ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Heidenreich, Tobias; Eberl, Jakob-Moritz; Tolochko, Petro; Lind, Fabienne; Boomgaarden, Hajo G.

Article — Published Version

My Voters Should See This! What News Items Are Shared by Politicians on Facebook?

The International Journal of Press/Politics

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Heidenreich, Tobias; Eberl, Jakob-Moritz; Tolochko, Petro; Lind, Fabienne; Boomgaarden, Hajo G. (2022) : My Voters Should See This! What News Items Are Shared by Politicians on Facebook?, The International Journal of Press/Politics, ISSN 1940-1620, Sage, Thousand Oaks, CA, Iss. OnlineFirst Articles, pp. --, https://doi.org/10.1177/19401612221104740

This Version is available at:
https://hdl.handle.net/10419/259790

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by-nc/4.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Supplementary Information

Section A1. Data Matching

While for the media, all URLs were available in their natural full-form, URLs included in the Facebook data required further preprocessing. Here, we first extracted all shared URLs from posts classified as "links", indicating that the purpose of the status post was to share an external source, such as an article. In addition, we screened "shared posts" (i.e., posts that simply share content from another account on the platform) as well as other types of status posts (i.e., including only text, text and video, or text and photo) for URLs. In an iterative process, URLs were unshortened (e.g., for short links such as tinyurl) and cleaned from superfluous fractions used to trace users reading an article coming from social media. Investigating the URL roots of content shared by political actors on Facebook, we find that 14,935 status posts reference one of the outlets in our news media sample. Matching the news media data and status posts of Facebook of political actors, we furthermore identify 6,030 social media messages as linking to one of the articles in our sample (40%). While some status posts did not provide any reference to external sources (i.e., an URL) at all, other URLs were non-matchable due to broken URLs (e.g., outdated short links), references to sub-pages of an outlet (e.g., specifically for sports or recipes) or URLs leading to articles that were never visible on the main page of an outlet or not long enough to be part of our gathering process for the media data.

Section A2. Polylingual Topic Modeling

Since we are dealing with articles in several languages (German, Spanish, and English), a special topic modelling technique had to be chosen. We decided for Polylingual Topic Models (PLTM; Mimno et al. 2009) as this type of topic model can work with the original text data which avoids expensive full translations of the N = 623,919 news items in the

corpus (Pruss et al. 2019). The model derives multilingual topics (a multilingual topic is basically a set of word lists, where each word list represents the topic version for a different language) and probabilities per article for each of these multilingual topics.

To deal with the multilingualism of the news article, PLTM requires training documents, from which the alignments between languages are learnt. These training documents (so called tuples) consist for example of direct translations of documents which share the languages and ideally the domain (here news texts) of the actual text corpus to be modeled. To this end, we used the "Global Voices Corpus" (Prokopidis et al. 2016) from which we extracted N = 2,403 tuples, i.e. news articles for which a German, Spanish, and English translation was available. These tuples were used to train PLTM topic models. The final selected model was then used to infer topic probabilities for the N = 623,919 news items in the corpus. The PLTM is implemented in MALLET (McCallum 2002). We followed the steps outlined in previous work for document pre-processing and hyperparameter optimization. We used only lemmatized nouns for the training corpus. Varying the number of topics (K = {5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100}), we derived 15 different models. The final model selection was guided by two automated metrics and, subsequently, the chosen model was assessed using systematic manual assessment as well automated procedures. First, we used MTA (Boyd-Graber and Blei 2012; Pruss et al. 2019) and NPMI (Lau et al. 2014) metrics to determine the matching translation accuracy and the pointwise mutual information, respectively. The model for K = 20 topics was selected as it yielded the highest MTA and NPMI values (for an overview of the topics and top words please see Table A2). Second, we used the selected model to infer topics for the full corpus of news articles. Subsequently, the topics of this model were labeled individually by three researchers fluent in German, Spanish, and English. For a final assignment of topic labels, all top words and topics were discussed amongst the researchers. Third, the selected model was further validated via manual

evaluation tasks. For this purpose, two coders that did not see the topics or top words before were asked to read 142 news articles from the corpus (47 German, 47 Spanish, and 48 English articles), and sort them in thematic groups. The coders should combine articles in groups that appear to write about closely related topics and label these topics subsequently. The 142 news articles that both coders read were a subset of the full news article corpus, and chosen as the PLTM had assigned highest topic probabilities per topic for these articles. The 142 news articles, therefore, form in fact subgroups of articles that should deal with the same topics. As we saw a huge overlap in the groups as assigned via PLTM and the sorting and labeling 'solution' by the two coders, we concluded that PLTM was useful to identify cross-lingual topics. This positive assessment of the PLTM was furthermore underlined with another evaluation method. For the articles with the highest probabilities per topic, we compared the PLTM topic label for the most prevalent topic with the section of the respective articles as assigned by the media outlet's sections. This subject information, however, was available for only some outlets.

Section A3. Partisanship of Audiences

The variable was computed based on the question "[t]here are several parties in [COUNTRY], each of which would like to receive your vote. Using the scale of 0 to 10, how likely is it that you would ever vote for each of the following parties?". Respondents with a score higher than 7 were assigned to be supporters of the respective party. One respondent could be assigned to be a supporter of more than one party.

Tables and Figures

Table A1. Matched media and Facebook data.

Country	Media Outlet	Media Articles	Unique References to this Media Outlet as found in the Facebook Communication of Political Actors	Matched unique References to this Outlet as found in the Facebook Communication of Political Actors (in %)
Germany	Bild.de	20,516	630	158 (25.08%)
	faz.net	27,449	762	335 (43.96%)
	spiegel.de	15,299	1,306	549 (42.04%)
	sueddeutsche.de	21,833	713	196 (27.49%)
	taz.de	11,201	324	177 (54.63%)
	welt.de	26,269	1,854	794 (42.83%)
	zeit.de	13,205	632	305 (48.26%)
Spain	20minutos.es	20,048	48	15 (31.25%)
	abc.es	27,532	99	37 (37.37%)
	eldiario.es	18,291	454	246 (54.19%)
	elmundo.es	14,865	218	92 (42.2%)
	eplais.com	28,339	267	161 (60.3%)
	larazon.es	21,414	22	12 (54.55%)
	lavanguardia.com	28,637	91	20 (21.98%)
UK	bbc.co.uk	9,923	1,381	430 (31.14%)
	dailymail.co.uk	109,528	138	97 (70.29%)
	express.co.uk	34,701	67	15 (22.39%)

	mirror.co.uk	26,295	542	152 (28.04%)
	telegraph.co.uk	28,299	213	70 (32.86%)
	theguardian.com	26,861	1,685	696 (41.31%)
	thesun.co.uk	68,384	92	61 (66.3%)
	thetimes.co.uk	25,030	267	42 (15.73%)
Total		623,919	11,805	4,660 (39.48%)

Note. It is possible that an individual news article was published on Facebook multiple times.

Table A2. Parties

Country	Party
Germany	Alternative for Germany (AfD)
	Christian Democratic Union/Christian Social Union (CDU/CSU)
	Free Democratic Party (FDP)
	Alliance‘90/Greens (Greens)
	The Left (Left)
	Social Democratic Party of Germany (SPD)
Spain	Citizens (Citizens)
	United We Can (Podemos)
	People's Party (PP)
	Spanish Socialist Workers’ Party (PSOE)
UK	Conservative Party (Conservatives)
	Labour Party (Labour)
	Liberal Democrats (LibDems)
	Scottish National Party (SNP)

Table A3. Topics identified with PLTM.

Topic ID	Topic Label	Top Words English
1	Blogging	blog, post, voice, media, word, article, time, blogger, story, author
2	Culture, Nature and Traditions	animal, tree, tradition, holiday, celebration, tourist, church, tiger, nature, time
3	Culture (Art)	artist, strike, music, song, art, cartoon, train, campaign, hunger, video
4	Culture (TV / Video / Games)	video, film, documentary, movie, message, event, game, team, flag, scene
5	Development	country, people, region, project, development, percent, population, technology, report, access
6	Digitalization	internet, media, user, service, network, government, content, news, platform, citizen
7	Economy	company, business, market, oil, power, industry, electricity, price, energy, product
8	Education	student, community, school, project, people, book, education, culture, event, life
9	Elections	election, party, opposition, vote, people, leader, power, politician, result, president
10	Environment & Climate Change	water, food, climate, farmer, change, rice, disaster, pollution, resident, air
11	Family and Domestic Violence	woman, child, girl, mother, family, violence, victim, daughter, parent, husband
12	Health	people, health, hospital, doctor, map, center, team, disease, effort, volunteer
13	Infrastructure	people, time, city, life, friend, house, death, night, hour, street
14	Law / Human Rights	rights, law, freedom, journalist, activist, prison, crime, court, authority, charge
15	Local	community, worker, land, people, village, government, resident, province, district, town
16	Migration	family, refugee, country, life, permission, people, border, camp, story, child

17	Politics	country, government, crisis, issue, change, time, situation, debate, response, term
18	Security	police, protest, people, protester, street, movement, activist, video, demonstration, government
19	Social Media	twitter, tweet, news, people, user, hashtag, comment, media, photograph, netizen
20	War	war, attack, conflict, peace, force, people, revolution, regime, army, report

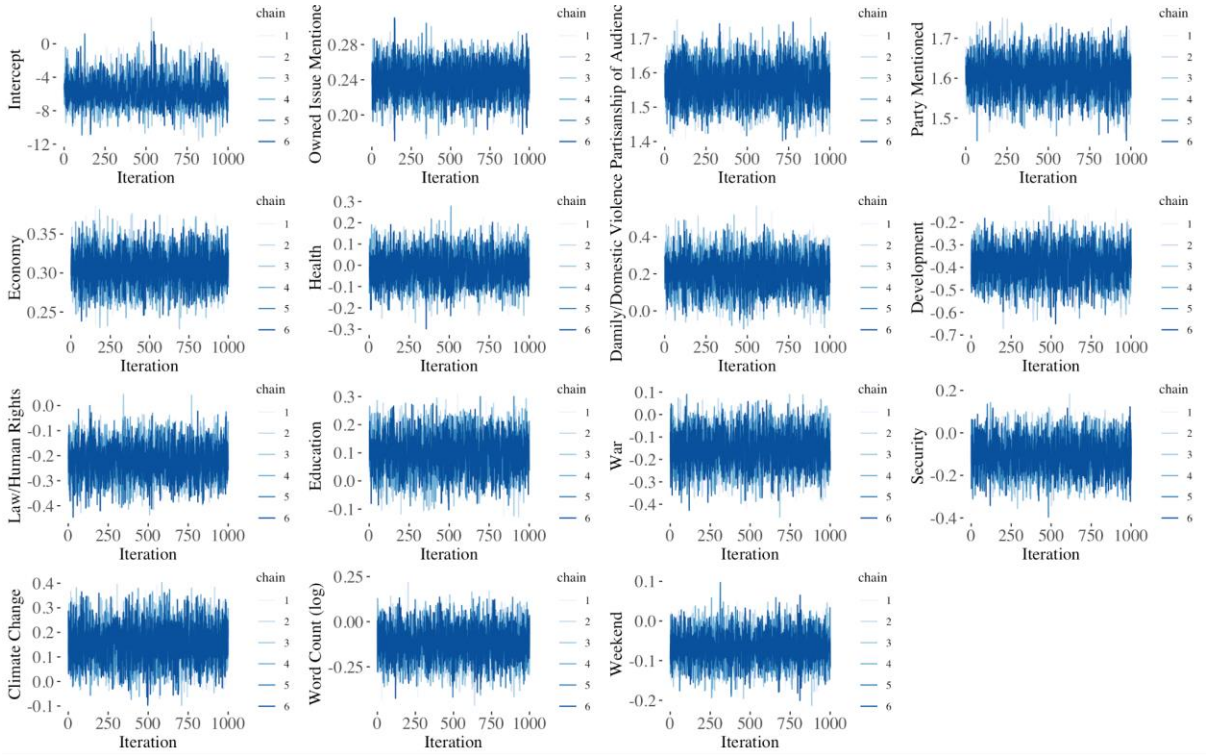
Notes. Ten top words for English language are displayed; Topics 1, 2, 3, 4, and 15 were labeled but identified as boilerplate topics as they are, according to top words and highest scoring articles, too ambiguous to implement for further analysis; Topics 5, 7, 8, 10, 11, 12, 14, 16, 18, and 20 were identified as distinct policy issues and therefore implemented for further analyses.

Table A3. Posterior parameter estimates.

	Posterior Value	95% Credibility Interval
Intercept	-6.04	[-8.48 -2.35]
Party Mentioned	1.52	[1.61 1.69]
Owned Issue Mentioned	0.24	[0.21 0.27]
Partisanship of Audiences	1.57	[1.47 1.67]
Topics (ref. Migration)		
Development	-0.22	[-0.35 -0.09]
Economy	-0.01	[-0.15 0.13]
Education	-0.16	[-0.31 -0.01]
Environment & Climate Change	-0.12	[-0.29 0.05]
Family / Domestic Violence	-0.39	[-0.53 -0.25]
Health	0.21	[0.03 0.38]
Law / Human Rights	0.10	[-0.03 0.20]
Security	0.15	[0.01 0.30]
War	-0.10	[-0.25 0.04]
Controls		
Word Count (log)	0.31	[0.26 0.35]
Weekend	-0.07	[-0.14 0.01]
Random Effects		
Article level σ (Levels: 35612)	0.03	[0.01 0.11]
Country level σ (Levels: 3)	1.92	[0.41 5.66]
Party level σ (Levels: 14)	1.15	[0.77 1.99]
Outlet level σ (Levels: 22)	1.22	[0.89 1.77]
N	160,794	

Notes. Parameter estimates from the Stan model. All parameters have *R hat* of 1, indicating good convergence.

Figure A1. Trace plots for MCMC posterior draws.



Notes: Trace plots indicate a good mixture of the 6 Markov Chains.

Figure A2. Posterior Distribution of Model Parameters. German Cases Only.

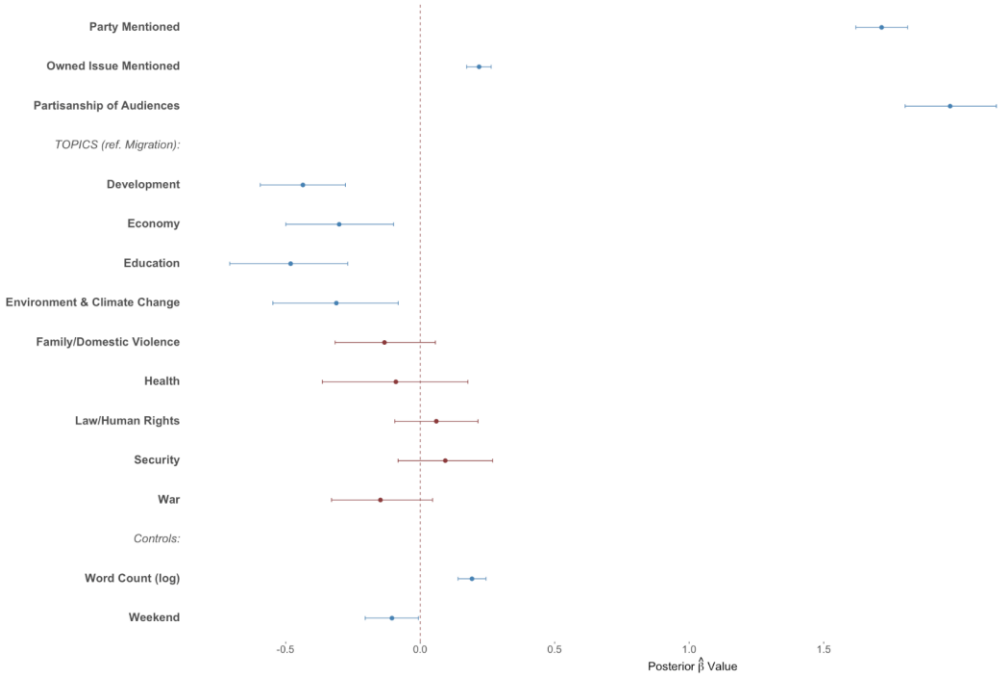


Figure A3. Posterior Distribution of Model Parameters. Spanish Cases Only.

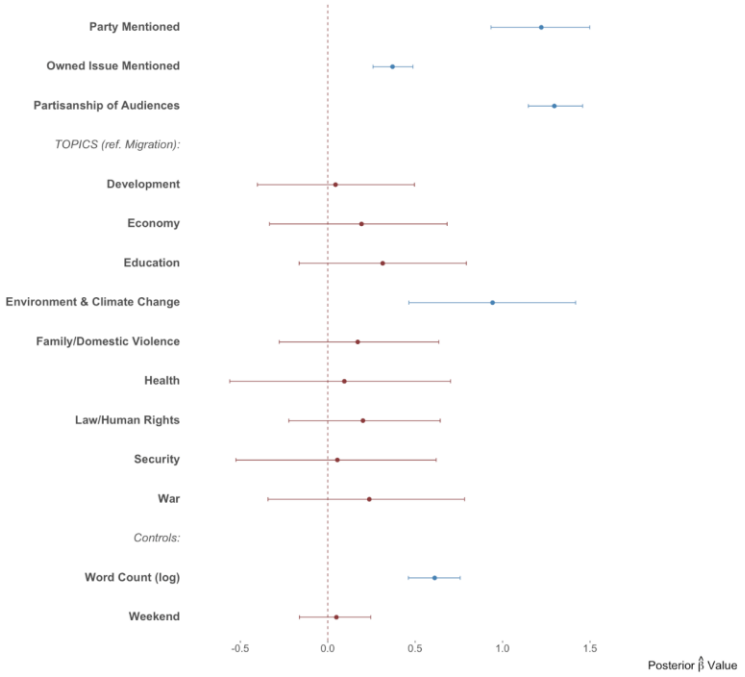


Figure A4. Posterior Distribution of Model Parameters. UK Cases Only.

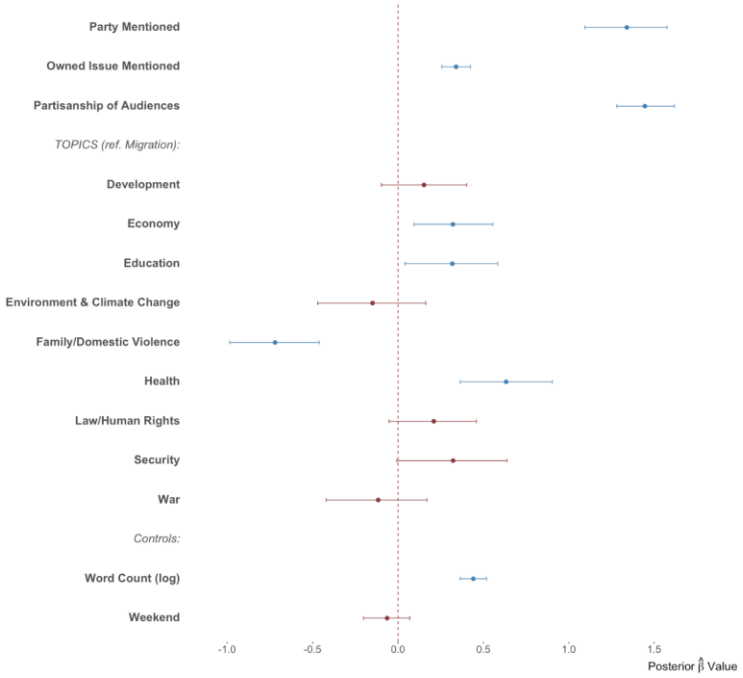
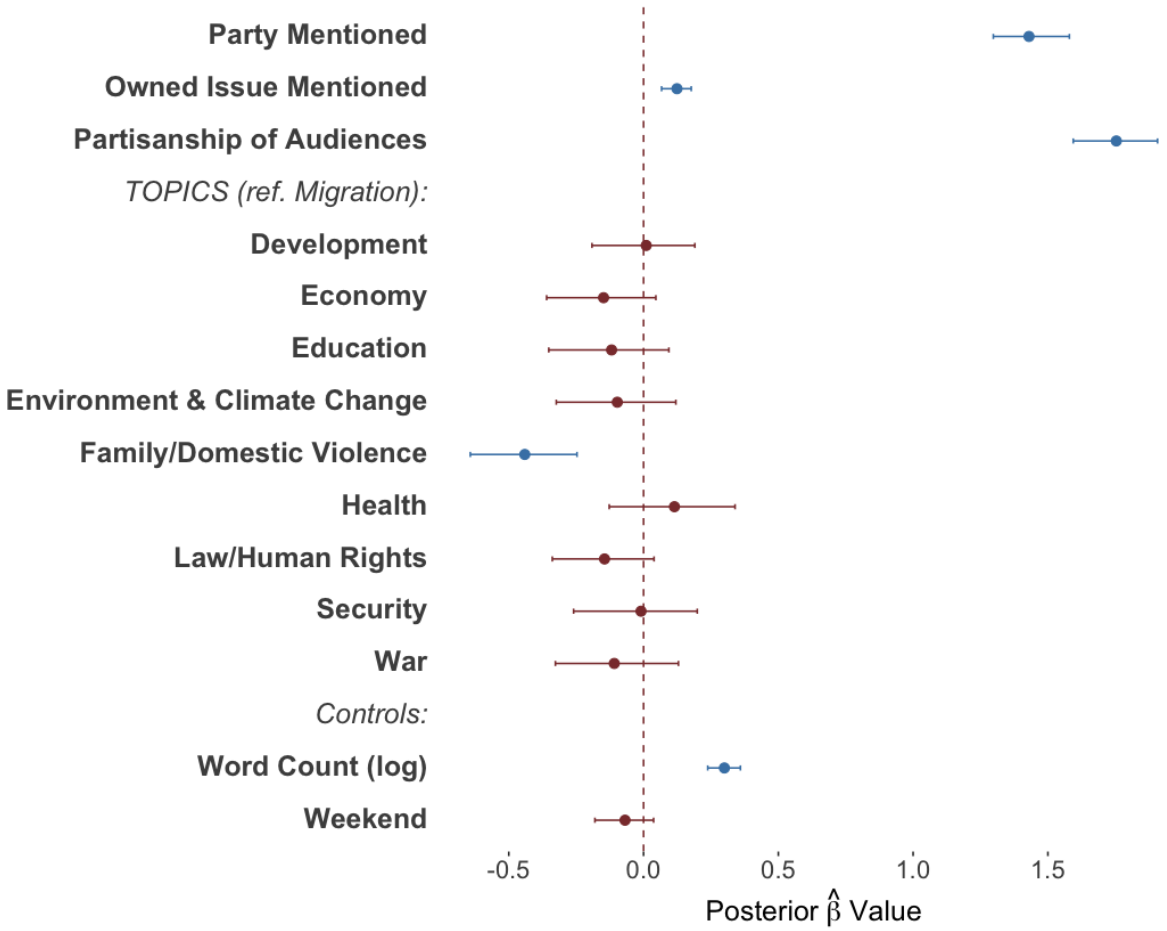


Figure A5. Posterior Distribution of Model Parameters. Robustness model with political news only.



Supplementary Information: References

Boyd-Graber, J., and Blei, D. (2012). Multilingual topic models for unaligned text. In D. McAllester (Ed.), *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (pp. 75-82). AUAI Press.

Lau, J. H., Newman, D., and Baldwin, T. (2014). Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In S. Wintner, S. Goldwater, & S. Riezler (Eds.), *Proceedings of the fourteenth conference of the European chapter of the association for computational linguistics* (pp. 530-539). Association for Computational Linguistics.

McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit.* [Computer software].

Mimno, D., Wallach, H., Naradowsky, J., Smith, D. A., and McCallum, A. (2009). Polylingual topic models. In P. Koehn & R. Mihalcea (Eds.), *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 880-889). Association for Computational Linguistics.

Prokopidis, P., Papavassiliou, V., and Piperidis, S. (2016). Parallel global voices: a collection of multilingual corpora with citizen media stories. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, …, & S. Piperidis (Eds.), P*roceedings of the tenth international conference on language resources and evaluation* (LREC'16) (pp. 900-905). European Language Resources Association (ELRA).

Pruss, D., Fujinuma, Y., Daughton, A. R., Paul, M. J., Arnot, B., Albers Szafir, D., and Boyd-Graber, J. (2019). Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PloS one, 14*(5), 1-23.