



# Artifacts of Crisis: Textual Analysis of Euromaidan

Thomas Magelinski<sup>1</sup>(✉), Zachary K. Stine<sup>2</sup>, Thomas Marcoux<sup>2</sup>,  
Nitin Agarwal<sup>2</sup>, and Kathleen M. Carley<sup>1</sup>

<sup>1</sup> Carnegie Mellon University, Pittsburgh, PA 15213, USA  
{[tmagelin](mailto:tmagelin@cs.cmu.edu),[kathleen.carley](mailto:kathleen.carley@cs.cmu.edu)}@cs.cmu.edu

<sup>2</sup> University of Arkansas at Little Rock, Little Rock, AR 72204, USA  
{[zkstine](mailto:zkstine@ualr.edu),[txmarcoux](mailto:txmarcoux@ualr.edu),[nxagarwal](mailto:nxagarwal@ualr.edu)}@ualr.edu

**Abstract.** We analyze three textual data streams to characterize the change that occurred during the Ukrainian revolution of 2014. These data streams include legislative bill text, posts on Ukrainian political blogs, and Twitter data. Each stream provides a different perspective: politicians, local citizens, and global citizens. It is apparent that bill production stalled early on in the demonstrations, and that the post-revolution government quickly began voting on bills. Topic analysis of blogs and tweets revealed growing interest in Ukraine following the march on the legislature. Interest in Ukraine eventually overtook that of the conflict in the Middle East, before dying back down in the following month. Our results suggest that a stall in bill production may be an early indicator of dysfunction in the government, while spikes in Twitter activity can be seen almost immediately after the event. This effect is true for blogs as well, although for a prolonged period, implying a more detailed discourse about the event.

**Keywords:** Verkhovna Rada · Factions · Social media · Euromaidan

## 1 Introduction

In November of 2013, a wave of protests overtook Ukraine. These protests are now called The Euromaidan. The Euromaidan culminated in the Ukrainian revolution of February 2014, when violence reached the Ukrainian legislator's doorstep, and the president fled to Russia. At the heart of the movement were negotiations on the EU association agreement. Due to its placement between the EU and Russia geographically, Ukrainian politics are of great significance to the

---

This material is based upon work supported by the Office of Naval Research Multidisciplinary University Research Initiative (MURI) under award number N00014-17-1-2675. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Office of Naval Research. Additionally, Thomas Magelinski was supported by an ARCS foundation scholarship.

rest of the world. Methods for characterizing and understanding transformative events such as The Euromaidan play an important role for national security.

More generally, groups of aligned politicians, or “factions” are of great interest in political science. While politicians may technically be aligned with their party, faction analysis can uncover their true political allies. This gives a more data-driven picture of a country’s political landscape. Due to Ukraine’s geopolitical importance and complex party structure, it is a powerful case-study for developing methods to understand factions.

Much of the prior work on Ukrainian factions has been focused on detection of disruptive events in faction structure or beliefs. The prior work uses roll call voting data from the Verkhovna Rada, Ukraine’s legislature, to analyze both how parliamentarian’s political positions change, as well as how their political alliances change. This analysis combined with change detection leads to known dates where the Rada has undergone change.

Knowing that a change has occurred is a good first step, but it opens the door to many more questions: What sparked it? Are there early indicators that a change is coming? Did the public react to the change? These and other questions arising cannot be answered using roll call votes alone, but are key to understanding what factors affect change in factions. In this work, we look to answer some of these questions. To do so, we use three sources of media collected around the revolution of 2014: legislative bill text, Ukrainian blog posts, and Twitter data. These forms of media were selected in order to take a deeper dive into the inner workings of the parliament, the thoughts of the Ukrainian people, and a more global perspective of the situation. We show that topic analysis on each of these data streams sheds light on different aspects of changes in factions.

## 2 Prior Work

Various models have been developed to explain the occurrence of nonviolent uprisings (including grievance approaches, resource mobilization theory, modernization theory, and political opportunity approaches), which have been shown to have varied explanatory power when comparatively analyzed [4]. In the case of Ukraine, legislative change points have been studied in Ukraine through two main mechanisms in prior literature [9,14]. First, through changes in relationships between the politicians. Relationships between MPs can be quantitatively studied using a network science perspective by creating a co-voting network. This network uses link weights to encode the frequency that a pair of MPs co-vote, or cast the same vote on a bill. Then, community detection algorithms uncover groups of MPs, or unofficial “factions.” Faction detection uncovers many relationships that cut across party lines, and give a sense of the true political landscape [8].

The MP relationship network in the Verkhovna Rada, is inherently dynamic, as alliances are constantly changing. Thus *dynamic* community detection must be used to discover when factions are disrupted. [12]. One such method is time segmentation, which involves segmenting a dynamic network into a series of

static networks such that the community structure within time segments is relatively static, and the change between segments is maximized [7]. Applying this methodology to the 8th convocation of the Verkhovna Rada showed a massive shift in faction structure following the revolution in February of 2014 [9].

The second mechanism for analyzing political change in Ukraine is considering the bills themselves [9, 14]. A traditional way of analyzing roll-call bill data is through ideal-points, wherein both bills and politicians are placed on a political spectrum from conservative to liberal [5, 6]. Dynamic ideal points models, then, can uncover changes in political ideology [3, 10]. When applied to the Verkhovna Rada, it was seen that ideological shifts occurred on February 2014, during the revolution that also saw change in faction structure [9].

Beyond just votes, bills themselves contain quite valuable data. Each bill is tied to a committee, and the bill text can be analyzed to understand which issues it addresses. Topic modeling, through Latent Dirichlet Allocation (LDA), provides an automated way of analyzing bill text [1]. From the uncovered topics, bill *novelty* can be calculated, which measures uniqueness of a bills topic mixture, given those that have come before it [14]. Changes in novelty, then, can provide a different type of change occurring in the legislature.

While all of the previously described methods are important for understanding when change occurs, different techniques must be used to understand the implications of each change. To this end, we turn to topic analysis. Topic analysis of bills already informed some of the change-points though bill novelty, but the topics themselves have not yet been deeply studied. Additionally, to understand the implications of major political events, one must look beyond just the legislature and into public discourse, which can be found on blogs and Twitter. Through these three mediums, bills, blogs, and tweets, we can see changes in actual political action, how it is received locally, and how it is discussed globally.

## 3 Methodology and Data

### 3.1 Legislation

The legislation produced by the Ukrainian parliament, the Verkhovna Rada, serves as an important record of the political goals and priorities of the parliament over time. The Verkhovna Rada makes the text of registered bills publicly available through their website<sup>1</sup>. We download the text of each available bill from the Verkhovna Rada website for convocations 5–8, resulting in 35,112 total Ukrainian documents, which will serve as training data for topic modeling.

We preprocess bill text by first lowercasing and tokenizing each document. We reduce the vocabulary size by excluding word types based on three criteria: the 100 most frequent word types, word types that occur in more than 25% of the collection of documents, and word types that occur in fewer than ten documents are stopped. This results in a vocabulary size of 36,566 word types. We then train topic models via LDA [1], as implemented in [11], with 30, 50, and

<sup>1</sup> <https://rada.gov.ua>.

100 topics. We find that many of the key topics exist across the three models, but report our results using the 50-topic model due to its interpretable topics which are well-balanced in terms of specificity and generality. Our topic labels are qualitatively determined by examining both high probability words in each topic, as well as exemplar documents that feature the topic.

We next examine bills that were voted on in the parliament before, during, and after the 2014 revolution that began in mid-February of that year. Only eight bills are available in January, with another nine available in December. Therefore, we include November 2013 through January 2014 to represent the period before the revolution. Giving 82 bills in the period before the revolution, 73 bills in February 2014, and 80 bills in March 2014.

To compare differences in bill topics between the three time periods, we assign each document to the dominant topic with the highest probability in the document's topic distribution after ignoring two purely stylistic topics. By organizing the bills in each time period by their topic assignments, we can interpret the changing legislative priorities across each period.

### 3.2 Blogs

In order to gain insight on social media activity related to the Ukrainian revolution, we collect blog posts discussing political topics relevant to Ukraine. Blog sites were collected identified from various sources such as Twitter, geofencing, Google, etc., using relevant keywords and hashtags. We perform manual relevancy checks on identified blogs to ensure they are relevant, active, and public. Finally, we collect data using the Web Content Extractor (WCE) tool<sup>2</sup>.

The range of collected blog posts spans from February 2004 to December 2018 for a total of 168,121 English blog posts. We perform a rigorous data-cleaning routine to remove noise and standardize all publication dates. All text is lowercased, tokenized, and words are removed using a standard list of English-language stopwords. The processed blog posts are used to train a topic model with 50 topics via LDA [1] using the implementation in [11].

With the topic model, we analyze blog posts published before, during, and after the Ukrainian revolution from January 1st to March 31st, 2014, which comprises 2,438 blog posts from 19 unique blog sites. We examine the 20 most frequent dominant topics from these posts and focus our analysis on six highly relevant topics. In order to analyze the change in narratives during this time period, we calculate the average topic proportions for the relevant topics across all blog posts published on a given day and visualize the resulting topic streams.

### 3.3 Twitter

Twitter data comes in the form of tweets, or short user posts, capped at 280 characters. This data could be analyzed in a similar way to blogs, but one feature opens the door for a more interpretable analysis: hashtags. Hashtags allow users

<sup>2</sup> <http://www.newprosoft.com/>.

to self-label the content of their tweet. Since hashtags are searchable, and are advertised by the platform through the “trending” category, users who use them have potential to reach more users [13]. Thus, the use of hashtags is pervasive throughout twitter, and can be leveraged to understand topics [15].

Here, we use a simple method to uncover twitter topics through the hashtag co-occurrence network. First, a network is created with hashtags represented as nodes. Then, connections between a pair of hashtags are drawn as a link in the graph when both hashtags occur in the same tweet. Link weights are used to reflect the total number of tweets that a pair of hashtags occurred in. Given the short nature of tweets, this is a strong notion of association.

Since individual hashtags are a specific label that a user gives their tweet, large groups of closely related hashtags can be understood as a topic. For example, if a group of hashtags is #rada, #ukrainepolitics, #legislature, the overall topic is Ukrainian politics. At scale, we can uncover these groups through community detection algorithms. Perhaps the most popular community detection algorithm is Louvain grouping, which attempts to maximize modularity [2]. It has gained popularity due to its success in finding empirically validated groups, and its ability to scale to large datasets. Thus, we proceed using Louvain grouping of the hashtag co-occurrence network as topics in our Twitter data.

**Table 1.** Data summary for Twitter data surrounding the revolution of 2014.

Month	Tweets	Users	Hashtags	Links	Topics
January	38105	17810	2994	22753	6
February	71763	40877	6669	38941	4
March	40945	25784	2811	14714	3

Twitter Data was collected from January to March 2014, using its API. The dataset was broken up into 3 periods, each 1-month long: before, during, and after. Only topics appearing in more than 5% of the tweets were considered in analysis. A summary of the dataset is provided as Table 1.

## 4 Results

### 4.1 Legislation

From examining the most frequent topics in the bills from each of the three time periods, we find stark changes between the frequent topics from the period before the revolution and both February and March of 2014. Notably, the most frequent dominant topics from the period before the revolution do not appear to reflect the sociopolitical shifts underway. These bills concern the commemoration of certain anniversaries (topic 30), budgets (topic 18), pensions (topic 31), and crime (topic 3) (see Table 2).

**Table 2.** Most frequent dominant topics from bills voted on before the revolution.

Dominant topic	Bills	Proportion of bills in period
Topic 30, Commemorations	19	0.2317
Topic 18, Budgets and funds	5	0.0610
Topic 31, Pensions	5	0.0610
Topic 02, Releases and Dismissals	4	0.0488
Topic 03, Offenses	4	0.0488

As stated in Sect. 3.1, votes were only taken on 9 and 8 bills during December 2013 and January 2014 respectively. Careful readings of these bills reveal potential antecedents for the drastic changes observed later in February 2014. These bills include a resolution of no confidence in the Cabinet of Ministers of Ukraine (bill 3692), the formation of an investigative commission on the actions of law enforcement agencies against protesters (bill 3832), and the legal protection of protesters (bill 3787) among other protest-related bills.

**Table 3.** Most frequent dominant topics from bills voted on during February 2014.

Dominant topic	Bills	Proportion of bills in period
Topic 47, Executive services and agencies	14	0.1918
Topic 11, Administrative-executive	7	0.0959
Topic 01, Administrative-parliamentary	6	0.0822
Topic 02, Releases and dismissals	6	0.0822
Topic 41, Constitutional and human rights	6	0.0822
Topic 33, Conflict, violence, security	5	0.0685
Topic 22, Early terminations	4	0.0548

During February 2014, we find that topic 47 accounts for almost 20% of the bills voted on during this time period, which reflect the remaking of the Ukrainian government including various appointments of new ministers and commissioners, the appointment of Arseniy Yatsenyuk as the new Prime Minister, the formation of a new Cabinet of Ministers, and others (see Table 3).

Each of the topics shown in Table 3 reflect different aspects of the establishment of a new government in response to the revolution. Notably, the topic 2 bills in February are about the dismissal of various government officials while the topic 22 bills concern the early terminations of MPs taking up new positions outside of parliament. Also notable is that three of the five bills dealing with topic 33 address violence carried out against protesters. The two remaining bills concern state security with one referencing Russia.

The most frequent dominant topics in bills voted on after the revolution in March 2014 share several similarities with bills voted on in the month prior. Bills with dominant topics 22, 2, and 47 all continue similar actions seen in February to remake the Ukrainian government. However, several changes are clear. Notably, each of the four bills with dominant topic 33 now concern security threats stemming from Russia, highlighting a shift in priorities from conflict involving protests (as seen in February). Additionally, two of the bills with dominant topic 36 reference Crimea. A summary of the most frequent dominant topics from bills during March 2014 is provided in Table 4.

**Table 4.** Most frequent dominant topics from bills voted on during March 2014.

Dominant topic	Bills	Proportion of bills in period
Topic 22, Early terminations	10	0.1250
Topic 02, Releases and dismissals	9	0.1125
Topic 47, Executive services and agencies	6	0.0750
Topic 33, Conflict, violence, security	4	0.0500
Topic 34, Parliamentary hearings	4	0.0500
Topic 36, Territorial and local issues	4	0.0500

## 4.2 Blogs

In Table 5, we present the six topics that are both highly frequent and relevant. These topics have been manually selected because of their high proportion within the corpus and because they are specific political topics. The table gives the number of blog posts in which each topic is dominant in addition to the proportion of documents, and some of the most frequent and relevant words to justify the topic's label.

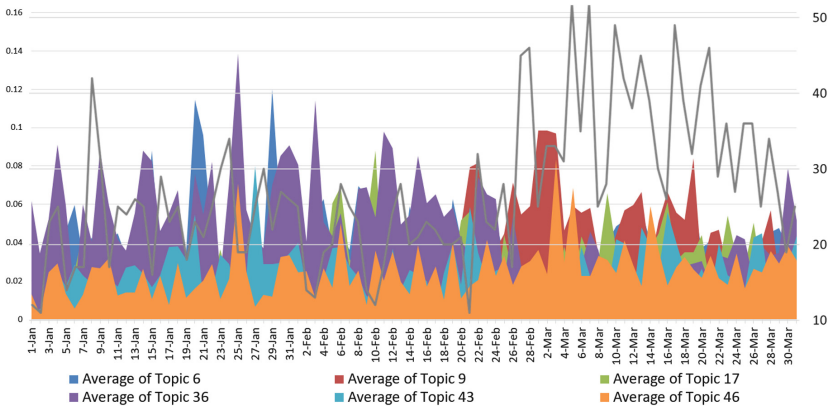
**Table 5.** Most frequent dominant topics from blog posts January-March 2014.

Dominant topic	Word 1	Word 2	Word 3	Word 4	Word 5	Count	Proportion
Topic 6: Iran	Iran	Iranian	Political	Regime	Deal	118	0.0484
Topic 46: War	War	Military	Nato	States	United	47	0.0193
Topic 43: Elections	Party	Political	Election	President	Elections	41	0.0168
Topic 17: America	America	Americans	American	Clinton	Trump	109	0.0447
Topic 9: Ukraine	Ukraine	Ukrainian	Poroshenko	Crimea	Kyiv	125	0.0513
Topic 36: Syria	Syria	Islamic	Turkey	Terrorist	Damascus	164	0.0673

In Fig. 1, we show the average topic proportions of the six topics of interest for each day along with the number of blog posts published on that day. From this,

we find that the blog discourse in January 2014 is dominated by discussions about the Middle East, specifically Iran (topic 6) and Syria (topic 36). War-related discourse (topic 46) is present throughout.

We see this pattern continue until mid-February, when Ukraine-related discussions (topic 9) begin to dominate and continue to do so throughout much of early March. Several peaks of war-related discourse (topic 46) occur in early March as well. Notably, the daily number of blog posts is elevated throughout much of late February and early March, with posting frequency gradually declining in late March. By late March, the six topics appear more evenly mixed ending with a rise in Syria-related discussions (topic 36).



**Fig. 1.** Average daily topic distribution of Blog posts between January and March 2014 with daily blog count. The left vertical axis shows each topic’s average proportion, while the right vertical axis gives the number of posts on each day represented by the line not corresponding to a topic.

### 4.3 Twitter

The benefit of analyzing Twitter is its wide audience. Given Twitter’s global user base, it is expected that conversations span many geo-political issues. This is seen in the topics obtained in our three time periods. Other than Ukrainian unrest, one of the largest developing stories of 2014 was the Syrian civil war. Specifically, there was increase in conflict with Islamist groups in the region.

Both discussion of Ukraine and the Middle East are highly prevalent in the topic groups. In January and March, there is a topic dedicated to Middle Eastern political discussion. This topic had top hashtags “syria”, “iran”, “iraq”, and was present in 20.7% and 16.6%, of tweets in the respective months. In February, the discussion was split. First into a small topic containing hashtags like “iran” and “iraq”, occupying 7.7% of tweets. Other hashtags like “syria” entered the main topic, which was focused on Ukraine, and was present in 56.3% of tweets.

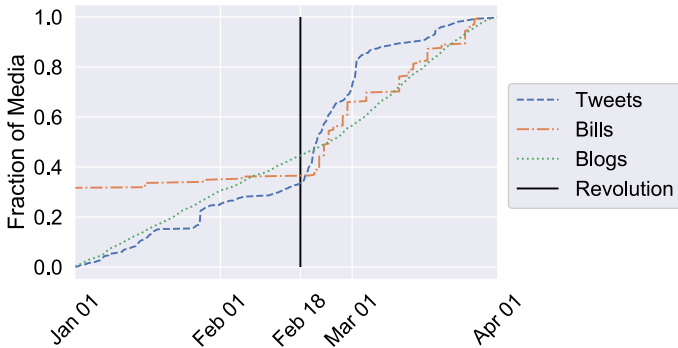


While the Middle Eastern topics were a large part of the conversation, topics relating to Ukraine were a larger fraction and increased significantly during the protests. We could expect this given the increase in total tweets when the first protests broke out on February 18th, as shown in Fig. 2. Ukrainian topics were present in 35.1%, 56.3%, and 34.1% of the tweets in each month, with top hash-tags: “ukraine”, “kiev”, “euromaidan” (English and Ukrainian), and “russia”.

This analysis demonstrates the power of Twitter analysis in understanding geo-political events in real time. The activity plot in Fig. 2, show that Twitter user’s activity can respond almost immediately to a significant event. At the same time, topic analysis shows that not only does the volume of tweets change, but the relative amount of discussion of different events does as well.

#### 4.4 Data Stream Timing

After completing analyses on the different data streams, a key question remains: how do the streams fit together? One way of answering this is by seeing when exactly changes occurred across modalities. To visualize this, Fig. 2 shows the cumulative fraction of media produced at a given time. Thus, the slope of each line is the rate at which new bills, blogs, or tweets are being voted on or created.



**Fig. 2.** Cumulative fraction of the data streams over time. As previously noted, data collection for bills goes back further than January.

First, we see that bill production has stalled months before the protests met the Rada. Prolonged inactivity in the legislature, then, may be in early indicator that tensions are rising. Posting frequencies from blogs and Twitter data do not seem to have a recognizable “early indicator” property in the same way.

After the event, Twitter reacts first, with an immediate jump in activity. This combined with the topic analysis shows that this is an increase in tweets specifically about the revolution. So, while it might be difficult to use Twitter to *predict* events such as this, the platform seems very appropriate for detecting and analyzing event once it has taken place.

Next, bill activity is kick-started, breaking the stalemate where no laws could be voted on. There is a slight but prolonged uptick in blog activity, showing it is less sensitive to external events than Twitter but provides a detailed discourse about the event. Since the protests had started in November, Ukrainian citizens were likely invested in the situation from the beginning.

## 5 Conclusions and Future Work

Through a multi-modal topic analysis of the 2014 Ukrainian revolution, we have constructed a multifaceted view into language characterizing the event, both from within the Ukrainian government itself and from English-language discourse about the event through Twitter and blogs. Within Ukrainian legislation, we find that the vast majority of bills voted on during February and March of 2014 directly concern the formation of a new post-revolution government following a conspicuous drop in the number of bills voted on during December 2013 and January 2014. From English-language discourse on Twitter and various blogs, we find that attention to Ukraine breaks through prevailing concerns about the Middle East and the war taking place in Syria. Among the blogs analyzed, posts about Syria dominate until late February when posts concerning Ukraine become dominant throughout early March 2014. We find that a similar pattern occurs in hashtag usage on Twitter—attention given to the Middle East is replaced by attention to Ukraine during February 2014.

Taken together, these findings provide a rigorous and in-depth characterization of the 2014 Ukrainian revolution. Future research will focus on how tools of analysis used in this work can directly tie into faction analysis. The lack of MP presence on social media makes this challenging. However, the overlap in topic meaning and trends between sources point to way forward: tie factions to topics in bill text, and then study similar topics arising in social media.

## References

1. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**(Jan), 993–1022 (2003)
2. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**(10), P10008 (2008)
3. Caughey, D., Warsaw, C.: Dynamic estimation of latent opinion using a hierarchical group-level IRT model. *Polit. Anal.* **23**(2), 197–211 (2015)
4. Chenoweth, E., Ulfelder, J.: Can structural conditions explain the onset of non-violent uprisings? *J. Conflict Resolut.* **61**(2), 298–324 (2017). <https://doi.org/10.1177/0022002715576574>
5. Clinton, J., Jackman, S., Rivers, D.: The statistical analysis of roll call data. *Am. Polit. Sci. Rev.* **98**(2), 355–370 (2004)
6. MacRae, D., Jr.: *Dimensions of Congressional Voting: A Statistical Study of the House of Representatives in the Eighty-First Congress*. University of California Press, Berkeley (1958)

7. Magelinski, T., Carley, K.M.: Community-based time segmentation from network snapshots. *Appl. Netw. Sci.* **4**(1), 1–19 (2019). <https://doi.org/10.1007/s41109-019-0136-1>
8. Magelinski, T., Cruickshank, I., Carley, K.M.: Comparison of faction detection methods in application to Ukrainian parliamentary data. In: 2018 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation, Washington DC (2018)
9. Magelinski, T., Hou, J., Mylovanov, T., Carley, K.M.: Detecting disruption: identifying structural changes in the Verkhovna rada. In: Thomson, R., Bisgin, H., Dancy, C., Hyder, A. (eds.) SBP-BRiMS 2019. LNCS, vol. 11549, pp. 194–203. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-21741-9\\_20](https://doi.org/10.1007/978-3-030-21741-9_20)
10. Martin, A.D., Quinn, K.M.: Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. *Polit. Anal.* **10**(2), 134–153 (2002)
11. Řehůřek, R., Sojka, P.: Software framework for topic modelling with large corpora. In: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, ELRA, Valletta, Malta, pp. 45–5, May 2010. <http://is.muni.cz/publication/884893/en>
12. Rossetti, G., Cazabet, R.: Community discovery in dynamic networks: a survey. *ACM Comput. Surv. (CSUR)* **51**(2), 1–37 (2018)
13. Small, T.A.: What the hashtag? A content analysis of Canadian politics on Twitter. *Inf. Commun. Soc.* **14**(6), 872–895 (2011)
14. Stine, Z.K., Agarwal, N.: A quantitative portrait of legislative change in Ukraine. In: Thomson, R., Bisgin, H., Dancy, C., Hyder, A. (eds.) SBP-BRiMS 2019. LNCS, vol. 11549, pp. 50–59. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-21741-9\\_6](https://doi.org/10.1007/978-3-030-21741-9_6)
15. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic sentiment analysis in Twitter: a graph-based hashtag sentiment classification approach. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp. 1031–1040 (2011)