Issue Communication by Political Parties on Twitter

Stiene Praet
Applied Data Mining Research Group, Department of Engineering Management, University of Antwerp

Walter Daelemans
Computational Linguistics and Psycholinguistics Research Group, Department of Linguistics, University of Antwerp

Tim Kreutz
Computational Linguistics and Psycholinguistics Research Group, Department of Linguistics, University of Antwerp

Peter Van Aelst
Media, Movements and Politics Research Group, Department of Political Sciences, University of Antwerp

Stefaan Walgrave
Media, Movements and Politics Research Group, Department of Political Sciences, University of Antwerp

David Martens
Applied Data Mining Research Group, Department of Engineering Management, University of Antwerp

ABSTRACT

In political science, the theory of issue ownership is increasingly used to study party competition and voting behavior. According to this theory, political parties focus their communication on specific issues for which they have a reputation of competence. Existing studies on issue ownership mainly analyze party manifestos, campaign ads and press releases using dictionary-based methods. Since Twitter is more and more used both as a medium for political communication, and as a journalistic source, we propose a new data mining-based methodology to discover the topics that political parties focus on in their communication on Twitter. Our analysis introduces data mining concepts in this domain, from topic modeling to text classification, and is performed on seven Flemish political parties in the winter of 2017. We answer two main questions: which topics do parties communicate about on Twitter that discriminate them from other parties, and how consistent is their communication on these topics? The results show that parties indeed communicate about the issues they own, in line with the existing theory about issue ownership. However, most parties seem to trespass and tweet about other issues too, with extreme left and right parties communicating much more consistently about their issues than the moderate parties. Furthermore, our study reveals additional complexities in party communication on Twitter, including event-driven communication, stylistic differences and the influence of party characteristics, which could have important implications for political scientists and journalists using social media data.

CCS CONCEPTS

• **Applied computing → Sociology**; *Document analysis*; • **Information systems** → Document topic models; Content analysis and feature selection; Dictionaries;

DSJM Workshop at KDD '18, August 2018, London, United Kingdom

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

KEYWORDS

Issue ownership, Political parties, Twitter, Topic detection, Text classification

ACM Reference Format:
Stiene Praet, Walter Daelemans, Tim Kreutz, Peter Van Aelst, Stefaan Walgrave, and David Martens. 2018. Issue Communication by Political Parties on Twitter. In *Proceedings of KDD Workshop on Data Science, Journalism and Media (DSJM) (DSJM Workshop at KDD '18).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In politics, *issue ownership* substantially influences political competition and voter behavior. The theory of issue ownership states that a specific party is considered by the public at large as the most competent and/or most committed to deal with a specific issue [27]. A classic example are green parties who are considered by voters in many countries as the best party to deal with the policy issue of environmental protection. The voters trust green parties more than any other party to do a good job with regard to the environment. The theory states, supported by empirical evidence [4], that if owned issues are high among the priorities of the voters, chances increase that voters will vote for the owning party.

Hence, it is in a party's interest to make sure that the issues it owns are high on the priority list of voters. That is why parties tend to focus on the owned issues in their communication: by talking about it, they hope to impact the importance of the issue on the 'public agenda'. This way, issue ownership is supposed to not only affect the voters (and their voting) but also parties' own communicative behavior. The evidence with regard to issue ownership and party communication are more mixed, though. While some studies find that the parties indeed focus on their issues, others show that parties 'trespass' frequently and also address issues owned by their competitors [6]. The findings of previous work seem to be dependent on the type of communication channel one looks at. Previous work mainly analyzed party manifestos, campaign ads and press releases [25].

An important communication channel has largely been neglected in existing studies on issue ownership: social media. Especially Twitter is more and more used by politicians and political parties to communicate with voters and media [26]. Politicians use Twitter to inform their audience and influence the public perception on their party [15]. Journalists, at their turn, tend to use social media messages to keep up with what is going on in society, complementary to offline news sources [18]. In sum, voters are influenced by the issues addressed in political tweets, both directly and through journalistic intervention.

In this work, we analyze issue ownership and communication by Flemish political parties on Twitter and try to answer two main questions: 1) which topics do parties communicate about that discriminate them from other parties and 2) how consistently do they communicate about these topics? We define topics in function of words and built classification models to find the most discriminative topics per party. The main methodological contributions of this paper are twofold: 1) we develop a new method, based on text mining and classification techniques, to analyze political party communication on issues and 2) we apply this method to Twitter data in Flanders to test the theory of issue ownership for this communication channel.

2 RELATED WORK

Studies about political communication and issue ownership in social media channels are missing in the current literature. In what follows, we provide a brief overview of existing work on issue ownership using Twitter and other textual sources. Thereafter, we summarize established text classification methods in the field of political science and for the classification of tweets.

Issue ownership. Guo and Vargo analyze issue ownership on Twitter during the 2012 United States presidential election [11]. Tweets from media and citizens are collected that mention Obama or Romney. When one of the 16 predefined issues (health, education, labor, etc.) is mentioned in a tweet together with one of the election candidates, a link is created in the 'issue ownership network'. Additionally, sentiment analysis shows whether the candidate is perceived as competent or not in handling the issue. This study tries to draw conclusion on the *voter's perception* of issue ownership, whereas in our work we aim to examine *party communication* about issue ownership on Twitter. Another important difference is that the Flemish political system is much more fragmented, with seven major parties compared to only two (liberals and conservatives) in the United States, leading to increased complexity and granularity of the results.

To define issues in political texts, researchers often refer to the Comparative Agendas Project (CAP) codebook, consisting of 21 major topics and more than 200 subtopics [1]. Sevenans et al. manually compiled a Dutch dictionary of indicator words for each of the 21 CAP topics and showed it performs relatively well for topic classification [23]. We will use this dictionary in this work. Issue communication has been analyzed before in party manifestos, campaign ads and press releases [25] and was largely based on manual encoding of the documents or counting word frequencies. These methods have the limitation that they are based on the *frequency* of communication about a certain issue. If all parties talk a lot about a certain issue, it is not inherent to a particular party's communication and will therefore not influence the priorities of voters in favor of a particular party according to issue ownership theory. We try to solve this problem by applying classification models that discriminate between political parties based on Twitter-content. The models automatically identify the most discriminative topics per party.

Text classification in politics. Grimmer & Stewart argue that the understanding of language to know what political actors are saying and writing is central to the study of politics [10]. Yet, the *volume* of existing political texts does not allow for the manual reading and interpretation of all these documents. Automated content methods however, can make the systematic analysis of large-scale text collections possible. For the classification of political texts typically two methods are considered: *dictionary methods*, based on the relative frequency of predefined keywords in a document and *supervised learning methods* where the algorithm learns to classify documents into categories using a labeled training set [10]. In this work, a dictionary-based method will be used to assign words to predefined expert topics. Supervised learning will be used to train classifiers on labeled data to predict political party from the text of a tweet.

An important limitation of dictionary methods is of course that they depend on the quality of the predefined keywords and that dictionaries are of limited length, thus unable to capture all possible words related to a certain topic. This is especially an issue when working with short texts such as tweets, as the probability for dictionary words to appear in such a short text is low [28]. To overcome the drawbacks of dictionaries, supervised learning has become popular for political applications such as topic classification [8], sentiment analysis [21], measuring ideological positions [24] and predicting political affiliation [5]. However, the use of supervised learning to draw conclusions on issue ownership is new.

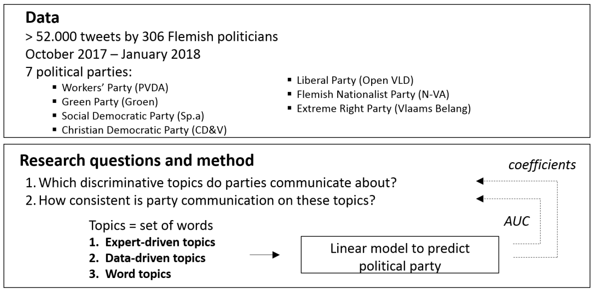
Tweet classification. Often-used methods for text classification are Logistic Regression, Support Vector Machines and Naive Bayes [21]. Also recently, different variations of neural networks have been proposed for text classification [17]. Instead of representing documents as a Bag of Words (BoW), feature processing can be done using topic modeling techniques such as Latent Dirichlet Allocation (LDA) [29], (Probabilistic) Latent Semantic Analysis ((p)LSA) [13], Non-negative Matrix Factorization (NMF) [9] and variants thereof. Albeit useful to discover hidden topic structures in the data, topic detection techniques do not always improve classification performance, especially when working with short texts [5]. Generally, classification of short texts has been shown to be more challenging than large documents because of the data sparseness. Several methods have been proposed to semantically expand the texts using large background corpora such as Wikipedia or other external databases [14, 22] or the results returned by a web search engine (e.g. Google) [2]. Word embeddings capture both the semantic and syntactic context of words and can therefore be used to find semantically similar words [19]. We will extend the Dutch CAP dictionary using word embeddings trained on political texts.

3 METHODOLOGY

We have collected tweets from seven Flemish political parties and their elected politicians. Every tweet is composed of words and has a label attached indicating from which political party the tweet stems. Topics are defined in function of words and classification

[1]http://www.comparativeagendas.net/pages/master-codebook

Figure 1: Overview of our methodology to investigate issue communication by political parties on Twitter.



models are built to predict the author's political party based on the topics of a tweet. The specifications of the trained models are investigated to draw conclusions on issue communication per political party. The more discriminative the topics (e.g. by looking at the coefficients of a linear model), the more a party communicates about this topic on Twitter, compared to other parties. The higher the discriminative power of the model, in terms of Area Under the Receiver Operating Characteristic Curve (AUC), the more consistent the party's communication or in other words, the more they stick to their issues. The steps of our methodology are further elaborated in the sections below and represented in Figure 1.

3.1 Data collection

For a time period of three months between October 2017 and January 2018, we collected tweets from the official Twitter accounts of the seven main Flemish political parties: the Workers' Party (PVDA), the Green Party (Groen), the Social Democratic Party (Sp.a), the Christian Democratic Party (CD&V), the Liberal Party (Open VLD), the Flemish Nationalist Party (NVA) and the Extreme Right Party (Vlaams Belang) and all their party members elected in the national or regional parliament including cabinet ministers and party leaders. This resulted in a dataset of more than 52.000 tweets by 306 individual politicians and 7 parties.

3.2 Preprocessing of tweets

Since the main interest of this research is to see how word usage in tweets might relate to political topics, we aim to reduce the event-specific information the tweets contain. Through intensive preprocessing we also want to reduce the noise that is common to social media texts [12].

Frog is a feature-rich natural language toolkit for Dutch [3] that is suitable towards most of our ends. Using Frog, tweets were first split into tokens and non-alphanumeric characters were removed. For Twitter specifically, this means that hashtags lost their '#'-prefix and were then handled as any other word. The use of user mentions, numbers and URLs in tweets is commonplace and might be informative for certain political topics; numbers playing an important role in financial news for example. However, we are not interested in the specific user, number or URL since it is unlikely that we can generalize from these. For that reason, these tokens were replaced with distinct placeholders.

Similarly, we argue that specific named entities (NE) in tweets are less informative to detect general political topics. Using these words as features will most likely cause our system to model events that occurred in the three months of our data collection and are indicative of a certain political party, rather than the more general political topics that would be comparable to the expert dictionary. However, when it comes to named entities, the type of entity can still be informative for our purposes. Frequent mentioning of locations, for example, could be more indicative of topics like foreign affairs or defense, while frequent occurrence of organizations and products could relate to national economy. Fortunately, Frog allows for fine-grained tagging of named entities. We distinguish six types of named entities, namely: locations, persons, organizations, products, events and miscellaneous, and replace them with their respective placeholders. To assess how named entities would influence our results, we have also repeated the same experiments for the data with named entities included.

Lastly, we reduce word variations by lemmatizing the remaining tokens. We are only interested in the lemma form of words because we aim to model their relatedness to political topics, regardless of their inflectional form. The expert dictionary also contains lemma forms, which makes for easy comparison.

3.3 Defining topics

Before the actual modeling can start, tweets are transformed to a topic representation. For this, topics need to be defined in function of words. This will be done in three different ways, ranging from top-down to bottom-up. Finally, we assess how well these defined topics reflect the 21 CAP topics (Table 1).

3.3.1 *Expert-driven topics.*

In the first method, topics are defined by words that are carefully chosen by domain experts. We will use the Dutch CAP dictionary compiled by Sevenans et al [23].

The dictionary maps keywords to their respective political topics and aims to be very precise, with keywords having a very distinct meaning and low probability to be present in one of the other topics. For analysis of short social media texts such as tweets, in which very few words are present, this precision is less important and coverage with the expert dictionary is of more concern. To extend the indicator words in the original dictionary, we use word embeddings trained on a large corpus of political social media data [16]. The word embeddings encode a numerical vector per word, which contains the point-wise mutual information (PMI) with other words in the corpus. Using these vectors, we can find candidate words that are semantically similar to the keywords already present in the dictionary, using a cosine-similarity of 0.6 or higher. The candidates were then manually inspected and filtered to contain only words that extend coverage of the expert topics without clearly impairing their delineation. Using word embeddings in this way, we were able to extend the keywords from an average of 87 per expert topic to 157 per expert topic. Based on the extended dictionary, every tweet in our collection is transformed to the 21 CAP topics and classification models are built on this representation. A drawback of this expert-driven approach however, is that even with the extended dictionaries the coverage remained very low which resulted in many zero input features.

Table 1: Overview of the 21 CAP topics [23].

Code	Topic
t100	Macroeconomics
t200	Human rights
t300	Health
t400	Agriculture
t500	Labor and employment
t600	Education
t700	Environment
t800	Energy
t900	Immigration
t1000	Transportation
t1200	Law and crime
t1300	Social welfare
t1400	Community development and housing
t1500	Banking, finance and domestic commerce
t1600	Defense
t1700	Space, science, technology and communications
t1800	Foreign trade
t1900	International affairs and foreign aid
t2000	Government operations
t2100	Public lands and water management
t2300	Culture and arts

3.3.2 Data-driven topics. To define topics in function of words based on the data, two well-known topic modeling techniques were applied: Latent Dirichlet Allocation (LDA) and Non-negative Matrix Factorization (NMF). The basic assumption behind LDA is that each documents contains a random mixture of topics, where each topic is represented by a distribution over words [1]. LDA is a probabilistic graphical model that tries to infer the underlying topics present in a collection of documents based on the observed words. NMF is applied in multiple domains to decompose a non-negative matrix into two non-negative matrices. In the context of topic modeling, the term-document matrix is represented by two matrices, one containing the topics and one containing the coefficients to approximate the original matrix as close as possible [20]. After building the topics from the data, the tweets are represented by the data topics as input for our classification models.

In our setting, the classification models based on the topics produced with NMF achieved higher discriminative power than with LDA, which is why we will report results using the NMF topics. We have tried both the political tweet collection and a larger background collection to built the topics. The larger collection consists of more than 160.000 tweets, stemming from not only Flemish politicians but from Flemish media channels and political journalists as well. Unfortunately, it did not lead to more interpretable or more accurate results than topic detection on the political tweets only. With 10-fold cross validation optimizing classification performance, the number of topics was set to 300, which is considerably higher than the 21 expert topics. Our data-driven topics are thus much more specific than the expert topics.

3.3.3 Word topics. The final method uses a basic bag of words approach and represents each topic by one word only. Thus, the number of topics in this setting is equal to the number of words in our collection (81.653). Topics (words) are transformed into a numerical matrix using term frequency-inverse document frequency (tf-idf). Classification models are built on this numerical matrix.

3.3.4 Interpretability of the defined topics. We defined topics in three ways, ranging from top-down to bottom-up with different levels of intuitiveness. We define interpretability as the extent to which the topics correspond with the 21 CAP topics.

(1) In a first attempt we approximate the interpretability of our defined topics by comparing the words in our topics to the words in the extended CAP dictionary. This comes down to the *precision* of our topics, i.e. how many of the words in our topic, also appear in the dictionary? This measure will be referred to as Int_{dict}.
(2) Only an extremely small part of the words used in tweets does also appear in the CAP dictionaries, even though they have been expanded with pre-trained word embeddings, resulting in a very low precision. Therefore we introduce another metric for interpretability based on the opinion of *domain experts*. We asked domain experts to manually select the words in our topics that belong to one of the 21 CAP topics. This results in a second measure for precision (referred to as Int_{DE}), that is less objective than the one-on-one match with existing dictionaries but does account for words that are not (yet) included in the dictionary. For example, the word 'glyphosate' (the chemical substance found in Roundup weedkiller) is clearly related to the topic Environment but is not included in the extended CAP dictionary.

3.4 Model building

Per political party, a classification model is built to predict whether the author of the tweet belongs to the political party or not, based on the topics of the tweet. From the results of the classification models we want to gain insights into which topics are most discriminative for each of the seven parties. For this reason we choose to work with Logistic Regression, as the coefficients of this model are straightforward to interpret. Moreover, the discriminative power of the Logistic Regression model showed higher or similar to the other classifiers in our benchmark, including (Multilayer) Perceptron, Logistic Regression with L1 or L2 penalization, Linear Regression, Decision Tree, Random Forest, Linear Support Vector Machine with L1 or L2 penalty, Multinomial and Bernoulli Naive Bayes, for the three different settings. A finding that is in line with the benchmark study of De Cnudde et. al. [7], comparing 11 classification techniques on 43 fine-grained behavioral data sets and showing that Logistic Regression with L2 regularization is the best performing technique overall in terms of AUC. More specifically, we have used Logistic Regression with L2 regularization and the liblinear solver from the scikit learn module in Python.

The last 20% of the tweets in our dataset was used as a separate out-of-time holdout set to report the discriminative power of each model. The model parameters were optimized per model using 10-fold out-of-time cross validation, also called 'evaluation on a rolling forecasting origin'. In our setting, the training data was split in 10 folds and first the 10th fold is used as a validation set while the previous folds are used for training, then the 9th fold is used for validation, etc. The regularization parameter (C) was optimized in a range from 0.001 till 1000. For the data topic-based models, C is simultaneously optimized with the number of topics, ranging from 0 till 400 with a stepsize of 50.

4 RESULTS

In the following sections, we present the results of our Twitter study on political communication in Flanders. The first question "which topics do parties communicate about that discriminate them from other parties", is answered by looking at the top three most discriminative topics per party. If the theory about increased communication about a party's owned issues is true, the three most discriminative topics per party should be topics owned by the political party. Otherwise, this could point into the direction of the alternative theory about 'trespassing' issues. The discriminative power of the models provides us with an answer to the second question "how consistently do parties communicate about these topics". We will start first with the evaluation of our three different topic definitions.

4.1 Comparison of topic definitions

The classification models are built on topics defined in three different ways, based on expert-driven topics, data-driven topics and words. When comparing these three approaches (Table 2), a trade-off between discriminative power of the classifiers and interpretability of the results is discovered. The word-based model achieves the highest classification performance while the expert topics offer the most direct interpretation of the CAP topics (Figure 2).

As mentioned before, the overlap between the words used in tweets and the words from the CAP dictionary is small. Therefore, when we defined topics in terms of the CAP dictionary, a large portion of the tweets could not be assigned any topic and received a zero input vector. This resulted in a low average AUC for the classification models based on expert topics, which was 57% for both the models with and without NE. On the other hand, results are 100% interpretable as the topics are constructed top-down from the CAP dictionary itself.

The average weighted AUC for the data topic-based models is 62% and 60% for the data with and without NE respectively. Per party we look at the 3 most discriminative topics (each represented by 15 words) and match them with the most corresponding CAP topic (an example for the green party is shown in Table 4). The dictionary-based interpretability is 7% without NE and 5% with NE. As discussed in Section 3.3.4 we also asked domain experts to manually count the topic words belonging to the CAP topics. We only did this for the results without NE. This way, the interpretability was increased to 21%.

With an average weighted AUC of 65% without NE and 70% with NE, the models based on words achieve the highest classification performance. For these models we showed the first 45 most discriminative words and matched the 3 most corresponding CAP topics to these words (an example is shown in Table 5). The average weighted interpretability, measured in terms of dictionary precision is 3% for both with and without NE. After consulting domain experts, the interpretability was set at 53%. It is striking that the words were easier to interpret by a domain expert than the data-topics, while for the dictionary interpretability it was the other way around.

4.2 Which discriminative topics do parties communicate about?

For every party, the most discriminative topics (i.e. the topics with the highest coefficients in our linear model) are the topics the party communicates significantly more about then their competitors. The topics per party (Table 3) differ slightly for our three different settings but generally we can state that they are consistent with what we would expect in Flanders: At the left spectrum, the Workers' - and Socialist Party (PVDA, Sp.a) mainly communicate about the topics Labor and employment, Macroeconomics, Social welfare and Health and the Green Party (Groen) talks about Environment and Energy while at the right spectrum, the Flemish Nationalist Party (NVA) talks about Government operations and the Extreme Right Party (Vlaams Belang) about Law and crime, Immigration and Civil rights. At the center, the Christen Democratic Party (CD&V) and Liberal Party (Open Vld) seem to communicate about several topics including Education, Macroeconomics, Health and Civil rights.

Tresch et. al. [25] summarized several studies on issue ownership perceptions of Flemish respondents. We base ourselves on this study to compare the owned issues per party in Flanders (i.e. the voter's perception) to the party's issue communication on Twitter. According to these results, the extreme parties (PVDA-left and Vlaams Belang-right) focus mainly on their owned issues in their Twitter communication. The central parties on the other hand, talk about several topics (leading to less consistent results over our three different settings) and seem to 'trespass' their owned topics. For example, the topic Energy is not owned by the Liberal Party Open Vld but they do have a minister for Energy, which might be the reason for heavy communication. Also, opposition parties communicate about topics as a reaction on government decisions. The topic Defense is not owned by the socialist party Sp.a but in the period of data collection they criticized the government decision to buy fighter planes.

To compare the parties to each other based on their discriminative topics, we also performed hierarchical clustering on the coefficients of the models in our three settings (Figure 3). When the topics were defined before the model building, left and right parties are clustered together. However, when we built the models on the words, providing the most fine-grained information, the (small) differences in communication between opposition parties (PVDA, Groen, Sp.a, Vlaams Belang) and government parties (NVA, open VLD, CD&V) become apparent.

4.3 How consistently do parties communicate?

To assess how consistently parties communicate about their issues we explore the discriminative power of the models per party (Table 2). We assume that high AUC indicates consistent communication for the considered party. As we already suspected from looking at the topics per party in the previous section, the extreme left party (PVDA) is the most consistent in their communication about topics. The models for the extreme left Workers' Party PVDA achieve the highest AUC in all three setting. For the expert-based model the AUC is 68%, meaning that they clearly communicate

Table 2: Classification performance and interpretability of the expert-driven, data-driven and word topics.

Party	Expert-driven topics			Data-driven topics			Word topics		
	AUC	Int_{dict}	Int_{DE}	AUC	Int_{dict}	Int_{DE}	AUC	Int_{dict}	Int_{DE}
PVDA	68%	100%	100%	68%	7%	22%	72%	7%	29%
Groen	54%	100%	100%	58%	11%	40%	64%	4%	40%
Sp.a	58%	100%	100%	60%	9%	47%	66%	4%	44%
CD&V	55%	100%	100%	58%	11%	22%	62%	2%	38%
Open Vld	58%	100%	100%	61%	11%	41%	66%	5%	40%
NVA	57%	100%	100%	61%	2%	0%	65%	1%	42%
Vlaams Belang	59%	100%	100%	61%	11%	42%	70%	7%	47%
Weighted Average	**57%**	**100%**	**100%**	**60%**	**7%**	**21%**	**65%**	**3%**	**53%**

Table 3: The discriminative topics of Flemish political parties on Twitter. Topics printed in bold are owned by the party [25]. A '/' indicates that the topic could not be brought back to one of the CAP topics.

Party	Expert-driven topics	Data-driven topics	Word topics
PVDA	1. **Labor and employment** 2. **Energy** 3. **Macroeconomics**	1. / 2. **Labor and employment** 3. **Macroeconomics**	1. **Social welfare** 2. Civil rights 3. **Labor and employment**
Groen	1. **Environment** 2. **Social welfare** 3. Science and technology	1. **Energy** 2. **Social welfare** 3. Science and technology	1. **Health** 2. **Environment** 3. Civil rights
Sp.a	1. Defense 2. Foreign trade 3. **Macroeconomics**	1. Government operations 2. **Health** 3. **Macroeconomics**	1. **Macroeconomics** 2. Defense 3. **Labor and employment**
CD&V	1. **Education** 2. Environment 3. Foreign trade	1. **Macroeconomics** 2. International affairs 3.**Health**	1. Law and crime 2. Community development 3. Agriculture
Open Vld	1. Energy 2. Community development 3. Culture and arts	1. **Civil rights** 2. **Civil rights** 3. **Macroeconomics**	1. Energy 2. Health 3. **Macroeconomics**
NVA	1. Labor and employment 2. **Government operations** 3. **Law and crime**	1. / 2. / 3. /	1. Energy 2. / 3. /
Vlaams Belang	1. **Immigration** 2. **Civil rights** 3. **Government operations**	1. **Government operations** 2. **Civil rights** 3. **Government operations**	1. **Civil rights** 2. **Immigration** 3. **Law and crime**

about the CAP topics, compared to for example the Christen Democratic Party CD&V (AUC 55%). For all other parties, AUC rises substantially when models are built on all the words instead of predefined topics. This could indicate that their communication is more complex and not reducible to predefined topics. When we look at the word-based models, also the Extreme Right Party (Vlaams Belang) seems to communicate more consistently then the central parties. Two remarks must be made to nuance our conclusions based on AUC: (1) the differences in AUC between parties are relatively small and (2) with the data-driven topic definitions we model other characteristics of party communication rather than just the topics they tweet about. A clear example can be found in the data-driven topics from the Flemish nationalists party NVA. The topics we found do not contain any content information but have grouped stylistic information about NVA's communication,

Figure 2: A comparison of our three methods on both evaluation criteria shows a clear trade-off between interpretability and discriminative power. The word-based model achieves the highest classification performance while the expert topics offers the most direct interpretation of issues.

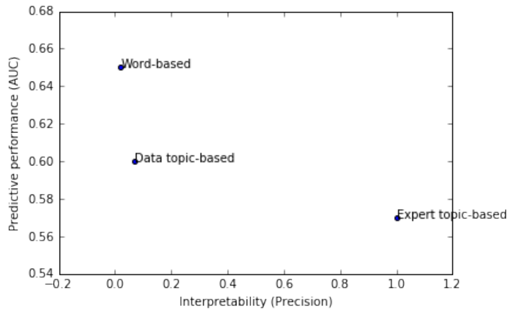
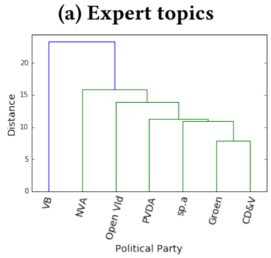
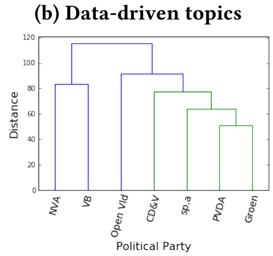


Figure 3: Hierarchical clustering on the coefficients of the models shows that for both models based on topics, right and left parties are clustered together. For the word-based models however, opposition parties are clustered together.

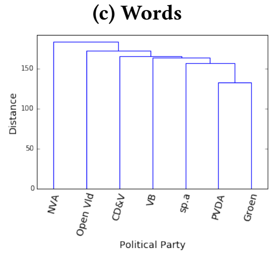
(a) Expert topics



(b) Data-driven topics



(c) Words



which seems to be rather formal with the use of neutral words such as 'point of view' or 'news item'. Also the third NVA-topic consisted of English words (all other topics are in Dutch) and was discriminative for NVA as it is the only party tweeting in English.

5 DISCUSSION AND FUTURE RESEARCH

Using three different definitions of topics we tried to discover which topics are specific to a political party's communication on Twitter. A trade-off between classification performance and interpretability of the three definitions was discovered. The AUC of expert topic-based models is low, suggesting that a lot of information is lost by trying to reduce political communication on Twitter to predefined topics, which is the standard method in political science. AUC for the most fine-grained topics (i.e. based on the words) is highest, especially when including named entities. From examination of the most discriminative words it was clear that communication on Twitter is largely event-driven, with parties talking about and reacting to certain events that are limited in time. This is an important difference compared to for example party manifestos where the party's long-term goals, views and ambitions are communicated [25]. Because of these additional complexities in Twitter communication, together with the very short texts that are used, dictionary-based methods turn out to be insufficient to study this communication channel. Data-driven topics are able to capture more detailed information (including stylistic differences) but are more difficult to interpret. Political scientists and journalists should consider this trade-off when studying political communication on Twitter.

According to our results, especially the extreme parties communicate clearly about the issues they own. Other parties seem to trespass and also communicate about other issues. We have suggested some alternative explanations such as opposition reactions or a minister concerned with the specific issue, which can be further validated by political scientists. Secondly, the expansion of topic analysis with sentiment analysis would offer a much more complete view upon political communication. Political parties might communicate about the same subject but with an opposite sentiment. By looking at the AUC of our models we drew conclusions on the consistency of political communication. Twitter is a much more personal communication channel than manifestos or press releases and individual politicians are free to tweet what they want. This would explain the rather low AUCs. An interesting issue for future research might be to look into how aligned party members are in their communication, and investigate party member characteristics (popularity, seniority, etc.) to explain the differences.

6 CONCLUSION

We propose a new data mining-based methodology to discover the topics that political parties communicate about on Twitter. Our analysis in terms of expert and data-driven topics reveals that overall, political parties do indeed talk about the expected issues owned, confirming the existing theory on issue ownership and communication. However, our evaluation in terms of discriminative power (AUC) and interpretability of the topics reveals several interesting nuances: except for the extreme parties, most parties seem to trespass and tweet about several other issues too. The low discriminative power of the models indicate that communication on Twitter is less well-considered than communication via offline media channels. Additional complexities in party communication on Twitter that are revealed include: event-driven communication (e.g. about local news), stylistic differences (e.g. the use of the English language), and the influence of party characteristics, such as opposition versus ruling parties, and left versus right. A further investigation of these findings on Twitter and other communication channels, as well as a validation thereof in different countries are interesting opportunities for future research. Finally, we hope that this work will motivate the broader use of data mining in political science, and will spark further multidisciplinary collaborations to reveal insights into communication in the political world.

Table 4: The most discriminative topics for the Green Party (Groen) when using the data topic-based models.

Party	Most discriminative topics	CAP topics
Groen	1. green, energy, renewable, light, rapid, amongst others, target, cover, red, congress, alternative, flow, fetch, sustainable, launch	1. Energy
	2. time, unemployment benefit, urgent, old, minute, avoid, limit, number, action, long, train, money, plead, policy, think	2. Social welfare
	3. show, committee, golden, study, itcanbedifferent, tip, file, dialogue, clear, government, result, painful, joint, especially, faction	3. Science and technology

Table 5: The most discriminative words for the Green Party (Groen) when using the word-based models.

Party	Most discriminative words	CAP topics
Groen	itcanbedifferent, congress, glyphosates, doer, welfare guarantee, BE invest, whistle blower, waiting list, new hope, on behalf of, air pollution, unworthy, fuckthesideline, craving, hormone disruptor, punk, hurt, food, silenced, living together, audit, surrounding, air quality, agree, glyphosate, presence, longlivepolitics, ships, podcast, climate policy, healthy, wave, sexual, supplement, grave, deposit, soil, youth aid, demonstrating, progressive, moon, green, psychic	1. Health 2. Environment 3. Civil rights

REFERENCES

[1] David M Blei and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*. ACM, 113–120.

[2] Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka. 2007. Measuring semantic similarity between words using web search engines. *www* 7 (2007), 757–766.

[3] Antal van den Bosch, Bertjan Busser, Sander Canisius, and Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. *LOT Occasional Series* 7 (2007), 191–206.

[4] Ian Budge and Dennis Farlie. 1983. *Explaining and predicting elections: Issue effects and party strategies in twenty-three democracies.* Taylor & Francis.

[5] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on.* IEEE, 192–199.

[6] David F Damore. 2005. Issue convergence in presidential campaigns. *Political Behavior* 27, 1 (2005), 71–97.

[7] Sofie De Cnudde, David Martens, Theodoros Evgeniou, and Foster Provost. 2017. A benchmarking study of classification techniques for behavioral data. (2017).

[8] Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto. 2017. Cross-lingual classification of topics in political texts. In *Proceedings of the Second Workshop on NLP and Computational Social Science.* 42–46.

[9] Derek Greene and James P. Cross. 2016. Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach. *CoRR* abs/1607.03055 (2016). arXiv:1607.03055 http://arxiv.org/abs/1607.03055

[10] Justin Grimmer and Brandon M Stewart. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis* 21, 3 (2013), 267–297.

[11] Lei Guo and Chris Vargo. 2015. The power of message networks: A big-data analysis of the network agenda setting model and issue ownership. *Mass Communication and Society* 18, 5 (2015), 557–576.

[12] Bo Han and Timothy Baldwin. 2011. Lexical Normalisation of Short Text Messages: Makn Sens a #Twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11).* Association for Computational Linguistics, Stroudsburg, PA, USA, 368–378. http://dl.acm.org/citation.cfm?id=2002472.2002520

[13] Thomas Hofmann. 1999. Probabilistic latent semantic analysis. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence.* Morgan Kaufmann Publishers Inc., 289–296.

[14] Xia Hu, Nan Sun, Chao Zhang, and Tat-Seng Chua. 2009. Exploiting Internal and External Semantics for the Clustering of Short Texts Using World Knowledge. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09).* ACM, New York, NY, USA, 919–928. https://doi.org/10.1145/1645953.1646071

[15] Nigel A Jackson and Darren G Lilleker. 2009. Building an architecture of participation? Political parties and Web 2.0 in Britain. *Journal of Information Technology & Politics* 6, 3-4 (2009), 232–250.

[16] Tim Kreutz and Walter Daelemans. 2018. Enriching general sentiment lexicons for domain-specific polarity classification. *Manuscript in preparation* (2018).

[17] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent Convolutional Neural Networks for Text Classification.. In *AAAI*, Vol. 333. 2267–2273.

[18] Sophie Lecheler and Sanne Kruikemeier. 2016. Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources. *New Media & Society* 18, 1 (2016), 156–171.

[19] Quanzhi Li, Sameena Shah, Xiaomo Liu, and Armineh Nourbakhsh. 2017. Data Sets: Word Embeddings Learned from Tweets and General Data. *arXiv preprint arXiv:1708.03994* (2017).

[20] Derek O'Callaghan, Derek Greene, Joe Carthy, and Pá Cunningham. [n. d.]. An analysis of the coherence of descriptors in topic modeling. ([n. d.]).

[21] Debjyoti Paul, Feifei Li, Murali Krishna Teja, Xin Yu, and Richie Frost. 2017. Compass: Spatio Temporal Sentiment Analysis of US Election What Twitter Says!. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1585–1594.

[22] Xuan-Hieu Phan, Le-Minh Nguyen, and Susumu Horiguchi. 2008. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web.* ACM, 91–100.

[23] Julie Sevenans, Quinn Albaugh, Tal Shahaf, Stuart Soroka, and Stefaan Walgrave. 2014. The Automated Coding of Policy Agendas: A Dictionary Based Approach (v. 2.0.). In *7th annual Comparative Agendas Project (CAP) conference, Konstanz, June.* 12–14.

[24] Yanchuan Sim, Brice DL Acree, Justin H Gross, and Noah A Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.* 91–101.

[25] Anke Tresch, Jonas Lefevere, and Stefaan Walgrave. 2017. How parties' issue emphasis strategies vary across communication channels: The 2009 regional election campaign in Belgium. *Acta Politica* (2017), 1–23.

[26] Chris J Vargo, Lei Guo, Maxwell McCombs, and Donald L Shaw. 2014. Network issue agendas on Twitter during the 2012 US presidential election. *Journal of Communication* 64, 2 (2014), 296–316.

[27] Stefaan Walgrave, Jonas Lefevere, and Anke Tresch. 2012. The associative dimension of issue ownership. *Public Opinion Quarterly* 76, 4 (2012), 771–782.

[28] Cäcilia Zirn, Goran Glavaš, Federico Nanni, Jason Eichorts, and Heiner Stuckenschmidt. 2016. Classifying topics and detecting topic shifts in political manifestos. (2016).

[29] Cäcilia Zirn and Heiner Stuckenschmidt. 2014. Multidimensional topic analysis in political texts. *Data & Knowledge Engineering* 90 (2014), 38–53.