Data Mining Wikipedia Disasters and Emergencies Traffic

Or Elroy
Faculty of Computer Science
and Information Systems
College of Law and Business
Ramat Gan, Israel
elroyo@uw.edu

Abraham Yosipof†
Faculty of Computer Science
and Information Systems
College of Law and Business
Ramat Gan, Israel
avi.yosipof@gmail.com; aviyo@clb.ac.il

ABSTRACT

Wikipedia is a popular social network for sharing supported information, acting as de-facto open-encyclopedia, and as such, is a solid ground to inspect the perception and interest of different populations over time. Moreover, people look for information on social media in times of disasters and emergencies, and Wikipedia provides reliable information. This research investigated the traffic of seven representative disasters and emergencies events on Wikipedia over time. We analyzed the traffic with respect to different languages and used various statistical tests, including correlation and cross-correlation between the traffic of the same pages in different languages. In addition, the Granger causality test was used to determine whether the activity of one page can be used to forecast that of others. Finally, we used the ARIMA model to build a predictive time series model for each page to predict future lags.

CCS CONCEPTS

• Information Systems → Wikis

KEYWORDS

Emergences, Disaster, Wikipedia, CORE

ACM Reference format:

Or Elroy and Abraham Yosipof. 2022. Data Mining Wikipedia Disasters and Emergencies Traffic

1 INTRODUCTION

Social networks are increasingly popular and globally used and can be a sound source of information. Wikipedia is a popular social network, recently nominated as the fifth most visited website [1], for the sharing of generally accurate and supported information. Such information includes documentation of high-impact natural disasters and manmade incidents [2]. The actual traffic generated on certain events' pages on Wikipedia is a solid ground for the perception of different populations' preferences and areas of interest at different points in time. Wikipedia's use is primarily as an online, user-contributed data store, comprised of over 185 million encyclopedia pages in 325 different languages. The collaborative nature in which anyone can easily contribute to its articles, with or without sufficient or even minimal knowledge and understanding of a topic, has been key to Wikipedia's success but has also increased uncertainty in the reliability of the content published [3].

Some measures are in place to prevent malicious editing, spam included, of articles while preserving the delicate balance between being collaborative and open for all, but these are not perfect and more can be done [4]. Proper public review and providing references to allow the users to verify the content generally increase the level of confidence in the content of the article [5], but these are less available for articles that are not frequently visited or with lower interest, which makes them more susceptible. The ability to extract data regarding interest in certain pages brings an opportunity to investigate the interest of different populations in different topics, the change thereof over time, and the information flow between different communities that speak certain languages or live in different countries [6, 7]. Evaluation of emergencies and disasters' pages traffic on Wikipedia can provide valuable information and insights about the population. A sudden increase in interest in disasters and emergencies often indicates a recent evolvement or the unfolding of new details.

In this research, we collected the traffic data, including the number of views and edits for seven major natural disasters and emergencies, in several languages (see table 1). We investigated the data of each page over time. We evaluated the population's interest in each event over time and investigated the connection between different languages and different emergencies. A description of the seven events is provided in table 1. The events were chosen as representative [8] case studies of disasters and emergencies for the sCience & human factOr for Resilient sociEty (CORE), Horizon 2020 project (https://euproject-core.eu). The events took place between 2009 and 2021 within a relatively narrow time frame, except

† Corresponding author

for the covid-19 pandemic that began in late 2019 and has been ongoing.

Event	Country	Year
L'Aquila earthquake	Italy	2009
Manchester Arena bombing	UK	2017
Aude river flooding	France	2018
Visakhapatnam gas leak	India	2020
Tōhoku earthquake and tsunami	Japan	2011
Covid-19 pandemic	Global	2019
2021 European floods	Europe	2021

Table 1: The emergencies and disasters events.

2 RELATED WORKS

Analysis of Wikipedia's traffic data has proven to be productive for research and has been at the center of various research, each constructive in its way. Previous research used Wikipedia traffic for building a model for electoral prediction.[9] In the case of disasters and manmade incidents Kanhabua et al.,[2] analyzed long-term dynamics of Wikipedia as a global memory place for high-impact events.

Some research concentrated on tools and software dedicated to providing better accessibility to the data and assessing its reliability. Roy et al.,[10] investigated information asymmetry in Wikipedia by introducing the WikiCompare. The WikiCompare is a browser plugin that aims at the differences between pages on the same topic in different languages by providing readers with a comprehensive overview of topics by incorporating missing information from Wikipedia pages in other languages [10]. Vardi et al., [6] introduced WikiShark, an online tool that analyzes Wikipedia traffic and trends by extracting the data from the Wikipedia API.

2 DATA COLLECTION

We created a fast and simple yet powerful system to periodically reach for the respective Wikipedia API's endpoints to collect updated views and edits data for each event. The data collection is straightforward and ultimately consists of a few requests to Wikipedia's API. The data is then preprocessed to fit a relational database for local storage, and consistent time series are created for statistical analyses and predictive models.

To support future research, the system was built with expandability in mind. New events, or pages in other languages for existing ones, can be included easily. In addition, we intend to instantiate a free, publicly accessible website including the raw data to increase researcher collaboration in the future.

3 DATA SETS & DATA ANALYSIS

We collected the traffic data for the seven representative case studies in multiple languages. For each case study, we used traffic data of the case study's page in its official language, e.g., Italian for an earthquake in Italy, and usually English as well as a global language (see table 2). For covid-19, we used the data of five languages: Chinese, English, German, Italian and French.

Event	Languages
L'Aquila earthquake	Italian, English
Manchester Arena bombing	English
Aude river flooding	French
Visakhapatnam gas leak	English, Hindi
Tōhoku earthquake and tsunami	Japanese, English
Covid-19 pandemic	English, German, French, Italian, Chinese
2021 European floods	German, English

Table 2: The languages of the pages inspected per event.

The data can be sliced and displayed in infinite possible ways. This section shall provide some examples for peaks in interest and possible explanations for their cause interest.

The L'Aquila earthquake took place on April 6, 2009 in Italy, with a magnitude of 5.9 on the Richter scale, with thousands of foreshocks and aftershocks. 308 casualties have been reported, making this the deadliest earthquake in the region since 1980. Figure 1 displays the daily views' frequency for the L'Aquila earthquake page on Wikipedia in Italian. A major peak in interest was observed on August 24, 2016 (marked with a square, figure 1), following another earthquake in the same region at a magnitude of 6.2 on the moment magnitude scale.

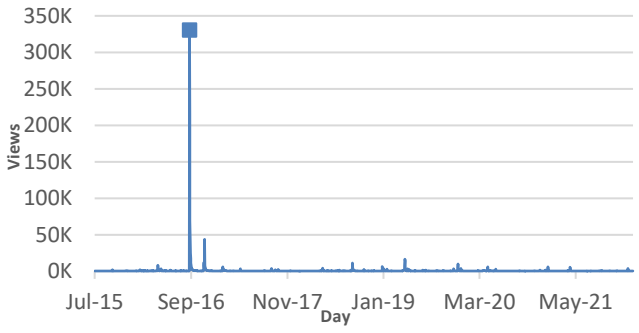

Figure 1: Daily views frequency for the L'Aquila earthquake page in Italian.

The Manchester Arena bombing is a terror attack that took place on May 22, 2017 in the United Kingdom, carried by an

Islamist suicide bomber. 23 casualties and over a thousand injuries have been reported, with hundreds more suffering from psychological trauma, making it the deadliest terror attack in the UK.

Figure 2 displays the monthly views frequency for the Manchester Arena bombing page on Wikipedia in English. A recurring, seasonal peak in the frequency can be observed in May of every year since 2017 (marked with squares, figure 2), the month the attack took place. The substantially larger peak of May 2018 (marked with a triangle, figure 2), may be attributed to increased media coverage of the event after precisely one year of its occurrence, with a series of media reports with updates about the victims and claims of misconduct of journalists when approaching victims' families [11].

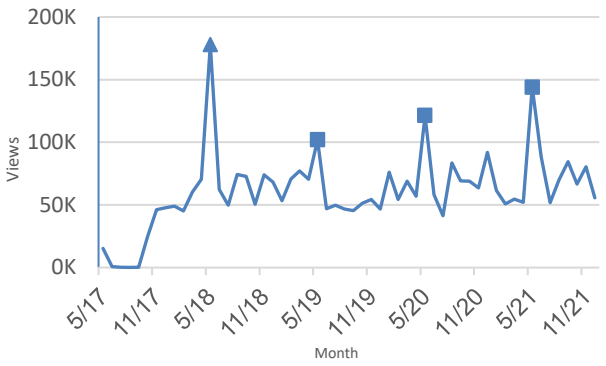

Figure 2: Monthly aggregated views frequency for the Manchester Arena bombing page in English.

The Aude River flooding took place in 2018 in France, caused by heavy thunderstorms leading to 7 meters rise in the height of the river. This was reported to have been the highest level of the river since 1891. At least 14 casualties have been reported.

The peak of October 22, 2019 (marked with a square, figure 3) is likely related to some 120 mm of rain falling in less than 3 hours, causing a river that runs through a city in southern France to rise by almost 4.3 meters by the next day, to a dangerous level [12].

January 22, 2020 has also seen a [13]substantial peak (marked with a triangle, figure 3), attributed to Storm Gloria that hit Spain and southern France between January 17, 2020 and January 25, 2020, causing multiple fatalities and damage. Forecasters have said the storm was the worst to hit the region during the winter period since 1982 [13]. The peak of September 8, 2021 (marked with circle, figure 3), is likely related to nearly two months' worth of rain that fell within just a few hours in southwestern France [14].

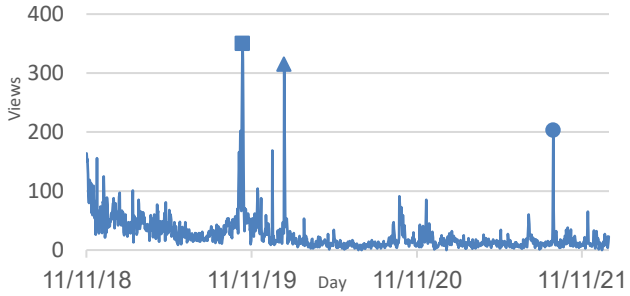

Figure 3: Daily views frequency for the Aude River flooding page in French.

The Visakhapatnam gas leak took place on May 7, 2020 in India, as a result of an industrial accident at a chemicals plant. 13 casualties have been reported, and over a thousand more reported to be sick after being exposed to the gas. The Visakhapatnam gas leak page in Hindi has seen steady level of interest with an average daily frequency of slightly over 27 views per day since the leak occurred. Figure 4 displays the daily views' frequency for the Visakhapatnam gas leak page in Hindi.

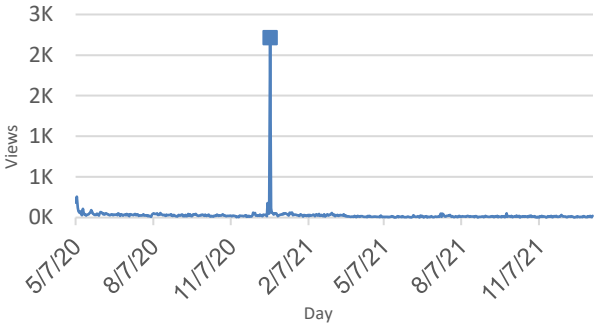

Figure 4: Daily views frequency for the Visakhapatnam gas leak page in Hindi.

An extraordinary peak was observed on December 23, 2020 (marked with square, figure 4), with over 2,200 views in a single day, which may be attributed to a second leak of Ammonia gas at the same place on that date. Media reports about the second leak described public concern and fear that it would be as harmful as the first gas leak [15]. The same peak is observed at the daily views' time series of the English page for the incident.

The Tohoku earthquake and tsunami took place on March 11, 2011 in Japan, with a magnitude of 9.1. A tsunami followed the six-minutes earthquake, along with over a dozen thousand aftershocks. Nearly 20 thousand casualties have been, over 6

thousand injuries and more than 2,500 missing have been reported. Figure 5 displays the daily views' frequency of the Tohoku earthquake and tsunami page in Japanese.

Peaks in interest are seen every year on the same day of the actual event, March 11 (marked with x, figure 5). Other peaks are likely to be associated with subsequent similar events. A couple of peaks in November 21 and November 23, 2016 (marked with squares, figure 5), may be related to another earthquake near Fukushima, Japan, with a magnitude of 6.9 [16]. The more significant peak of February 13, 2021 (marked with a triangle, figure 5), is likely related to a magnitude 7.1 earthquake near Fukushima, Japan, on that day [17].

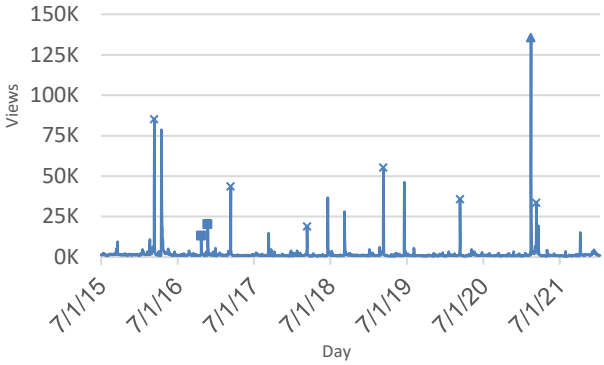

Figure 5: Daily views frequency for the Tohoku earthquake and tsunami page in Japanese.

The covid-19 pandemic is an ongoing pandemic started in caused by the SARS-CoV-2 virus that has been spreading globally. The pandemic is reported to have caused over 360 million confirmed cases. More than 5.6 million deaths have been reported, but the real amount is estimated between 13.4 and 22.7 million deaths. Figure 6 displays the daily views' frequency of the covid-19 pandemic English page.

An exceptional peak in traffic was observed on November 27, 2021 (marked with a square, figure 6), with roughly double the views of the days before and after that date. The peak can be attributed to a press release of the World Health Organization on November 28, 2021, stating that the Omicron variant was designated on November 26, 2021 with inherent uncertainty of the disease [18].

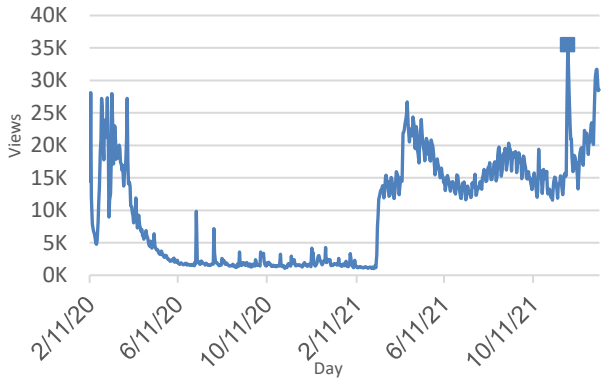

Figure 6: Daily views frequency for the covid-19 pandemic page in English.

The 2021 European floods occurred in July 2021, affecting several European countries throughout the continent. As a result, 242 casualties have been reported, and damages caused by the floods are estimated at a minimum of €10 million.

3.1 Time series: Granger Causality and Correlation

The Granger Causality test is a statistical model to test whether one time series may be useful for forecasting another time series in different lags [19]. In this work, we used the frequency of each of page's view count to create a daily frequency time series. We used the Granger causality model to determine if one time series (X) can be of use to forecast future behavior of other time series (Y). A p-value score of less than 0.05 is considered to be useful [20].

We applied the Granger causality test to determine the feasibility of predicting the views' frequency of any of the pages we inspected, using any of the other pages. Table 3 displays the significant results of the Granger causality test, except for the daily views' frequency of the Covid-19 page.

Time series X	Time series Y	P-value (lag 1)
EU floods (DE)	EU floods (EN)	<0.01
EU floods (DE)	Visakhapatnam (HI)	<0.01
EU floods (EN)	EU floods (DE)	<0.001
EU floods (EN)	Visakhapatnam (HI)	<0.01
L'Aquila (EN)	L'Aquila (IT)	<0.001
L'Aquila (IT)	Visakhapatnam (HI)	<0.001
L'Aquila (IT)	Visakhapatnam (EN)	<0.01
Manchester (EN)	Visakhapatnam (HI)	<0.05
Tohoku (JP)	Tohoku (EN)	<0.01
Visakhapatnam (HI)	Visakhapatnam (EN)	<0.05
Visakhapatnam (HI)	EU floods (DE)	<0.01
Visakhapatnam (HI)	EU floods (EN)	<0.001

Table 3: Significant results of the Granger causality test on daily views frequency time series at one lag.

The results suggest that views frequency of the English page for the EU floods, the German page for the EU floods, and the Hindi page for the Visakhapatnam gas leak, may all be useful for the prediction of each other. The Visakhapatnam gas leak page in Hindi may further be useful for predicting the English counterpart. The L'Aquila flooding page in English was found to be a significant predictor of the frequency of the Italian one. The Italian page may be useful for predicting the Visakhapatnam gas leak pages in both English and Hindi. The

Manchester Arena bombing page may be useful for the prediction of the Visakhapatnam page in Hindi. The Tohoku page in Japanese is possibly useful for the prediction of its English counterpart.

These indications of feasibility set the ground to use traffic time series of different types of disasters and emergencies to forecast others.

We further tested the correlation and cross-correlation between pages discussing the same events in different languages to support the Granger causality test results. The correlation between the EU floods page in German and English counterpart was found to be significant and strong (r=0.88). The cross-correlation results suggest that the pages are cross-correlated between lags -2 and +2. The strongest cross-correlation is between the German page at time *t* and the English page at time *t+1*, with a coefficient of 0.93. This result suggests that a higher views frequency on the German page at time *t* leads to a higher views frequency one day later (*t+1*) on the English page. This means that the English page views frequency follows the German page the next day

However, the cross-correlation between the English page at time *t* and the German page at time *t+1* was weaker, with a coefficient of 0.71. This result suggests the opposite of the previously stated result but with a lower cross-correlation coefficient.

The two-way cross-correlation supports the Granger causality test's results, according to which both pages are likely to be useable for the prediction of each other.

The correlation between the L'Aquila earthquake pages in Italian and in English was found to be strong and significant (r= 0.96). The cross correlation between these pages was found between lags -1 and 1, with low cross-correlation coefficients between 0.36 and 0.56.

The correlation between the Tohoku earthquake and tsunami pages in English and Japanese was found to be 0.59. The cross-correlation between these pages was found between lags -1 to +1, with low cross-correlation coefficients between 0.27 and 0.37.

The Visakhapatnam gas leak's pages in Hindi and English are weakly correlated, with a correlation coefficient of 0.27. In addition, no cross-correlation was found.

For the covid-19 pandemic's pages, we tested the Pearson correlation between the views' frequency time series of pages in multiple languages. Table 4 displays the Pearson correlation coefficients for the daily views' frequency time series of the covid-19 pandemic page on Wikipedia in five languages.

	DE	**EN**	**FR**	**IT**	**ZH**
DE	1				
EN	0.04	1			
FR	0.93**	0.17**	1		
IT	0.77**	0.11**	0.88**	1	
ZH	0.77**	0.25**	0.90**	0.81**	1

** p-value <0.01

Table 4: Pearson correlation coefficients between the covid-19 page in different languages.

The views' frequency time series of the covid-19 pandemic's pages in four out of five different languages, namely German (DE), French (FR), Italian (IT), and Chinese (ZH), are strongly correlated with each other. English (EN) has low and significant correlation with French, Italian, and Chinese and is not correlated with German (DE).

A possible explanation is that geographically distanced countries that speak the same language, e.g., the United Kingdom and Australia, were affected by the pandemic at different times, thus affecting the reliability of the results. Another explanation is that the information flow for the pandemic starts with English, and other languages come later.

Additionally, we investigated the cross-correlation between the English page and pages in other languages at different lags. The English page was not found to be cross-correlated at any reasonable lag with any of the other languages. Following the correlation and cross-correlation results for the traffic of the Covid-19 pages, we inspected the Granger causality results between all five pages. We hypothesized that the English covid-19 page could be used to predict other languages. Table 5 displays the Granger causality results where the covid-19 English page views in different lags is used as a predictor for covid-19 pages views in French, German, Italian and Chinese.

Time series Y	**Lag 1**	**Lag 2**	**Lag 3**	**Lag 4**
German	>0.05	>0.05	<0.05	<0.05
French	>0.05	>0.05	>0.05	<0.001
Italian	<0.001	<0.001	<0.001	<0.001
Chinese	>0.05	>0.05	<0.05	<0.05

Table 5: The p-value of the Granger causality test where the frequency of the Covid-19 English page may be used to predict that of other languages in different lags.

The Granger causality test's results indicate that the English page is, in fact, tightly related to the covid-19 pandemic's pages in other languages. The results were significant for Italian at lag 1, German and Chinese at lag 3, and French at lag 4. Ultimately the English page is not correlated with the other languages because it leads the way, and the others follow, some faster than others.

Future works may attempt to determine the cause for the correlation between Covid-19 pages in different languages and the lack of correlation between others.

3.2 Predictions: ARIMA

ARIMA is an Auto-Regressive Integrated Moving Average model that aims to forecast future lags in a time series. ARIMA model looks at the history of the time series and assumes that the history, with an error term, can be used to forecast future values. We used the ARIMA model to build a predictive time

series model for each daily views' frequency time series. We used AutoARIMA to find the optimal order for each model and made slight adjustments to achieve better results.

The model was trained on the first 80% of each event time series and tested on the last 20% of the same time series. The prediction focused on one future lag in every step, after which the actual observation was added to the model, repeatedly. We used the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) to evaluate the results of each model on the test data. Table 6 displays the order and statistical evaluation results for each model.

Event	Order	RMSE	MAPE
Aude River (FR)	1,1,2	15.130	>1
Covid-19 (DE)	2,1,1	235.63	0.11
Covid-19 (EN)	1,1,3	1910.9	0.82
Covid-19 (FR)	5,1,3	240.25	0.16
Covid-19 (IT)	0,1,0	200.33	0.10
Covid-19 (ZH)	2,1,3	218.45	0.11
EU floods (DE)	4,1,0	169.67	0.18
EU floods (EN)	0,1,0	84.890	0.16
L'Aquila (EN)	1,0,0	85.947	0.45
L'Aquila (IT)	1,1,2	446.70	0.28
Manchester (EN)	1,1,1	1720.3	0.20
Tohoku (EN)	1,1,2	2660.8	0.18
Tohoku (JP)	0,1,4	6441.8	0.55
Visakhapatnam (EN)	1,1,1	38.096	0.16
Visakhapatnam (HI)	0,1,1	6.2760	0.61

Table 6: Orders and statistical evaluation of the ARIMA models on test sets.

The ARIMA models provided mixed results. For example, the models for the L'Aquila earthquake in English and the Visakhapatnam gas leak in Hindi provided poor results with a MAPE of 0.45 and 0.61, respectively.

While the Visakhapatnam gas leak pages in English and the L'Aquila earthquake in Italian ARIMA models returned good predictions with a MAPE of 0.16 and 0.28, respectively.

The ARIMA models for the pages of the European floods provided good results for both the German and English pages, with a MAPE of 0.18 and 0.16, respectively. The ARIMA model for the Manchester Arena bombing behaved similarly, with a MAPE of 0.2.

The covid-19 pages, in all languages, except for English, provided excellent models with a MAPE score of 0.1 for Italian, 0.11 for German and Chinese, and 0.16 for French. The model for the English page, however, performed poorly with a MAPE of over 0.8.

The Aude River flooding had a very high MAPE, pointing at extremely bad results. However, since most of the predictions were not that bad, the statistical evaluation was likely thrown off by a couple of extremely high, sudden peaks in traffic for that page, combined with the fact that the time series is not long enough to compensate for those peaks. This resulted in a MAPE of more than 1. Figure 7 compares the predictions provided by the ARIMA model for the test data of the daily views' frequency time series of Aude River page in French, with the actual observations.

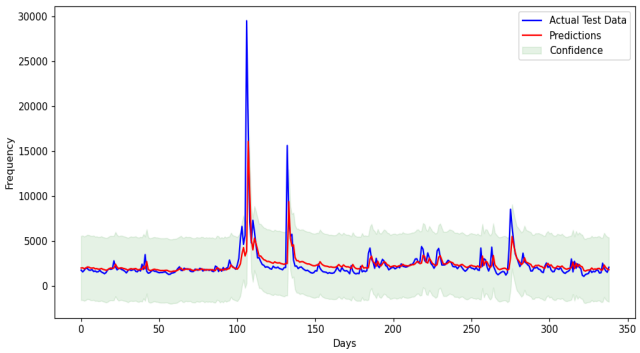

Figure 7: ARIMA prediction results for daily views' frequency compared to actual test data of the Aude River page in French.

The results for the Tohoku earthquake and tsunami pages were mixed. The Japanese page did poorly with a MAPE of 0.55. The English page did well, despite experiencing a high and sudden peak in interest similar to that which might have caused the statistical evaluation of the Aude River model to be thrown off. In this case, the time series was a few years longer, which may have compensated for the abnormalities and provided a MAPE of just 0.18. Figure 8 compares the predictions provided by the ARIMA model for the test data of the daily views' frequency time series of Tohoku page in English, with the actual observations.

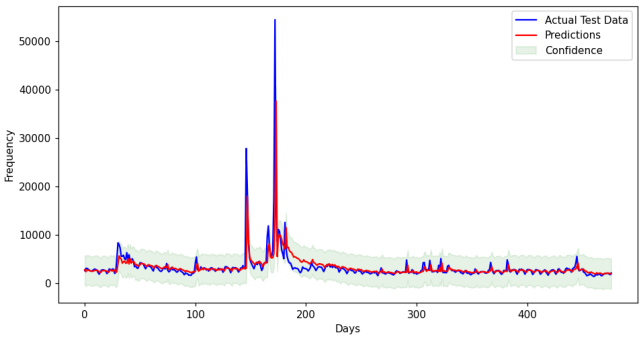

Figure 8: ARIMA prediction results for daily views' frequency compared to actual test data of the Tohoku page in English.

4 CONCLUSION

Wikipedia typically hosts pages relevant to multiple communities in multiple languages. With the assumption that users are more likely to view pages in their native language and

depending on their physical location, we collected traffic data of seven representative disasters and emergencies case studies on Wikipedia.

We analyzed the data to find abnormalities, such as peaks in interest. Then, we provided possible explanations that could assist future researchers in better understanding the unfolding of the public interest in emergencies and disasters.

We then tested the different time series with the Granger causality test for an indication of whether one time series may be useful for the prediction of another and received solid results. In addition, we tested the correlation between the different languages for the same event and further tested for cross-correlation at different lags. We found that cross-correlation differs between pages, where some are correlated, some are cross-correlated in different lags, and some are not cross-correlated.

Finally, we built predictive models for the events page views over time using the ARIMA model, providing good predictive time series models

Future works could use machine learning algorithms, such as Long Short-Term Memory, for predicting the views' frequency of one disaster or emergency by others.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101021746, CORE (science and human factor for resilient society).

REFERENCES

[1] Bracciale, L., Loreti, P., Detti, A. and Blefari Melazzi, N. Analysis of Data Persistence in Collaborative Content Creation Systems: The Wikipedia Case. *Information*, 10, 11 (2019), 330.

[2] Kanhabua, N., Nguyen, T. N. and Niederée, C. *What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia*. IEEE, City, 2014.

[3] Adler, B. T., Chatterjee, K., Alfaro, L. d., Faella, M., Pye, I. and Raman, V. Assigning trust to Wikipedia content. In *Proceedings of the Proceedings of the 4th International Symposium on Wikis* (Porto, Portugal, 2008). Association for Computing Machinery, [insert City of Publication],[insert 2008 of Publication].

[4] Green, T. and Spezzano, F. Spam Users Identification in Wikipedia Via Editing Behavior. *Proceedings of the International AAAI Conference on Web and Social Media*, 11, 1 (05/03 2017), 532-535.

[5] Lewoniewski, W., Węcel, K. and Abramowicz, W. *Analysis of references across Wikipedia languages*. Springer, City, 2017.

[6] Vardi, E., Muchnik, L., Conway, A. and Breakstone, M. *WikiShark: An Online Tool for Analyzing Wikipedia Traffic and Trends*. Association for Computing Machinery, City, 2021.

[7] Jung, C., Hong, I., Sáez-Trumper, D., Lee, D., Myung, J., Kim, D., Yun, J., Jung, W.-S. and Cha, M. *Information flow on COVID-19 over Wikipedia: A case study of 11 languages*. Association for Computing Machinery, City, 2021.

[8] Yosipof, A. and Senderowitz, H. Optimization of Molecular Representativeness. *Journal of Chemical Information and Modeling*, 54, 6 (2014/06/23 2014), 1567-1577.

[9] Yasseri, T. and Bright, J. Wikipedia traffic data and electoral prediction: towards theoretically informed models. *EPJ Data Science*, 5, 1 (2016/06/18 2016), 22.

[10] Roy, D., Bhatia, S. and Jain, P. Information asymmetry in Wikipedia across different languages: A statistical analysis. *Journal of the Association for Information Science and Technology* (2021).

[11] BBC News. 2018. Manchester Arena attack: Bomb 'injured more than 800'. Retrieved January 24, 2022 from https://www.bbc.com/news/uk-england-manchester-44129386

[12] Floodist. 2019. France – Major Flooding in South After 120mm of Rain in 3 Hours. Retrieved January 30, 2022 from https://floodlist.com/europe/france-south-flood-herault-october-2019

[13] The Watchers. 2020. Storm Gloria: Death toll rises to 13, over 1 500 evacuated in France after two months' worth of rain in 48 hours. Retrieved January 30, 2022 from https://watchers.news/2020/01/24/storm-gloria-spain-france-january-2020/

[14] Floodist. 2021. France – Record Rainfall Floods Streets in Lot-Et-Garonne. Retrieved from https://floodlist.com/europe/france-floods-agen-lotetgaronne-september-2021

[15] India TV News. Andhra Pradesh: Ammonia gas leak at private company in Visakhapatnam. Retrieved January 27, 2022 from https://www.indiatvnews.com/news/india/andhra-pradesh-ammonia-gas-leak-at-visakhapatnam-private-company-706636

[16] United States Geological Survey. 2016. Manitude 6.9 Earthquake off Japan. Retrieved January 30, 2022 from https://www.usgs.gov/news/featured-story/magnitude-69-earthquake-japan

[17] Wikipedia. 2022. 2021 Fukushima earthquake. Retrieved January 30, 2022 from https://en.wikipedia.org/wiki/2021_Fukushima_earthquake

[18] World Health Organization. 2021. Update on Omicron. Retrieved January 24, 2022 from https://www.who.int/news/item/28-11-2021-update-on-omicronRetrieved

[19] Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424-438.

[20] Wilms, I., Gelper, S. and Croux, C. The predictive power of the business and bank sentiment of firms. *European Journal of Operational Research*, 254, 1 (2016-10-01 2016), 138 - 147.