U.S. Immigration in the Trump Era: Using Text Analysis to Link Public Sentiment and Policy Change

Leslie J. Root and Monica J. Alexander
Extended abstract submitted to PAA 2019

Abstract

Previous research finds that public sentiment on immigration is influenced to a large extent not by changes in immigrant stocks and flows, but by specific highly visible events (e.g., a terrorist attack) and media coverage. However, public opinion data is available only at infrequent intervals, making it difficult to establish a clearer relationship between specific news events or changes in the news environments and migration sentiment. In this paper, we explore the utility of natural language processing in analyzing the U.S. migration discourse in the first two years of the Trump presidency, connecting national news coverage to public sentiment through the use of Twitter data. We suggest that this connection, and the polarized sentiment it engenders, could have significant implications for migration policy in the near future.

Background

Of the major types of demographic events – birth, death, and migration – the last of these is perhaps the most extensively influenced by the non-demographic. In addition to the social and economic determinants of migration, political fortunes such as changing visa regimes, border security measures, and natives' attitude to immigrants influence the ability and desire of potential immigrants to cross borders. Natives' attitude in particular can change migration patterns both directly (by changing the perceived appeal of the destination country) and, in democratic societies, by changing the government's attitude. Wilmoth and Ball (1992) connect news coverage of population issues to popular opinion, and from there to the creation of "an institutional structure for understanding and altering population-related events" in the "population bomb" era of the 1960s (p. 631). Given the attention paid to campaign promises to build a wall with Mexico and return jobs to native-born Americans, there is little reason to think that popular sentiment has become less relevant today.

Previous research finds that public sentiment on immigration is governed to a large extent by outside events and media coverage (Atwell Seate & Mastro, 2016; Brader, Valentino, & Suhay, 2008; Dunaway, Branton, & Abrajano, 2010; Kim, Carvalho, Davis, & Mullins, 2011; van Klingeren, Boomgaarden, Vliegenthart, Vreese, & H, 2015; Zúñiga, Correa, & Valenzuela, 2012). However, nationally representative survey data on this issue, like much survey data, is not particularly granular, and the effect of specific events (and specific media coverage) is not known.

The Trump administration has created a new context for the immigration discourse, in that it has made immigration one of its key policy priorities and has created a news climate in which immigration is a constant subject of media coverage. Motivated by evidence of Democratic voters becoming more open to immigration ("Shifting Public Views on Legal Immigration Into the U.S.," 2018) and some survey data indicating that even Republican voters are significantly less negative on immigration than the current administration's policies would suggest, the present research uses natural language processing to analyze large datasets of news coverage and social media to connect changes in public opinion to four major news events: the election of

Donald Trump, Executive Order 13769 (the "Muslim travel ban"), the proposed end of the Deferred Action for Childhood Arrivals program, and the family separation policy.

First, we demonstrate that the Trump administration is indeed a new context for the immigration discourse, in that the administration's focus on immigration and border security as a key policy priority and its high-profile attempts to restrict immigration have created a news climate in which immigration is a constant subject of media coverage.

We will then explore how patterns of public opinion about migration are linked to the news cycle and migration policy, by analyzing changes in sentiment and topics in covered a time series of Twitter data which covers the Trump era.

Data

We analyze articles from the ProQuest "U.S. Major Dailies" database, which includes *The Chicago Tribune, The Los Angeles Times, The New York Times,* the *Wall Street Journal,* and the *Washington Post.* Articles are drawn from the first two years of the Trump era, defined as beginning with Donald Trump's formal nomination as the Republican candidate for president in late July, 2016 and ending on July 31, 2018. All articles containing the terms "immigra*" (that is, immigrant, immigration, or other variants of these words) and "U.S." were included in the sample, for sample size of 15,644 articles. (U.S. was included to narrow the focus by reducing the number of articles exclusively discussing non-U.S. migration topics such as Brexit and the Syrian refugee crisis in Europe.) For comparative analyses, the same search was done on an equivalent period in the early Obama era (7/29/2008 to 7/31/2010), yielding 6462 articles.

We will obtain immigration-related tweets during the Trump era from Twitter's Historical PowerTrack API.[1] This tool allows for the extraction of tweets that include a combination of relevant keywords (e.g. "immigra*"), hashtags (e.g. #DACA, #Dreamers) and user accounts (e.g. @realDonaldTrump). We are also able to obtain the impressions of tweets (i.e. the number of retweets and likes) and the user's location, where specified. Access to the API is available for a fee, based on the amount of data extracted. This data collection process will be set up in the coming months.

Methods

In order to investigate words, topics and sentiment contained in news articles and tweets, the raw data need to be converted into a format which is able to be analyzed in a systematic way. In particular, we tabulated every unique word and the number of occurrences for each news article. Once the dataset is in this form, we used several text analysis techniques to extract key patterns in the data.

First, we use sentiment analysis to summarize the average sentiment or tone of a document. The general idea is to compared the words contained in a document to a preexisting lexicon of words, which have a sentiment score assigned to them. Once each word in a document is assigned a sentiment score, these scores can be averaged over a document to give a sense of its tone. There

---

[1] https://developer.twitter.com/en/docs/tutorials/choosing-historical-api.html

are several existing methods; we chose to use the AFINN sentiment lexicon[2], which assigns words a score between -5 (negative) and 5 (positive). Thus, the lower the score, the more negative the sentiment.

We also analyze the term frequency–inverse document frequency (td-idf) of words to highlight which words are important in news articles, and how these change over time. The td-idf measure increases as a word is mentioned more frequently within a document, but this increase is offset by the number of times a word appears across all documents, thereby down-weighting words that are more common in general. As such, the td-idf measure highlight words like 'Trump' or 'Mexico', rather than 'the' or 'he'.

In future, analysis will also focus on topic modeling. This technique is a "suite of algorithms that aim to discover and annotate large archives of documents with thematic information." (Blei, 2012, p. 77) Like the hand-coding of texts, topic modeling allows researchers to label texts by theme and, for example, analyze how themes co-occur, vary and change within or across sets of documents; but topic models allow for the analysis of much larger bodies of text than does hand-coding. Here, to estimate the probability distributions of words and topics, we will use structured topic modeling, which incorporates the assumptions of latent Dirichlet analysis (LDA - a simple type of topic model that assumes documents are comprised of multiple topics in different proportions), but allows the incorporation of document metadata, such as date, author, or publication source into the formal model (Roberts et al., 2014).

We used the 'tidytext' and 'stm' R packages for text analysis and topic modeling of the data. All analysis was performed using R version 3.4.4.

Descriptive analysis
As a share of all news coverage, immigration was significantly more dominant in the Trump era than in the Obama era, indicating that the news climate has indeed changed. Table 1 gives the total number of articles in the database for each time period. Immigration is mentioned in nearly three percent of articles in the Trump era – 3.2 times as frequently as in the Obama era. Figure 1 shows the number of immigration articles per day. Although the raw number does not account for differences in the overall number of news articles each day, the large difference in overall numbers between the two periods indicate that, proportionally, the Trump line can be reasonably supposed to be even higher than the Obama line.

	Articles on immigration	Total articles	Proportion of articles on immigration
Obama era	6462	698457	0.93%
Trump era	15644	528255	2.96%

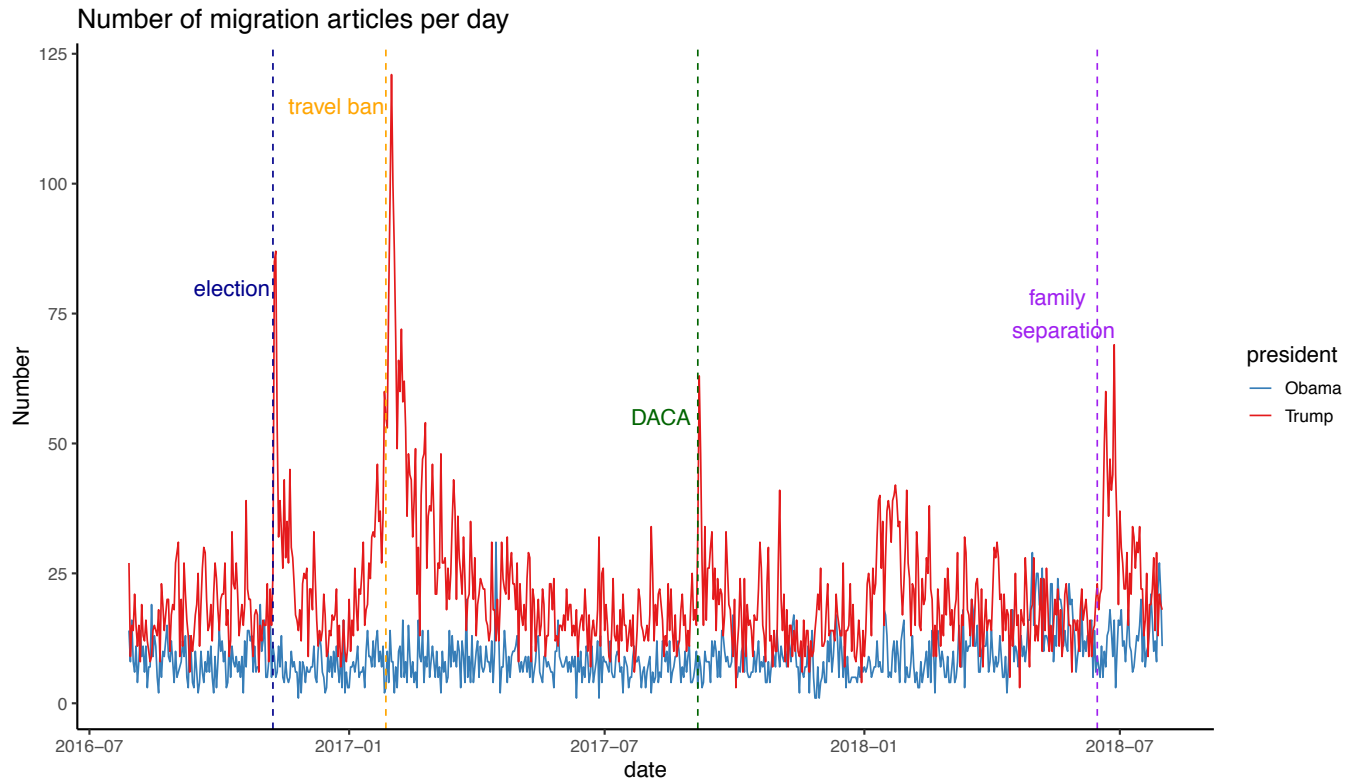Table 1. Number of news articles in the Obama and Trump eras.

Figure 1. Number of articles containing immigration per day, by era. Note that the Obama era articles have been shifted by eight years for comparison, and the labeled events relate to the Trump era.

While day-to-day coverage of immigration is somewhat higher in the Trump era, a significant portion of the difference is driven by major news events. These include the election (large one-day spike in November 2016), the Muslim travel ban (gradual increase in early 2017 to a maximum around February 1, when Executive Order 13769 was published), the plan to end the Deferred Action for Childhood Arrivals program (DACA, peaking with the announcement of the plan September 7, 2017 and dominating coverage again in January 2018 as the courts began to block the plan), and the family separation policy (peaking near the end of June, 2018).

One of the most obvious peaks in the Obama era, by contrast, on April 15, 2007, was due to a deportation stay granted to an accused former Nazi prison guard coinciding with the release of a report on undocumented immigrants by the Pew Hispanic Center.
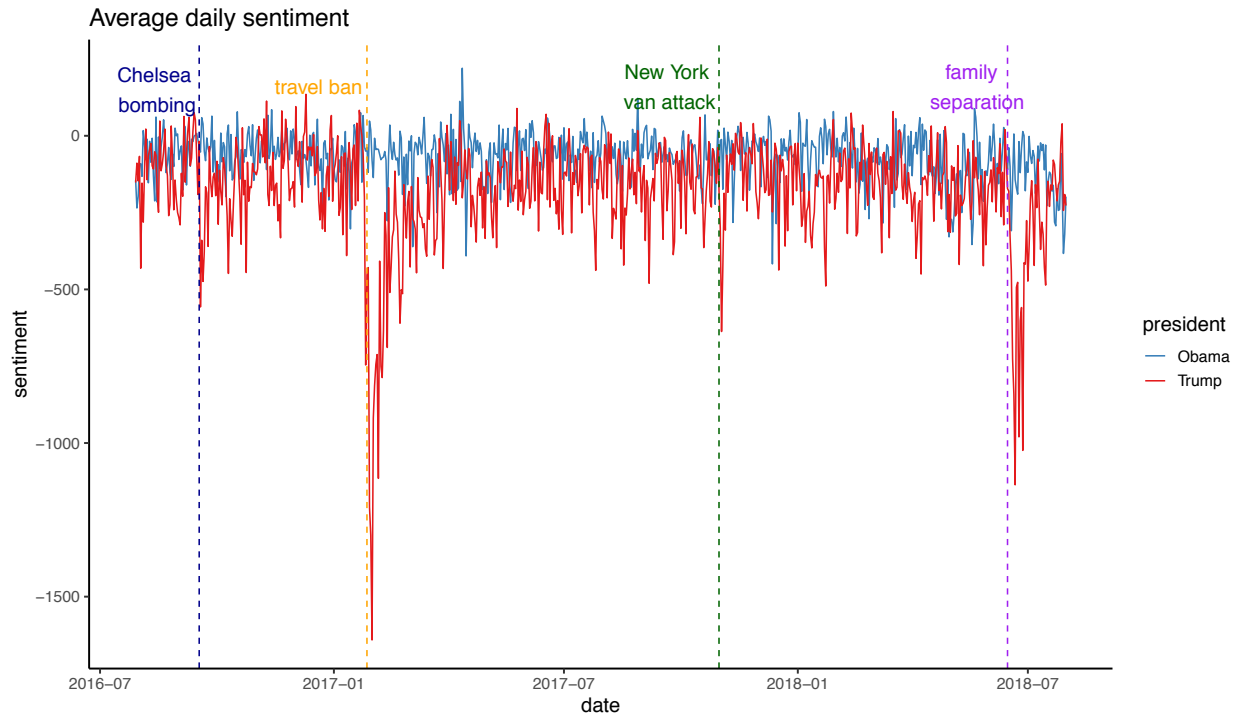
Figure 2. Average sentiment of daily migration news articles. Note that the Obama era articles have been shifted by eight years for comparison, and the labeled events relate to the Trump era.

Sentiment analysis also shows the stark contrast in migration-related reporting in the Trump era compared to the Obama era (Figure 2). In general, migration articles contain more negative words in the more recent presidential period. Sentiment analysis indicates, too, that the Trump era media environment on immigration is marked by unusual events. In this case, we see that sentiment is relatively constant, with the notable exception of coverage of the Muslim travel ban and the family separation policy. Other negative peaks occurred around two terrorist incidents (September 17-19, 2016 and October 31, 2017). Decrease in sentiment around terrorist incidents in the Obama era was not as stark – coverage of an attempted airplane bombing on December 25, 2009 and an attempted bombing in Times Square on May 1, 2010 show no dip in sentiment. However, it is difficult to make a direct comparison, because neither of these attacks resulted in any casualties, while the two incidents in the Trump period caused injuries and deaths, respectively.
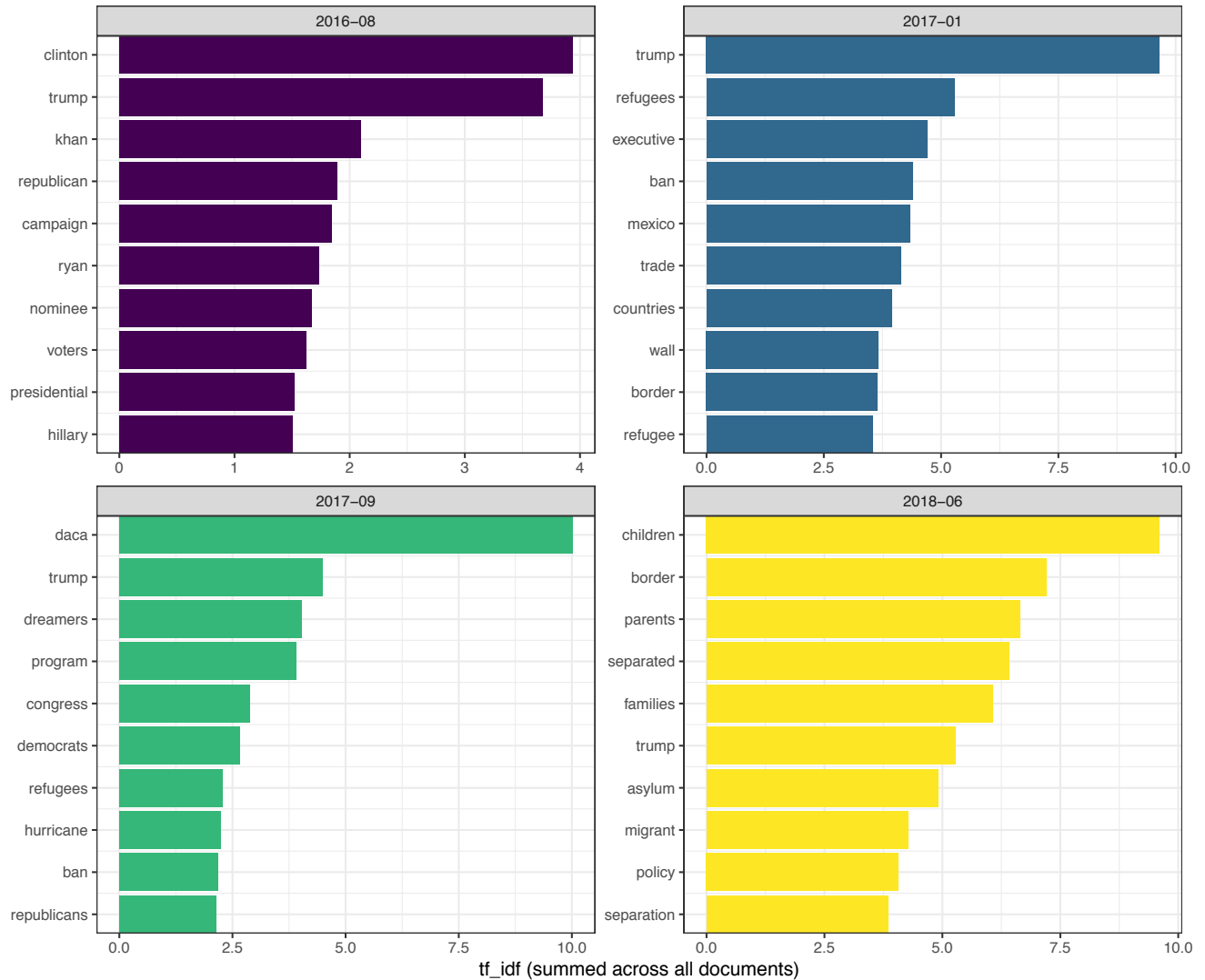
Figure 3. Most important words based on highest tf-idf score for four different months in the Trump era.

Figure 3 illustrates the utility of td-idf in extracting important words from text. Four select months are shown, and in each, the most important words provide a clear picture of one or two major news events that dominated the news cycle during that month. In August 2016, this was the presidential election; in January 2017, the travel ban and the proposed wall on the Mexican border; in September 2017, the proposed end of the DACA program and Hurricane Maria; and in June 2018, the family separation policy.

Further research
We have collected a dataset of over 21,000 migration-related news articles and parsed the information to get in a format that is suitable for text analysis. Our next stage of data collection will involve extracting past tweets over the period of the first two years of Trump's presidency. Analysis will then focus on looking at concurrent patterns in the two datasets. In particular, we are interested in seeing how sentiment in public opinion, as measured by social media data, changes in response to migration-related events and news coverage, and whether this varies over

time and by geographic location. In addition, we will use topic modeling to see what migration-related topics are discussed on social media and how well they correspond to new coverage and policy changes.

References

Atwell Seate, A., & Mastro, D. (2016). Media's influence on immigration attitudes: An intergroup threat theory approach. *Communication Monographs*, *83*(2), 194–213. https://doi.org/10.1080/03637751.2015.1068433

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. https://doi.org/10.1145/2133806.2133826

Brader, T., Valentino, N. A., & Suhay, E. (2008). What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration Threat. *American Journal of Political Science*, *52*(4), 959–978. https://doi.org/10.1111/j.1540-5907.2008.00353.x

Dunaway, J., Branton, R. P., & Abrajano, M. A. (2010). Agenda Setting, Public Opinion, and the Issue of Immigration Reform*. *Social Science Quarterly*, *91*(2), 359–378. https://doi.org/10.1111/j.1540-6237.2010.00697.x

Kim, S., Carvalho, J. P., Davis, A. G., & Mullins, A. M. (2011). The View of the Border: News Framing of the Definition, Causes, and Solutions to Illegal Immigration. *Mass Communication and Society*, *14*(3), 292–314. https://doi.org/10.1080/15205431003743679

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Shifting Public Views on Legal Immigration Into the U.S. (2018, June 28). Retrieved September 4, 2018, from http://www.people-press.org/2018/06/28/shifting-public-views-on-legal-immigration-into-the-u-s/

van Klingeren, M., Boomgaarden, H. G., Vliegenthart, R., Vreese, D., & H, C. (2015). Real World is Not Enough: The Media as an Additional Source of Negative Attitudes Toward Immigration, Comparing Denmark and the Netherlands. *European Sociological Review*, *31*(3), 268–283. https://doi.org/10.1093/esr/jcu089

Wilmoth, J. R., & Ball, P. (1992). The Population Debate in American Popular Magazines, 1946-90. *Population and Development Review*, *18*(4), 631–668. https://doi.org/10.2307/1973758

Zúñiga, H. G. de, Correa, T., & Valenzuela, S. (2012). Selective Exposure to Cable News and Immigration in the U.S.: The Relationship Between FOX News, CNN, and Attitudes Toward Mexican Immigrants. *Journal of Broadcasting & Electronic Media*, *56*(4), 597–615. https://doi.org/10.1080/08838151.2012.732138