Getting the Agenda Right: Measuring Media Agenda using Topic Models

Damir Korenčić
Department of Electronics
Ruđer Bošković Institute
Zagreb, Croatia
damir.korencic@irb.hr

Strahil Ristov
Department of Electronics
Ruđer Bošković Institute
Zagreb, Croatia
ristov@irb.hr

Jan Šnajder
University of Zagreb
Faculty of Electrical
Engineering and Computing
Zagreb, Croatia
jan.snajder@fer.hr

ABSTRACT

Agenda setting is the theory of how issue salience is transferred from the media to media audience. An agenda-setting study requires one to define a set of issues and to measure their salience. We propose a semi-supervised approach based on topic modeling for exploring a news corpus and measuring the media agenda by tagging news articles with issues. The approach relies on an off-the-shelf Latent Dirichlet Allocation topic model, manual labeling of topics, and topic model customization. In preliminary evaluation, the tagger achieves a micro F1-score of 0.85 and outperforms the supervised baselines, suggesting that it could be successfully used for agenda-setting studies.

Categories and Subject Descriptors

H3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

General Terms

Design, Measurement, Experimentation

Keywords

Agenda setting; Agenda measuring; News media analysis; Topic modeling; Document tagging

1. INTRODUCTION

Agenda setting is the effect of salience transfer from the media to the media audience [21, 20]. More precisely, agenda setting refers to how the *media agenda* (the set of issues that get media coverage) affects the *public agenda* (the set of issues that the public considers important). Agenda setting typically refers to news media, but in a broader sense it may refer to other types of agenda, such as a political agenda [16]. Agenda-setting studies offer valuable insights into the role of the mass media and the way they shape public opinion.

A crucial part of an agenda-setting study is the *measuring* of the media agenda, carried out by first defining a set of agenda *issues*. The issues are defined depending on the study of interest and may be at different level of abstractness (general issues such as *foreign policy* or *civil rights*, or specific issues such as *building construction and land use*). The conceptual space covered by the set of issues may vary from wide-coverage categorizations [18] to single-item agendas [32, 29]. Once the issues have been defined, their salience is typically measured by sampling the news articles, tagging them with issues, and counting the number of articles for each issue. The process of tagging, referred to as *coding* in social science literature [7], is usually done manually. Coding is a labor intensive task and thus hinders analyses of large news collections.

Recently, there has been a growing interest in the use of computational tools for efficient and large-scale analysis of media agenda. Particularly suited for this purpose are topic models [2], statistical text mining models for unsupervised discovery of topic distributions across documents, where topics are represented as probability distributions over words. The topics inferred by such models often correspond well to the issues on the agenda. However, the topics may be of poor quality. Moreover, standard topic models may be inadequate if one wants to focus on a predefined set of issues – a setup common for agenda-setting studies. A number of topic models have been proposed to address the problem of topic customization [11, 12].

In this paper we present our initial work on media agenda measuring using the standard Latent Dirichlet Allocation (LDA) topic model [2]. Using an off-the-shelf LDA tool, we aim at obtaining high-quality topics that correspond closely to issues of interest, enabling a more precise media agenda analysis. Unlike previous work, which use the inferred topics directly for agenda measuring, we first customize the model topics to match the agenda issues and then use these topics to tag articles with issues. We propose a semi-supervised approach to achieve this. Our approach covers both agenda issue definition via exploratory corpus analysis and agenda measuring by tagging the articles with issues. We perform preliminary experiments on issues related to *human and civil rights*. Although our findings are preliminary, we believe the approach is promising for any kind of agenda-setting research involving media texts.

2. RELATED WORK

Our work relates to work on topic-based media agenda analysis. Chuang et al. [3] use LDA for a large-scale analysis of the Media Cloud corpus. Koltsova and Koltcov [15] use LDA to discover and analyze the public agenda of the Russian blogospehere. After determining the optimal number of topics using an extrinsic method, they categorize the topics into higher-level categories. Grimmer [6] uses a custom topic model to measure the agendas of the US senators press releases and Quinn et al. [25] use a custom topic model with time-varying topics to measure the agendas of US Senate speeches. Nguyen et al. [23] propose a custom model that jointly learns a hierarchy of topics and a continuously-valued ideological polarity variable, and apply it to the analysis of congressional debates. Kim et al. [13] train a Hierarchical Dirichlet Process model on news articles and user comments, and compare the set of news topics inferred by the model to the public agenda (how users share and comment on news). Kok et al. [14] propose an approach where document clusters are labeled with agenda issues. To the best of our knowledge, their work is the first work on computational agenda measurement.

Our work differs from the above-mentioned approaches in two important respects. First, we customize model topics to match the agenda issues and we do not optimize the number of topics. Instead, we train several models, label the topics with *themes* (conceptual topics) and use them to construct customized topics matching the issues. Secondly, unlike previous work, which measures the agenda by aggregating per-document topic proportions, we measure the agenda by tagging the news articles with issues. This mimics the coding process commonly employed in agenda-setting studies and also lends itself to a straightforward validation against a human-annotated set.

We resort to the standard LDA model, while for topic customization we rely on the technique based on topic priors [8]. A number of more sophisticated topic customization techniques have been proposed [11, 12], but we opted for a simple and readily available approach.

3. DATASET

We conduct our experiments on political articles from mainstream US news outlets. To build the corpus, we first collected the URLs from Google News feed for two weeks and then ranked them by Alexa Rank[1] for US (traffic originating from the US). We then chose the top 25 outlets from this list, crawled them for feeds, and manually selected news feeds containing political news. After removing the outlets without purely political news feeds, we obtained the following list of 19 outlets: *Bloomberg Politics*, *CBS News*, *CNBC*, *CNN*, *Daily News*, *Fox News*, *Houston Chronicle*, *International Business Times*, *NBC News*, *Reuters*, *SFGate*, *The Atlantic*, *TheBlaze*, *The Guardian*, *The Huffington Post*, *The New York Times*, *The Wall Street Journal*, *The Washington Post*, and *Time*.

We used a custom-made application for collecting the URLs and downloading the articles, and the Newspaper[2] tool for scraping. The corpus contains articles published between Jan 26th 2015 and Apr 13th 2015. After inspecting a random sample of articles from the corpus, we decided to remove very short texts and errors (missing page errors, video and photo gallery pages, etc.). A simple heuristics that we found to work well was to filter out texts with less than 40 alphanumeric tokens. The final dataset contains $24,532$ news articles.[3]

4. AGENDA MEASURING

Our approach to agenda measuring comprises two main steps: *theme discovery* and *issue tagging*. The first step serves to identify the themes covered by the corpus and link them to topics. A theme can be thought of as a coherent topic situated between model-inferred topics and agenda issues. In the second step, we use the obtained themes to define the issues of interest and to build a custom model for tagging articles with these issues.

4.1 Theme discovery

4.1.1 Topic modeling

We use the LDA model [2] coupled with a fast online learning algorithm [9], available as part of the Gensim package [26]. For text preprocessing, consisting of stop-word and non-word removal, lemmatization, and subsequent stemming, we used the NLTK toolkit [1].

We train the models with $T = 50$ and $T = 100$ topics. Following [5], we set hyperparameter $\alpha = 50/T$, while we set $\beta = 0.01$. To optimize online learning hyperparameters [9], we conduct a grid search using the perplexity [2] as the measure of model quality. The chosen parameters values are $S = 1000, \tau_0 = 1000, \kappa = 0.5$.

4.1.2 Topic–theme labeling

Model-inferred topics ideally aggregate thematically coherent texts. In practice, however, the topics are often noisy and their quality depends on the model hyperparameters. Furthermore, even when hyperparameters are optimized, the quality may vary because of stochasticity induced by sampling and random initialization. The quality of the topics may be assessed by measuring the alignment between the inferred and reference topics [4].

To obtain less noisy and more robust topics, we introduce an intermediate layer of topics that we call *themes*. A theme is a topically coherent subset of text from the corpus that refers to a concept, an event, or an entity. In other words, a theme is to a human mind what topic is to a topic model: an ideal, noise-free, topically coherent unit. As noted by [4], mapping topics to concepts is a many-to-many mapping. Typically, a news article will be focused on one main theme, whereas opinion columns will often discuss a number of themes with equal salience.

To acquire the themes, we train the topic models with different random seeds, pool the topics from all the models, and then inspect and label the topics with discovered themes. The underlying idea is that, by looking at several models,

[1]http://www.alexa.com

[2]http://newspaper.readthedocs.org

[3]The dataset can be obtained from http://bit.ly/AGENDADATASET

Issue / *Theme* / Model.Topic
violence against women
women rights
M1.T43: victim, gender, identity, abuse, pregnant
sexual assaults
M0.T29: student, abortion, sexual, victim, men
M11.T25: men, sexual, girl, assault, female
LGBT rights
transgender
M1.T43: victim, gender, identity, abuse, pregnant
gay rights
M1.T24: religious, gay, marriage, indiana, freedom
M10.T54: marriage, couple, gay, same-sex, judge

Table 1: Example of topics (Model.Topic) labeled with themes (italics), and of issues (bold) related to these themes. High-probability words for each of the topics are displayed. Note that the topic M1.T43 is a mixture of two themes. In general, the topic–theme and theme–topic relations are many-to-many relations.

we get a better theme coverage as we compensate for the deficiencies of the individual models. We trained five models: three with 50 topics and two with 100 topics.

The labeling was carried out by two annotators. One 50-topic model was labeled by both annotators, after which the labeling conventions were discussed and revised. The remaining 300 topics were split among the two annotators who then labeled them separately. The themes were introduced dynamically and the annotators maintained a shared list[4] of themes to ensure label consistency. The annotators labeled the topics by inspecting, for each topic, the list of most probable words, list of article titles sorted by topic proportion, and, optionally, the full text of articles. Among these, article titles are by far the most useful piece of information as they can be inspected quickly and in the majority of cases represent a good summary of the articles – a property typical of good news headlines.

The labeling of 350 topics resulted in 134 themes. Of the 350 topics, 189 (54%) matched exactly one theme, 121 (34.6%) were a mix of two or more themes or a mix of one theme plus some noise, while the remaining 40 (11.4%) topics were noise (topics containing random articles or non-content words). Some themes were detected in all or most of the models, while others were detected in only a few models. Table 1 lists example topics and the themes they are labeled with.

4.2 Issue tagging

4.2.1 Defining the issues

An agenda-setting study defines a set of issues of interest. To simulate such a setting, we inspected the themes derived in the previous step and chose the following 12 issues: *civil rights movement*, *LGBT rights*, *police brutality*, *Chapel Hill attack*, *reproductive rights*, *violence against women*, *death penalty*, *surveillance*, *marijuana*, *gun rights*, *net neutrality*, *vaccination*. Issues related to human and civil rights are often the subject of agenda-setting studies [28, 27, 31]. Nine of these issues are related to a single theme each, and three are related to two themes each. While the chosen issues seem to be well distinguishable from each other – it might even be possible to distinguish them based on a small set of discriminative keywords – the challenge lies in distinguishing these issues from other, similar issues not in this list. Most of the above issues have at least one such similar issue; e.g., surveillance is similar to cybersecurity, violence against women is similar to general women's rights, and vaccination is similar to other health issues. Two issues and related themes and topics are presented in Table 1.

4.2.2 Customizing the model

We proceed with training a model for tagging the articles with issues. The topics of this model are customized to correspond as closely as possible to the chosen issues. Following [8], we manually define a list of issue-related seed words and use these to construct a prior for the issue-matching topic. The customization step is necessary to turn topic models into a reliable agenda measuring tool; without customization, learned topics are random and in general one cannot expect them to correspond to issues.

For each issue, we obtain the seed words using the following heuristic procedure. First, we select the issue-related themes and take the topics labeled with these themes. Next, we rank the words in each topic by their discriminativeness (defined as the product of word probability and inverse document frequency). Then, for each word, we inspect the articles ranked by tf-idf. Finally, if among 20 articles with the highest rank, 80% or more are about the issue in question, we add the word to the seed list for this issue. We continue adding words until 300 words are examined or 10 words are selected. This resulted in a list of 5–10 highly discriminative seed words per issue. As in the topic-theme labeling step, we rely on well formed article titles to speed up the procedure.

After we select the seed words, we translate them into Dirichlet priors for model topics. We accomplish this by defining two parameters: a prior pr for the non-seed words and a probability P_s for the seed words. To compute the prior vector, we first assign the value of pr to the non-seed words, while for the seed words we set a value pr_s such that expectation for each seed word is equal to P_s. Based on preliminary experiments, we set $P_s = 0.03$ and $pr = 0.001$.

4.2.3 Model optimization

To optimize (and later evaluate) the issue taggers, we manually labeled a dataset of 2800 articles from our corpus. We used 200 articles for calibration, 1600 as the development set, and 1000 as the test set. The Cohen's kappa coefficient of inter-annotator agreement on the calibration set is $\kappa = 0.93$, which according to [17] is considered a perfect agreement.

Besides the model for article tagging, we need a decision scheme for making the actual tagging decisions. We experiment with two schemes. The first scheme (single-label) tags the article with the issue corresponding to the topic with the highest proportion. The second scheme (multi-label) tags the article with the highest proportion issue and with all the issues with proportion above a threshold t.

To construct an optimal LDA tagger (a combination of a

[4]The list of themes can be obtained from
`http://bit.ly/AGENDADATASET`

seeded LDA model and a decision scheme), we perform a grid search over the hyperparameters that define a customized model and the hyperparameters that define a decision scheme. We consider $pr \in \{0.0005, 0.001, 0.002, 0.005\}$ for the non-seed word priors, $P_s \in \{1\%, 3\%, 6\%, 12\%\}$ for the seed word probability, and $T \in \{50, 100, 120, 150\}$ for the number of topics. Both decision schemes are considered, as well as the values of $t \in \{7\%, 10\%, 13\%, 15\%\}$ for the threshold. For the model learning algorithm hyperparameters, we use the values obtained by the grid search from Section 4.1.1. We rank the taggers by micro F1-score and, among the taggers within 0.01 from the optimal result, choose the one with the highest macro F1-score (a multi-label tagger with threshold set to 10%). To allow for a fair comparison against the supervised model, only one half of the development set is used for calculating the F1-scores (i.e., for optimizing the LDA tagger).

4.3 Tagging evaluation

We test four LDA taggers: a single-label tagger (Single) and a multi-label tagger (Multi), both optimized (Opt) and non-optimized (Nopt). The non-optimized taggers are not optimized on the development set. Instead, the number of topics is set to $T = 100$, and the priors are hand-optimized based on the inspection of the inferred topics. We test the taggers on the labeled test set comprised of 1000 documents.

We compare the LDA taggers against a supervised SVM tagger that labels documents with issues using one binary classifier per issue. To this end, we use the Linear SVM implemented in the scikit-learn [24] toolkit. We represent the documents as bag-of-words vectors with standard tf-idf weights computed as $\mathit{tfidf}(w, d) = (1 + \log \mathit{freq}(w, d)) \times \log(N_{doc}/N_{doc}(w))$ for word w and document d. We optimize class weights hyperparameters to account for the class imbalance, performing five-fold cross-validation on the complete development set and using F1-score as the objective function.

The final SVM tagger is constructed from classifiers trained on the complete development set. As noted above, the LDA taggers are optimized on one half of the development set. This way we can compare taggers with comparable amount of effort invested in their construction. For the LDA taggers, both seed words and 800 labeled documents are used, while for the SVM tagger 1600 labeled documents are used. The justification for this is that the time to label 800 documents roughly equals the time needed to construct seed word sets. We also evaluate the SVM tagger by training it on varying number of documents and averaging the micro-F1 score on the test set over five samples. The learning curve, shown in Fig. 1, indicates that the SVM tagger could improve with additional training documents.

Table 2 shows the tagging performance on the test set. The best LDA model is the optimized multi-label LDA tagger (Multi Opt). It outperforms the SVM tagger in terms of F1 and recall, but its precision is lower. SVM outperforms the other LDA taggers in terms of F1-score but its recall is the lowest among all taggers. Table 3 shows the per-issue results for the best performing LDA tagger and the SVM tagger. While precision is consistently high for the SVM tagger, its recall suffers on issues with a small number of articles.

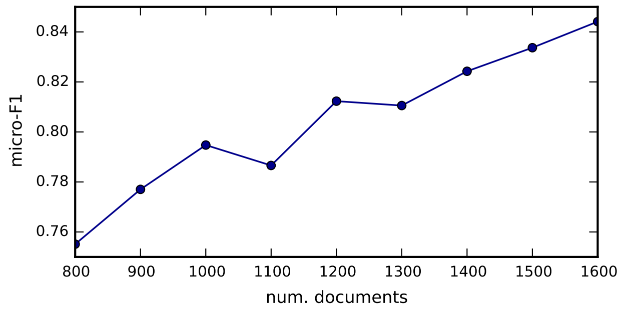

Figure 1: SVM tagger learning curve.

	Micro			Macro		
Tagger	P	R	F1	P	R	F1
Single Nopt	0.75	0.80	0.77	0.67	0.75	0.69
Single Opt	0.78	0.79	0.79	0.65	0.68	0.65
Multi Nopt	0.67	0.89	0.77	0.63	0.85	0.70
SVM	**0.99**	0.74	0.84	**0.92**	0.59	0.68
Multi Opt	0.80	**0.91**	**0.85**	0.74	**0.89**	**0.80**

Table 2: Averaged tagging performance

4.4 Aggregation

In social sciences, one is often interested in obtaining aggregate proportions of categories in an article collection, instead of correctly categorizing the individual articles [10]. This is also true for the agenda setting studies, where issue proportions are often used to model the media agenda [21, 30]. To test the LDA tagger in this setup, we calculate aggregate proportions on the test set and compare the results with the method from [10] (henceforth *Readme*). Readme directly approximates class proportions, without relying on individual article categorization. For a set of mutually exclusive categories whose probabilities sum up to one, the model is trained on labeled articles and outputs category proportions for a set of unlabeled articles. Experiments in [10] show that, in the case of multiple categories, Readme outperforms SVM classifiers with various kernels. The method is implemented in the freely available Readme package.[5]

As we deal with multilabel classification, we use Readme to

[5] `http://gking.harvard.edu/readme`

		Multi Opt			SVM		
Issue	#	P	R	F1	P	R	F1
civil rights movement	11	0.67	0.91	0.77	1.00	0.45	0.62
LGBT rights	60	1.00	0.93	0.97	0.98	0.95	0.97
police brutality	15	0.72	0.87	0.79	1.00	0.67	0.80
chapel hill	1	1.00	1.00	1.00	1.00	1.00	1.00
reproductive rights	7	0.42	0.71	0.53	0.00	0.00	0.00
violence agst. women	8	0.67	0.75	0.71	1.00	0.25	0.40
death penalty	5	0.80	0.80	0.80	1.00	0.40	0.57
surveillance	4	0.50	0.75	0.60	1.00	0.50	0.67
gun rights	7	0.78	1.00	0.88	1.00	0.29	0.44
net neutrality	6	0.67	1.00	0.80	1.00	0.67	0.80
marijuana	12	0.85	0.92	0.88	1.00	0.92	0.96
vaccination	15	0.83	1.00	0.91	1.00	1.00	1.00

Table 3: Per-issue tagging performance

calculate, for each issue, the proportion of articles belonging to this issue. We compare Readme against Multi Opt LDA tagger and the SVM tagger on the test set. We train Readme on the development set with default parameters and average the results over five runs since the algorithm is randomized. The results are shown in Fig. 2. Following [10], to quantitatively compare the results, we calculate Mean Absolute Proportion Error (MAPE) and Root Mean Squared Error (RMSE). MAPE is 0.18 for Optimal LDA tagger, 0.49 for SVM tagger, and 1.70 for Readme. RMSE is 0.003 for Optimal LDA tagger, 0.006 for SVM tagger, and 0.01 for Readme. These results indicate that, in the case of highly unbalanced binary categories, LDA tagger can outperform other methods.

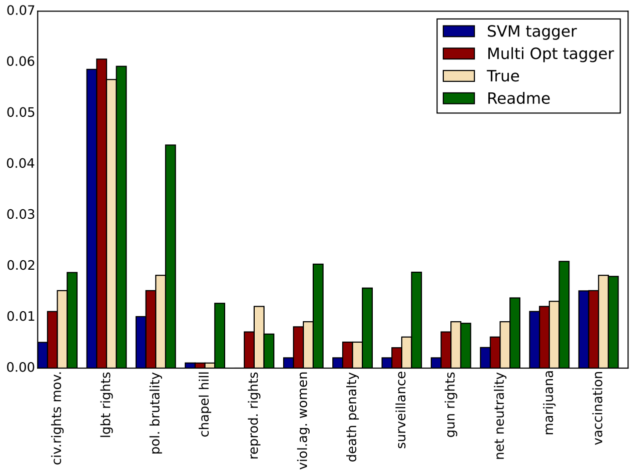

Figure 2: Approximations of the test set issue proportions.

4.5 Qualitative validation

A natural way to validate topic models used for agenda measuring is to search for correlation between real world events and peaks in a topic salience as it varies over time [25, 6]. Since our method tags documents with issues, we can measure issue salience by the number of articles tagged with that issue. We use the *police brutality* issue for validation. To find the events of interest, we observe the number of articles per day and inspect the articles in the periods of increased activity. We found that each of the peaks correlates with at least one major event. Figure 3 shows the article counts for the entire timespan of our corpus, together with the correlated events.

5. CONCLUSION AND FUTURE WORK

We proposed a semi-supervised method for measuring the media agenda using the standard LDA model. The first step consists of discovering the corpus themes and of labeling the topics with themes. This step facilitates the corpus exploration and the definition of a set of issues. Once the set of issues is defined, the method automates the coding process using model topics customised to match the issues. In our preliminary evaluation, the optimized issue tagger achieves the micro F1-score of 0.85 and outperforms supervised baselines on tasks of tagging documents with issues and calculating the aggregate issue proportions.

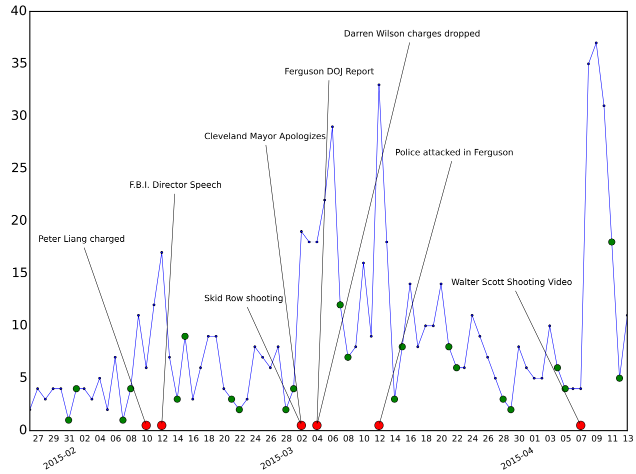

Figure 3: Number of articles tagged with *police brutality* issue and real world events relating to the issue. Days of the weekend are marked with green color.

Our method requires manual theme discovery and topic labeling, construction of seed word sets and labeling of a development set. For the corpus used in the experiment, the effort for these steps was 43 person-hours (35 for theme-topic labeling, 4 for labeling 800 articles, and 4 for constructing the seed word sets).

The majority of effort was devoted to theme discovery. The themes enable the analyst to explore the corpus and define issues without relying on a predetermined set of issues – a useful property in cases when the analyst does not have an overview of the media content. The method also relies on themes to link the issues with topics, so that the issue-specific seed words can be selected from topic words. However, our intuition is that, the larger the corpus, the less feasible the theme discovery approach becomes. For this reason, automatic theme discovery and creation of topic priors directly from the issues should be explored.

Other research directions include developing tools to facilitate steps of human theme discovery, topic labeling, and seed words construction. More advanced topic models that enable topic customization [11, 12], construction of higher quality topics [19], and inclusion of other document features in the learning process [22] should also be explored. Further evaluation is needed to compare the method with supervised and semi-supervised taggers in terms of tagging performance and the amount of human effort needed to construct the models.

6. REFERENCES

[1] S. Bird. NLTK: The Natural Language Toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 69–72. Association for Computational Linguistics, 2006.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *the Journal of Machine Learning Research*, 3:993–1022, 2003.

[3] J. Chuang, S. Fish, D. Larochelle, W. P. Li, and

R. Weiss. Large-Scale Topical Analysis of Multiple Online News Sources with Media Cloud. *NewsKDD: Data Science for News Publishing, at KDD*, 2014.

[4] J. Chuang, S. Gupta, C. Manning, and J. Heer. Topic Model Diagnostics: Assessing Domain Relevance via Topical Alignment. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 612–620, 2013.

[5] T. L. Griffiths and M. Steyvers. Finding Scientific Topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.

[6] J. Grimmer. A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis*, 18(1):1–35, 2010.

[7] J. Grimmer and B. M. Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, pages 1–31, 2013.

[8] D. Hall, D. Jurafsky, and C. D. Manning. Studying the History of Ideas using Topic Models. In *Proceedings of the conference on empirical methods in natural language processing*, pages 363–371. Association for Computational Linguistics, 2008.

[9] M. Hoffman, F. R. Bach, and D. M. Blei. Online Learning for Latent Dirichlet Allocation. In *advances in neural information processing systems*, pages 856–864, 2010.

[10] D. J. Hopkins and G. King. A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1):229–247, 2010.

[11] Y. Hu, J. Boyd-Graber, B. Satinoff, and A. Smith. Interactive Topic Modeling. *Machine learning*, 95(3):423–469, 2014.

[12] J. Jagarlamudi, H. Daumé III, and R. Udupa. Incorporating Lexical Priors into Topic Models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 204–213. Association for Computational Linguistics, 2012.

[13] Y. Kim, S. Kim, A. Jaimes, and A. Oh. A Computational Analysis of Agenda Setting. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 323–324. International World Wide Web Conferences Steering Committee, 2014.

[14] Y. H. Kok, A. Goh, and D. Holaday. Agenda: A Tool for Agenda Setting Research. *Information Services and Use*, 19(4):265–276, 1999.

[15] O. Koltsova and S. Koltcov. Mapping the Public Agenda with Topic Modeling: The Case of the Russian Livejournal. *Policy & Internet*, 5(2):207–227, 2013.

[16] G. M. Kosicki. Problems and Opportunities in Agenda-Setting Research. *Journal of communication*, 43(2):100–127, 1993.

[17] J. R. Landis and G. G. Koch. The Measurement of Observer Agreement for Categorical Data. *biometrics*, pages 159–174, 1977.

[18] J. K. Lee and R. Coleman. Testing Generational, Life Cycle, and Period Effects of Age on Agenda Setting. *Mass Communication and Society*, 17(1):3–25, 2014.

[19] W. Li and A. McCallum. Pachinko Allocation: DAG-Structured Mixture Models of Topic Correlations. In *Proceedings of the 23rd international conference on Machine learning*, pages 577–584. ACM, 2006.

[20] M. McCombs. The Agenda-Setting Role of the Mass Media in the Shaping of Public Opinion. In *Mass Media Economics 2002 Conference, London School of Economics*, 2002.

[21] M. E. McCombs and D. L. Shaw. The Agenda-Setting Function of Mass Media. *Public opinion quarterly*, 36(2):176–187, 1972.

[22] D. Mimno and A. McCallum. Topic Models Conditioned on Arbitrary Features with Dirichlet-Multinomial Regression. *arXiv preprint arXiv:1206.3278*, 2012.

[23] V.-A. Nguyen, J. L. Boyd-Graber, and P. Resnik. Lexical and Hierarchical Topic Regression. In *Advances in Neural Information Processing Systems*, pages 1106–1114, 2013.

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[25] K. M. Quinn, B. L. Monroe, M. Colaresi, M. H. Crespin, and D. R. Radev. How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, 54(1):209–228, 2010.

[26] R. Řehůřek and P. Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA.

[27] M. Roberts, W. Wanta, and T.-H. D. Dzwo. Agenda Setting and Issue Salience Online. *Communication Research*, 29(4):452–465, 2002.

[28] B. Sayre, L. Bode, D. Shah, D. Wilcox, and C. Shah. Agenda Setting in a Digital Age: Tracking Attention to California Proposition 8 in Social Media, Online News and Conventional News. *Policy & Internet*, 2(2):7–32, 2010.

[29] S. N. Soroka. Media, Public Opinion, and Foreign Policy. *The International Journal of Press/Politics*, 8(1):27–48, 2003.

[30] T. Takeshita. Agenda-Setting Effects of the Press in a Japanese Local Election. *Studies of Broadcasting*, 29:193–216, 1993.

[31] D. Weaver and S. N. Elliott. Who Sets the Agenda for the Media? *Journalism Quarterly*, 62(1):87–94, 1985.

[32] J. P. Winter and C. H. Eyal. Agenda Setting for the Civil Rights Issue. *Public opinion quarterly*, 45(3):376–383, 1981.