CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17805

A multi-platform framework for nowcasting social phenomena: a case study for food insecurity

Bia Carneiro[1], Giuliano Resce[2], Giulia Tucci[3], Giosuè Ruscica[2], Nicola Caravaggio[2] Laura Fanelli[2], Agapito Emanuele Santangelo[2], Pietro Cruciata[2]

[1]Bioversity International, Italy, [2]Department of Economics, University of Molise, Italy, [3]International Center for Tropical Agriculture, Brazil.

How to cite: Carneiro, B.; Resce, G.; Tucci, G.; Ruscica, G.; Caravaggio, N.; Fanelli, L.; Santangelo, A. E.; Cruciata, P. 2024. A multi-platform framework for nowcasting social phenomena: a case study for food insecurity. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17805

Abstract

Given the growing significance of internet-based information flows, this research proposes a conceptual framework that integrates digital platforms to nowcast social phenomena, applied to the context of food security monitoring in the Global South. Building on the foundations of Digital Methods and online issue mapping, our research objective is to establish a multi-modal, multi-media model that monitors events from different perspectives to identify potential early warning signals arising from the data, ultimately informing policy actors and supporting early action. We apply three analytical processes: social listening, media monitoring and search interest analysis. Exploratory analysis on data from Zimbabwe point to the feasibility of the models applied to identify food security dimensions in text and search engine data. Further analysis is needed to interpret converging and diverging trends across the data streams, and their implications to food insecurity early warning.

Keywords: *digital platforms; digital methods; nowcasting; food security*

1. Introduction

Given the growing significance of internet-based information flows (Carneiro et al., 2022), this research proposes a conceptual framework that integrates digital platforms to nowcast social phenomena, applied to the context of food security monitoring in the Global South. The issue is pertinent because despite years of decrease, there is a current upsurge in food insecurity, and addressing this challenge requires comprehending the various factors contributing to it (Balashankar et al., 2023; Queiroz et al., 2021). However, while hazards to food security like climate shocks, conflicts, or market disturbances are extensively monitored through traditional quantitative and qualitative means, the application of digital platforms analysis can enhance localized knowledge about potential areas of concern.

2. A framework to nowcast food insecurity

Figure 1 presents a systematization of our proposed framework. Our research objective is to establish a multi-modal, multi-media model that monitors events from different perspectives to identify potential early warning signals arising from the data, ultimately informing policy actors and supporting early action. The framework relies on the foundations of the digital methods epistemology, which seeks to explain social phenomena through online dynamics (Rogers, 2013) and specifically on the approach of online issue mapping (Rogers, 2015). Online issue mapping brings forth a set of digital techniques for the detection, analysis and visualisation of topical affairs. It leverages on textual, visual and network analysis to understand how issue are formed in digital environments. Its interdisciplinary approach aims to answer questions such as 'is this topic an issue?', 'who are its actors?' 'what are its animating concerns?', and 'where are the issues happening (media, institutional locations, geography)?'.

It has been recognized that online spaces offer opportunities for analysing and visualising contemporary issue dynamics due to several aspects: issue traces are accessible online; the analytics leverage on the dynamic features and affordances of online media; and digital platforms provide data (for instance, metadata, links, hashtags, mentions, etc.) that can be structured for systematic analysis (Carneiro et al., 2022).

To achieve our aim, we apply three analytical processes. First, social listening and media monitoring aim to cover different dimensions of public discourse, at different timespans. While the social media dimension provides insights into the interests, concerns, and opinions of the general public, news media enables insights into "traditional" event coverage, with more established actors and discourses. Thirdly, as the Google search engine is the most visited website in the world, understanding what people are searching in it can reveal the level of interest in a particular topic. Effectively, the predictive potential of Google Trends supports social listening and media monitoring for nowcasting, as evidenced in the literature (Choi & Varian, 2012). The next section describes the final levels of the framework, pertaining to platforms and data sources.

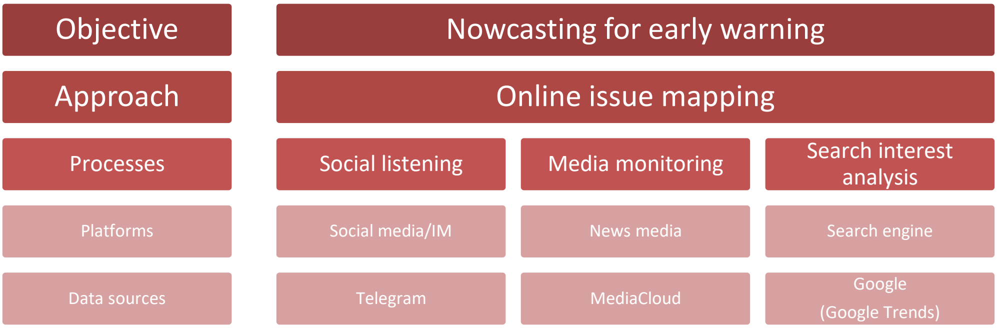

Figure 1. A multi-platform framework to monitor strategic issues

3. Data and methods

3.1 Data sources

For the social listening component, Telegram has been selected. Launched as a messaging app in 2013, Telegram is a currently a digital platform that allows users to create and subscribe to broadcasting channels and create and join discussion groups. It currently has 800 million active users on average, per month. Most importantly, Telegram provides an open API, which emables the extraction of publicly available data from channels and groups.

Media monitoring is performed through data provided by Media Cloud, an open source platform for media analysis aimed at academic research[1]. The platform monitors more than 60 thousand online media sources, with stories processed daily. It also contains curated country collections, enabling granularity in local news collection.

Lastly, search engine data is collected though Google Trends, adapting the daily economic sentiment index (DESI) approach proposed by Eichenauer et al. (2022).

3.2. Data collection

A dedicated software was developed to extract, filter, and visualize Telegram data available on the platform's API. The data collection begins with the creation of curated lists of relevant Telegram groups and channels, which are uploaded to the tool and their content can then be queried, exported and manipulated.

[1] https://www.mediacloud.org

To capture diverging perspectives, data collection covers content generated within our countries of interest (i.e. based on groups and channels from the country in question), as well as content that mentions the country in external or global groups.

News media is queried through Media Cloud's API following a similar structure, where national and local news sources from the countries in question are collected, as well as news stories from global sources that mention the country in the headline.

For Google Trends, the first step of data collection involved determining the main predictors of food insecurity. For this, we leveraged in prior work that used natural language processing to classify food security dimensions in publicly available reports by the USAID's Famine Early Warning Systems Network (FEWS NET). A LASSO regression was applied on the classification results from the NLP model to identify the predictors of food insecurity . The top ten positive features were selected, after excluding those disease-related, such as Covid-19 or Ebola, due to their time or geographic specificity. For each country of interest, a long-run frequency-consistent daily trend of food insecurity was constructed.

3.3. Data analysis

Each of our sources require analytical approaches that leverage on their affordances. Text mining models developed for topic classification and sentiment analysis of food security reporting by FEWS NET were adjusted and combined with visual analysis and machine learning models to monitor and identify the prevalence of food insecurity drivers from diverse perspectives and timescales. Determining the continuous associations and dynamics between drivers supports the identification of potential food insecurity hotspots.

Supervised text mining is applied to all textual data using the previously developed analytical framework and taxonomy for detection of food security-related topics. Leveraging on visual media disseminated on Telegram, the BLIP algorithm (Junnan Li et al., 2022) is applied to generate image captions that describe any images shared on posts. These captions are then combined with the text from the post body, and topic classification is applied. In addition, sentiment analysis is performed using the Syuzhet package (Jockers, 2015). TF-IDF analysis is also performed to identify significant and emerging terminology. Specifically for news media, existing Media Cloud algorithms for entity recognition are used to detect people, organizations, and geographic coverage of news stories.

For search interest, the final time series constructed for each country represents a synthetic search interest (SSI) index for food insecurity based on Google Trends data. In the final step of our procedure, weights were applied to allow for cross-country comparison. The first weight has been constructed by considering the worldwide Google Trends interest of each topic while the second one relied on yearly data of internet penetration (WB, 2024) (inverse) to compensate for the digital divide among countries.

4. Case study: Preliminary results from Zimbabwe

Zimbabwe was selected as the pilot country to test the application of this framework. The results presented here are descriptive in nature as the analysis is ongoing and further interpretation is needed.

670,562 news stories were collected through Media Cloud 2018-2023. Text mining was applied to the headlines. Figure 2 shows the most prevalent topics in news stories from Zimbabwe, and figure 3 presents the aggregated sentiment for these stories.

A list of Telegram groups and channels was curated, as well as a list of Africa-level groups and channels. Four lists of Zimbabwean national and local news sources were used for news data collection, as well as news stories from global English language news sources that mention Zimbabwe. In total, 113,279 posts were collected between 2017-2023. Figure 4 shows the most prevalent topics in groups and channels from Zimbabwe, and figure 5 presents the aggregated sentiment for these posts. The synthetic search interest index (SSI) for food insecurity in Zimbabwe is presented in figure 6.

Creating a nowcasting tool to detect food insecurity and food shocks requires the integration of the diverse data streams (as represented in Figures 2-6) into a unified analysis platform. Time-series analysis enables to track trends over time, employing models that can handle seasonality, trends, and irregular patterns to predict imminent shocks or stressors. Anomaly detection algorithms can be applied to identify sudden changes in the data that deviate from historical patterns, which could indicate emerging crises. Real-time changes in the data can be visualized through an integrative dashboard.

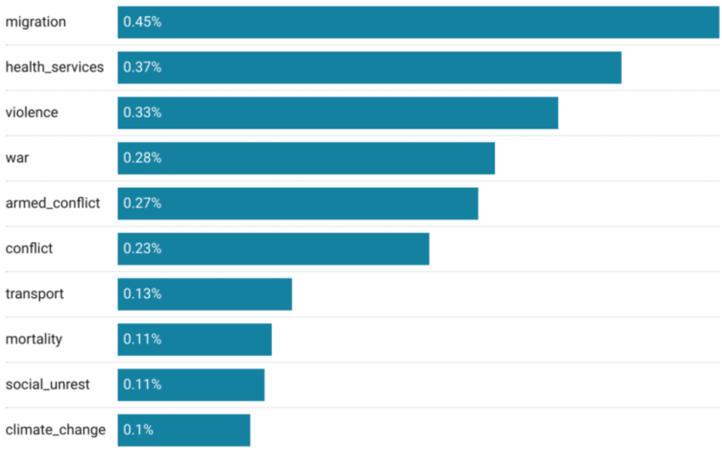

Figure 2. Average percent of headline words dedicated to each topic. Covid-19 has been excluded.

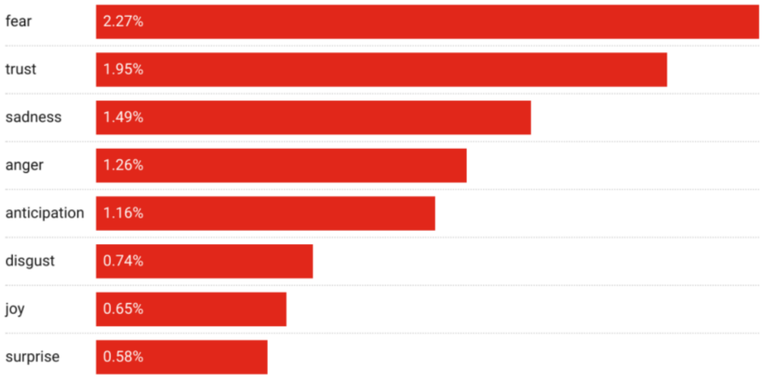

Figure 3. Percent of headlines that registered each emotion.

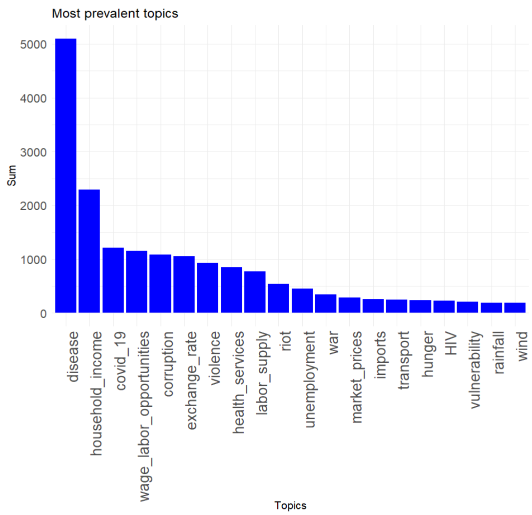

Figure 4. 20 most prevalent topics in Telegram groups and channels from Zimbabwe. Groups related to cryptocurrencies and "buy and sell" groups have been excluded.

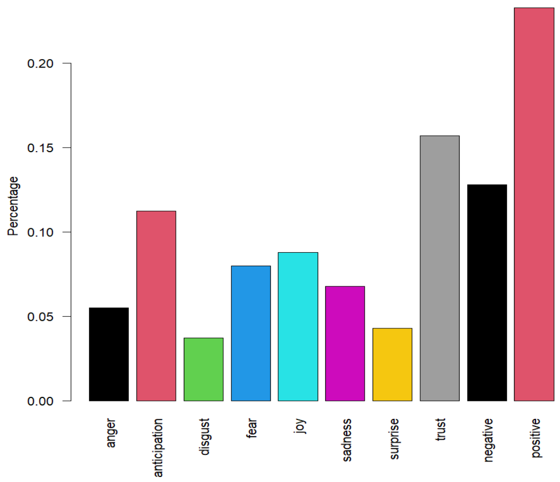

Figure 5. Aggregated sentiment in Telegram groups and channels from Zimbabwe. Groups related to cryptocurrencies and "buy and sell" groups have been excluded.

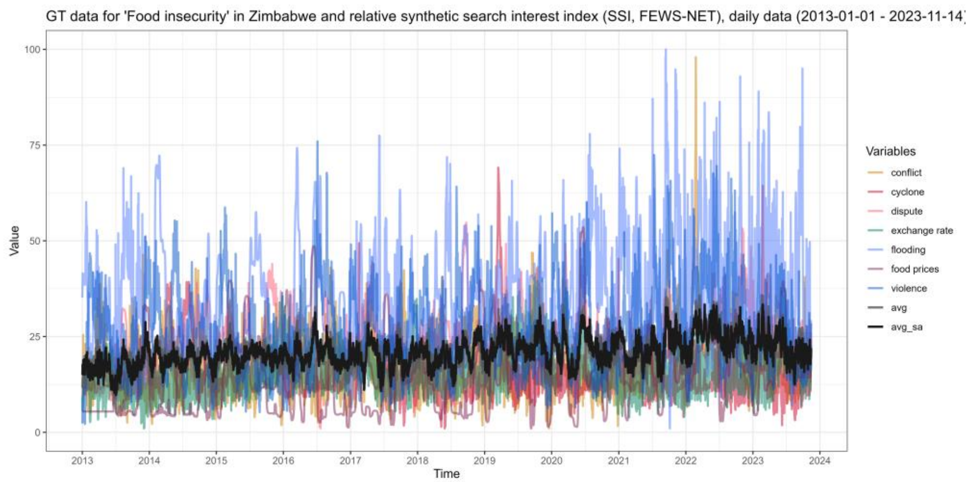

Figure 6. Synthetic search interest index (SSI) for food insecurity in Zimbabwe.

5. Conclusion

The aim of the multi-platform framework for nowcasting is to enable high frequency monitoring of potential food insecurity concerns at the country level, based on machine learning analysis of textual and image data from digitally native sources. Exploratory analysis on data from Zimbabwe point to the feasibility of the models applied to identify food security dimensions in text and search engine data. Further analysis is needed to interpret converging and diverging trends across the data streams, and their implications to food insecurity early warning.

Acknowledgements

This work was carried out with support from the CGIAR Initiative on Climate Resilience, ClimBeR, and the CGIAR Initiative on Fragility, Conflict, and Migration. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders/.

References

Balashankar, A., Subramanian, L., & Samuel P. Fraiberger. (2023). Predicting food crises using news streams. Science Advances, 9(eabm3449). https://doi.org/10.1126/sciadv.abm3449

Carneiro, B., Resce, G., & Sapkota, T. B. (2022). Digital artifacts reveal development and diffusion of climate research. Scientific Reports, 12(14146). https://doi.org/10.1038/s41598-022-17717-8

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. Economic Record, 88, 2–9.

Junnan Li, Dongxu Li, Caiming Xiong, & Steven Hoi. (2022). BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. ArXiv. https://doi.org/10.48550/arXiv.2201.12086

Jockers, M. L. (2015). Syuzhet: Extract Sentiment and Plot Arcs from Text. https://github.com/mjockers/syuzhet

Queiroz, C., Norström, A. V., Downing, A., Harmáčková, Z. V., De Coning, C., Adams, V., Bakarr, M., Baedeker, T., Chitate, A., & Gaffney, O. (2021). Investment in resilient food systems in the most vulnerable and fragile regions is critical. 2(8), 546–551.

Rogers, R. (2013). Digital Methods. The MIT Press.

Rogers, R. (2015). Digital Methods for Web Research. In R. A. Scott & S. M. Kosslyn, Emerging trends in the social and behavioral sciences: an interdisciplinary, searchable, and linkable resource.

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google Trends. Economic Inquiry, 60(2), 694–705.

WB. (2024). World Development Indicators. Washington, D.C., US. (World Bank. Retrieved from: https://databank.worldbank.org/source/world-development-indicators [Accessed February 1, 2024])