

# Powering Blogosphere Analytics with BlogTracker: COVID-19 Case Study

Abiola Akinnubi, Nitin Agarwal, Mainuddin Shaik, Vanessa Okeke, Ayokunle Sunmola

*COSMOS Research Center*

*University of Arkansas at Little Rock*

Little Rock, AR, USA

{asakinnubi, nxagarwal, mxshaik, vokeke, seayokunle}@ualr.edu

**Abstract**—Tracking and analyzing large datasets from various blog collections has become a common challenge for analysts, social scientists, information scientists, behavioral scientists, industries, and government agencies. Since Blogging has expanded exponentially, with various platforms now incorporating Blogging as an extension of their websites and applications, Blogging has since become where people share thoughts and seek to influence and push narratives to diverse sections of the public, be it customers, followers, amongst others. The blogosphere is a virtual community of blogs that exists together with influence spanning beyond the borders of countries and regions and is not restrictive in the length of what bloggers can post; this allows various authors to generate content since blogs provide an avenue to shape narratives and influence due to their unlimited capabilities and are free from government regulation and controls. Blogs create a treasure trove of information to be analyzed and generate meaningful insights through various visualization mechanisms and information science approaches. These exciting opportunities come with the difficulty of tool availability that frees analysts from repeated procedures. Despite some tools developed in the early years of Web 2.0 specifically for gathering information on blogs, many of these tools have either seen a hiatus in development or gone extinct, while other emerging tools have become mainly commercialized. With the trove of large scale data available in the blogosphere, there is a need to have a readily available public tool to analyze large scale blog data. We develop the BlogTracker application to help researchers, analysts, information scientists and the general public in understanding narratives, keyword trends, how influences are building up within a topic of interest and who are the various influential bloggers and people contributing to these various topics, we make BlogTracker readily available via <https://btracker.host.ualr.edu> to solve the challenge of blogosphere analytics. We demonstrate a few salient features in the BlogTracker application using COVID-19 as an example case study. We examined 383 unique blog sites with over 800,000 blog posts and millions of comments. We demonstrate BlogTracker’s capability to provide intuitive visualization and help analysts make sensemaking.

**Index Terms**—Blogs, BlogTracker, Social media, COVID-19, Vaccines, Topic, Blogosphere, Narrative Analysis

This research is funded in part by the U.S. NSF (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, IIS-1110868), ONR (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), AFRL, ARO (W911NF-20-1-0262, W911NF-16-1-0189), DARPA (W31P4Q-17-C-0059), Australian DST (2020-106-094), Arkansas Research Alliance, and Jerry L. Maulden/Entergy Endowment at the UA Little Rock. 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.

## I. INTRODUCTION

The blogosphere has been a virtual community with deliberation and conversation around various topics of interest since narratives, influence, and sentiments around hot topics formation have migrated to online communities from the traditional media and physical interactions. This has given birth to the increase in engagement in the blogosphere by various users, i.e., bloggers and commenters or followers, since this is free from the long historical censorship and regulation of traditional media in many parts of the world. Since traditional public media pushes the narratives of the primary stakeholders like governments and private individuals who founded this media and their political and socio-cultural beliefs, this gave birth to the adoption of blogospheres and bloggers whose contents and followership have been the driving force of this virtual community of the web. Blogging allows authors to share an infinite amount of information and content, be it video, text or images, thereby becoming a powerful tool in forming narratives and influencing audiences who share similar beliefs and societal values. In order to study the infinite amount of data generated in the blogosphere daily, analysts have had to put together many makeshift tools or several third-party libraries, code and software. The difficulty in finding free publicly available resources capable of following events and activities in the blogosphere or analyzing various discussions, extracting narratives, visualizing narratives, detecting keywords hot topics, highlighting bloggers and authors with influence and understanding sentiments that each blog post and blogger push. We developed BlogTracker Application to cover various aspects of online behavioral studies and with the capability to reduce the workload that analysts traditionally go through in understanding what is happening around the world. In [1], we demonstrated some capabilities of BlogTracker but focused mainly on influence campaigns without expanding the scope to study various aspects of the blogosphere. This work expands our earlier work [1] by demonstrating how the BlogTracker Application can power the blogosphere analysis. We use COVID-19 as a use case to demonstrate BlogTracker’s capabilities by explicitly showing how COVID-19 narratives were formed—and visualized, among other critical salient features that BlogTracker can perform, Thereby allowing analysts to focus on more important things than hacking various

makeshift solutions when studying the blogosphere. Compared to other tools that have either gone into extinction or not offer public accessibility, or restricting their usage to paid licensed customers. BlogTracker has emerged to solve these various challenges and many more. paid licensed customers. BlogTracker has emerged to solve these various challenges and many more.

## II. LITERATURE REVIEW

Over the years, there has been an increase in blogosphere study, but new insights emerge daily in our ever-changing world with more studies. Bloggers and audiences have had to react to new information daily. Studies on various aspects of the blogosphere like narratives, topic modelling, content diversity, blog posting frequency, sentiment analysis have all been part of the complete, exhaustive analysis carried out on the blogosphere. Mining significant scale opinion for blogs has its challenges, especially with the scale at which blog posts are churned out daily by bloggers due to its unregulated nature and the bloggers being the traditional regulator of their blog sites themselves, the works of Nikos et al. [2] involved developing a tool for social, news and blog data but howbeit with emphasis on business. Their tool called "PaloPro" focuses on the platform as a service for business or brand monitoring. Cuong et al. (2012) [3] explore learning analytics for learning blogosphere using two approaches: blogosphere structural analysis and content analysis, with the help of a tool called Mediabase and eTwinning Network by applying social network analysis methods leveraging these tools in studying the blogosphere while using the Mediabase for content analysis. Shaik et al. (2021) [4] explored the Australian blogosphere using a multi-method analytical framework to cover various aspects of the Australian blogosphere. Shaik et al. (2021) [4] analyzed over 20,066 blog posts and 10,113 comments between 2019 and 2020. COVID-19 discourse absorbs much of the blogger's attention during the study period. Agarwal et al. [5] studied the Influential bloggers in a virtual community by aggregating bloggers of similar interests [5] treated these bloggers as a virtual community in the blogosphere. [5] observed that similar and influential bloggers communities showed more clusters, [5] also discussed the concept of long-tail on the relationship of the influential blog site. He concludes that influential blog sites are few, given the nature of the blogosphere.

A study by Hussain et al. (2018) [6] on analyzing the shift in narrative explored the concept of narratives with a focus on migrants to Europe; this allowed the authors to use targeted sentiment analysis to study how narratives have shifted toward migrants in the European blogosphere with over 9000 blog posts studied from 2005 to 2016. Hussain et al. [6] had used the migrant's crisis in the E.U. to understand how narratives are weaponized by extensively monitoring citizens' sentiment change towards migrants and understanding how the shift of narratives occurred in the E.U. blogosphere.

The works of Al-Ani et al. [7] explored the concept of counter-narratives in the Egyptian blogosphere during the Arab

upspring by using topic modelling to analyze blog data authored from 2005 to 2011; their work was able to explore and look at blogs across the interaction of societal, personal, and revolutionary blog topics for six years. Their works found that bloggers could organize counter-narratives against government narrative push. With the government controlling the general or traditional public media, bloggers could sieve out government narratives by generating social commentary through blogs that allow citizens to find alternative information. Egyptians used blogs as free speech platforms despite government actions, where opposition to Mubarak's authoritarian regime was gradually fostered, legitimized, and organized. The work of Al-Ani et al. [7] was able to explore how blogs could be a suitable mechanism in counter-power wielding since blogs are neither dependent nor regulated by authorities like the mainstream media since blogs are not susceptible to traditional gatekeepers and character limitations of microblogging platforms.

Bandeli et al. [8] proposed a framework using NLP techniques in identifying actors and actions and POS tagging since social media users can create narratives and spread information using blogs. The framework by Bandeli et al. [8] can be used in building effective counter-narratives to reduce the impact and spread of propaganda campaigns. Although Bandeli et al. [8] had limitations with some grammar rules that needed to be updated if the sentence was complex, their work demonstrated the person, organization, and location by relating this to the sentence and events in the blogosphere.

## III. METHODOLOGY

In studying the blogosphere, we considered available tools that provide analyst capability to leverage solutions developed from years of various research and better visualization with no cost to the Analyst. At the same time, there were either underdeveloped, non-frequent updated, non-robust or expensive paid tools. We then proposed a BlogTracker tool for end-to-end social data analysis of the blogosphere. Blogtracker allows users to follow blogs of interest and track various hot social discussions. BlogTracker is an application that allows users to perform various analytics and insight mining from blog data that the user chooses to track or follow using the information or keyword the user intended to track. It is a state-of-the-art tool capable of providing results in areas not limited to cluster analysis, sentiment analysis, narrative analysis, topic modelling, keyword trend and influence analysis, amongst other features, since the blogosphere is an entire virtual universe of its own shaped by community participatory of bloggers aka Authors, commenters and in some situations by policymakers. This then mandates studying various topics discussed within this virtual global community and how they connect to what is happening worldwide since the blog provides influencers and opinion-makers with the ability to shape narratives with the volume of content, and any local or international laws do not restrict it. We have developed several modules in our BlogTracker application that analysts can leverage to analyze what is going on in the blogosphere's virtual world. In showing the capabilities of the BlogTracker application, we will discuss

and demonstrate how various features of the BlogTracker application helps analysts in achieving this and understanding the entire blogosphere.

#### IV. BLOGTRACKER

##### A. Data Gathering and Technical Architecture

Our BlogTracker web app is developed based on our capability to ingest extensive data set from blog sites worldwide with a resilient data pipeline in the background to handle data traffic and structure this data properly into a temporary data store. Which is then ingested by threaded data post-processing and preprocessing layer capable of crunching massive dataset in a matter of minutes which is developed on Java. Due to the nature in which these data set are retrieved. We then layered our application on both MySQL database and NoSQL document database to serve the elastic search layer for faster data fetch and retrievals. Our backend Figure 1 and Figure 2 seats Java and spring boot which provides a structured way of architecting an application capable of crunching large datasets retrieved from the blogosphere, with Nodejs and express js serving as our frontend architecture for rendering data visualization and analysis for the users Fig 1, Fig2 shows a high-level overview of our technical architecture and how the BlogTracker application is served to the end-user.

##### B. Narrative Analysis

To extract and analyze narratives on the blog data in the Tracker of study, we adopt our framework for narrative analysis as explained in the work of Hussain et al. [9] and shown in Fig 3. This feature allows users to see the top ten keyword entities extracted. These keywords, along with each associated narrative, can then be viewed along with the associated blog post. The tool also enhances user capability by allowing users to search for keywords associated with a tracker; the search returns the keyword match associated with the user search criteria, thereby enabling the users to visualize narratives.

##### C. Influence Analysis

Analyzing and mining social networks can provide new insights into how people interact with and influence each other and why their ideas and opinions on different subjects can spread in social circles [10] by examining how information, experiences, ideas, and innovations propagate across blogs and bloggers who have the most influence in the blogosphere. Our approach to influence analysis hinges on finding bloggers and topics of influence. This feature Fig4 helps identify a blogger or blog post's influence on the blogosphere.

##### D. Blogger And Blog Portfolio

In our approach to generating and having bloggers' portfolios, we aimed to decode the history and patterns that blogs and bloggers could have in shaping the data collected around the COVID-19 dataset in the blogosphere. We sought to generate what overall sentiment blogs have and bloggers. We also create a weekly breakdown of how these blogs and

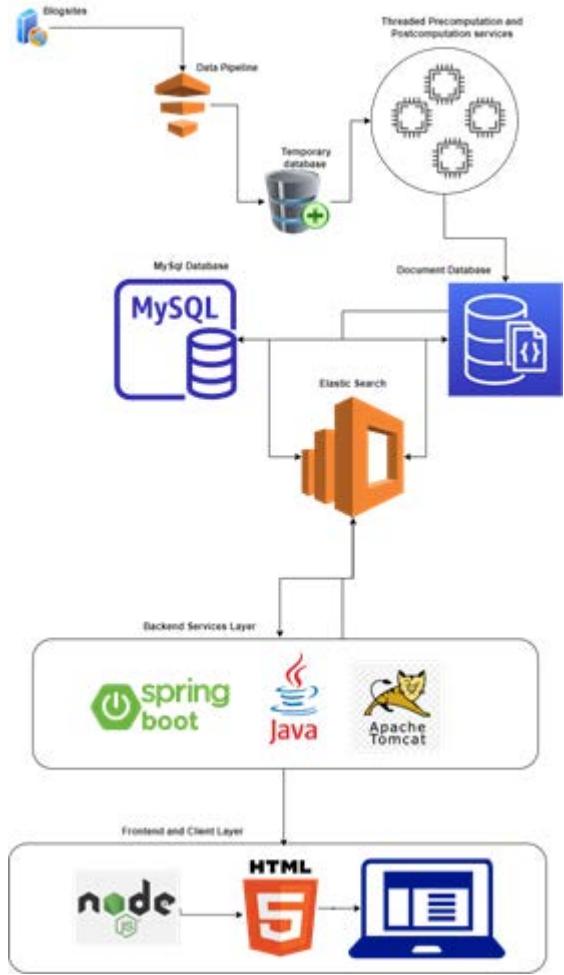


Fig. 1. High-level architecture Overview of the BlogTracker Application

bloggers generate content around interest in the blogosphere. We then generate the historical posting of bloggers and blogs and provide insights to some questions like "What is the posting pattern?", "Do blogs or bloggers post more at the beginning, middle or end of the year?", "How often do blogs and bloggers post?". Fig 5 and Fig6 shows blogs and bloggers that posted more during the cause of the coronavirus pandemic. Blogger and Blog Portfolio can study Daily, Monthly, and Yearly patterns to derive more accurate assumptions from blog posts.

##### E. Sentiments Analysis and Emotion Analysis

We used our sentiment and emotion analysis feature during this study, which is built on various social and behavioral Likert scales. This analytical tool helps display the trend of sentiments categorized into positive and negative sentiments with various emotion ratings like joy, sadness, trust, and anticipation on over 383 blogs for the selected period in the COVID-19 tracker Fig 7 shows the dashboard obtained from our BlogTracker tool.

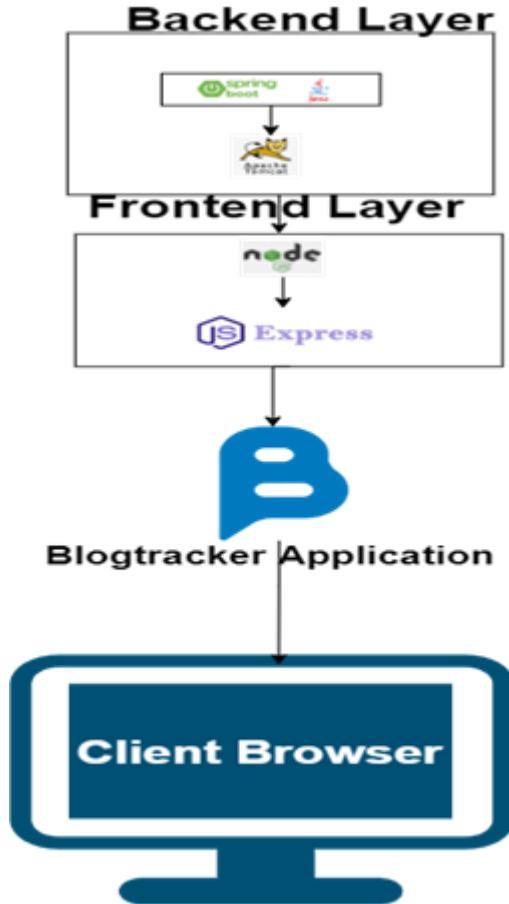


Fig. 2. A high-level overview of the BlogTracker Application

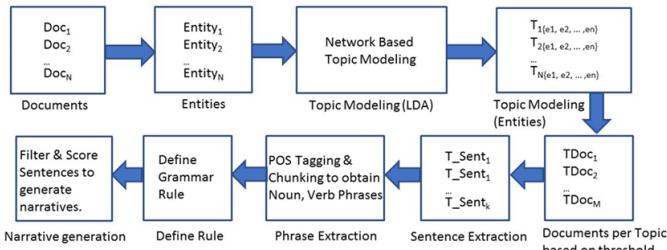


Fig. 3. Framework to extract narratives adapted after Hussain et al. (2021) [9]

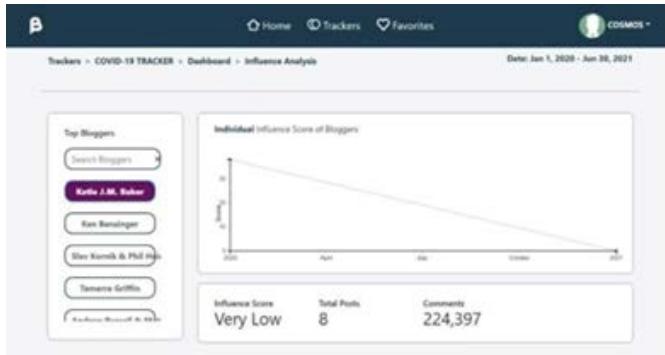


Fig. 4. Influence Analysis of Selected Bloggers on COVID- 19



Fig. 5. BlogTracker's Blog Portfolio

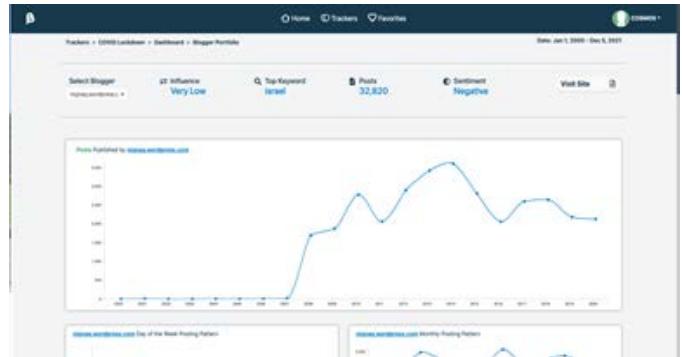


Fig. 6. BlogTracker's Blogger Portfolio

#### F. Topic Distribution

To identify the important subjects and the important themes are, we developed a solution around the Latent Dirichlet Allocation (LDA). This topic distribution determines how topics differentiate from each other Fig 8. We also limit our ranking of the generated categorical topics to 10 topics due to huge data volumes. Topic modelling involves discovering hidden patterns in text corpora through identifying topics present in a text object. Thus, assisting better decision making. It is an unsupervised approach used for finding and observing a bunch of words (usually called "topics") in large clusters of texts. A good topic model should result in – "pencil", "teacher",

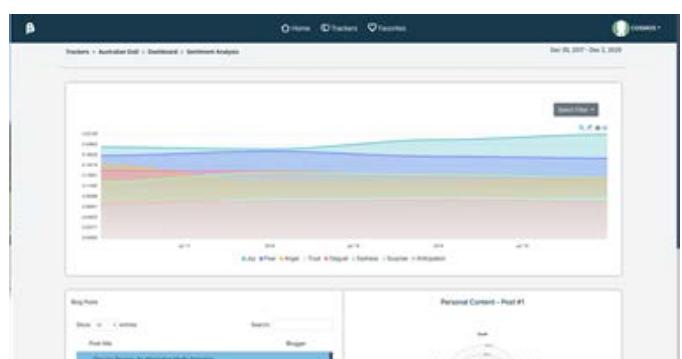


Fig. 7. BlogTracker's Sentiments Analysis and Emotion Analysis



Fig. 8. BlogTracker's Topic Distribution

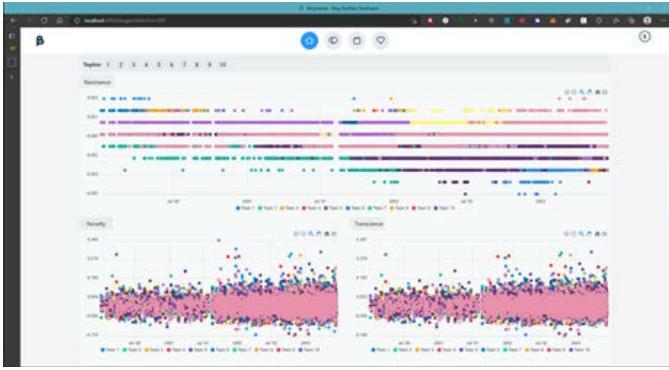


Fig. 9. Image of content diversity analysis capability of the BlogTracker Application

"book", "study" for a topic – Academic, and "rice", "bean", "wheat" for a topic – "Food" [11].

#### G. Content Diversity Analysis

Our application extends the capability of blogger portfolio and blog portfolio by using the computed LDA topic distribution data to compute for Novelty, Transience and Resonance, which was first defined in Barron et al. [12] and the works of Stine and Agarwal [13]. This analysis helps analysts identify shifting discursive priorities within a blog over time. Fig 9 shows an example visualization within the BlogTracker application to understand the blogosphere's shifting discursive.

#### H. Keyword and Trending Hot Topics

Keyword trend refers to the overall popularity of a search term compared to other searches. Keyword Trends are rated on a scale of 0 to 100, with 100 representing the highest popularity. With this information, users can see and analyze how a keyword grew over time, the circumstances that affected the growth and the motives behind bloggers that influenced that growth. Fig 10 shows an example of keyword trend obtained from the BlogTracker application.

### V. ANALYSIS AND FINDINGS CASE STUDY: COVID-19

To show the capability and how BlogTracker is used in studying the blogosphere to track the COVID-19 discourse,

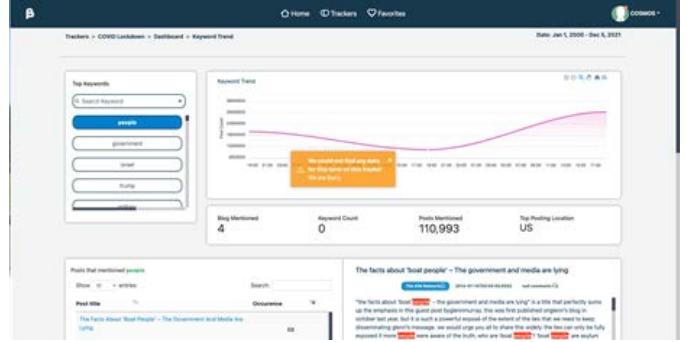


Fig. 10. BlogTracker's Keyword Trend

COVID-19 TRACKER	STATISTICS
Blog Sites	383
Blog Post	852106
Number of Bloggers	64,520
Number of Comments	11,392,673

TABLE I  
KEY STATISTICS OF THE COVID-19 TRACKER

we look at a tracker that contains a total of 383 unique blog sites and a total of 852,106 blog posts, the number of bloggers contained in this specific Tracker is 64,520 with a total of 11,392,673 comments. Table I shows this tracker information, while Fig 11 shows how this information is presented to analysts in the BlogTracker application. Fig 12 shows the dashboard summarized information of the entire blogosphere of the selected Tracker with 383 unique blog sites we were able to Tracker under this Tracker, with the information presented in a way that gives analysts a concise view but detailed of the hot discussion under a topic that was discussed during the COVID-19 pandemic. The dashboard provides some other key information like the top posting location where the topics were discussed, the top language used in generating posts, top narratives and their extracted entities, top bloggers, and much other key information.

#### A. Cluster Analysis

We used K-mean clustering from the studied blog site data and selected only the top 10 clusters for visualization with Fig 13, showing cluster 1 as the dominant cluster and statistics associated with cluster 1 selected. As we can see from the first cluster visualization in Figure 13, the top posting location for the blogs in this cluster is the United States, and it has a total of 275 bloggers mentioned in over 128,000 blogs; these statistics are data obtained over the time of February 8, 2022, till February 11, 2022. The U.S. is the top posting location where bloggers have contributed the most during the study period. While the posting location may be where the blog site server is hosted, we use the location captured as the posting location that has contributed the most to the topics discussed in cluster 1. We also presented the top keywords that shape the various discussions in cluster 1 in the word cloud and how these common terms overlapped in Fig 13.



Fig. 11. Shows Dashboard statistics at the top with summary statistics of information from the COVID-19 Tracker

During the pandemic, words like health, people, Canada, public, American, governments, presidents and many more were the centre of attraction as the dominant keywords in cluster 1. We also show some of the top blog posts in Figure 13 for analysts to visualize the blog post from each cluster and correlate this with the various keywords are shown in Figure 13. An analyst can also view the blog post content itself, and This table gives researchers the flexibility to view all blog posts associated with a selected cluster and do further analysis using other tools on the BlogTracker application.

### B. Posting Frequency

We also track and analyze the posting pattern of various blog sites and authors using our feature called posting frequency. With the posting frequency Fig 14, we observed the global-news.ca as one of the top contributors with a peak posting trend of globalnews.ca peaking between 2013 to 2020 before a steady decline afterwards, demonstrating that this blog site seems to have lesser contribution around COVID-19 after the global pandemic slowed down. We also compared the posting trend performance of globalnews.ca to other top blogging sites and noticed similar behavior, although globalnews.ca's peak posting period outperformed other top sites like inforwars.com

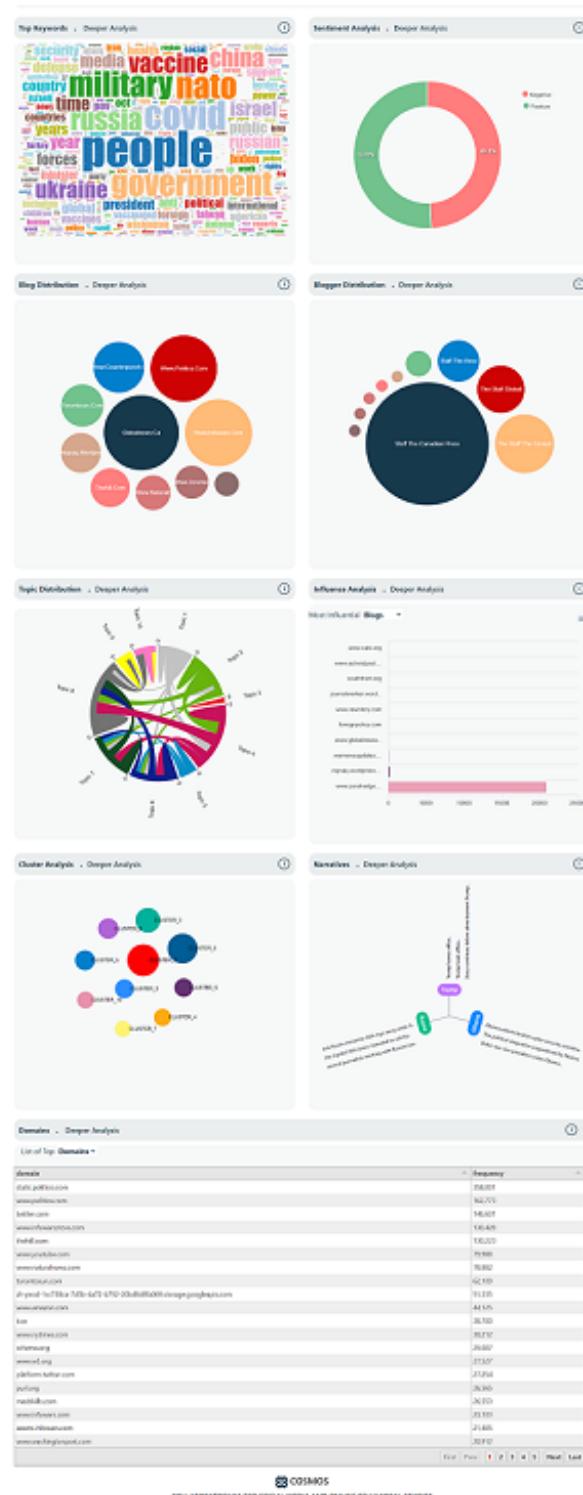


Fig. 12. BlogTracker Dashboard showing the concise yet detailed visualization of the COVID-19 Tracker.

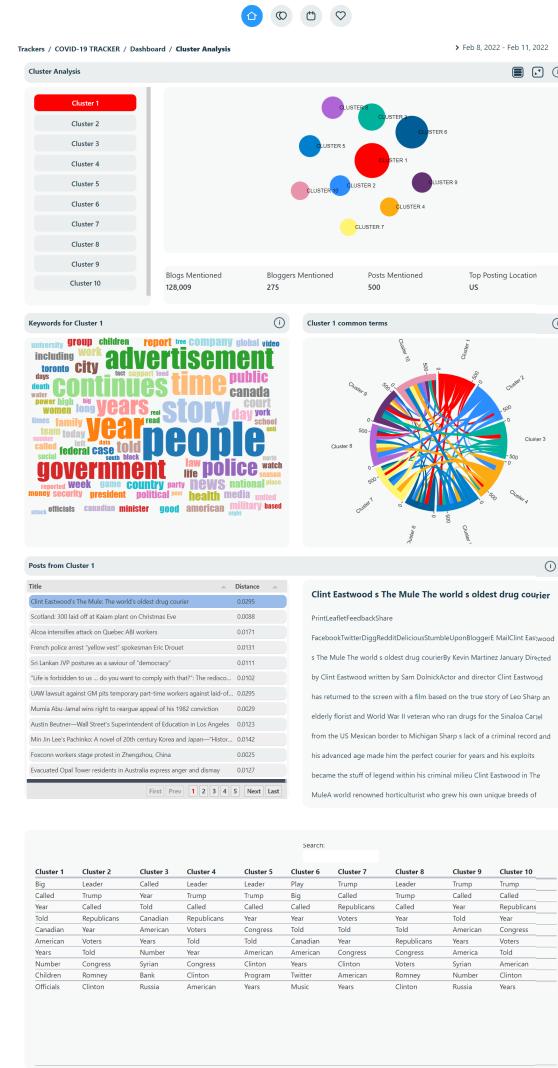


Fig. 13. Showing 10 clusters and keywords in the word cloud and common terms chord overlaps in various clusters

and counterpunch.org. However, when backtracking the visualization in Fig 15, which shows these comparisons, we noticed that counterpunch.org seems to have been an early blogger around pandemic and health-related discussions. From the statistics tab of the selected blog post below, we see that the overall sentiment is positive for globalnews.ca; included in the statistics is the number of posts on the selected blog site, 106,224, the number of comments on the blog site, 964,393 and the top keyword from this unique blog site is seen to be 'advertisement'. The top keywords for this posting trend of the selected Tracker were People, Government, Trump, Covid, Police, Cases, etc.; these keywords and entities as shown in Fig 14.

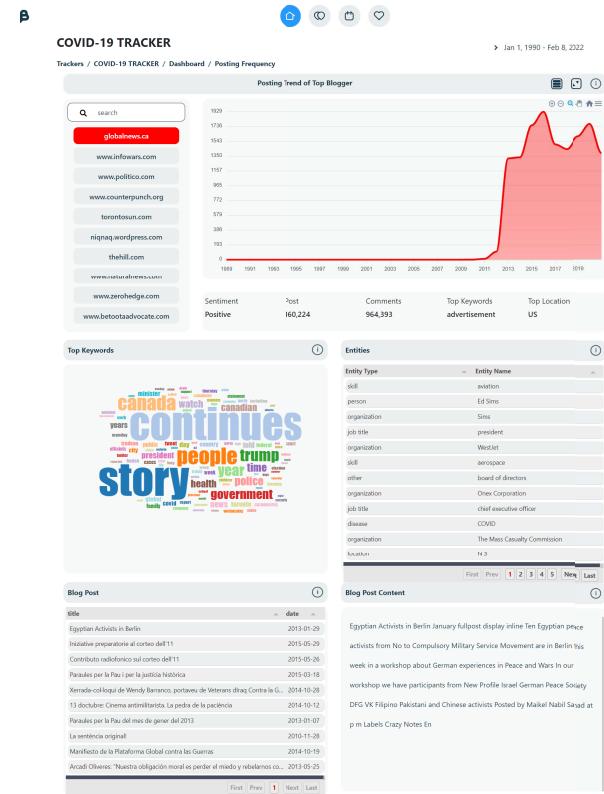


Fig. 14. Showing posting trends and dashboard statistics for globalnews

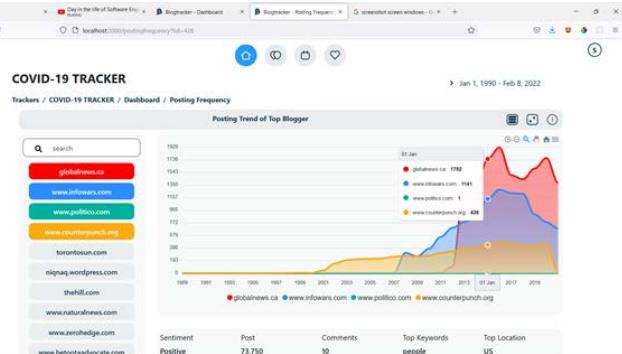


Fig. 15. Showing comparison of various top blog site posting trends.

### C. Sentiment Analysis

For sentiment analysis on the selected Tracker, we have presented here how BlogTracker can allow analysts to have a user-friendly visualization of various Likert scale emotions variables, and we took into consideration some key human emotions for which we compute emotion distribution; these emotion scales are name; Joy, Fear, Anger, Trust, Disgust, Surprise, Anticipation and Sadness Fig 16. Gauging the various blog posts in the COVID-19 Tracker lets us see which blog post and date show each Likert scale emotion. We also have a feature on BlogTracker that measures the toxicity and emotion score of various blog posts and the author. We visualize these



Fig. 16. Emotion Distribution visualization using the Likert scale emotions variables for rating reactions.

using the radar chart Fig 16 and allow analysts to compare these scores on the rather chart. Analysts can also compare two different blog posts to see how they perform. Using the radar chart at the bottom section of Fig 16 visualization, we observed that the emotions of the two compared blog posts are negative emotions symbolic of the effects of a pandemic and reaction towards a pandemic.

#### D. Narrative Analysis

The narrative capability of BlogTracker enables analysts to see how each entity and the respective blog title and narratives that the blog post author is influencing to be visualized easily. We used our narrative extraction model described in (Fig 3) adapted after Hussain et al. (2021) [9]. These entities, narrative and associated blog posts were then visualized using our intuitive developed narrative visualization interface as presented Fig 17 and [9]. We also provide an analyst with the ability to group these entities, merge narratives, and edit these narratives. Also, the keywords of these narratives were highlighted in the blog post, and Analysts can also sort these narratives and entities by criteria like relevance, date and alphabetical order.

#### E. Influence Analysis

We measured how influence is formed for the selected Tracker; our BlogTracker application was used to compute the top ten bloggers that influence the discus of the COVID-19 Tracker. We then measured and visualized their influence and the posting pattern of these top bloggers that are dominant influencers within the selected Tracker. We provide a section that shows the statistics like how strong their influence is, the top keyword used by these bloggers, total comments the post by the blogger has generated over the years as seen in Fig

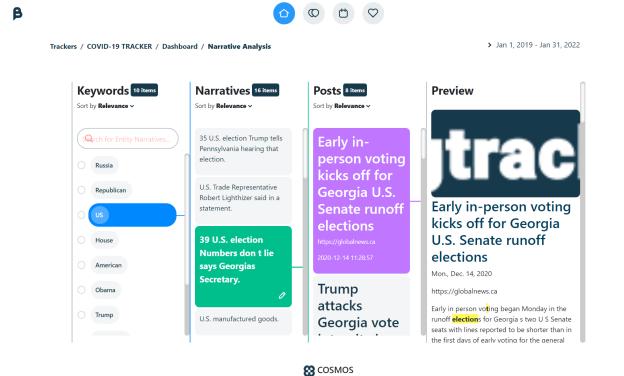


Fig. 17. Narrative visualization and highlighted narratives with the blog post content

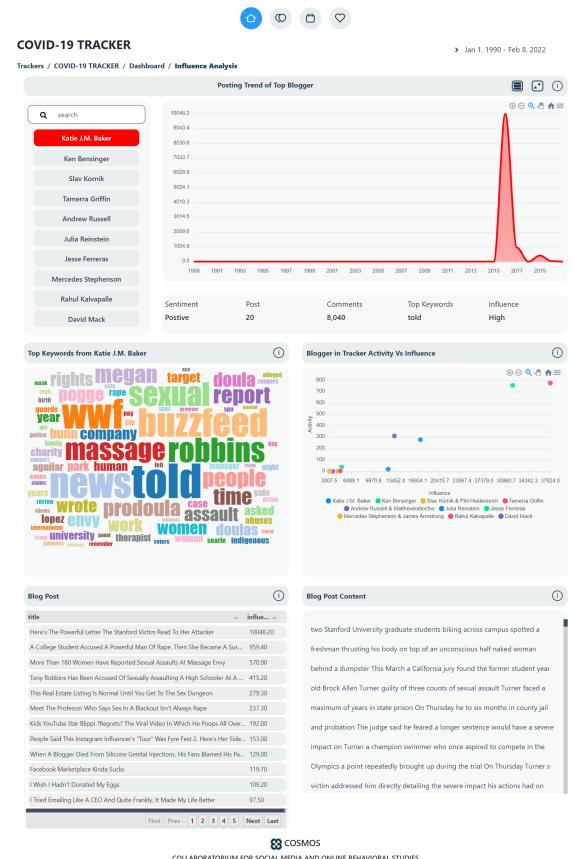


Fig. 18. Influence analysis dashboard, keyword cloud, influence activities and statistics of a selected blogger

18 which also shows the statistics of a selected blogger 'Katie J.M Baker', in the COVID-19 Tracker, has a positive sentiment attached to her blog sites, her blogs over 8,000 comments we also show the top keywords Fig 18. Our tool also allows analysts to compare and contrast multiple bloggers' influences and visualize these bloggers' statistics.

## F. Top Keyword Trends

We also used our keyword trend feature capability in this case study, and we were able to follow dominant keywords in the selected COVID-19 Tracker. According to the statistics computed by the BlogTracker tool, the top posting location for the selected keyword 'Trump' in Fig 19 is seen to be the United States, with over 165,000 posts mentioning the selected keyword. Fig 19 also shows the trend of the selected keyword in the COVID-19 Tracker. The keyword trend analysis feature in the BlogTracker tool is also able to highlight various blog posts in which the selected keyword (Fig 19) occurs dominantly along with a side panel of the blog post content and the keyword highlighted with this blog post content, thereby helping the Analyst to emphasize these sets of words with Fig 19 showing a comparison trend chart of keywords placed side by side.

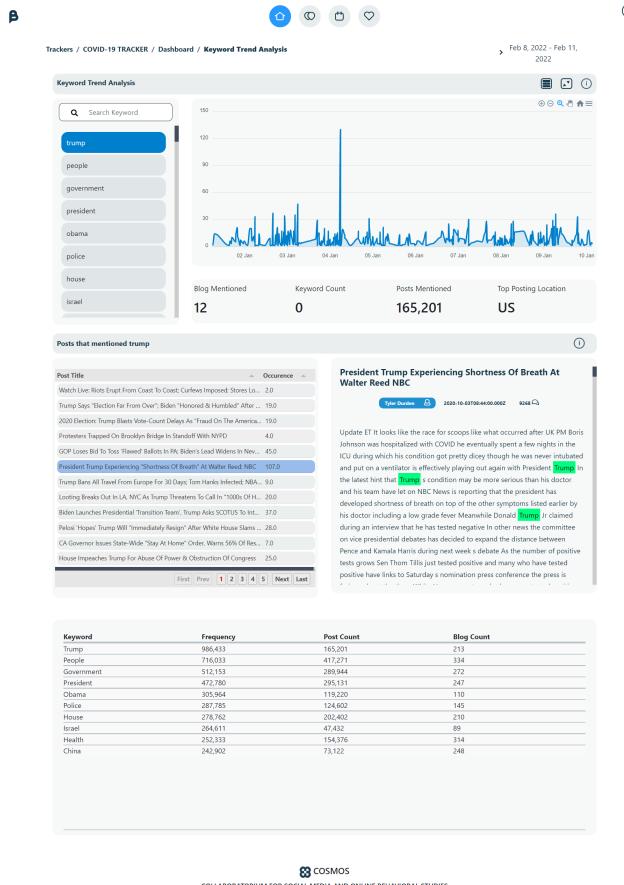


Fig. 19. Keyword Trend visualization showing top keywords and posts mentioned

## G. Topic Distribution Analysis

Furthermore, we showcase the capability of the BlogTracker tool in extracting and analyzing engagement topics. We generated the top ten topics and then tabulated these topics and their associated keywords to help analysts have in-depth insights into what words were dominant in each topic. Fig 21 shows

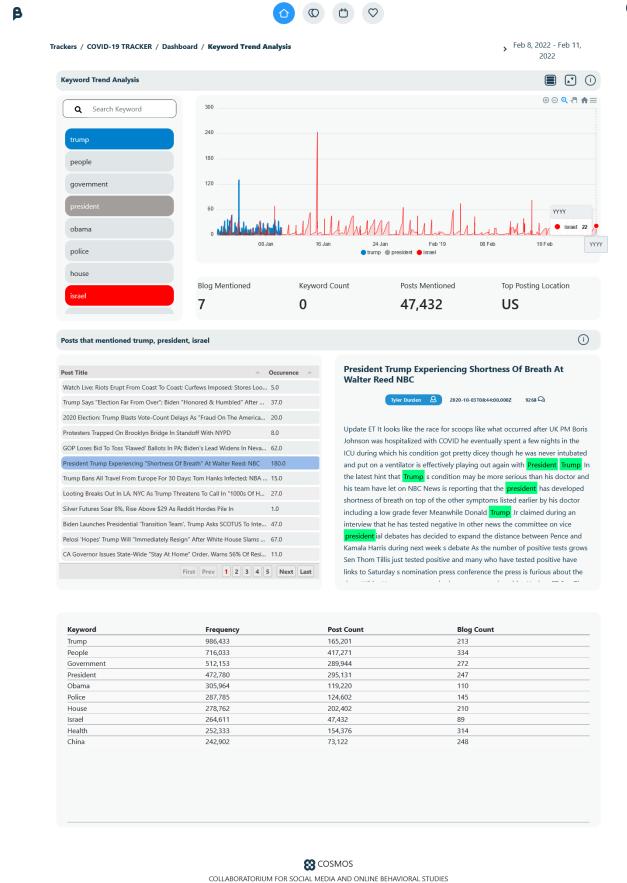


Fig. 20. Visualization showing compared selected keyword trends

the visualization of the topics distribution trend chart, while a card on Fig 21 also show the keywords word cloud of a selected topic and a chord diagram showing topic overlaps of how each blog post mentioned these keywords. In the selected topic, Topic 1 from the COVID-19 Tracker, we see the statistics show that the distribution of the total blogs in this topic is 9.71 percent of the total blogs in this Tracker, with a total of 1,146 bloggers and an associated 4,518 posts, we also aptly see that the top blogger in this Topic 1 is 'VD'. Furthermore, on careful analysis of the 4,518 posts in Topic 1. Our BlogTracker tool was also able to group these dominant keywords under their respective topics in a table format with words like "Trump", "vaccine", "Canada", and other trending words that were extracted from the various blog posts under study.

## VI. CONCLUSION AND FUTURE WORKS

In this paper, we focused on the capability of the BlogTracker tool to meet the yearnings of analysts who have had limited resources in studying the blogosphere in a centralized version where applications capable of analyzing various aspects of social and behavioral science may be unavailable, with also show how BlogTracker is capable of finding narratives and making the visualization interactive to the user, we also demonstrated how we were able to use BlogTracker

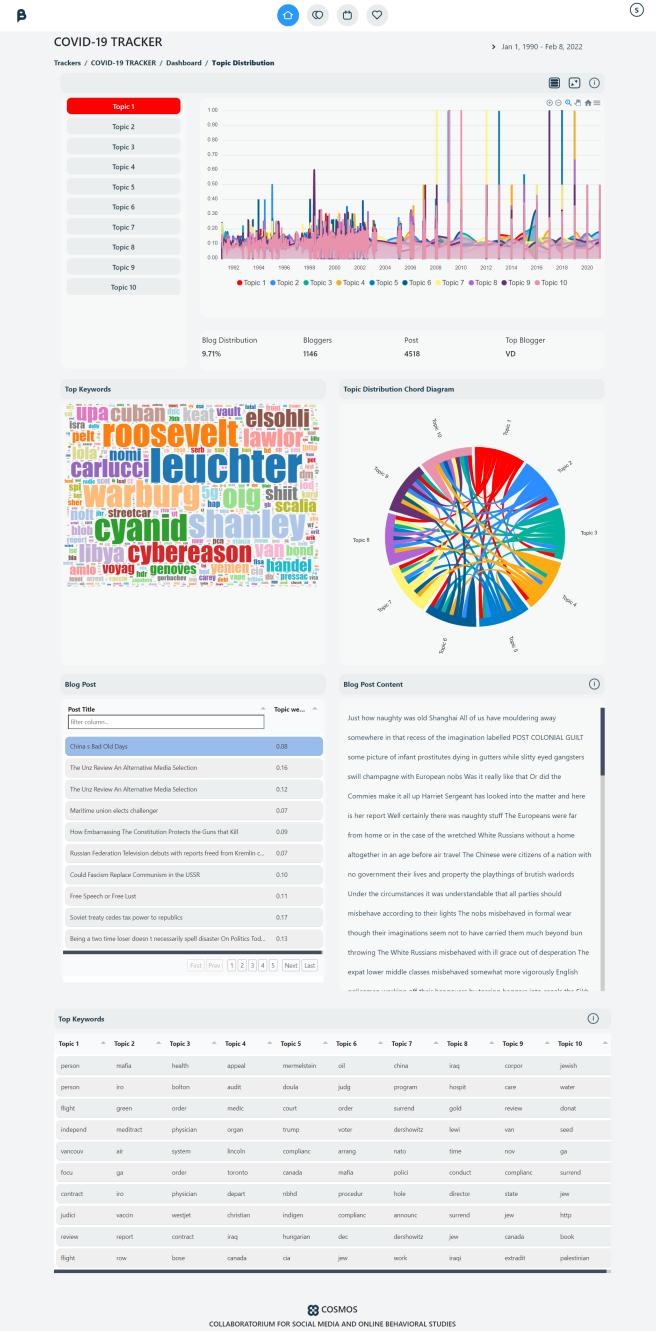


Fig. 21. Visualization and statistics of Topic Distribution of COVID-19 Tracker

in extracting important topics, sentiments and other opinion and user emotion activities towards a blog post. To show the capabilities of the BlogTracker tool in crunching large dataset and helping analysts fast-track research and behavioral studies, we used the COVID-19 Tracker and examined 852,106 blog posts from over 380 unique blogs sites. We were able to show some salient features of the BlogTracker application <https://btracker.host.ualr.edu>. Our work also solved the need for ready and easy access tools backed by years of research and continuous learning.

Future works will entail tracking the network of activities around deviant behaviors and how the blogosphere is affected by bots posting hyperlinks to various misinformation blog sites. We hope to have capabilities like this incorporated into BlogTracker to further extend the system's capabilities by understanding how narratives are shaped and how bots may have influenced narratives, and many other emerging issues.

## ACKNOWLEDGMENT

This research is funded in part by the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121, N00014-21-1-2765), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189), 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] A. Akinnubi, N. Agarwal, Z. Stine, and S. Oyedotun, "Analyzing online opinions and influence campaigns on blogs using blogtracker," in *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. Association for Computing Machinery*, New York, NY, USA, 2021, p. 309–312.
- [2] N. Tsirakis and V. Poulopoulos, "Panagiotis tsantilas, iraklis varlamis, large scale opinion mining for social, news and blog data," *Journal of Systems and Software*, vol. 127, 2017, iSSN 0164-1212. [Online]. Available: <https://doi.org/10.1016/j.jss.2016.06.012>.
- [3] p. Cuong, M. Derntl, Y. Cao, and R. Klamma, "Learning analytics for learning blogospheres," *Learning Analytics for Learning Blogospheres*, vol. 7558, p. 258–267, 2012.
- [4] M. Shaik, M. Hussain, Z. Stine, and N. Agarwal, "Developing situational awareness from blogosphere: An australian case study."
- [5] N. Agarwal, C. Sci, Eng, H. Liu, and L. Tang, "Identifying the influentials in the blogosphere," 2022.
- [6] M. Hussain, K. K. Bandeli, S. Al-khateeb, and N. Agarwal, "Analyzing shift in narratives regarding migrants in europe via blogosphere," 2018.
- [7] B. Al-Ani, G. Mark, J. Chung, and J. Jones, "The egyptian blogosphere: A counter-narrative of the revolution," p. 17–26, 2012.
- [8] K. K. Bandeli, M. Hussain, and N. Agarwal, "A framework towards computational narrative analysis on blogs," 2020.
- [9] M. N. Hussain, H. A. Rubaye, K. K. Bandeli, and N. Agarwal, "Stories from blogs: Computational extraction and visualization of narratives," in *4th International Workshop on Narrative Extraction from Texts (Text2Story) co-located with the 43rd European Conference on Information Retrieval. March*, Lucca, Italy, 2021, p. 28 –.
- [10] S. Peng, G. Wang, and D. Xie, "Social influence analysis in social networking big data: Opportunities and challenges," 2017.
- [11] S. Bansal, "Beginners guide to topic modeling in python and feature selection," 2020, retrieved February 9, 2022, from. [Online]. Available: <https://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/>
- [12] A. Barron, J. Huang, R. Spang, and S. DeDeo, "Individuals, institutions, and innovation in the debates of the french revolution," *PNAS*, vol. 115, no. 18, p. 4607–4612, 2018.

[13] Z. Stine and N. Agarwal, "Characterizing the language-production dynamics of social media users," *Soc. Netw. Anal. Min.*, vol. 9, 60, 2019.