, Christian Rohrdantz, Florian Mansmann, Daniela Oelke, and Daniel A. Keim. Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008. In *Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009), IUI2009, Feb 8 2009, Sanibel Island, Florida, USA*, 2009.

Albert Weichselbraun, Stefan Gindl, and Arno Scharl. Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 1053–1060, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0717-8. doi: 10.1145/2063576.2063729. URL http://doi.acm.org/10.1145/2063576.2063729.

Susan Welch. Are women more liberal than men in the u.s. congress? *Legislative Studies Quarterly*, 10(1):125–134, 1985.

Yorick Wilks and Mark Stevenson. The grammar of sense: Using part-of-speech tags as a first step in semantic disambiguation. *Natural Language Engineering*, 4:135–143, 5 1998. ISSN 1469-8110. doi: null. URL http://journals.cambridge.org/article_S1351324998001946.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. HLT/EMNLP-2005*, pages 347–354, 2005.

Ian H. Witten, Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham. Weka: Practical machine learning tools and techniques with Java implementations. In Nikola Kasabov and Kitty Ko, editors, *Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems*, pages 192–196, 1999. Dunedin, New Zealand.

Arjumand Younus, M. Atif Qureshi, Fiza Fatima Asar, Muhammad Azam, Muhammad Saeed, and Nasir Touheed. What do the average twitterers say: A twitter model for public opinion analysis in the face of major political events. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 618–623, 2011. doi: 10.1109/ASONAM.2011.85.

Jing Zhang and Qinke Peng. Constructing chinese domain lexicon with improved entropy formula for sentiment analysis. In *Information and Automation (ICIA), 2012 International Conference on*, pages 850–855, 2012. doi: 10.1109/ICInfA.2012.6246900.