

# A Computational Analysis of Online Political Discourse on Ukrainian-Russian Blogosphere

Mainuddin Shaik

COSMOS Research Center  
University of Arkansas – Little Rock, Arkansas, USA  
mxshaik@ualr.edu

Nitin Agarwal<sup>1,2</sup>

<sup>1</sup>COSMOS Research Center  
University of Arkansas – Little Rock  
<sup>2</sup>International Computer Science Institute, University of  
California, Berkeley  
nxagarwal@ualr.edu

**Abstract**— This research addresses the critical problem of how online political discourse, particularly in the Ukrainian-Russian blogosphere, shapes and is shaped by geopolitical conflicts from 2014 to 2020. Utilizing a comprehensive computational framework, the research analyzes over 26,000 English-language blog posts, employing advanced methods such as topic modeling, influential analysis, sentiment analysis, and network analysis to uncover coordinated information campaigns and the propagation of pro-Russian narratives. The study presents new results, revealing the complex interplay between narrative intensity, sentiment, and toxicity in response to key geopolitical events. These findings underscore the influential role of blogs in modern political communication, providing crucial insights into the dynamics of online influence. Ultimately, the research concludes that understanding these dynamics is essential for developing strategies to navigate and counteract the evolving digital landscape in politically charged environments.

**Keywords**-Social Media; Blogs; Ukraine; Russia; Politics; Online Discourse.

## I. INTRODUCTION

Social media has undoubtedly become a vital channel for communication in today's digital landscape, enabling the seamless exchange of information and ideas. These platforms empower individuals to share their perspectives, disseminate knowledge, and shape discussions. Social media has played a crucial role in initiating and executing various cyber-campaigns, ranging from awareness initiatives to actions that threaten democratic principles.

Blogs, in particular, have evolved significantly since their inception, transforming from personal diaries to powerful tools for expressing opinions, conveying information, and shaping narratives. The blogosphere has become a critical medium for content framing, allowing individuals to share their thoughts without limitations. Blogs have become essential tools for understanding and interacting with political actors in our interconnected online world, making them a vital component of online narrative campaigns.

The work in [1] provided a multidimensional analysis of YouTube communities in the Indo-Pacific region, focusing on geopolitical discourse related to China, the Uyghur crisis, and COVID-19. Unlike this study, which analyzes blogs in the Ukrainian-Russian context, their work emphasizes video-based platforms, engagement dynamics, and community

behaviors to understand online influence and misinformation trends. Together, these studies reveal how both video and text-based platforms contribute to the propagation of geopolitical influence, highlighting distinct yet interconnected dynamics in shaping public perception and online behavior during politically charged situations.

This research focuses on English-language blog discourses related to Ukraine, aiming to identify prevailing viewpoints on Ukrainian matters between 2014 and 2020. Ukraine's unique geopolitical position between Russia and the European Union makes it a significant case study, with profound implications for the global community. This period witnessed pivotal political events in Ukraine, including the Euromaidan protest movement and the Russian annexation of Crimea. (leading to an ongoing conflict in eastern Ukraine) underscore this period's significance [2].

While studying the Ukrainian-language blogosphere would offer a direct view of Ukraine's internal viewpoints, this research focuses on understanding the perceptions of and engagement with Ukrainian issues within the English-language blogosphere. This approach allows us to ascertain what aspects of Ukraine are salient from an English-language viewpoint, such as Ukraine's role in US politics.

The study examines blog data through text analysis, author relationships, and patterns such as topic discussions and URL sharing. It detects underlying themes in blog texts and their temporal changes, evaluates sentiment and toxicity in language, and uncovers a coordinated anti-Ukraine narrative, diverse stories from the Donetsk and Donbas conflicts, political extremism from Euromaidan, and significant coverage of right-wing nationalism. The research provides a comprehensive understanding of the blogosphere's role in shaping public discourse on Ukraine. The remainder of this paper is organized as follows: Section II presents a literature review covering the evolution of blogs from personal diaries to influential political platforms, the impact of blogs in political discourse, and the dynamics of political blogging with a focus on Ukrainian geopolitical contexts. Section III describes the methodology, including the data collection strategies and computational analysis techniques such as topic modeling, sentiment analysis, influence metrics, and network analysis. Section IV provides the results, detailing the topic models related to Ukrainian political news, international relations, and domestic affairs, alongside their sentiment,

toxicity, and influence and elaborates on the network analysis, focusing on the relationships between entities, influential actors, and core networks of bloggers. Finally, Section V discusses the key findings, highlighting the prevalence of pro-Russian narratives, and concludes with broader implications for understanding public discourse in the Ukrainian-Russian context.

## II. LITERATURE REVIEW

The literature review examines how blogs transitioned from early 2000s democratic discussion spaces to significant political influence platforms, exploring the motives behind political blogging, its conversational dynamics, and communication styles. It concludes with studies on Ukraine's geopolitical discourse, underscoring the blogosphere's impact on political dialogue.

### A. Impact of Blogs

Authors in [3] observed an intriguing trend from 2007 to 2013 wherein police bloggers in the United Kingdom discontinued their blogging activities and indirectly influenced their counterparts within the community to follow suit. This shift signifies the tangible consequences and potential accountability associated with blogging activities in real-world scenarios. This implies that bloggers could face repercussions for sharing information or expressing their opinions online, which might inadvertently curb the exercise of free speech.

Numerous studies have explored the influence of blogs on democratic processes and on pivotal, geographically specific events, such as elections worldwide. These examinations encompass countries including Australia [4], South Korea [5], Sweden [6], Greece [7], Kuwait [8], Indonesia [9], and Singapore [10]. A common conclusion is that blogs often serve as insulated echo chambers, reflecting and amplifying existing viewpoints rather than promoting diverse discourse.

Several studies have employed a sociological lens to scrutinize the political potential of blogs, utilizing research tools such as self-reporting surveys, user and expert interviews, and public data. In contrast, our study seeks to offer computational solutions to address our research questions, providing a fresh perspective on the dynamics and implications of blogging in the political sphere.

### B. Blogs and politics

Focusing on the political implications of blogs, [11] unveiled the rapid transformation of online communities and foresaw blogs evolving from democratic tools into personality-driven platforms. They further highlighted a preference among public figures to use Twitter as a soapbox, especially when toxic online communities gravitate towards alternative platforms promoting free speech (Gab, Parler, etc.). Echoing these insights, [12] suggested that blogging might be more appealing to ideologies valuing individualism. Their study in Sweden found a notable pattern: right-wing content gravitated towards candidate-centric blogs, while the left favored collective discourse.

### C. Access to and evolution of blogs

Researchers [11], [13], [14] observe that, despite initial technical barriers, the advent of user-friendly platforms now enables virtually anyone with a subscription to create and manage blogs. They highlight that blogs gain popularity during significant events, like elections, conflicts, or terror attacks, due to their immediacy. This trend has stimulated the rise of citizen journalism. [14] further concludes that in Malaysia, the blogosphere has fostered greater democratic access. [12], [15], [16] underscore the growing role of blogs in political campaigning, particularly during election cycles. [16] studying the EU 2009 Parliament elections, notes varied communication strategies among these blogs.

### D. Political potential of blogs for Members of Parliaments (MPs)

These authors [11] probe the potential political role of blogs, questioning whether they are suitable tools for politicians to stimulate public dialogue or disrupt party systems. They, along with [5], note a cautious approach to new technology among many members of parliament (MPs), particularly in Korea, while [17] found that right-wing politicians are generally quicker to adopt such innovations. These observations align with our research findings, which reveal a minority of MPs dominating the online dialogue, with the majority largely absent. This data deficiency has presented significant challenges when attempting to employ computational solutions and network algorithms to analyze elected officials' online presence. The work in [11] also highlights the prominent role of advocates in driving blog usage, especially during election periods, suggesting that many MPs may have bypassed blogging in favor of direct engagement on social media platforms.

### E. Nature of MP blogs and political

Reference [11] observed that, in most instances, MPs typically have minimal involvement in either the creation or the daily management of their blogs. Consequently, these blogs often revolve around the MPs, rather than being authored by them, frequently featuring press articles about them. The possibility of facing repercussions for publishing controversial content also discourages engagement with blogs. Their analysis of out-links showed that bloggers primarily linked to other blogs sharing their partisan views or personal interests, with no indication of blogs bridging political factions or divisions. Yet, they did find instances of collaborative efforts among specific groups of MPs with other bloggers. They also observed that blogs fare exceptionally well with traditional media personality cults and are particularly efficient during election campaigns due to their niche but highly attentive audience. Given the ever-changing landscape of social media, it is likely to morph significantly by the next election cycle. Therefore, serious consideration should be given to burgeoning social media platforms that bear a resemblance to Parler.

### F. General blogs studies covering politics

Blogs often provide a space for readers to interact with the authors' entries, fostering discussions between bloggers and

their audience. This capacity for engagement and collaboration paves the way for direct communication between the readers and the writers, making blogs a viable platform for political participation and discourse [18]. In this context, blogs have been recognized as significant catalysts for political mobilization [19]. Blogs effectively sidestep the mediating role traditionally assumed by journalists. They empower citizens and provide a conduit for politicians across the spectrum to manage and manipulate the messages communicated [6]. As [20] argues, weblogs perform unique politically oriented functions with potential ramifications on the political landscape. They can influence the direction of public discourse from the ground up—a concept known as "public agenda-setting"—and thereby shape both online and offline public deliberations. For instance, bloggers often present facts, arguments, and analyses that are either overlooked or underrepresented in traditional media outlets. This allows them to question and influence the narrative of political news coverage directly. The substantial growth of the U.S. political blogosphere in recent years has prompted the academic community to reevaluate the process of agenda-setting [21]. Moreover, bloggers have the potential to conduct "independent investigations," revealing political scandals or controversies and thereby contributing to greater transparency [20].

#### G. Ukrainian blog studies on politics and public affair discourse

Political blogging as a subject of influence and agenda-setting has been a recurrent theme in literature [22]. Bloggers often emerge as influential agenda-setters, as proposed by [23]. [24] further emphasized this role, illustrating how bloggers amplified elite narratives in the context of the 2016 Dutch referendum on Ukraine. Their study highlighted bloggers as intermediaries in the strategic deployment of narratives to legitimate foreign policy. Studies exploring war narratives between Ukraine and Russia found that bloggers on both sides often employed a troll-like discourse, marked by excessive patriotism, emotive language, and derogatory rhetoric aimed at discrediting opponents [25], [26]. This included practices such as prosecution, blaming, threatening, and negative forecasting.

In a study focused on the Ukrainian blogger community during the 2004 Orange Revolution, it was found that the bloggers were a mix of activists and professional journalists. Their content provided a unique blend of personal online diaries and alternative news that mainstream media often overlooked or ignored [27]. This study also suggested that internet-using Ukrainians were more likely to be proactive online citizens, engaging in activities ranging from forwarding political emails to participating in online chat discussions. [28] further investigated the role of blogs as sources of agenda and opinion setting in Ukraine and Russia. They discovered that topic modeling could identify public agendas, their composition, structure, salience, and evolution, even without prior knowledge of the issues being addressed. Coupled with methods to track attitudes, this approach generated a comprehensive representation of self-initiated public opinion, analogous to traditional opinion polls.

#### H. Other computational studies of blogs

Previous research by [29] suggested that while blogs could serve as platforms for discourse, their capacity to enrich public debate was variable. They observed heightened partisanship as elections approached, indicating that public debates within these spaces were significantly influenced by political biases. The New Zealand blogosphere, they found, was typically dominated by a handful of unofficial blogs that overshadowed official media blogs. These researchers posited that the audience for these blogs primarily consisted of those already engaged in political activities, suggesting a limited potential for igniting grassroots movements among the wider population. Building on this earlier work, [30] examined the blogosphere during the 2008 New Zealand election cycle, focusing on the previously identified dominant blogs. Their content analysis revealed a marked negativity in the campaigning period, substantiating findings from their 2005 study. Despite declarations of independence, the popular blogs they analyzed were ideologically aligned with certain campaigns. These blogs expressed strong positive sentiments about the issues that favored their preferred party and harsh criticisms of their opponents [29]. According to Hopkins, while blogs were initially portrayed as venues for policy debate and political mobilization, they have evolved into arenas for robust debate. Participating in these energetic exchanges, Hopkins argued, seems to be the main draw for many commenters. Rather than engaging in political debate per se, they appear to view the discourse as a form of entertainment or competitive sport. Further examination of real political events in Greece by [7] suggests that blogs form clusters around major blogs that share affiliations or are established information sources. They found that these blogs network through interconnected links, resulting in self-segregated clusters, a phenomenon corroborated by Adamic and [12], [31], [32].

### III. METHODOLOGY

In this section, we detail the data collection process and the analytical techniques used, including topic modeling, influence scores, sentiment, toxicity, and network analysis.

#### A. Data Description

For this study, we collected relevant blog posts as described in the collection methodology section, including the post's title, publication date, author, post text, and comments. We collected data for seven years starting from January 1, 2014, to December 31, 2020. Total statistics are reported in Table I. We extracted the title, date, author, post of the article, and comments.

TABLE I. DATA STATISTICS

Total posts	130,668
Unique blog domains	2,111
Unique authors	20,980
Comments	1,487,700

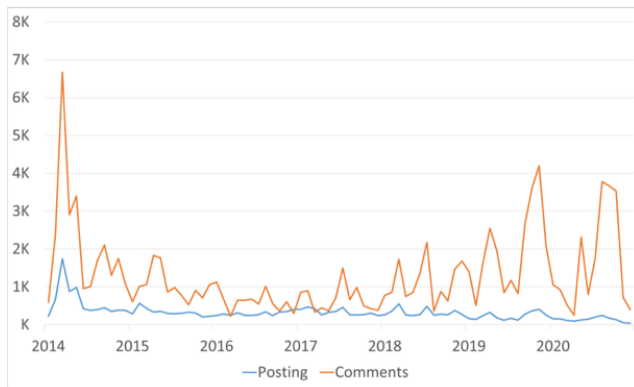


Figure 1. Monthly frequency of filtered blog posts and comments.

From this corpus, we leveraged topic modeling to identify the most relevant blog posts and whittled our corpus down to the numbers reported as a yearly breakdown of post and comment frequencies in Figure 1.

### B. Data Collection

The collection of these blog posts and comments was based on a set of keywords provided by our collaborators at Arizona State University. Data collection was performed with three separate sets of keywords.

The first set relates to discussion surrounding political and geopolitical arguments around the Russian hybrid war in Crimea, eastern Ukraine, and Russia. For this set, a sample keyword for collection would be: “(Russia Kremlin) + (Ukraine totalitarian ethnic-tension discrimination desertion)”. The second set of keywords was related to key events surrounding the Ukrainian revolution, post-Euromaidan movement. An example set of such keywords includes “Ukrainian Revolution, Maidan, Kiev, Ukraine”. The third set of keywords were the names of the Ukrainian political parties, party members, parliament members, and other politicians. An example set of such keywords includes “abdullin, oleksandr, rafkatovych, abramovfedir, mykhaylovyh, communist party of ukraine”.

We initially targeted known blogging sites such as blogspot.com, wordpress.com, and livejournal.com. We then followed a four-step process shown in Figure 2 to collect valid blog URLs that hosted content about the topics of interest.

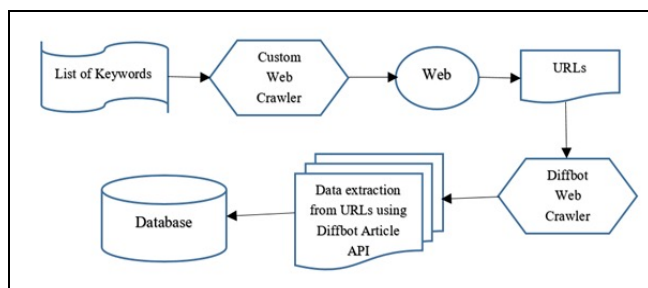


Figure 2. Data collection flow

#### 1) First step:

We deployed a custom private web crawler built on a Python and Scrapy framework [33]. The tool queries the Google Custom Search API through Google’s custom search engine and collects URLs of websites whose articles or bodies are a match for the required keywords. The URL search was conducted by querying the Google platform using a first layer of simple Boolean operations, then a proximity terms search to filter for relevant content.

#### 2) Second step:

Additional blog links were obtained by crawling the blogs presented in the blogrolls from the platforms already collected to target sites outside of the initially queried domains.

#### 3) Third step:

We extracted the blogs and media links from the body of the article, also referred to as out-links, to discover and crawl further websites discussing topics of interest.

#### 4) Fourth step:

Finally, we obtained tweets from Twitter during the Ukrainian revolution period and extracted blog links, leveraging approximate geographic coordinates to locate tweets from both Ukraine and Russia. The collected valid URLs are then processed through Diffbot, a commercial platform that extracts the required data elements from any given articles using “Article Extraction API” [34].

### C. Topic Modeling

In this research, we employ Latent Dirichlet Allocation (LDA) to implement topic modeling, a statistical approach within the field of natural language processing. It categorizes a document set into various abstract “topics”. Each “topic” signifies a set of words that describe a general underlying theme. For each unique blog post (or “document”), we assign a probability indicating its association with a particular topic. The LDA topic model was chosen for its widespread application and proven efficacy [35]. A set of N stopwords are eliminated post-training, a practice which is demonstrated to yield comparable results to pre-training removal [36].

To visualize narrative evolution and extract lexical meaning from our corpus, we designed a processing pipeline. Each document in the corpus is processed, identifying text field, an identification key, and publication date. For topic modeling, we utilized the LDA model as implemented in the Python tool Gensim, following the recommended methodology, pre-processing best practices [37], and best stopwords practices [36]. We found 20 different topics yielded the most relevant outcomes.

Upon training the models, the documents were date-ordered and a Numpy matrix was computed, assigning each document a score for each topic. A document with a high enough score was considered part of that topic. This facilitated the plotting of a chronological graph for each topic and narrative patterns.

### D. Influence

In order to understand the importance of a given blog post within the larger blogosphere, we calculated a measure of blog

influence proposed by Agarwal et al. [38], [39]. Calculating this influence measure for blog posts requires constructing a directed graph in which blog posts are nodes and edges indicate a link in one blog post to another post. For each blog post in the graph, its influence reflects four variables: Recognition,  $\iota$ , which represents other posts which link to this post (i.e., the post's in-links); Activity generation,  $\gamma$ , which represents how many comments the post received; Novelty,  $\theta$ , which represents the posts that the current post links to (i.e., the post's out-links); and Eloquence,  $\lambda$ , which represents the length of the blog post.

In this graph, influence can be thought of as flowing between nodes such that a node's influence is in part a function of the influence of the nodes that it is connected to. In this directed influence graph, the flow of influence through blog post  $x$  is given by a flow function,  $f(x)$ , reflecting the incoming influence of post  $x$  minus its outgoing influence. The influence of post  $x$  is then defined by

$$I(x) = w(\lambda) \cdot (W_{com}\gamma_x + f(x)) \quad (1)$$

where,  $w(\lambda)$  is the weight of the post's length and  $w_{com}\gamma_x$  is the weight of the post's comments. The influence flow of post  $x$  is given by:

$$f(x) = w_{in} \sum_{m=1}^{|\iota|} I(x_m) - w_{out} \sum_{n=1}^{|\theta|} I(x_n) \quad (2)$$

where,  $w_{in}$  and  $w_{out}$  are the weights of the incoming and outgoing influence. The first term reflects a weighted sum of the incoming influence to  $x$  from which the second term, reflecting a weighted sum of outgoing influence, is subtracted.

To analyze the relationships between the influence of blog posts and the topics of blog posts, we combined influence scores with the corresponding topic probabilities of each blog post by multiplying them together. Since each topic is present within each blog post with non-zero probability, this results in a topic-influence score for all topics in each post. To get a sense of the relationship between topics and influence within some time period, we calculate the mean topic-influence score for all posts within the period for each topic.

#### E. Sentiment & Toxicity Analysis

For each blog post, the topic model assigns a distribution value from 0 to 1, depending on the model's certainty that a given document belongs to a particular topic. For sentiment as well as toxicity, only documents with a distribution of 0.8 or higher were retained in order to concentrate on only the documents most relevant to each topic. From each of these subsets of documents, we then compute the yearly average sentiment and toxicity scores.

**Sentiment Analysis with Linguistic Inquiry and Word Count (LIWC).** In this study, we employed the Linguistic Inquiry and Word Count (LIWC) software [20], [40] to extract sentiment scores from blog posts, analyzing emotional, cognitive, and structural components of language. Focusing on affective processes, it identified positive and negative

emotions, including anxiety, anger, and sadness, providing scores for these emotions in the analyzed content.

**Toxicity.** Online disinhibition, or toxic behavior such as cyberbullying, disrupts social norms, causing harm that extends beyond the incident and affects the community [41], [42]. Individuals may engage in this behavior for various reasons like boredom, entertainment, or to express frustrations [43], [44], [45]. While [42] some theories suggest certain personality traits predispose individuals to toxicity, others argue that anyone can exhibit such behavior under specific conditions. To measure toxicity, Google's Perspective API, a Convolutional Neural Network model, is used [46], a Convolutional Neural Network (CNN) model trained to identify "toxic" content. Trained on diverse toxic discourse elements, it assigns a probability score to content, with higher scores indicating greater toxicity [47], [48]. This model is particularly effective in identifying and analyzing toxic online content, including YouTube comments [49].

#### F. Network Analysis

Blog analysis critically involves observing influence networks and understanding information operations through network analysis. This method quantifies online influence, pinpointing users who significantly impact opinions in the blogosphere. Users and their interactions, like comments or shares, are modeled as nodes and edges in a graph, offering a nuanced view of influence dynamics [50]. Key measures like centrality identify influential users, with factors like engagement levels and social media metrics (retweets, favorites) indicating influence [51]. For example, users engaging many people with their publications could be classified as influencers due to a high in-degree within their commenting network. Additional metrics might encompass the total number of retweets, favorites, or mentions received by a user [52], [53]. This study employs social network analysis to gauge online influence, particularly among bloggers with shared topic interests. Tools like Gephi are utilized to analyze multi-modal networks and aid in identifying core bloggers and entities within blog discussions [54].

## IV. RESULTS

This section is structured into two primary topic models, one centered on Ukrainian political news and international relations, and the other on domestic affairs and conflicts. Within each of these broad categories, a secondary topic modeling was conducted to further segment the blog activities. For each of the top ten resultant topics, the section provides an in-depth exploration, encompassing both the socio-political context and computational data analysis, including sentiment and toxicity metrics. The discussion for each overarching topic begins with a high-level summary of key findings, lists all identified topics, and details the selection process based on relevancy and activity levels before delving into an individual examination of each specific topic.

A. Topic Models Set #1 - Ukrainian Political News & Neighbor Relations

TABLE II. SUBTOPICS SUMMARY OF SET#1

Subtopic Summary	ID	Topic Presence	Influence	Sentiment	Toxicity	Relevance
Ukrainian Corruption	0	Low activity	Varying	Average	Average	Not Retained
Ukrainian Conflicts	1	Average	Varying	Average	Average	Not Retained
International Affairs	2	Inactive	Varying	Varying	Varying	Irrelevant Narrative
Ukrainian Presidential Elections	3	Low activity	Varying	Negative	Average	Not Retained
Ukrainian News	4	Active	Growing	Varying	Varying	Retained
US Political Personalities	5	Low activity	Went down	Average	Average	Irrelevant Narrative
Ukrainian Neighbors	6	Active	Varying	Average	Average	Retained
Ukrainian Extremism	7	Active	Varying	Average	Average	Retained
Euromaidan	8	Low activity	Went down	Negative	Average	Not Retained
Ukrainian Parliament Elections	9	Low activity	Varying	Average	Average	Not Retained

Table II consolidates a detailed summary of each subtopic related to the primary topic, encompassing their primary

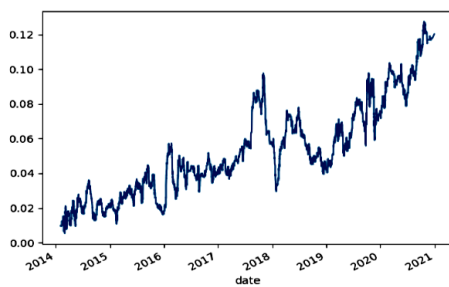


Fig. 3a. Probability of Topic appearing within blog posts over time – Topic 4.

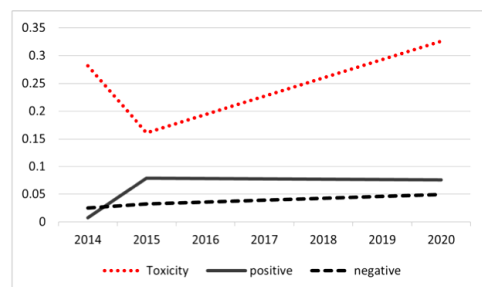


Fig. 3c. Sentiment distribution of blog post comments – Topic 4.

Figure 3. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 4.

subjects, time-bound metrics, and inclusion status in the final report.

1) Ukrainian News Topic (#4):

The trend in Topic 4, as depicted in Fig. 3a, reflects major geopolitical events and shifts in international discourse from 2017 to 2020. In 2017, discussions focused on the Ukraine war and Russian involvement, particularly during peak interest in July and August. In 2018, the narrative shifted to the impact of U.S. sanctions on Russian oligarchs, highlighting escalating U.S.-Russia tensions.

The dialogue in 2019 centered on the MH17 incident, with controversial claims by F. William Engdahl about Ukraine's potential involvement, though these lacked solid evidence and faced criticism. In 2020, the focus was on political narratives, notably Joe Biden's actions in Ukraine and Andriy Derkach's alleged meddling in the U.S. Presidential Election, underlining concerns about foreign influence in U.S. politics. Sentiment and toxicity trends (Fig. 3b & 3c) mirrored these events, showcasing the connection between international incidents and public sentiment shifts. The influence score (Fig. 3d) showed a linear rise with significant spikes in 2019 and 2020, corresponding to the MH17 incident and discussions around the U.S. election and Russian interference, emphasizing the importance of these events in shaping public discourse and sentiment.

2) Ukrainian Neighbors Topic (#6):

Topic 6, depicted in Fig. 4a, traces the narrative trends related to NATO and Eastern European countries, with significant activity spikes in 2014, 2017, 2018, and 2020. The year 2014 saw pro-Russian blogs justifying Russia's stance during the Ukraine crisis, promoting pro-Russian sentiments,

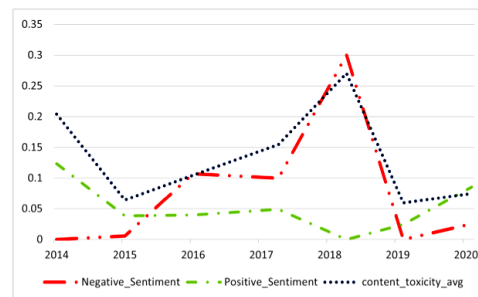


Fig. 3b. Sentiment distribution of blog post content – Topic 4.

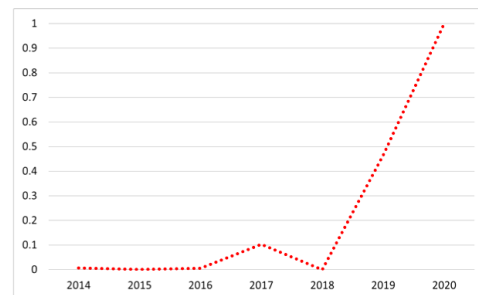


Fig. 3d. Influence of blog posts – Topic 4.

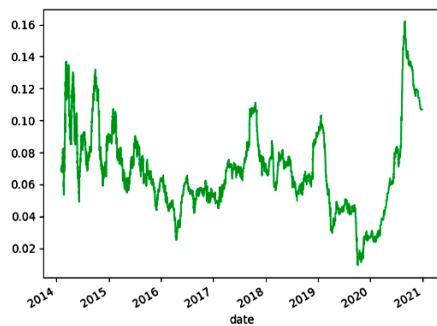


Fig. 4a. Probability of Topic appearing within blog posts over time – Topic 6.

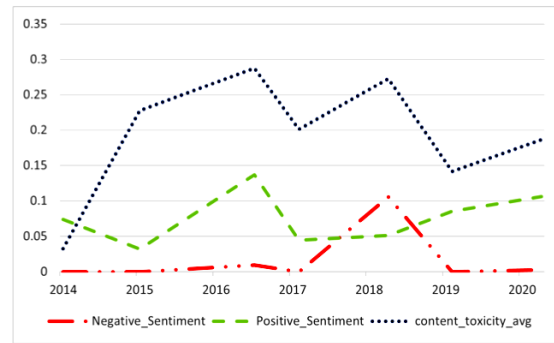


Fig. 4b. Sentiment distribution of blog post content – Topic 6.

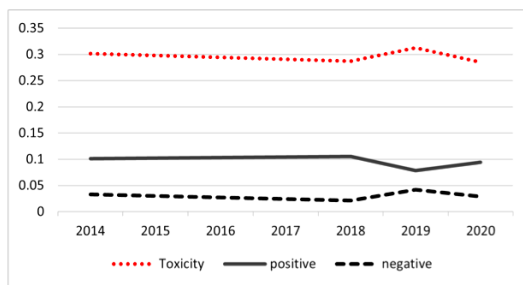


Fig. 4c. Sentiment distribution of blog post comments – Topic 6.

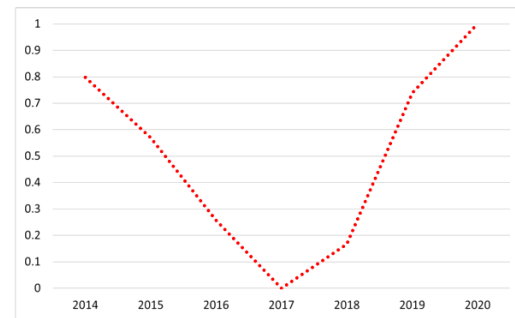


Fig. 4d. Influence of blog posts – Topic 6.

Figure 4. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 6.

and expressing heightened toxicity against the West. In 2017, the focus shifted to the complex diplomatic relations between Ukraine, Belarus, and Poland, influenced by the Belarus revolution and NATO's backing of Ukraine, leading to mixed sentiment levels and toxicity. The year 2018 marked an increase in pro-Russian content criticizing NATO and Ukraine, amplified by 'Bridgeblogging,' intensifying negative sentiment and toxicity toward these entities.

In 2020, narratives centered around the contentious Belarusian presidential elections, with pro-Kremlin blogs accusing the Lublin triangle of inciting a coup in Belarus, fueling negative sentiment and toxicity. Furthermore, disinformation campaigns post-election exacerbated this negativity. Blog engagement dynamics underscored the evolving relationships between Russia, Ukraine, and their allies, such as Austria's favorable portrayal in 2016 and the critical stance toward Russia and its allies by 2018, resulting in increased negative sentiment and toxicity. Sentiment and toxicity trends, as shown in Fig. 4b & 4c, were relatively stable, though toxicity slightly overpowered positive sentiment. Critique of the U.S.'s military support to Ukraine correlated with a dip in positive sentiment and a surge in negativity and toxicity. The influence score, illustrated in Fig. 4d, exhibited notable fluctuations, aligning with key narratives and events. Posts with minimal impact, like "THE VINEYARD OF THE SAKER" in 2015, led to a downward trend, whereas influential posts, particularly from 2018 to 2020, notably "A well-known political scientist called the results of the possible removal of Lukashenka by the opposition from power," drove the trend upwards.

### 3) Ukrainian Extremism Topic (#7):

Topic 7, represented in Fig. 5a, primarily focused on the emergence and discussion of right-wing nationalism and nationalist movements in Ukraine, especially post-Euromaidan, with activity peaking between 2014 and 2015 and then stabilizing with a slight decline towards late 2019 before rebounding in 2020. The narrative was heavily centered on the rise of nationalism, shifts towards Nazism in Ukraine, and the placement of nationalists in significant government roles. This discourse was primarily driven by two distinct blogger groups, one endorsing nationalism and the other advocating for communism, resulting in a diverse range of sentiment and toxicity levels.

In 2017, as seen in Fig. 5b, a surge in negative sentiment was noted, mainly due to discussions that brought up historical conflicts like the Cold War-era tensions, the Ukrainian famine, and the Holodomor, the latter inciting intense negative reactions due to its tragic historical significance. The overall sentiment and toxicity trends in blog comments, illustrated in Fig. 5c, maintained consistency, with a noticeable increase in toxicity post-2017. The influence scores of Topic 7, as shown in Fig. 5d, were marked by considerable fluctuations.

Blogs from 2016 to 2017, particularly those like "Complete failure of 'alternative Russia' in Ukraine," which featured extensive external links and high engagement, led to an upsurge in influence scores. On the contrary, 2019 saw a sharp decline in influence, largely due to blogs such as "Why the presidential election in Ukraine will not change anything" and "Freeland Responds to Putin: Liberalism Will Prevail! (Nazis Will Help)," which showed a significant gap in social

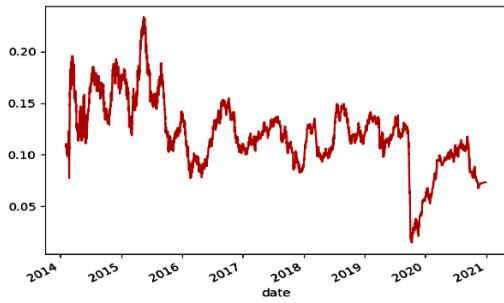


Fig. 5a. Probability of Topic appearing within blog posts over time – Topic 7.

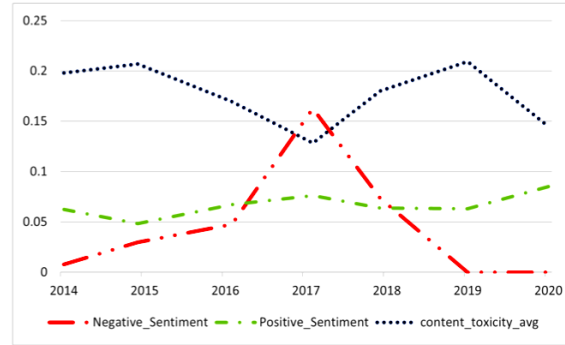


Fig. 5b. Sentiment distribution of blog post content – Topic 7.

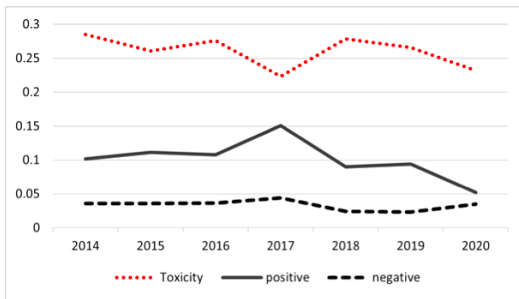


Fig. 5c. Sentiment distribution of blog post comments – Topic 7.

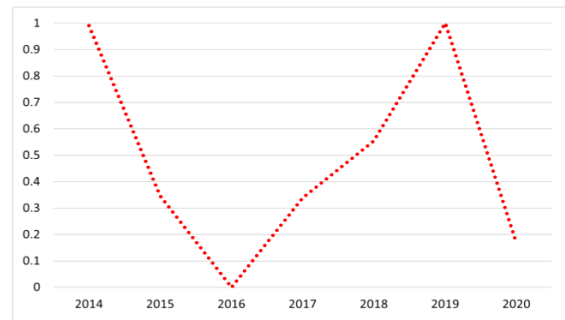


Fig. 5d. Influence of blog posts – Topic 7.

Figure 5. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 7.

influence and engagement, with the latter lacking both engagement and visual context.

### B. Topic Models Set #2 – Domestic Affairs

Table III presents a comprehensive summary of each topic related to the primary subject, focused on their central themes, time-based metrics, and the decision regarding their inclusion in the final report.

TABLE III. SUBTOPICS SUMMARY OF SET#2

Subtopic Summary	ID	Topic Presence	Influence	Sentiment	Toxicity	Relevance
<b>Donbas</b>	<b>0</b>	<b>Active</b>	<b>Average</b>	<b>Growing</b>	<b>Varying</b>	<b>Retained</b>
NATO & Syria	1	Inactive	Varying	Varying	Varying	Irrelevant Narrative
Pro-Russia	2	Inactive	Varying	Average	Varying	Not Retained
US/Russia/China Relations	3	Average	Growing	Average	Average	Irrelevant Narrative
US Investigations on Russia	4	Varying	Varying	Average	Average	Irrelevant Narrative
<b>Russian Medias</b>	<b>5</b>	<b>Active</b>	<b>Varying</b>	<b>Average</b>	<b>Varying</b>	<b>Retained</b>
Religious Conflicts	6	Inactive	Varying	Average	Average	Not Retained

Belarus Scantions	7	Inactive	Average	Varying	Varying	Irrelevant Narrative
<b>Military</b>	<b>8</b>	<b>Active</b>	<b>Growing</b>	<b>Average</b>	<b>Average</b>	<b>Retained</b>
Russia-Ukraine	9	Average	Varying	Average	Varying	Not Retained

#### 1) Donbas Topic (#0):

Topic 0, shown in Fig. 6a, was particularly prominent in 2014, 2015, 2019, and 2020, with marked activity spikes in 2014 and 2015. The conversation in 2014 was dominated by the presidential election and eastern Ukrainian insurgents' declaration of independence. Many bloggers, especially pro-Russian ones, discussed the ceasefire agreement, often blaming the U.S. for escalating tensions with Russia. In 2015, the focus was on the "New Minsk negotiation," with a general sentiment of disappointment among bloggers. The topic also covered parliamentary riots, (de)centralization of power, and local elections in Donbas. The narrative resurfaced in 2019 and 2020, highlighting election tensions in the Donbas region during President Volodymyr Zelensky's term and his peace efforts with Russia.

Blog discussions in 2014 were rich in strategic government negotiations, sanctions, and rumors of potential civil unrest post-election. Joe Biden's visit to Ukraine was a significant event amidst escalating tensions. Negative sentiment, as depicted in Fig. 6b, intensified with discussions on Donbas's independence and recognition. Tensions escalated further with Ukraine cutting transport links to Donbas and Russia's proposed recognition of the region, exacerbating cold war rhetoric. Mid-2017 saw negative



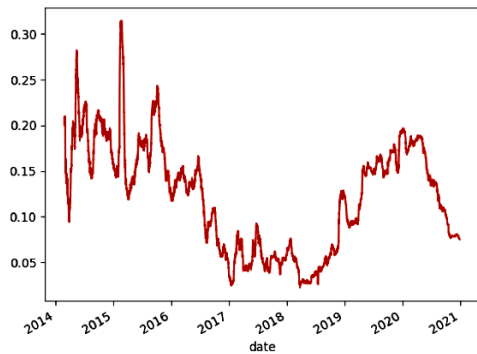


Fig. 6a. Probability of Topic appearing within blog posts over time – Topic 0.

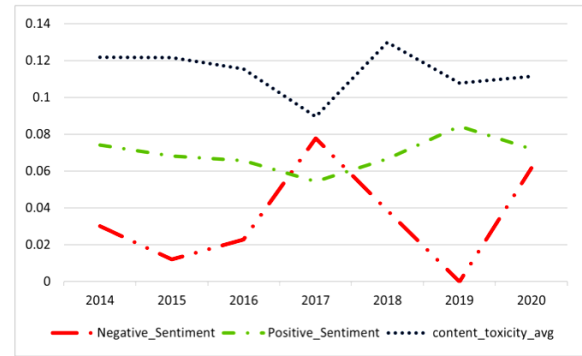


Fig. 6b. Sentiment distribution of blog post content – Topic 0.

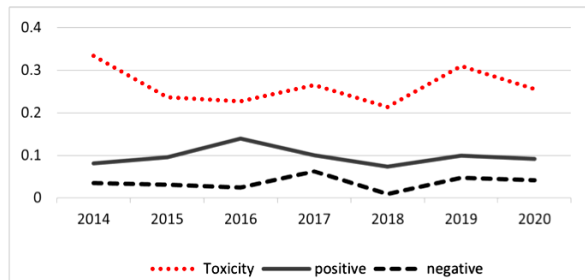


Fig. 6c. Sentiment distribution of blog post comments – Topic 0.

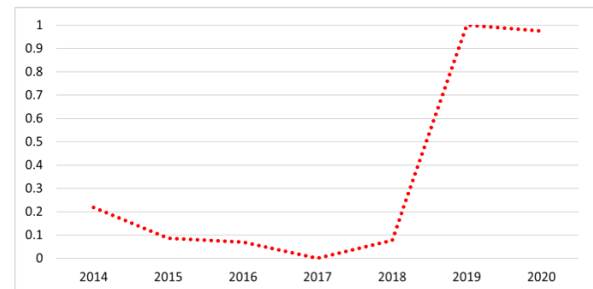


Fig. 6d. Influence of blog posts – Topic 0.

Figure 6. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 0.

sentiment peaking, mainly due to discussions about the violation of the Minsk accord (Fig. 6b & Fig. 6c). Sentiment and toxicity scores for blog posts, when capped at 0.5, were low and stable, as seen in Fig. 6b & Fig. 6c. The toxicity trend had minor fluctuations, with a notable rise in positive sentiment in 2016 linked to a blog post titled "Moveable Feast Cafe 2016/05/21 ... Open Thread."

The influence score trend in Fig. 6d remained low from 2014 to 2015 but saw an increase in 2018 to 2019, driven by popular posts like "Protests Rage as Ukraine's Zelensky Allows Election In Separatist-Controlled East" and "Zelensky Discusses Implementation of Normandy Summit Agreements with Merkel," both garnering high engagement.

### 2) Russian Medias Topic (#5):

Topic 5, illustrated in Fig. 7a, displayed a consistent trend from 2014 to 2020, with significant activity in 2015, 2017, and 2018. The 2015 narrative was heavily influenced by Russian propaganda, focusing on the alleged information war against Euromaidan by Kremlin-controlled media, with most content contributed by "stopfake.org". In 2017 and 2018, discussions pivoted to the concern in Moscow regarding potential uprisings similar to Ukraine's 2014 revolution, highlighting suggestions for followers of Alexei Navalny to protest. Despite some criticism of Putin's regime, the predominant theme was image management, with posts often portraying Russia as a peace-promoting entity and featuring Russian Hollywood actors.

Sentiment analysis, as shown in Fig. 7b, indicated a dominance of positive sentiment, mainly because the contributors were personal bloggers rather than established

media houses. The sentiment trend in Fig. 7c was stable and low, with generally neutral negative sentiment and consistent positive sentiment. Toxicity levels varied slightly, with periods of higher positive sentiment correlating with lower toxicity, like the 2016 post "The Globalists Love Gefilte Fish," which implied strong Russia-Jewish ties and demonstrated low toxicity.

Influence scores, depicted in Fig. 7d, were generally average and linear but showed a notable increase in 2019 followed by a decline in 2020. Posts like "Putin's Gaffe Divides Russian Society, Part II" in 2020 had a low influence score due to limited engagement, whereas the 2019 post "About the Ukrainophobia of Russian Nazis Or The Defeat Of The Liberal Point Of View Of The Conflict," with substantial views and comments, exhibited a significantly higher social influence.

### 3) Military Topic (#8):

Topic 8, as seen in Fig. 8a, was notably active around mid-2014 and early 2019, focusing on the conflict between the Ukrainian army and pro-Russian rebels, and notably on the Malaysian Airline MH17 crash. In 2019, discussions around MH17 resurfaced, probing the details of the MH17 report and related political deceit. Over the studied period, the toxicity trend, shown in Fig. 8b, was stable with a minor increase in 2019. Overall sentiment stayed neutral, with no significant shifts towards positive or negative extremes. The content was largely centered on the Russia-Ukraine conflict, discussing

military developments and geopolitical strategies of both countries.

Sentiment for blog comments, according to Fig. 8c, remained largely neutral, exhibiting low negative sentiment.

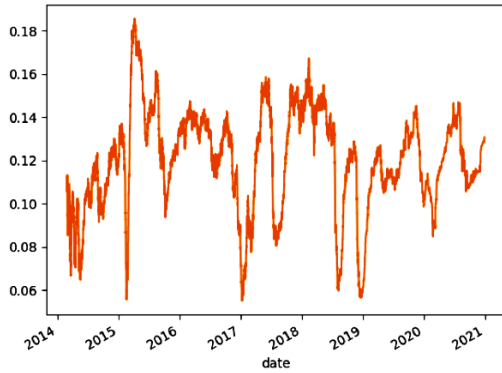


Fig. 7a. Probability of Topic appearing within blog posts over time – Topic 5.

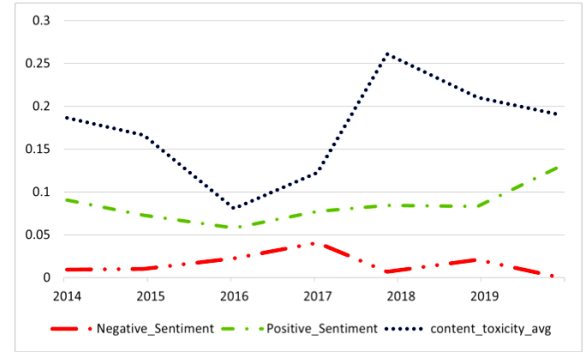


Fig. 7b. Sentiment distribution of blog post content – Topic 5.

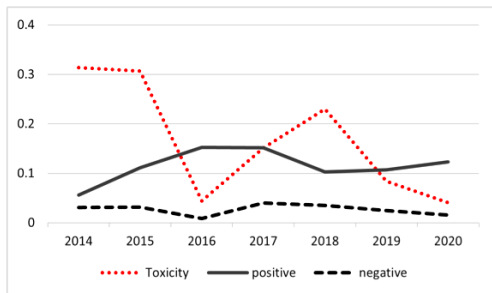


Fig. 7c. Sentiment distribution of blog post comments – Topic 5.

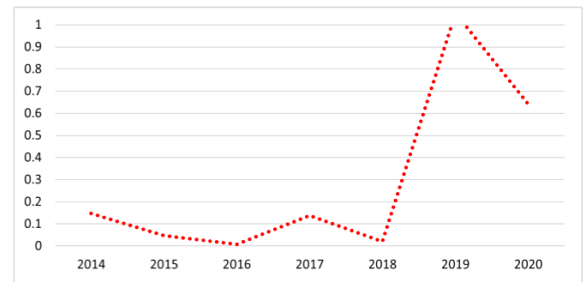


Fig. 7d. Influence of blog posts – Topic 5.

Figure 7. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 5.

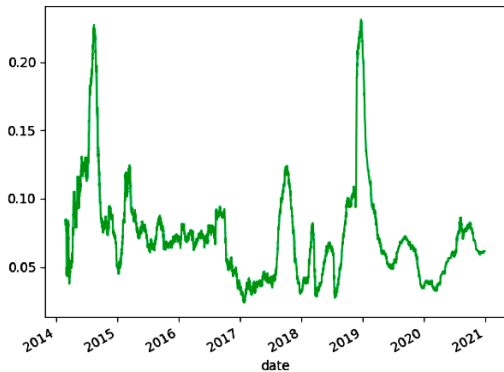


Fig. 8a. Probability of Topic appearing within blog posts over time – Topic 8.

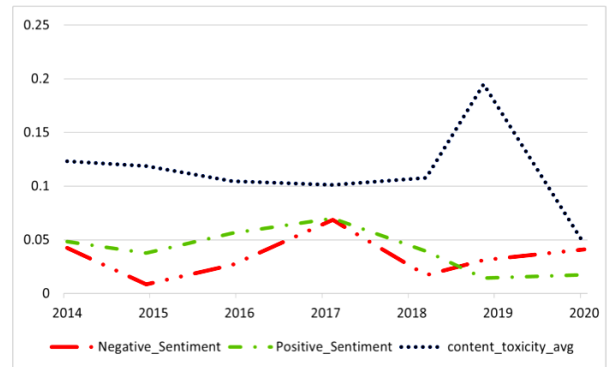


Fig. 8b. Sentiment distribution of blog post content – Topic 8.

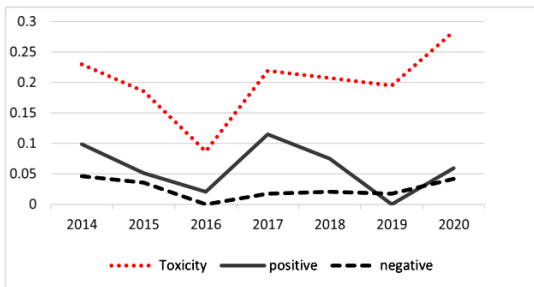


Fig. 8c. Sentiment distribution of blog post comments – Topic 8.

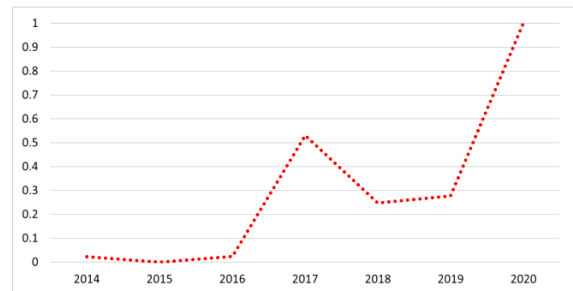


Fig. 8d. Influence of blog posts – Topic 8.

Figure 8. Topic probability, sentiment, and influence within blog posts and comments over time – Topic 8.

Positive sentiment and toxicity had similar patterns, though toxicity was slightly higher. A dip in both sentiment and toxicity in 2016 might relate to a neutral-toned blog post about Ukraine's National Bank issuing commemorative coins, which sparked more inquiries than emotional reactions. The influence of blog posts, depicted in Fig. 8d, showed a steady increase, with notable peaks during 2016–2017 and 2019–2020. A 2017 blog post, critically analyzing the validity of documents associated with the MH17 crash, stood out for its high influence, gaining over 20,000 views and 22 comments, and tags like CIA and Ukraine, marking its significant social impact.

The next section of the study is set to delve into network analysis, aiming to identify and understand the key players in the discussed topics and their relevance.

### C. Network Analysis

Our network analysis was partitioned into three distinct sections, each designed to answer a specific question:

- 1) What are the primary entities in the topics discussed within the blogosphere?
- 2) Who comprises the core network of bloggers sharing multiple common topics?
- 3) What sort of URLs do they commonly share within these topics?

We employed Term Frequency-Inverse Document Frequency (TF-IDF) to map the entities extracted from each blog post to their respective topics. This process unveiled intriguing clusters of entities that illuminated key insights into their unique connections within the blogs. The subsequent sections of this paper endeavor to answer these queries using the previously introduced data subsets.

#### 1) Topics and entities networks:

In Set #1, as illustrated in Fig. 9a, the network analysis identifies prominent entities within Topics 3 and 5, revealing both strong and weak connections with entities across other topics. This network highlights a diverse array of blog discussions focusing on Russian, Ukrainian, European, and U.S. politics, underscoring the unique nature of each topic's discourse, as well as shared thematic elements. Particularly, pro-Russian entities in Topic 7 were noted for promoting Russian involvement in Eastern Ukraine. In contrast, there was notable concern over Russian propaganda in other blog posts, accusing Kremlin-controlled media of launching a comprehensive information war against Euromaidan and the Ukrainian populace.

In Set #2, depicted in Fig. 9b, portrays entities with a pro-Russian inclination, primarily concentrating on Ukrainian politics. These entities show strong connections to pivotal figures such as Putin, Russia, and Ukraine. Analysis indicates these entities played central roles in discussions around the ceasefire between the Ukrainian government and pro-Russian separatist rebels, and the narratives concerning corruption among Ukrainian oligarchs. Additionally, other topics in this set shed light on key entities involved in discussions about economic sanctions, Europe, NATO, and Western relations, highlighting the breadth and interconnectedness of the discourse within these blog networks.

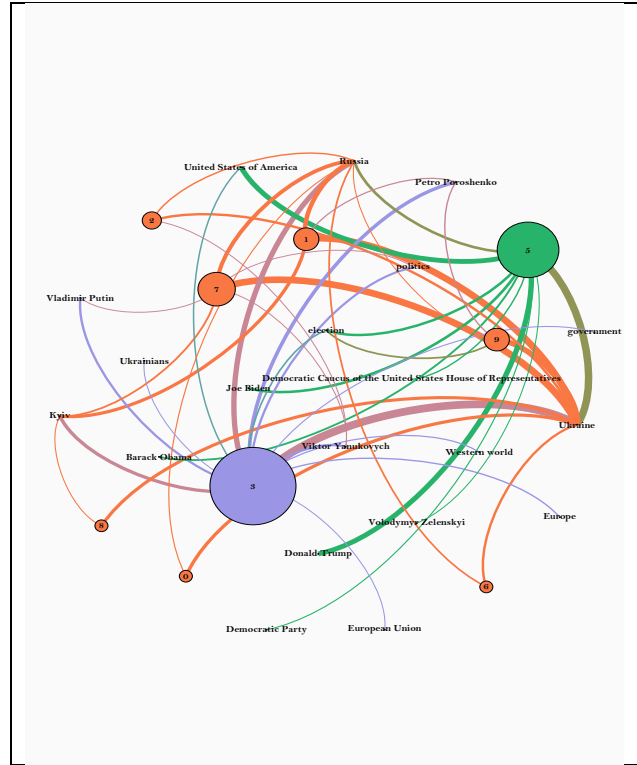


Figure 9a. Topics-entities network – Set#1.

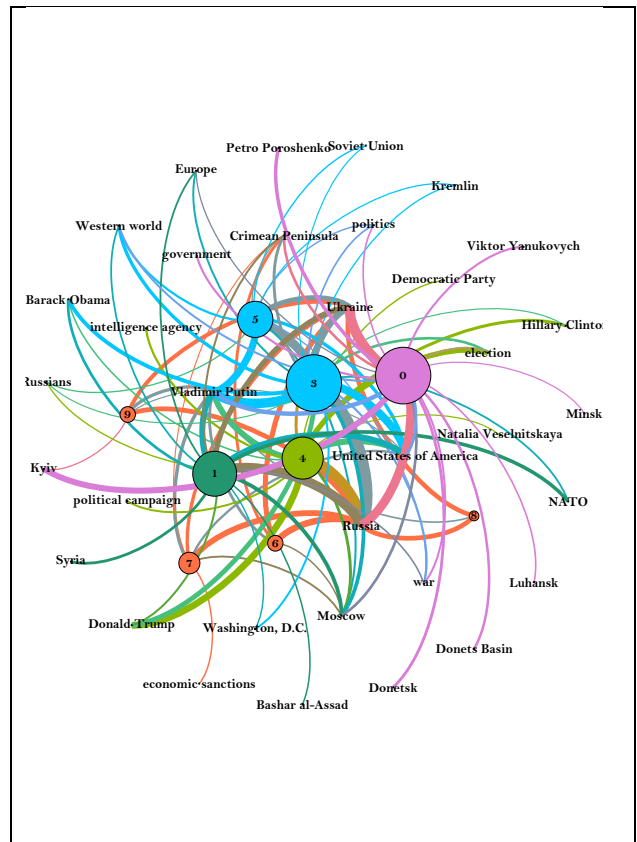


Figure 9b. Topics-entities network – Set#2.

2) *Network of core bloggers sharing common topics:*

The extraction of core bloggers was achieved by generating topic blogger networks first for each set. We then fold the two respective networks to identify core bloggers with common topics, as shown in Fig. 10a & 10b. A few bloggers, notably Tyler Durden, The Saker, RT, Consortiumnews, and Dailykos, appeared in both networks, displaying high centrality and extensive topic sharing with other bloggers. Many of these central bloggers are part of the Russian blogosphere, known for disseminating disinformation and contributing to anxiety within the blogging community.

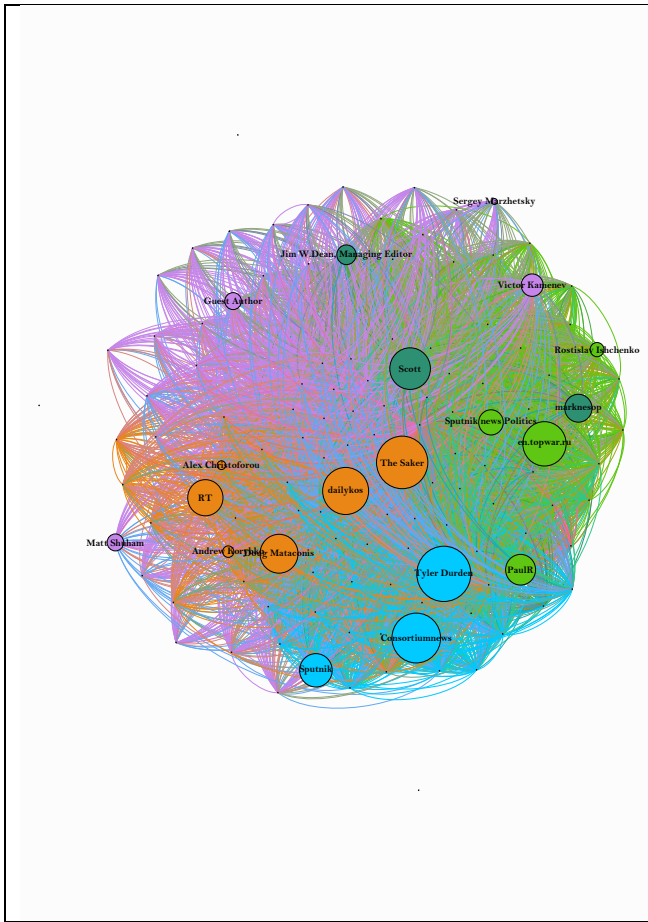


Figure 10a. Core blogger shared topics network – Set#1.

In these networks, the color of the edges represents the number of topics shared between bloggers. In Fig. 10a, orange edges represent the sharing of five topics, purple for four, green for three, and blue for two shared topics. In Fig. 10b, blue edges indicate six shared topics, orange for five, light green for four, dark green for three, and purple for two shared topics. These colors help visualize the extent of topic overlap and interconnectedness among the core bloggers within the networks.

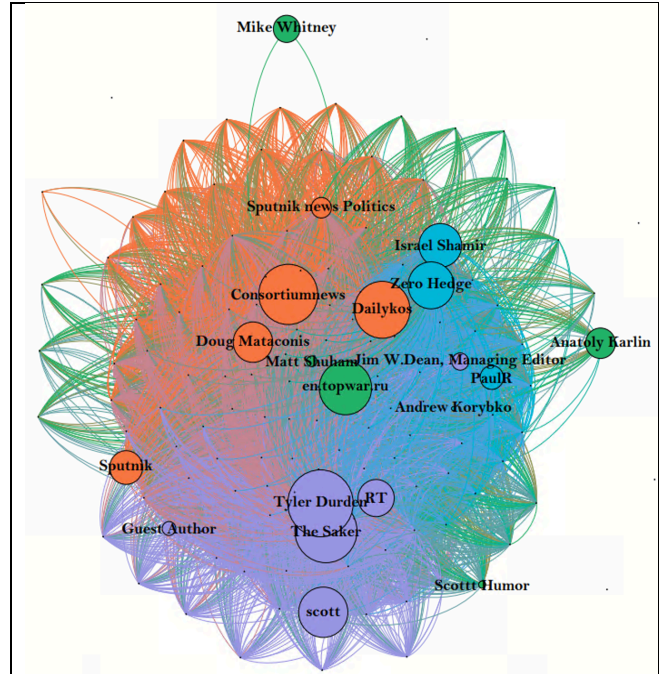


Figure 10b. Core blogger shared topics network – Set#2.

3) *Common URLs networks:*

The network analysis of hyperlinks within blog posts provided insights into latent blogging communities, revealing shared interests in specific narratives among various authors, even without direct interaction. In Fig. 11a & 11b, clusters of URLs are mapped, showing how topics (as nodes) are interconnected by URLs (as edges). Both sets demonstrate centralization of core URLs across all six topics. Set #1 identified 184 common URLs (Fig. 11a), displaying a mix of pro-Ukraine and pro-Russian leanings, with a predominance of pro-Russian bias. Set #2 identified 125 common URLs (Fig. 11b), with about 90% stemming from pro-Russian websites, mostly featuring imagery related to the Ukraine-Russia conflict, indicative of potential propaganda efforts.

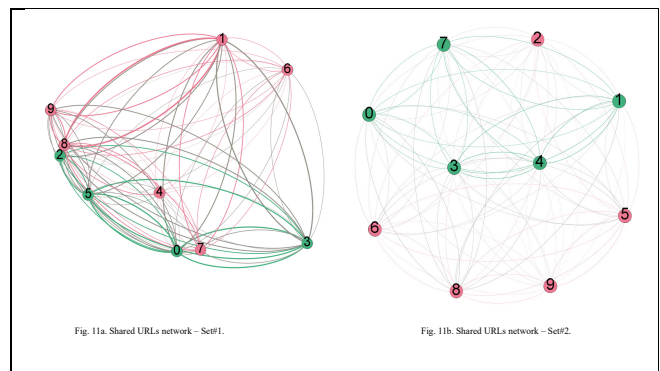


Figure 11. Topics and shared URLs networks.

Nodes in Figs. 11a & 11b represent topics, color-coded in green and red based on modular similarity among URLs, with

edges indicating shared URLs. Notable patterns emerged from the analysis: Both sets used web links with images from the 2014 Ukraine war, suggesting that pro-Ukraine and pro-Russian blogs employed these visuals to reinforce their war narratives and evoke emotional responses. Information dissemination extended beyond blogs to social media platforms like Twitter, Facebook, Reddit, and WhatsApp, mainly through 'global.ca', a Canadian website promoting pro-Russian narratives and disinformation. The majority of blogs in both sets displayed a pro-Russian bias, indicating that the pro-Russian blogosphere predominantly drove information dissemination, promoting Russian viewpoints, while pro-Ukraine links were comparatively less frequent.

## V. DISCUSSION AND CONCLUSION

We developed and presented a computational framework to assess online political discourse dynamics, revealing trend insights and sentiment to inform situational awareness and influence strategies. Network analysis identified key online entities for strategic engagement and counter-narratives. Our study analyzed blog content related to Ukrainian politics and conflicts, identifying key trends, narratives, and sentiments. We found a predominance of pro-Russian narratives across different topics, varying in intensity and influenced by geopolitical events and societal factors.

Topic probability plots highlighted real-world events, and sentiment analysis provided insight into perceptions and attitudes within the blogosphere. Positive sentiments were dominant, despite instances of toxicity and negativity. Network analysis revealed significant commonalities and interconnections among bloggers and entities, with certain bloggers having high centrality and influence in the discourse.

The analysis also uncovered potential hidden communities where authors shared common interests in specific narratives. Core URLs shared across topics often emanated from pro-Russian sources, potentially molding the discourse. Information dissemination extended beyond the blogosphere to social media platforms, expanding the sphere of influence.

Our topics demonstrated their interconnected nature, with geography, news, and conflict connecting through international relations and cultural space. We predicted and found similar peaks in activity and common trends across different topics. All topics showed a drastic increase in influence scores in recent years.

Our study underscores the complex dynamics of the blogosphere in shaping narratives and influencing public opinion in the context of the Ukraine conflict. It provides valuable insights into how sentiments, discussions, and key entities intermingle to form a broad and multifaceted discourse. These results can offer a foundation for further research and understanding of digital discourse in politically charged environments.

Key findings of this study include:

- Predominance of pro-Russian narratives across different topics
- Variation in narrative intensity influenced by geopolitical events and societal factors

- Positive sentiments dominant despite instances of toxicity and negativity
- Significant commonalities and interconnections among bloggers and entities
- Potential hidden communities where authors share common interests in specific narratives
- Expansion of information dissemination beyond the blogosphere to social media platforms
- Interconnected nature of geography, news, and conflict through international relations and cultural space.

## ACKNOWLEDGMENT

THIS RESEARCH IS FUNDED IN PART BY THE U.S. NATIONAL SCIENCE FOUNDATION (OIA-1946391, OIA-1920920), U.S. OFFICE OF THE UNDER SECRETARY OF DEFENSE FOR RESEARCH AND ENGINEERING (FA9550-22-1-0332), U.S. ARMY RESEARCH OFFICE (W911NF-23-1-0011, W911NF-24-1-0078), U.S. OFFICE OF NAVAL RESEARCH (N00014-21-1-2121, N00014-21-1-2765, N00014-22-1-2318), U.S. AIR FORCE RESEARCH LABORATORY, U.S. DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (W31P4Q-17-C-0059), ARKANSAS RESEARCH ALLIANCE, THE JERRY L. MAULDEN/ENTERGY ENDOWMENT AT THE UNIVERSITY OF ARKANSAS AT LITTLE ROCK, AND THE AUSTRALIAN DEPARTMENT OF DEFENSE STRATEGIC POLICY GRANTS PROGRAM (SPGP) (AWARD NUMBER: 2020-106-094). ANY OPINIONS, FINDINGS, AND CONCLUSIONS OR RECOMMENDATIONS EXPRESSED IN THIS MATERIAL ARE THOSE OF THE AUTHORS AND DO NOT NECESSARILY REFLECT THE VIEWS OF THE FUNDING ORGANIZATIONS. THE RESEARCHERS GRATEFULLY ACKNOWLEDGE THE SUPPORT.

## REFERENCES

- [1] U. Onyepunuka and T. Marcoux, Mainuddin Shaik, Mayor Inna Gurung and Nitin Agarwal. "A Multidimensional Analysis of YouTube Communities in the Indo-Pacific Region," in *The Twelfth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS 2022)*, October 16-20, 2022, Lisbon, Portugal.
- [2] "Ukraine crisis | Euromaidan, Crimea, & War in the Donbas | Britannica," Accessed: Jan. 25, 2024. [Online]. Available: <https://www.britannica.com/topic/Ukraine-crisis>
- [3] S. Pedersen, S. Burnett, R. Smith, and A. Grinnall, "The impact of the cessation of blogs within the UK police blogosphere," *New Technology, Work and Employment*, vol. 29, no. 2, pp. 160–176, 2014, doi: 10.1111/ntwe.12028.
- [4] L. Kirchhoff, T. Nicolai, A. Bruns, and T. Highfield, "Monitoring the Australian blogosphere through the 2007 Australian federal election," in *Communication, Creativity and Global Citizenship: Refereed*

*Proceedings of the Australian and New Zealand Communication Association Conference 2009*, ANZCA, 2009, pp. 982–1005.

- [5] H. W. Park and R. Kluver, “Trends in online networking among South Korean politicians—A mixed-method approach,” *Government Information Quarterly*, vol. 26, no. 3, pp. 505–515, 2009.
- [6] M. Karlsson and J. Åström, “The political blog space: A new arena for political representation?,” *New Media & Society*, vol. 18, no. 3, pp. 465–483, 2016.
- [7] K. Zafiroopoulos, V. Vrana, and D. Vagianos, “Bloggers’ community characteristics and influence within Greek political blogosphere,” *Future Internet*, vol. 4, no. 2, pp. 396–412, 2012.
- [8] J. Nordenson, “We want five: Kuwait, the Internet, and the public sphere,” Master’s Thesis, 2010.
- [9] M. Jacky, “Bloggers and deliberative democracy in Indonesia’s blogosphere,” *Asian Social Science*, vol. 11, no. 28, p. 15, 2015.
- [10] N. Pang and D. Goh, “Can blogs function as rhetorical publics in Asian democracies? An analysis using the case of Singapore,” *Telematics and Informatics*, vol. 33, no. 2, pp. 504–513, 2016.
- [11] M. Francoli and S. Ward, “21st century soapboxes? MPs and their blogs,” *Information Polity*, vol. 13, no. 1–2, pp. 21–39, 2008.
- [12] J. Åström and M. Karlsson, “Blogging in the shadow of parties: Exploring ideological differences in online campaigning,” *Political communication*, vol. 30, no. 3, pp. 434–455, 2013.
- [13] M. A. M. Sani and K. T. Zengeni, “Democratisation in Malaysia: The impact of social media in the 2008 general election,” in *18th Biennial Conference of the Asian Studies Association of Australia, di Adelaide*, 2010, pp. 5–8.
- [14] V. Vrana and K. Zafiroopoulos, “Rural tourism lodgings’ websites: a comparative study among Mediterranean countries,” *International Journal of Tourism Policy*, vol. 2, no. 1–2, pp. 89–106, 2009.
- [15] B. Burroughs, “Kissing Macaca: Blogs, narrative and political discourse,” *Journal for Cultural Research*, vol. 11, no. 4, pp. 319–335, 2007.
- [16] L. Vesnic-Alujevic, “Communicating with voters by blogs? Campaigning for the 2009 European Parliament elections,” *Discourse & Communication*, vol. 5, no. 4, pp. 413–428, 2011.
- [17] M. Williams, *The impact of radical right-wing parties in West European democracies*. Springer, 2006.
- [18] E. P. Baumer, M. Sueyoshi, and B. Tomlinson, “Bloggers and readers blogging together: Collaborative co-creation of political blogs,” *Computer Supported Cooperative Work (CSCW)*, vol. 20, pp. 1–36, 2011.
- [19] B. Nilsson and E. Carlsson, “Swedish politicians and new media: Democracy, identity and populism in a digital discourse,” *New Media & Society*, vol. 16, no. 4, pp. 655–671, 2014.
- [20] E. Siapera, “The political subject of blogs,” *Information Polity*, vol. 13, no. 1/2, p. 51, 2008.
- [21] K. Wallsten, “Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs,” *Review of policy research*, vol. 24, no. 6, pp. 567–587, 2007.
- [22] K. D. Sweetser and E. Metzgar, “Communicating during crisis: Use of blogs as a relationship management tool,” *Public relations review*, vol. 33, no. 3, pp. 340–342, 2007.
- [23] A. O. Larsson and S. Hrastinski, “Blogs and blogging: Current trends and future directions,” *First Monday*, 2011.
- [24] S. Zhabotynska and V. Velivchenko, “New media and strategic narratives: the Dutch referendum on Ukraine–EU Association Agreement in Ukrainian and Russian Internet blogs,” *European security*, vol. 28, no. 3, pp. 360–381, 2019.
- [25] O. Baysha, *Miscommunicating social change: Lessons from Russia and Ukraine*. Lexington Books, 2018.
- [26] M. Komova and V. Yakovyna, “Identification of Marked Lexicon and Its Contextual Features in Social Networks,” in *COAPSN*, 2020, pp. 152–164.
- [27] J. Goldstein, “The role of digital networked technologies in the Ukrainian Orange Revolution,” *Berkman Center Research Publication*, no. 2007–14, 2007.
- [28] O. Koltsova and S. Koltcov, “Mapping the public agenda with topic modeling: The case of the Russian livejournal,” *Policy & Internet*, vol. 5, no. 2, pp. 207–227, 2013.
- [29] K. Hopkins and D. Matheson, “Talking in a crowded room: Political blogging during the 2008 New Zealand general election,” *Media International Australia*, vol. 144, no. 1, pp. 108–117, 2012.
- [30] K. Hopkins and D. Matheson, “Blogging the New Zealand election: The impact of new media practices on the old game,” *Political Science*, vol. 57, no. 2, pp. 93–105, 2005.
- [31] L. A. Adamic and N. Glance, “The political blogosphere and the 2004 US election: divided they blog,” in *Proceedings of the 3rd international workshop on Link discovery*, 2005, pp. 36–43.
- [32] A. S. Veenstra, “Examining Political Blog Use Across Generations: An Exploration of the 2008 US Election,” *Internet Research*, vol. 10, 2009.
- [33] M. El Asikri, S. Knit, and H. Chaib, “Using web scraping in a knowledge environment to build ontologies using python and scrapy,” *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 03, p. 2020, 2020.
- [34] H. Nigam and P. Biswas, “Web scraping: from tools to related legislation and implementation using Python,” in *Innovative Data Communication Technologies and*

- Application: Proceedings of ICIDCA 2020*, Springer, 2021, pp. 149–164.
- [35] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [36] A. Schofield, M. Magnusson, and D. Mimno, “Pulling out the stops: Rethinking stopword removal for topic models,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, short papers*, 2017, pp. 432–436.
- [37] R. Řeh\u00f1ek and P. Sojka, “Software framework for topic modelling with large corpora,” 2010.
- [38] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Identifying the influential bloggers in a community,” in *Proceedings of the 2008 international conference on web search and data mining*, 2008, pp. 207–218.
- [39] N. Agarwal, H. Liu, L. Tang, and P. S. Yu, “Modeling blogger influence in a community,” *Social Network Analysis and Mining*, vol. 2, pp. 139–162, 2012.
- [40] R. L. Boyd, A. Ashokkumar, S. Seraj, and J. W. Pennebaker, “The development and psychometric properties of LIWC-22,” *Austin, TX: University of Texas at Austin*, pp. 1–47, 2022.
- [41] M. Märtens, S. Shen, A. Iosup, and F. Kuipers, “Toxicity detection in multiplayer online games,” in *2015 International Workshop on Network and Systems Support for Games (NetGames)*, IEEE, 2015, pp. 1–6.
- [42] J. Suler, “The online disinhibition effect,” *Cyberpsychology & behavior*, vol. 7, no. 3, pp. 321–326, 2004.
- [43] S.-H. Lee and H.-W. Kim, “Why people post benevolent and malicious comments online,” *Communications of the ACM*, vol. 58, no. 11, pp. 74–79, 2015.
- [44] P. Shachaf and N. Hara, “Beyond vandalism: Wikipedia trolls,” *Journal of Information Science*, vol. 36, no. 3, pp. 357–370, 2010.
- [45] K. Varjas, J. Talley, J. Meyers, L. Parris, and H. Cutts, “High school students’ perceptions of motivations for cyberbullying: An exploratory study,” *Western Journal of Emergency Medicine*, vol. 11, no. 3, p. 269, 2010.
- [46] B. Rieder and Y. Skop, “The fabrics of machine moderation: Studying the technical, normative, and organizational structure of Perspective API,” *Big Data & Society*, vol. 8, no. 2, p. 20539517211046181, 2021.
- [47] X. Han and Y. Tsvetkov, “Fortifying toxic speech detectors against veiled toxicity,” *arXiv preprint arXiv:2010.03154*, 2020.
- [48] J. Pavlopoulos, J. Sorensen, L. Dixon, N. Thain, and I. Androutsopoulos, “Toxicity detection: Does context really matter?,” *arXiv preprint arXiv:2006.00998*, 2020.
- [49] A. Obadimu, T. Khaund, E. Mead, T. Marcoux, and N. Agarwal, “Developing a socio-computational approach to examine toxicity propagation and regulation in COVID-19 discourse on YouTube,” *Information Processing & Management*, vol. 58, no. 5, p. 102660, 2021.
- [50] I.-C. Moon and K. M. Carley, “Modeling and simulating terrorist networks in social and geospatial dimensions,” *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 40–49, 2007.
- [51] F. Riquelme, P. Gonzalez-Cantergiani, X. Molinero, and M. Serna, “Centrality measure in social networks based on linear threshold model,” *Knowledge-Based Systems*, vol. 140, pp. 92–102, 2018.
- [52] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, “Everyone’s an influencer: quantifying influence on twitter,” in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011, pp. 65–74.
- [53] E. Dubois and D. Gaffney, “The multiple facets of influence: Identifying political influentials and opinion leaders on Twitter,” *American behavioral scientist*, vol. 58, no. 10, pp. 1260–1277, 2014.
- [54] N. Aggrawal and A. Arora, “Visualization, analysis and structural pattern infusion of DBLP co-authorship network using Gephi,” in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, IEEE, 2016, pp. 494–500.