Extracting and Visualizing Narratives from Social Video Sharing Platforms

Hayder Al Rubaye, Muhammad Nihal Hussain, Thomas Marcoux, Sruthi Kompalli,
Gaurav Raj Thapa, Mayor Inna Gurung, and Nitin Agarwal
COSMOS Research Center, University of Arkansas at Little Rock, Little Rock, Arkansas, United States
email: {hkalrubaye, mnhussain, txmarcoux, skompalli, gthapa1, mgurung, nxagarwal}@ualr.edu

***Abstract*—Social video sharing platforms are intended to allow users to associate with similar people to share their beliefs and encourage democracy. Nevertheless, deviant actors have used them to subvert the system. Deviant groups collaborate on digital platforms to propagate fake news, misinformation, and disinformation because of anonymity and perceived lower individual risk. Agenda setting, content framing, and weaponizing narratives are strategically employed to radicalize mobs and cause hysteria on uncontrolled social video sharing platforms, such as YouTube, which give a richer space for content development. Recent events and protests orchestrated via those platforms highlight the essential need for systems that can detect these fringe ideologies from the start. In this work, we show how to use a narrative visualization tool to help analysts find major topics and related narratives. The tool is based on a previously released framework for extracting narratives from video posts and is open to the public.**

Keywords-narrative visualizations; narrative; social video platforms; narrative extraction.

I. INTRODUCTION

YouTube, being the largest video sharing platform, not only enables its users to collaborate and share; it also serves as a treasure trove of data to study user behaviors. However, the recent infodemic of fake news on social media platforms has spread to YouTube as well. Rogue crowds are raising their presence within such online communities, spreading fake narratives by exploiting the ability to hide their identities with a low chance of being discovered. Their goals are to stir up, change, and divide people. For instance, with the latest COVID-19 pandemic, frauds, false information, and conspiracy theories have increased dramatically [3]. False information is spreading like the pandemic, therefore, it was named infodemic or misinfodemic [8]. Since those fringe narratives arise within dark web communities and spread to mainstream media, tracking those false narratives is harder than watching the growth of the pandemic. Unfortunately, we cannot “lockdown” or deactivate the internet to stem the infodemic. We need to actively track misinformation while it evolves, by establishing opposing courses of action to rapidly halt the damage to our communities within those video sharing websites.

However, to be able to analyze those false narratives precisely and create opposing courses of action, we would need algorithms that have efficient narrative extraction abilities and visualizations that are engaging to help analysts deeply navigate through them. While many studies have focused on extracting narratives from a collection of text documents [4] [11], only few have developed visualizations that would increase user engagement. This paper proposes a solution by developing a narrative visualization tool that enables its users to analyze narratives from video titles, descriptions, and transcripts visually. The tool extracts narratives leveraging our published research [1] [10] on narrative extraction and is integrated with the Vtracker [17] application, providing users a tree-like structure to visualize narratives anchored around prominent keywords or keywords of interest. The tool allows users to provide feedback, thereby helping to improve the narrative extraction approach. The proposed tool is scalable, language-agnostic, and adaptive to other narrative extraction approaches. It further offers several customizations, discussed in Section 4, to improve the overall user experience.

In the next part of the paper, Section 2, addresses the literature related to narrative extraction and visualization. In Section 3, the narrative extraction framework is discussed. In Section 4, the developed narrative visualization tool is explained in detail. In Section 5, we examine a case study and derive our findings. In Section 6, we explore the challenges and limitations of the proposed visualization. Finally, in Section 7, we state our plans for the future and conclude.

II. LITERATURE REVIEW

Narrative is defined as “a spoken or written account of connected events”, researchers have conducted several studies to visualize and analyze those narratives to help discover interesting themes. In the next subsections, we summarize a few of those studies and try to draw conclusions on their findings:

A. Narrative Extraction

Venugopalan et al. [16], Sah et al. [15], and Wang et al. [18] from their studies have all proposed different approach to address the common problem of translating videos to sentences using a unified deep neural network. Venugopalan et al. [16] proposed a solution to narrow down (extract) the important actions, subject and object (e.g. a person playing a guitar video then the object would be guitar) of the video and incorporating neural networks throughout the pipeline of those actions, from pixels to phrases (words), which enables model network training and tuning. Sah et al. [15] approach for translating videos to sentences differed a bit from Venugopalan et al. [16] the approach was to summarize videos and then annotate it. The summarization of video was first done by capturing the frame-to-frame movement of the video and capturing the cinematography all blended in an innovative method to create a summary from lengthy recordings. The approach involves use of method which is Recurrent Neural Network

(RNN) (Method that involves sequences of frames) to capture the summary of the video and use the SumBasic(Method to generate text summaries) to summarize the textual annotations. Another study in the same field by Estevam et al. [5] demonstrated how to improve performance on the dense video captioning problem by not only utilizing visuals of the video but also take cues from the audio. The model created showed how audio and speech modalities may improve a dense video captioning model. Independent studies have shown the effectiveness of both super frame cuts and key frame selection models. The major difficulty faced by different models was creating the textual summary. Hosseinzadeh et al. [7] in his model performed two tasks at the same time: anticipatory captioning (to forecast the future outcome) and video description development (to create the text summary).

The model by Venugopalan et al. [16] have demonstrated the approach to generate sentences by first predicting a semantic role representation i.e. high-level concepts such as the actor, action and object. Then to use a template or statistical machine translation to translate the semantic representation to a sentence. This strategy creates superior sentences and additionally, it provides insights that utilizes deep neural network models that automatically generate descriptions for images. The framework developed by Wang et al. [18] works on an end-to-end Dense Video Captioning Framework with Parallel Decoding (PDVC), which approaches dense video captioning by segmenting the video precisely into a lot of event pieces. Without requiring a dense-to-sparse proposal generation and selection phase, PDVC generates a set of temporally localized sentences directly based on the segmented video and makes the two sub-tasks deeply interrelated and mutually promoted through the optimization and considerably simplifying the conventional “localize-then-describe”(create and store the object and use it as reference) process. The objective of Hosseinzadeh et al. [7] is to build a statement that expresses the film’s likely future event. It tackles the problem by first anticipating the next event in the semantic space of convolutional features, combining contextual information into those features, and then forwarding them to a captioning module. The framework works by studying the sequence of historical frames in Red, Green, and Blue (RGB) space; i.e., each pixel is broken down into to RGB color code format and it predicts the next frame(s) in RGB space. For example, a javelin throw will have an event of the javelin traveling in projectile motion. So, the model can forecast the frames and develop the description. Chen et al. [2] proposed a unified caption architecture that extracts unsupervised multimodal topics from data and feeds them into the caption decoder. Multimodal themes that have been mined are more semantically and visually cohesive than pre-defined subjects and can more accurately mimic the distribution of video topics. It helped in better content understanding in deep multimodal applications. The proposed model can be used to predict the latent topics of videos and then generate topic-oriented video descriptions with the topic guidance jointly in an end-to-end manner. Iashin et al. [12] tried to tackle the problem of video understanding by combining a multi-headed proposal generator with a bi-modal transformer that was unique. The captioning module was influenced by transformer architecture, notably how the attention module combines the information from both sequences. Each proposal head in the bi-modal multi-headed proposal generator is inspired by You Only Look Once (YOLO), a very efficient object detector. Estevam et al. [5] presented a technique for unsupervised semantic visual information learning that is based on the notion that complex events (e.g., minutes) can be deconstructed into smaller events (e.g., a few seconds) and that these simple events are shared across several complicated events. Using a clustering approach like K-means clustering, they recover their latent representation from a long movie divided into short frame sequences. As a result, a visual codebook is created (i.e., a long video is represented by a sequence of integers given by the cluster labels).

B. Narrative Visualization

Storytelling has a long list of merits that profit its recipients; it helps them memorize the information, relate to the content, and improve their comprehension. We live in an open data world that is huge and complicated and there is a need to integrate storytelling into visualizations to spread that information effectively. After reviewing the literature related to the subject, we have identified two main themes. The first theme is a group of studies that surveyed existing visualizations for factors affecting our comprehension of a narrative visualization. Figueiras [6] addressed the problem by conducting a focus group of 16 individuals to investigate the most effective narrative elements utilized in a collection of 11 professionally produced visualizations hosted over the web. Additional information about the focus group’s likability, comprehension, and navigation preferences for each visualization was also collected. The results of the investigation led the authors to consider introducing the narrative elements of: context, empathy, and time to enhance the previously mentioned visualizations. Context helps users get a better grasp of the data. Temporal structure, can enhance users feelings about the story flow. Empathy helps in memorization and joy.

Another study by Segel et al. [14] considered a review of the current design methodologies used by online story writers like journalists to build advanced data visualization systems that enhance traditional storytelling. In some cases, a whole story can be replaced with a visualization. Several case studies related to media outlets and data visualization research were examined to recognize the different types of narrative visualization, and how factors like interactivity and narrative flow can affect readers’ perception of the story. After researching several visualizations, the authors proposed their own set of procedures, to be followed by journalists and educational media to design promising narrative visualizations.

The aforementioned studies lacked several aspects like a wider user base for testing, additional research on other types of visualizations, and investigating other effective storytelling elements. There is a need to do extra investigation of author vs reader-driven elements. Author-driven components provide information and structure, while reader-driven components enable interactivity and exploration.

On the other hand, the second theme consists of a different group of studies dealt with the extraction and visualization of entities and their narratives from text. Hussain et al. [9] used the Latent Dirichlet Allocation (LDA) model to extract topics from prominent entities. They showed how to use a narrative visualization tool to help analysts identify important themes and linked narratives. The technology is available for public use through the Vtracker tool [17].

Kanjirangat et al. [13] discussed how machine learning techniques can be utilized to help automate the creation of visual summarizations of any given narrative text. They mentioned several techniques used to achieve the required results, for instance, natural language processing tools were used to recognize named entities, Density-based Spatial Clustering of Applications with Noise (DBSCAN) a clustering algorithm was used to investigate characters and their aliases, a statistical analysis technique was used to detect the relationships between some of the important characters, and an undirected graph was used to visualize those characters and their relationships as nodes and edges. Finally, special sentiment analysis techniques were used to calculate sentiments to decide the colors for those characters, nodes and their relationships. According to the authors, the machine learning techniques they used proved to be having a deeper depiction of a given novel, and a case study of a series of books was conducted to prove their theory.

Both of the above narrative visualizations need further user testing to evaluate their success in providing analysts with the proper visual elements to understand the narratives. Another issue is related to the accuracy of the proposed grammar rules. They can be improved to make them less biased and work for complicated sentences. Finally, it is worth mentioning that we did not find literature that completely relates to the work we are conducting for narrative extraction and visualization from social video posts.

III. RESEARCH METHODOLOGY

Using our formerly published approach here, [1] (see Figure 1), we begin by extracting narratives from a set of videos using their title, transcript, and accompanying description. To achieve that, we obtain the names of notable personalities or locations from the videos and then center our attention on discourses related to them. We do that by extracting named entities and calculating their rankings based on their importance using document frequencies within the list of videos. After that, a mixture of the resulting list of named entities and videos is fed into a network topic modeling module to discover topics connected to those extracted entities.

In addition to recognizing topics exclusive to specific entities, the network topic modeling module assists in spotting overlapping topics. For instance, after examining the latest world events, we can recognize multiple rising entities like Putin's war, and inflation. There are topics exclusive to Putin's war like nuclear attack, defending Europe, war crimes, and any misinformation related to that event. On the other hand, inflation has unique topics like the economic recovery, Biden's family inflation shield, interest rates, etc. Meanwhile, intersecting topics may include Putin's war effect on inflation and how oil and gas prices are going higher, as well as, disinformation concerning those two prominent entities. The overlapping of topics allows a closer look into the investigation and can help improve analysis.

After extracting a specific number of topics for each entity, we can now adjust the accuracy of the LDA model to get the proper number of topics. We find the optimal number of topics by computing a range of topics (10 to 100) and picking the model with the highest coherence score. The LDA model has two main variables that we can modify, the log-likelihood, and beta. Topic distributions - over documents and words, have correspondent priors in LDA, which are often indicated with alpha and beta and are denoted as hyper-parameters because they represent the parameters of the prior distributions. The log-likelihood for any dataset is computed for every repetition that includes topics like 1, 5, 10, 25, 50, and so on. In an ideal world, the LDA model is supposed to improve data analysis in every iteration over time. As successive iterations make minor changes to the model, this value will ultimately normalize. Low beta values, conversely, emphasize that each of the topics should have only a small number of prominent words. As a result, combining these hyper-parameters aids in selecting the number of themes.

When the topics are extracted from the discovered named entities, we reduce the size of the video postings to the most prominent contributors according to their distributions within those topics. Then, out of these video posts, we extract the sentences that reference the entity connected to the topic. After that, we extract noun phrases and verb phrases utilizing Natural Language Processing (NLP) methods like Part of Speech (POS) tagging, and chunking. Later, we extract narratives by defining grammar rules that catch empirically recognized patterns. Lastly, for each entity we combine narratives according to their resemblance, and rank those narratives based on how prevalent they are in the dataset. We only explored English video posts in this study. But, we are planning to use our methods on non-English text too, and see how effective it will be.

IV. NARRATIVE VISUALIZATION

After extracting the narratives from video titles, descriptions, and transcripts, they are represented visually by the Vtrakcer narrative visualization tool [17] (see Figure 2). The proposed narrative visualization tool enlists important entities in a column view of keywords, where each keyword has narratives related to it. The aim of the column layout is to enhance user experience by hiding data complexity and enabling users to center their attention on one keyword or group of keywords per time. For instance, if a user wants to explore which narratives belong to a specific keyword, they would select that keyword to show all the narratives related to it. At that point, they can also select a particular narrative

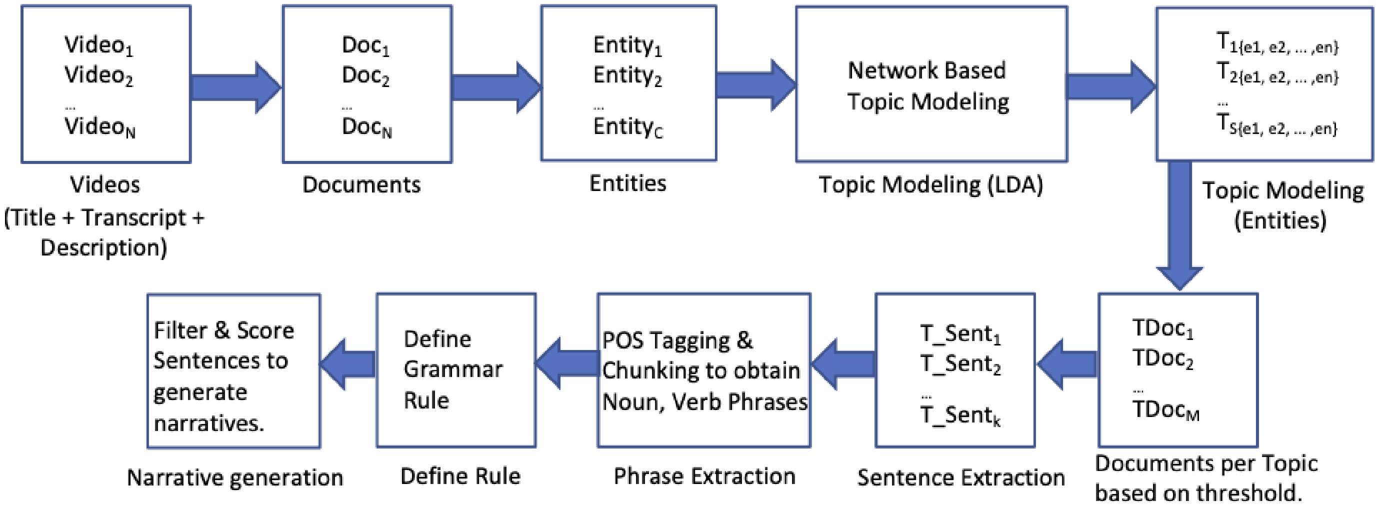
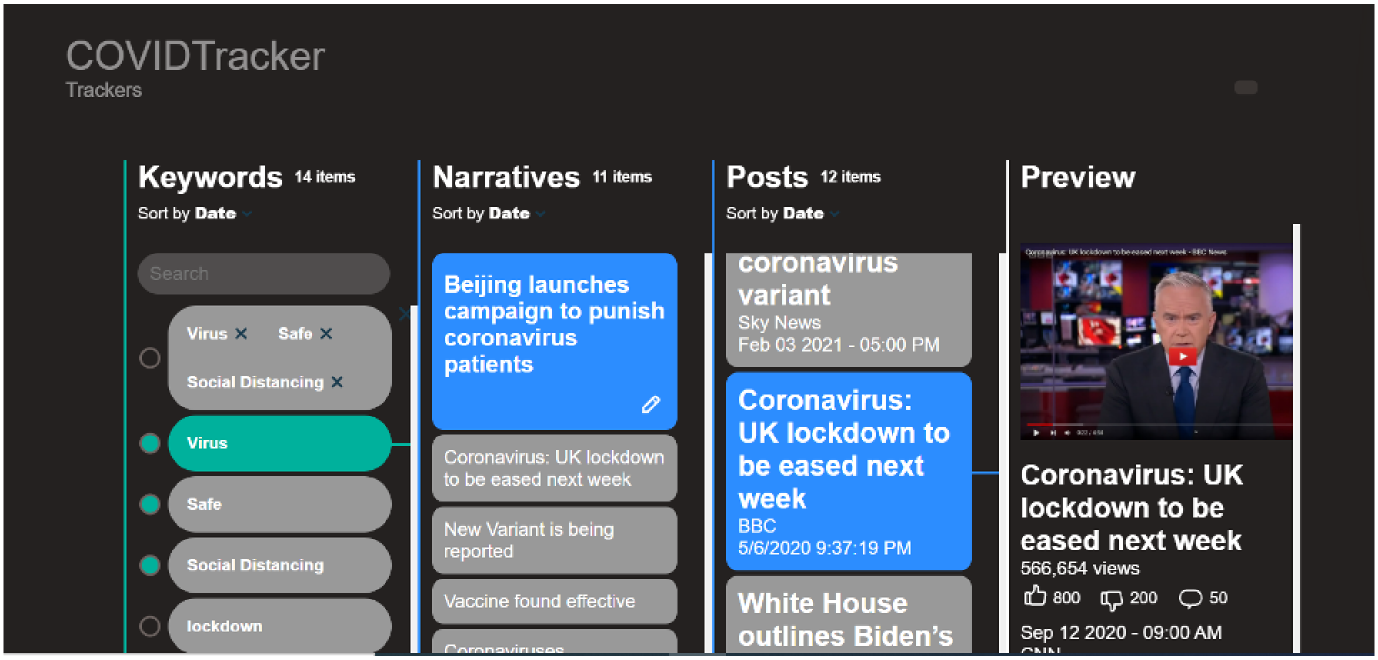


Figure 1. Framework to extract narratives from videos (Title, Transcript, and Description).



Beijing launches campaign to punish coronavirus patients

Coronavirus: UK lockdown to be eased next week

New Variant is being reported

Vaccine found effective

Coronaviruses

coronavirus variant
Sky News
Feb 03 2021 - 05:00 PM

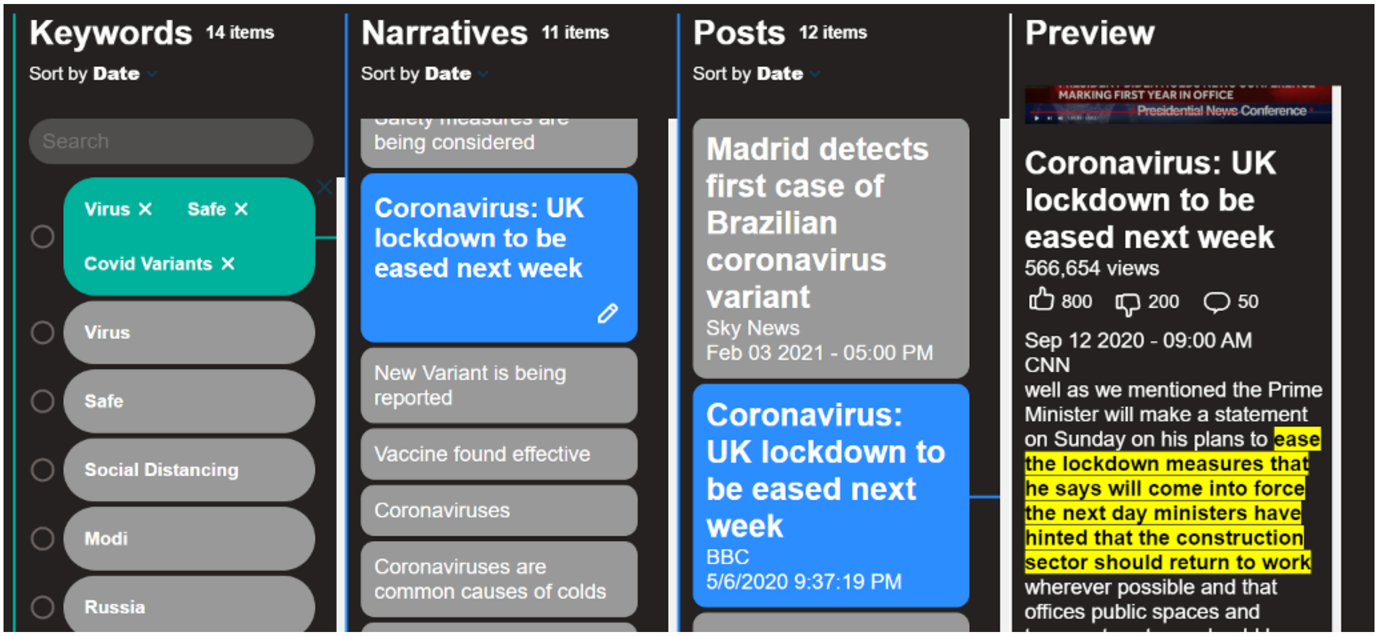
Coronavirus: UK lockdown to be eased next week
BBC
5/6/2020 9:37:19 PM

White House outlines Biden's

Coronavirus: UK lockdown to be eased next week
566,654 views
800 200 50
Sep 12 2020 - 09:00 AM
CNN

Figure 2. Narrative visualization tool in Vtracker application, showing multiple selected keywords to be combined into a collection to assist in spotting overlapping topics.

ISBN: 978-1-61208-979-9


Keywords 14 items
Sort by Date
Search
Virus ✕
Safe ✕
Covid Variants ✕
Virus
Safe
Social Distancing
Modi
Russia
Narratives 11 items
Sort by Date
Safety measures are being considered
Coronavirus: UK lockdown to be eased next week
New Variant is being reported
Vaccine found effective
Coronaviruses
Coronaviruses are common causes of colds
Posts 12 items
Sort by Date
Madrid detects first case of Brazilian coronavirus variant
Sky News
Feb 03 2021 - 05:00 PM
Coronavirus: UK lockdown to be eased next week
BBC
5/6/2020 9:37:19 PM
Preview
MARKING FIRST YEAR IN OFFICE
Presidential News Conference
Coronavirus: UK lockdown to be eased next week
566,654 views
800
200
50
Sep 12 2020 - 09:00 AM
CNN

well as we mentioned the Prime Minister will make a statement on Sunday on his plans to ease the lockdown measures that he says will come into force the next day ministers have hinted that the construction sector should return to work wherever possible and that offices public spaces and

Figure 3. Narrative visualization tool showing a collection of keywords being selected to reveal its top associated narratives.

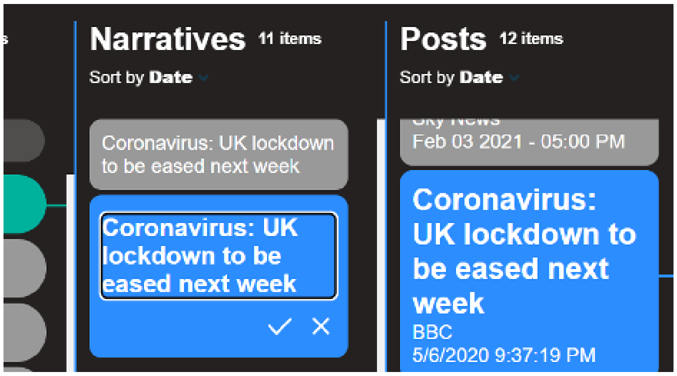

Narratives 11 items
Sort by Date
Coronavirus: UK lockdown to be eased next week
Coronavirus: UK lockdown to be eased next week
✓ ✕
Posts 12 items
Sort by Date
Sky News
Feb 03 2021 - 05:00 PM
Coronavirus: UK lockdown to be eased next week
BBC
5/6/2020 9:37:19 PM

Figure 4. Narrative visualization tool showing capability for a user to edit a narrative.

to list its related videos. The video can also be clicked to show its title, description, and transcript, if any. The tool also highlights where narratives have been referred to within the selected video data.

In addition to that, the Vtracker narrative visualization tool enhances user investigative powers by providing them with the ability to customize, search, and add new keywords. The tool will continually show all related keywords each time a user utilizes the search box. Once an item is selected from the search box drop-down list, it is added to the list of keywords that are being analyzed. Users can also investigate overlapping narratives that appear within multiple keywords by grouping them together, then selecting that group to view the related narratives. That can be done by enabling edit mode within the narrative visualization tool. In that mode the keywords list becomes editable where each keyword can be selected and combined to form groups of subjects. Groups can be deleted, and keywords can be removed from within a group (see Figure 3). Overlapping topics allows extracting compound narratives and keyword mixing.

User feedback can also be provided to improve the narrative extraction algorithm being used by the tool (see Figure 4). Users can edit a specific narrative and provide it as a suggestion for moderators to review and advance the extraction method. Additionally, the visualization tool is adaptive to change, the narrative extraction algorithm can be easily updated with any other future approach. Furthermore, the extracted narratives are pre-calculated and stored within the database, which makes the tool scalable to huge video sharing data sets.

Lastly, the visualization tool is coded with intense, lively, and vivid colors, where every color plays a specific role in enhancing user engagement. For instance, keywords were given a green color. Blue was used to represent narratives and videos. Whereas, yellow was used to highlight narratives within a video description to show its occurrence within the text.

V. FINDINGS

For our initial analysis, we had considered YouTube videos spanning from May to December of 2020 and extracted various narratives from those videos using topic modeling. COVID-19 was one of the most prominent entities, and the most dominant narratives were related to the control measures taken by the various nations to control the pandemic. For instance, the narratives related to easing of lockdown in the UK captured the keywords like “new scheme”, “the economic cost”, “hopefully”, “safe” leading us to conclude that the government of the UK is trying to ease the ongoing lockdown to minimize the further economic loss and are hopeful it will be safe. In contrast, we can find the negative narratives regarding steps taken by the Chinese government with the extraction of words like “Chinese”, “targeting”, “patients”, “leader” leading us to the narrative extracted that Bejing has launched a campaign to punish coronavirus patients. These are the narratives that escalated in May of 2020 while there were videos leaking in the media of Chinese police dragging people out of their homes. “Social distancing”, “Lockdown”, “economy”, etc. are some of the overlapping keywords found while looking for COVID-related videos.

VI. CHALLENGES AND LIMITATIONS

Although the narrative visualization tool in the presentation accurately recognizes narratives, the narrative extraction algorithm used in the above-mentioned technology has severe challenges and limitations. The application deals with large amounts of data, which is one of the major issues, as is integrating data from many sources. Because it is based on subject-based and empirical findings, the grammar rule needs to be updated. Because of the complexity of sentences and the language barrier, as well as many noun and verb phrase patterns, chunking may be a critical component of this framework, resulting in the rule’s failure. To be more effective, the narrative extraction algorithm must be enhanced. For example, it may have restrictions owing to the kind of words or phrases that the narrative library is exposed to, as well as translation and inference of what the video or blog is saying. It needs a particular level of precision to achieve it.

While narrative visualization allows analysts to customize it to help them identify narratives related to their keyword(s) of interest, it does have certain limitations. Furthermore, this type of research is prone to subjective bias, which must be avoided. Consequently, the proposed framework may be tested against several different research datasets. The various narratives that have been generated may not be completely true, but they can aid analysts in understanding. The only people that tested the User Interface (UI) were analysts and developers who were already familiar with the video tracker team. The usability of the user interface should be evaluated by a larger audience.

VII. CONCLUSION AND FUTURE WORK

The trend of using video-sharing social media platforms like YouTube is growing rapidly to share fake narratives. So, the advantage of using such a tool lies in its ability to extract those narratives from videos to identify the extreme and fringe ones being spread. This paper demonstrates that tool and its potency to identify videos with similar narratives which can be clustered to spot the channels, which can be helpful for the authorities to control circulation. The tool is based on the previously published narrative extraction algorithm [10] and has inherited the same level of accuracy, independence, and scalability but also it struggles to handle complex sentences and form grammatically correct sentences. However, the tool itself is independent of algorithms and can be adopted to better narrative extraction algorithms in the future. Also, it helps to collect user feedback as the tool lets users edit and save the extracted narratives.

Although the narrative visualization tool can extract simple and compound narratives, we are working to improve the extraction algorithm by making it faster, accurate, and platform-independent. Furthermore, we are also planning to include features to track and visualize the evolution of narratives to identify their origins as well as prominent events after which narratives evolve by merging, splitting, or even completely flipping, as observed in our previous study [9]. This can also help to detect intervals where a fringe narrative becomes dominant and vice versa.

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, N00014-22-1-2318), 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] K. K. Bandali, M. N. Hussain, and N. Agarwal. “A Framework towards Computational Narrative Analysis on Blogs.” In Text2Story@ ECIR, pp. 63-69. 2020.

[2] C. Shizhe, J. Chen, Q. Jin, and A. Hauptmann. “Video captioning with guidance of multimodal latent topics.” ArXiv:1708.09667 [Cs], Sep. 2017. http://arxiv.org/abs/1708.09667.

[3] “COVID-19 MISINFO | Home Page.” retrieved: Mar. 2022. https://cosmos.ualr.edu/covid-19.

[4] A. Dirkson, S. Verberne, and W. Kraaij. “Narrative detection in online patient communities” In Proceedings of Text2Story—Second Workshop on Narrative Extraction From Texts co-located with 41th European Conference on Information Retrieval (ECIR 2019) pp. 21-28. CEUR-WS.

[5] V. Estevam, R. Laroca, H. Pedrini, and D. Menotti. “Dense video captioning using unsupervised semantic information.” ArXiv:2112.08455 [Cs], Dec. 2021. http://arxiv.org/abs/2112.08455.

[6] A. Figueiras. “How to Tell Stories Using Visualization,” 2014 18th International Conference on Information Visualisation, 2014, pp. 18, doi: 10.1109/IV.2014.78.

[7] M. Hosseinzadeh and Y. Wang. “Video captioning of future frames.” In 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 979–88. Waikoloa, HI, USA: IEEE, 2021. https://doi.org/10.1109/WACV48630.2021.00102.

[8] "How misinfodemics spread disease - The Atlantic." retrieved: Feb. 2022. https://www.theatlantic.com/technology/archve/2018/08/how-misinfodemics-spread-disease/568921/.

[9] M. N. Hussain, K. K. Bandeli, S. Al-khateeb, and N. Agarwal. "Analyzing shift in narratives regarding migrants in Europe via blogosphere," In Text2Story@ ECIR, pp. 33-40. 2018.

[10] M. N. Hussain, K. K. Bandeli, H. Al Rubaye, and N. Agarwal. "Stories from blogs: computational extraction and visualization of narratives," In Text2Story@ ECIR, pp. 33-40. 2021.

[11] V. Kanjirangat, S. Mellace, and A. Antonucci. "Temporal embeddings and transformer models for narrative text understanding." ArXiv:2003.08811 [Cs], Mar. 2020. http://arxiv.org/abs/2003.08811.

[12] V. Iashin and E. Rahtu. "Multi-modal dense video captioning." In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 4117–26. Seattle, WA, USA: IEEE, 2020. https://doi.org/10.1109/CVPRW50498.2020.00487.

[13] V. Kanjirangat and A. Antonucci. "NOVEL2GRAPH: visual summaries of narrative text enhanced by machine learning." Text2Story@ ECIR (2019): 29-37., 2019.

[14] E. Segel and J. Heer. "Narrative visualization: telling stories with data." IEEE Transactions on Visualization and Computer Graphics 16, no. 6 (Nov. 2010): 1139–48. https://doi.org/10.1109/TVCG.2010.179.

[15] S. Sah et al. "Semantic text summarization of long videos" In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 989-997. IEEE, 2017.

[16] S. Venugopalan et al. "Translating videos to natural language using deep recurrent neural networks." ArXiv:1412.4729 [Cs], Apr. 2015. http://arxiv.org/abs/1412.4729.

[17] "Vtracker." retrieved: Feb. 2022. https://vtracker.host.ualr.edu/.

[18] T. Wang et al. "End-to-end dense video captioning with parallel decoding." In 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 6827–37. Montreal, QC, Canada: IEEE, 2021. https://doi.org/10.1109/ICCV48922.2021.00677.