

Fighting for Democracy: The Attempted Coup in Peru through the lens of TikTok

Gabriela Pinto^{1,2}, Keith Burghardt¹, Kristina Lerman¹, Emilio Ferrara^{1,2,3}

¹ Information Sciences Institute, University of Southern California, Marina Del Rey, CA, USA

² Department of Computer Science, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA

³ Annenberg School of Communication, University of Southern California, Los Angeles, CA, USA

gpinto@usc.edu, keithab@isi.edu, lerman@isi.edu, emilofe@usc.edu

Abstract

The events after the attempted coup in December 2022 consisted of protests in response to the decades of corruption within the Peruvian government. Several groups in underrepresented areas convened and exercised their right to protest and voice their opinion on social media. In reaction to the nationwide protests, clashes between the police and citizens led to numerous casualties. In this paper, we present a dataset of 43,697 TikTok video metadata published between November 20, 2022, and March 1, 2023, on the aftereffects of the attempted coup by former President Pedro Castillo. We further analyze the content within the metadata to gain insight into the discourse and reaction to the political crisis. The dataset will serve as a contributing source to motivate analyses that can benefit underrepresented groups in Peru.

Introduction

Peru has been through political instability for the past 20 years, which has resulted in a series of political scandals that led to the cycling of seven presidents in seven years. While Peru moved on from one president to the next, populations in the rural regions of Peru dealt with the inescapable fate of economic recession. Before the COVID-19 pandemic, the rural regions of Peru suffered from deep economic fragmentation, with more than two-thirds of the population surviving through informal employment, which has only worsened after the COVID-19 pandemic. After an attempted coup in December of 2022, economic instability for those residing in the rural regions only worsened as government support, such as access to medical care, education, and resources, decreased (Gumbiner 2023), leading to widespread protests and a subsequent government crackdown. While the political instability continues to persist, a lack of public social media data provides a perspective of the protesters and the residents within the urban and rural regions. Such data could be utilized for meaningful insight into online public discourse of a largely Spanish language region, including detection and analysis of information operations, misinformation, and police brutality.

The timeline for the coup and subsequent protests is as follows. In November 2022, Peru's Congress attempted to impeach President Pedro Castillo on the charges of leading

a criminal organization to profit off government contracts. In the face of imminent impeachment, Castillo attempted to dissolve Congress on December 7, 2022, but Congress removed Castillo instead. Former Vice President of Peru, Dina Boluarte, was subsequently sworn in as the new President of Peru. Consequently, a series of protests emerged throughout the nation driven by the distrust of political leaders, economic inequality that affects rural and indigenous areas, and extreme polarization. Soon after, newly sworn President Dina Boluarte declared a state of emergency, suspending the freedom of assembly (Turkewitz and Taj 2022; Cano 2022) for the people of Peru. Protests continued and called for President Boluarte's removal and vocalized the frustrations of the corrupted government that only worsened the economic inequality to those of the lower class. The protests gained the most traction in the provinces of Puno, Ayacucho, Arequipa, and Cusco, receiving ample support from Aymara and Quechua communities, who are usually underrepresented and currently negatively affected the most by Peru's economic inequality (Group 2024). Other groups joined in on the protests, including students, trade unionists, farmers, teachers, *frentes de defensa* (groups that defend regional interests), and *rondas campesinas* (rural patrol groups) to contribute to the call for early elections, Boluarte's resignation, closure of Congress, and the convening of an assembly to rewrite the constitution. Tensions between the protesters and Congress heightened when President Boluarte called the protesters "terrorists" (Tegel 2022).

On January 3, 2023, conservative groups organized the "Great March for Peace" to call for an end to the protests (Reuters 2023). Soon after, on January 9, 2023, 18 people were killed by the police during a protest in Juliaca, even though there was no evidence of weaponry carried by the protesters (McDonald and Tieffenthäler 2023). The civilians killed in the protest included three minors. January 19, 2023, was one of the protests that demanded Boluarte's resignation and the police raid at the National University of San Marcos that evicted protesters who took shelter at the university (Schmidt 2023; Collyns 2023).

The discussion of the political crisis quickly moved to TikTok, which has 20 million users in Peru (Bianchi 2024), out of 34 million total (United Nations Population Division 2024), providing a case study for the interactions between online activity and offline events. This sparked widespread

protests, with more than 2,370 demonstrations across the country, including 66 fatalities, in both urban and rural regions, according to data from the Armed Conflict Location and Event Data (ACLED) Project (ACLED 2023). The ACLED data was used in this paper because it contains information on the political demonstrations that occurred in Peru during the observed period. Furthermore, it resulted in at least 197 civilians and more than 200 police officers injured in the clashes. Possible research avenues that could be addressed from the presented work include analyzing content in a multilingual environment, multimodal analysis, and potentially enabling answers to social science questions, such as social media’s offline effects, for countries traditionally understudied by our research community. The contribution of this work provides an avenue for the research community to delve into answering research questions that can aid vulnerable communities in Peru. Political turmoil continues in 2024 with President Boluarte undergoing an “unlawful enrichment” investigation (Taj 2024), thus providing more motivation to delve into the online political discourse occurring in Peru. Our dataset contains metadata for 43,697 videos that have been shown to correspond to the political events included in the ACLED dataset. We are sharing the dataset publicly with the research community to apply state-of-the-art methods in creating an online space that doesn’t negatively affect vulnerable communities in Peru. Our data and code are available in the following link¹.

Related Work

Open data initiatives have enabled research on political turmoil and social unrest, predominantly using Twitter streams (Davis et al. 2016). The plethora of research on Twitter included multilingual content, with one of the studies centered on performing a descriptive analysis of the network topology of how participants of the Chilean 2011–2012 student movement self-organize and spread information (García et al. 2013). Similar analysis has been done on tweets discussing the #FreeYouth protests in Thailand, proving that Twitter was a vital tool for recruiting supporters, creating an open space for ideas to merge, and mobilizing pro-democratic sentiments among Thai Twitter users (Sinpeng 2021). As previous research has analyzed discussions on tweets in various languages related to political events outside of the U.S., we present this data as an avenue for future research to delve into such endeavors that include how protesters communicate on TikTok, but also as a source of analysis that will aid in the topics discussed that expose the discourse between police forces and Peruvian citizens offline.

Meanwhile, research on TikTok has been understudied but has included research into U.S. political communication (Medina Serrano, Papakyriakopoulos, and Hegelich 2020; Moir 2023), the spread of COVID-19 information (Li et al. 2021; Shang et al. 2021; Basch et al. 2021), the Brazilian Presidential Election (Lima et al. 2023), disinformation campaigns (Alonso López et al. 2021; Espinoza and Piña-García 2023), and climate change (Basch, Yalamanchili, and Fera 2022). Beyond political communication, previous work

has investigated the spread of hate on TikTok (Weimann and Masri 2023; Jacques et al. 2023). These studies, however, have highlighted limitations of TikTok data, questioning if the samples are representative, emphasizing the scarcity of non-English content during political crises in non-English speaking countries, or highlighting limitations in the data or the auto-generated speech transcripts that TikTok APIs produce (Medina Serrano, Papakyriakopoulos, and Hegelich 2020; Purushothaman et al. 2022; Li et al. 2021; Espinoza and Piña-García 2023). The dataset published in this paper supersedes in volume in comparison to the dataset used in the previous analysis conducted on TikTok and provides content related to the discourse that directly affects rural communities. The leading motivation for presenting the dataset is the lack of studies on political crises in South American countries, such as Peru, concerning disinformation campaigns, hate speech, and political-related content. Our dataset contains Spanish-speaking videos related to the attempted coup executed by former Peruvian President Pedro Castillo, which sparked nationwide protests that frequently occurred in provinces of the highest economic disparity, such as Cusco, Puno, and Arequipa. The metadata of the videos provides a clear view of the protests led by union leaders, teachers, students, and indigenous groups, which were soon met with altercations with the police. Datasets centered on these communities that are unrepresented by the research community could benefit from state-of-the-art techniques to gain a more profound scope of the country’s state to prevent more fatalities.

Data

Data Collection Method

We continuously queried the TikTok Research API,² for videos published between November 20, 2022, and March 1, 2023, and based on a set of keywords and hashtags, as shown in Table 1. The keywords were chosen to collect videos related to the former President, Pedro Castillo, and the current President of Peru, Dina Boluarte. Similarly, the hashtags were chosen based on the names and words that depicted words that generally depicted the country’s political state (e.g., state of emergency, coup). Our time range corresponded to the lead-up to, and the bulk of the protests after, the attempted coup.

Keywords	Hashtags
Castillo, Presidente Castillo, Pedro Castillo, Dina Boluarte, Presidente Boluarte, Boluarte	PedroCastillo, Castillo, Boluarte, pedrocastillo, pedrocastilloperu, pedrocastilloperú, golpedeestado, GolpeDeEstado, golpedeestadoperu, crisispoliticaenperu, crisispoliticaenperú

Table 1: List of keywords and hashtags queried from the API

The number of videos published and the demonstration recorded on ACLED during the queried time period are shown in Fig. 1. Noteworthy dates were also annotated to

¹https://github.com/gabbypinto/Peru_Coup_TikTok_Metadata

²<https://developers.tiktok.com/doc/about-research-api/>

provide clarity. These dates cover videos from 2 weeks before the political crisis until several months after protests reached their zenith. The collection of videos' features includes their unique identifier, the video's publication time and country code, the author's username, video caption (or formally labeled in the documentation as the "video description"), the unique identifier of the music (if any) in the video, the number of *likes*, *comments*, *shares*, and *views* the video received at the time of collection, the list of unique identifiers of the effects (if any) applied in the video, the list of associated hashtags, the unique identifier of its associated playlist (if any), and the audio transcripts (if any).

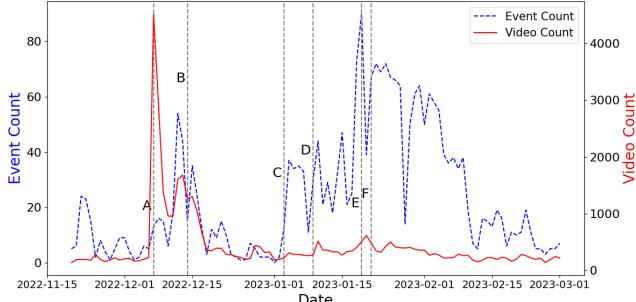


Figure 1: Timeline of events and volume of TikTok posts.

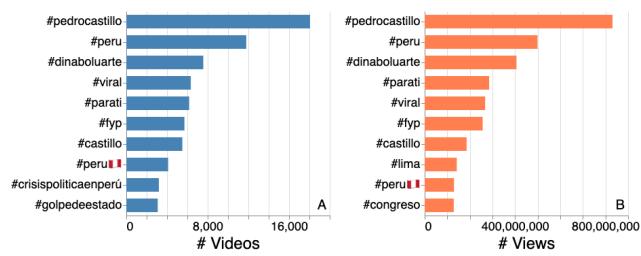


Figure 2: Top 10 most frequent and viewed hashtags. Left: number of videos containing the hashtag. Right: number of views for all videos containing that hashtag.

Attribute	Sum
No. Users	17,181
No. Likes	63,539,322
No. Views	1,396,069,669
No. Comments	6,264,238
No. Shares	6,964,547

Table 2: Statistics of the TikTok dataset.

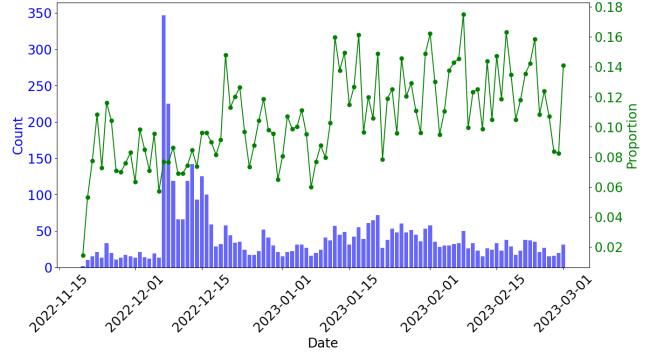


Figure 3: Number of TikTok videos with transcripts each day.

Data access

The dataset is publicly available and maintained on GitHub.³ In compliance with the TikTok Research API Terms of Service, we regularly update and prune our dataset, ensuring its alignment with current API offerings and adherence to TikTok's guidelines. Furthermore, our dataset only contains the video IDs to ensure user privacy in accordance with TikTok's Terms of Service.

Data Insights

Videos and Events. Our dataset currently has 43,697 videos posted during critical political events during the observed period. Most notably, Pedro Castillo's attempted coup and removal from office were followed by a series of demonstrations. The political unrest was initiated by protesters who vocalized the country's frustration with Peru's continuous political corruption and protests against the disproportionate police response. The range enables us to examine political discourse in relation to Peru's political environment one week before the attempted coup and after important events. The number of videos published per day and annotations of significant political events in Peru are shown in Figure 1.

Transcripts. Within our dataset, the TikTok API provided a small amount of TikTok-generated transcripts, potentially containing rich linguistic data. The number of videos with TikTok-generated transcripts based on the publication date and the proportion of videos with transcripts with respect to their publication date is shown in Figure 3, highlighting the scarcity of videos' transcripts. The percentage of videos with transcripts included, with respect to the publication date, ranges between 2 to 18%. Unlike shorter forms

³https://github.com/gabypinto/Peru_Coup_TikTok_Metadata

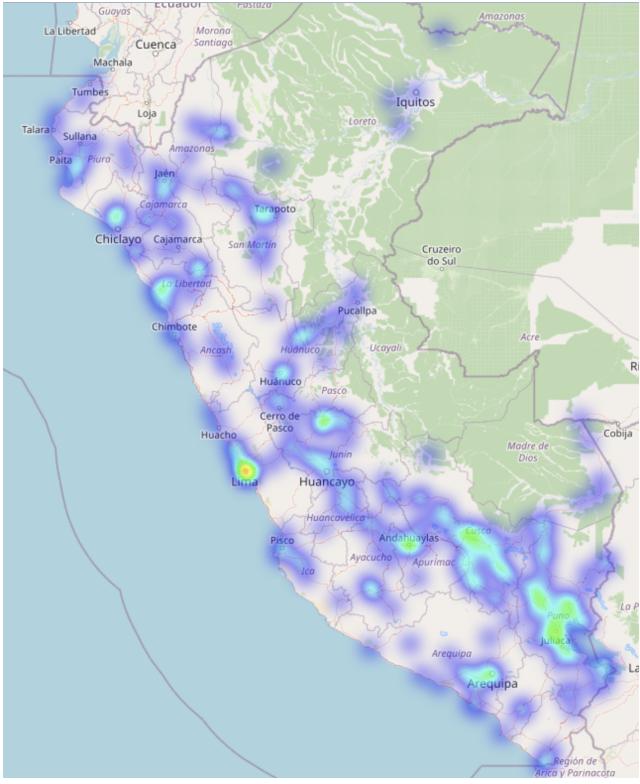


Figure 4: The locations of demonstrations mentioned in ACLED dataset (ACLED 2023).

of textual social media content such as tweets, the transcripts can provide more context and ample opportunity for linguistic analysis in future work.

Shares, Likes, Views, Comments, Users. Table 2 describes the total number of unique users, likes, views, comments, and shares across the entire dataset. While our dataset purely contains the videos’ metadata, the video’s ID can be used to collect user information and comments through the TikTok Research API for further analysis of the online discourse. In Figure 2, subplot A shows the most frequent hashtags, and subplot B shows the most viewed hashtags within our dataset.

Location of the Events. ACLED was utilized to analyze further the situation of the demonstrations that occurred within the country during the same time range included in the TikTok API call. Figure 4 illustrates the frequency of events with respect to region, where Lima, Puno, and Arequipa experienced the most activity. Similarly, we present the proportion of videos mentioning the locations shared with ACLED in the videos’ transcripts in Figure 5. The locations shared in the transcripts, and the ACLED data prove that the videos provide a strong representation of the regions where the most political demonstrations and casualties occurred, which include Puno, Arequipa, and Cusco provinces.

Content Analysis on TikTok Generated Transcripts. Given the linguistic richness within the videos’ transcripts,

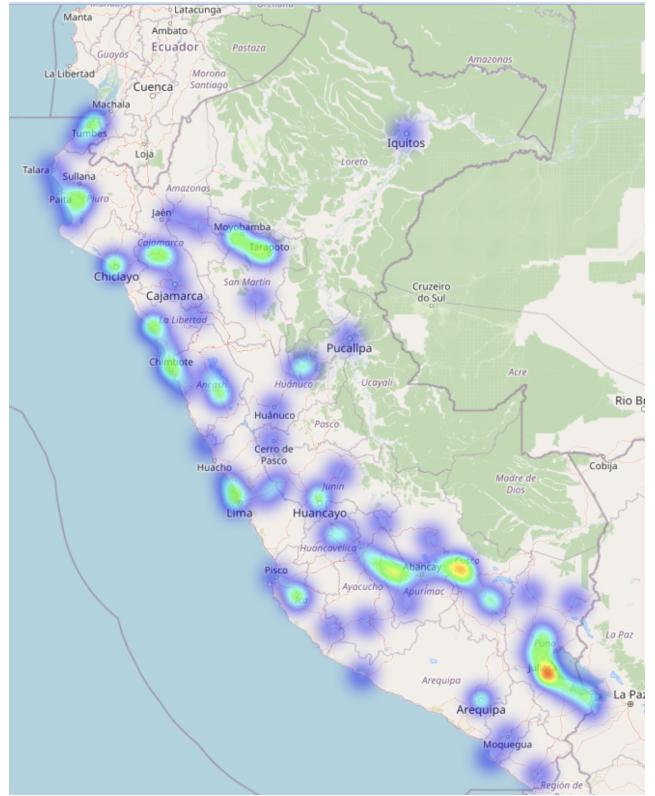


Figure 5: TikTok video locations. Percentage of TikTok generated transcripts that included locations mentioned in the ACLED data.

we performed a content analysis of 4,285 transcripts. The transcripts’ most common words indicate some references to Peru’s police enforcement, congress, government, armed forces, protests, violence, and terrorism. The verbiage within the most common words indicates online discussion mentioning violent altercations between the armed forces and civilians in Peru when attempting to exercise their right to assembly. The common words can be seen in the collection of bigrams shown in Figure 6. In fact, the effects of Castillo’s attempt to dissolve Congress on the people of Peru are even more prevalent with references to armed forces (*fuerzas armadas*), people of Peru (*pueblo peruano*), close Congress (*cerrar congreso*), human rights (*derechos humanos*), emergency government (*gobierno emergencia*), the Shining Path (*sendero luminoso*) political party of Peru, curfew (*toque queda*), the Peru Libre political party, and the raid at the National University of San Marcos.

In addition to word and bigram frequency, we analyze the topic distribution within the transcripts. Figure 7 illustrates the topics within the transcripts and the intersection of the topics with the frequent words and bigrams. The topics are ranked numerically within the plot, where one is the most frequent. The topics were generated using BerTopic (Groothedorst 2022) based on the sentence transformer model *paraphrase-multilingual-mpnet-base-v2 Semultilingual* for its embedding model (Reimers and Gurevych 2019). The



Figure 6: Most common bigrams within the transcripts of TikTok videos.

topics within the transcripts include mentions of terrorism, police forces, bombings, fires, the crime committed by Castillo, and protesting. Thus, providing motivation to analyze the online discussions between users on videos related to these topics, which emerged as the result of the political crisis.

The frequency of a select number of topics over time is shown in Figure 8, with annotations corresponding to the table in Figure 1. The topics shown in Figure 8 correspond to the topics mentioned in Figure 7. An increase in the selected topics corresponds to the pivotal events that occurred during the observed period. Videos published on December 7, 2022, and onwards have notable peaks in content related to Mexico. On December 7, 2022, which was the same day of Castillo's attempt to dissolve Congress and his attempt to flee to Mexico with his family (Wright and Buschschlüter 2022). Nevertheless, Castillo was detained in Peru at the time, while Mexico provided asylum for his family. Spikes of vandalism and protests, *vandalismo, protestas, marchas*, mentioned within the transcripts are observed around January 09, 2023, and January 19, 2023, which were days when 18 were killed during a protest in Juliaca and the protest in Lima demanding for President Boluarte's resignation, respectively. In addition, transcripts mentioning bombs and fires, *bomba, incendio*, also spiked during the same time period.

Discussion

The dataset presented in this article provides a glimpse into the aftereffects of the attempted coup in various regions of Peru. The mentioned locations in the transcripts correspond to the locations of the demonstrations seen on ACLED, thus providing the opportunity for future work to delve into how urban and rural regions experience the effects of the political crisis.

The topics embedded within the transcripts reflect the public opinion and discourse on the events that happened soon after the attempted coup committed by Pedro Castillo. Topics such as the emergency curfew were inflicted soon after Castillo was arrested. The Sendero Luminoso political party, which conducted a campaign of violence and destruction during the 80s and 90s, was also one of the mentioned topics. Further analysis could focus on the pres-

ence of opposing sides of the political spectrum and investigate how their online presence affects public opinion. Other notable topics include the police raid at the University of San Marcos, where numerous protesters were residing, and the bombings that killed civilians, both bystanders and protesters. Analyzing the online discourse can provide deeper insight into the government's opposition to the protesters, whether it takes the form of police brutality or the spread of misinformation from the opposing side.

Potential Applications. A plethora of applications for the presented dataset can be used to analyze online and offline events. The content analysis demonstrated evidence of discussion centered on the clashes between the police and civilians shown on TikTok. The topic analysis shown in Figure 7 on the transcripts indicates that the most prevalent topics include terrorists, bombs, fires, protests, police forces, and the University of San Marcos. Future work could *detect police brutality in a multilingual environment*, and even in a multi-modal environment when given the video.

While the analyzed transcripts are only provided for 4,285 out of the 43,697 videos, augmenting the number of transcripts through downloading and extracting the videos would provide a more expansive and descriptive analysis. Furthermore, downloading the TikTok videos would open a pathway for analyzing how the communities are depicted and represented in a multimodal environment. Communities shared between the transcripts and ACLED data, specifically Puno, Arequipa, Cusco, Ayacucho, and Juliaca.

The dataset can aid in answering various research questions on TikTok about the political state in Peru. *How different regions were affected by the political crisis? How does disinformation spread within the platform? Do outside countries influence affect the public opinion and discourse on TikTok?*

Conclusion

In summary, we present a collection of TikTok data for 43,697 videos that provide an overview of the online discourse in Peru after the attempted coup. To provide insight into the data, we include a metadata content analysis. Furthermore, this dataset bridges a significant research gap by focusing on non-English content and regions typically overlooked in social media studies. By sharing our dataset and codebase, we encourage further exploration into the impacts of digital platforms on societal and political dynamics.

Limitations. The dataset sheds light on online Peru’s political crisis discussions but has limitations. It doesn’t fully capture the sentiment on former President Castillo’s attempted coup, considering Peru’s 68.68% social media penetration and approximately 20 million TikTok users out of a 34.5 million population. Furthermore, the lack of video transcripts that contain rich linguistic data is significantly lacking from the API in general. We encourage the community to facilitate the availability of these transcripts as they provide a myriad of implications.

Ethical statement. In compliance with the TikTok Research API Terms of Service, we regularly update and prune

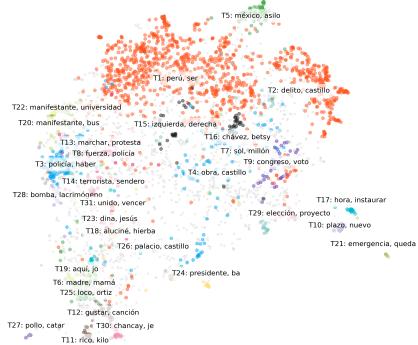


Figure 7: UMAP of the Top 30 Topics

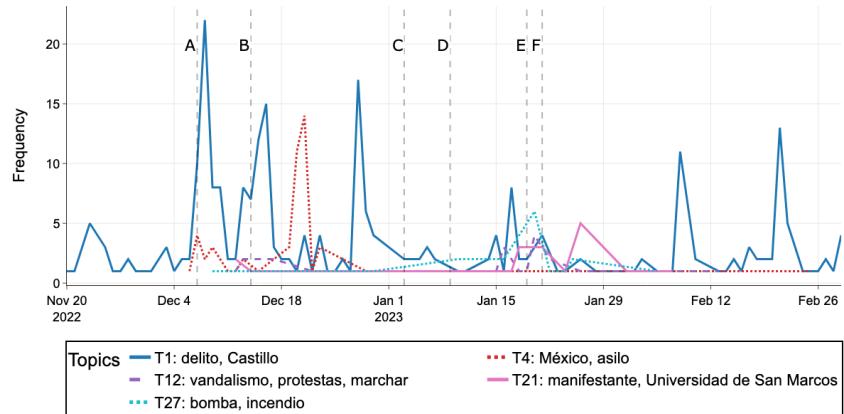


Figure 8: Topics over time and related events (cf., Figure 1)

our publicly available dataset, ensuring its alignment with current API offerings and adherence to TikTok’s guidelines. We only include video IDs in our public repository to ensure user privacy.

Acknowledgments. This work was supported in part by DARPA (contract no. HR001121C0169).

References

- ACLED. 2023. ACLED Data. <https://acleddata.com/>.
- Alonso López, N.; Sidorenko Bautista, P.; Giacomelli, F.; et al. 2021. Beyond challenges and viral dance moves: TikTok as a vehicle for disinformation and fact-checking in Spain, Portugal, Brazil, and the USA. *Anàlisi*.
- Basch, C. H.; Meleo-Erwin, Z.; Fera, J.; Jaime, C.; and Basch, C. E. 2021. A global pandemic in the time of viral memes: COVID-19 vaccine misinformation and disinformation on TikTok. *Human vaccines & immunotherapeutics*, 17(8): 2373–2377.
- Basch, C. H.; Yalamanchili, B.; and Fera, J. 2022. # Climate change on TikTok: a content analysis of videos. *Journal of community health*, 1–5.
- Bianchi, T. 2024. Social media usage in Peru - Statistics & Facts.
- Cano, R. 2022. Peru’s new government declares police state amid protests.
- Collyns, D. 2023. Police violently raid Lima university and shut Machu Picchu amid Peru unrest.
- Davis, C. A.; Ciampaglia, G. L.; Aiello, L. M.; Chung, K.; Conover, M. D.; Ferrara, E.; Flammini, A.; Fox, G. C.; Gao, X.; Gonçalves, B.; et al. 2016. OSOMe: the IJUNI observatory on social media. *PeerJ Computer Science*, 2: e87.
- Espinoza, A.; and Piña-García, C. A. 2023. Propaganda and Manipulation in Mexico: A Programmed, Coordinated and Manipulative “Pink” Campaign. *Journalism and Media*, 4(2): 578–598.
- García, C.; Chauveau, P.; Ledezma, J.; and Pinto, M. 2013. What can Social Media teach us about protests? Analyzing the Chilean 2011-12 Student Movement’s Network evolution through Twitter data. *arXiv preprint arXiv:1308.2451*.
- Grootendorst, M. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Group, I. C. 2024. Unrest on Repeat: Plotting a Route to Stability in Peru — Crisis Group — crisisgroup.org.
- Gumbiner, L. M. M. C., Lawrence; Camino. 2023. Rage and Repression in Peru: Restoring Citizen Trust and Public Order — usip.org.
- Jacques, E. T.; Basch, C. H.; Fera, J.; and Jones II, V. 2023. # StopAsianHate: A content analysis of TikTok videos focused on racial discrimination against Asians and Asian Americans during the COVID-19 pandemic. *Dialogues in Health*, 2: 100089.
- Li, Y.; Guan, M.; Hammond, P.; and Berrey, L. E. 2021. Communicating COVID-19 information on TikTok: a content analysis of TikTok videos from official accounts featured in the COVID-19 information hub. *Health education research*, 36(3): 261–271.
- Lima, J.; Santana, M.; Correa, A.; and Brito, K. 2023. The use and impact of tiktok in the 2022 brazilian presidential election. In *dg.o 2023*, 144–152.
- McDonald, B.; and Tieffenthaler, A. 2023. How Peru Used Lethal Force to Crack Down on Anti-Government Protests.
- Medina Serrano, J. C.; Papakyriakopoulos, O.; and Hegelich, S. 2020. Dancing to the partisan beat: A first analysis of political communication on TikTok. In *WEBSCI*, 257–266.
- Moir, A. 2023. The Use of TikTok for Political Campaigning in Canada: The Case of Jagmeet Singh. *Social Media+ Society*, 9(1): 20563051231157604.
- Purushothaman, V.; McMann, T.; Nali, M.; Li, Z.; Cuomo, R.; and Mackey, T. K. 2022. Content analysis of nicotine poisoning (nic sick) videos on TikTok: retrospective observational infodemiology study. *Journal of medical internet research*.
- Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *EMNLP*. Association for Computational Linguistics.
- Reuters. 2023. Conservative Peruvian Groups March for Peace Ahead of Fresh Protests.
- Schmidt, S. 2023. Protesters flood Peru’s capital to demand president step down.
- Shang, L.; Kou, Z.; Zhang, Y.; and Wang, D. 2021. A multimodal misinformation detector for covid-19 short videos on tiktok. In *2021 IEEE international conference on big data (big data)*, 899–908. IEEE.
- Sinpeng, A. 2021. Hashtag activism: social media and the# FreeYOUTH protests in Thailand. *Critical Asian Studies*, 53(2): 192–205.
- Taj, M. 2024. Police Raid Peruvian President’s Home, Looking for Rolex Watches — nytimes.com.

- Tegel, S. 2022. Protesters in Peru refuse to back down as political crisis deepens — washingtonpost.com.
- Turkewitz, J.; and Taj, M. 2022. Protesters in Peru Demand Justice for Ousted President Pedro Castillo (Published 2022) — nytimes.com.
- United Nations Population Division. 2024. Population, total - Peru.
- Weimann, G.; and Masri, N. 2023. Research note: Spreading hate on TikTok. *Studies in conflict & terrorism*, 46(5): 752–765.
- Wright, G.; and Buschschlütter, V. 2022. Peru crisis: Family of ousted leader Castillo leaves for Mexico — bbc.com.