Understanding the Rise of Twitter-based cyberbullying due to COVID-19 through comprehensive statistical evaluation

Sayar Karmakar
Univesity of Florida
Gainesville (FL)
sayarkarmakar@ufl.edu

Sanchari Das
University of Denver & Indiana University
Denver (CO), Bloomington (IN)
Sanchari.Das@du.edu

Abstract

The COVID-19 pandemic has created a challenging situation for everyone, sparking digital evolution due to stay-at-home restrictions to stop the spread. This has led to an uprise of digital presence, which many hypothesize has lead to a rise of cybersecurity attacks, including cyberbullying. To evaluate the significance of COVID-19 on cyberbullying reports, we collected $454,046$ *of publicly available tweets from Twitter by using MongoDB and Python libraries from January 1st, 2020–June 7th, 2020. We performed statistical analyses on the collected sample set to understand the situation from a quantitative perspective. We extracted tweets related to 27 unique keywords specific to cyberbullying, including online bullying, cyberbullying, Twitter bullying, and others. Due to the time-series' count nature, we propose a Bayesian estimation of this count data trends utilizing an autoregressive Poisson model. A simple change-point model fails to explain the subtle changes adequately. On the other hand, our Bayesian method clearly shows the upward trend beginning in mid-March, which is reportedly the time from which the stay-at-home orders were widespread globally. The pattern remains similar if we focus on one or more such keywords instead of the total count. We also provide a fine-grained lag based analysis of our model and contrast our methods with an alternative semi-Bayesian AR-ARCH model. Overall, such analysis shows somewhat conclusive evidence of the rise around the same time as COVID.*

1. Introduction

In this day and age, technology has become interwoven into nearly every aspect of daily life. One example of technology's push into most people's lives is the rise of social media. Social media seems to have eclipsed many aspects of human interaction, such as dating, networking, and sharing information. However, just as all in-person social interactions are not positive, neither are all social media interactions. Such digital evolution has also given rise to several cybersecurity attacks, including cyberbullying due to the ascent of online communication [1, 2]. The dynamics of face to face bullying have transcended beyond physical boundaries to the digital realm, targeting users despite geographical constraints [3, 4].

On social networking sites and applications cyberbullying is very prevalent, with 66% of cyberbullying episodes originating on these platforms [1]. Twitter gives individuals the opportunity to ever so often interact with outsiders (counting celebrities) [5]; however, this also leads to others mirroring and forging individuals online identities to trick users [6]. Verifying profiles only is successful for celebrities or people well-known within their field, making it difficult to accurately verify an individual's identity [7]. Additionally, there are bots which duplicate regular peoples profiles to engage in malicious activities such as follower fraud [8]. It can be especially challenging to recognize abusers when they are pretending to be someone they aren't. On top of this, the cyberbullying that takes place over Twitter's public platform has become very prevalent and a concern for many individuals [9].

Besides, given the current COVID-19 pandemic, many have expanded their use of social media to remain in contact with others while social distancing [10]. Regardless, there have also been reports of incivility originating from such platforms [11]. An sudden ascent in internet-based life - combined with children and young adults continually utilizing such stages - could lead to a concerning spike in cyberbullying [2]. Along these lines, we explicitly needed to see whether that is the situation, and answer the accompanying examination question:

RQ: How does a crisis, for example, a worldwide pandemic (COVID-19), impact cyberbullying?

Despite efforts on the technological side to reduce online harassment and bullying, little work has focused on cyberbullying from the user perspective. Thus, it is critical to understand the users' perspective on how COVID-19 may or may not have impacted cyberbullying through trend analysis. To better understand users' perspectives and COVID-19's impact

[1] https://www.pewresearch.org/internet/2014/10/22/part-2-the-online-environment/

[2] https://www.digitaltrends.com/news/coronavirus-cyberbullying-separationlearning/

URI: https://hdl.handle.net/10125/70923
978-0-9981331-4-0

on cyberbullying, we aggregated $454,046$ publicly available tweets related to cyberbullying to better understand the user experience online. As we hypothesized, there has been a significant increase in cyberbullying since the start of the pandemic, which showcases the impact of the crisis on cyberbullying trends. Our analysis provides details about how critical situations impact digital interactions, and provides details of the users' perceptions during such times. Our implications offer tangible solutions in helping the users to mitigate such threats online by analyzing the types of risks faced by online users every day.

In this paper we focus on first 5 months of 2020 with a goal to analyze if the later months showed more frequency of cyberbullying related keywords on twitter as compared to the first two months. We collected data containing certain keywords and then categorize them into daily counts. We first show that a simple visual analysis fails to divulge a lot about any particular change that has happened due to the pandemic. Reflecting upon the need of a more subtle analysis that takes care of dependence from time series and the obstacle of having a relatively small sample , we settled for a somewhat new method of Bayesian time-varying model from [12]. This helped us produce a clear increasing trend from Mid-March. It is Bayesian in the sense that we assume the parameters to be varying over time and assumed appropriate priors on them. This allowed us to construct posterior samples of these parameter function of time. Such flexibility in constructing posterior is very important for statistics inference literature as it gives some quantification of uncertainty. We also provide similar analysis for sub categories by combining a subset of keywords. This is to the best of our knowledge the very first such quantitative trend analysis on this type of data and also reveals a clear telling and negative effect of COVID-19 on cyberbullying. After discussing the impact of cyberbullying and some related works in Section 2, we discuss data collection, visual summary and a naive visual change-point analysis in Section 3. The prior change-point analysis as discussed in our preliminary work [13] summarizes tweets by their keywords and condenses them into their daily counts. We next motivate the Bayesian model and briefly discuss the posterior computation in Section 4. This Bayesian model details a clear upward trend in cyberbullying since the inception of COVID-19 as detailed in our earlier work as well [14]. The daily counts are analyzed in Section 5. We conclude by touching upon the main finds of this paper and some general discussions in Section 6.

2. Related Work

Cyberbullying is a major concern to the digital communication with various people being mentally impacted which can lead to several critical consequences. Cyberbullying has increased considerably given the advent of social media and billions of users being online everyday [4]. Additionally, because times of crisis can increase users' online presence and, as a result, cyberbullying, it is important to consider human factors to protect users during such situations– especially with the current pandemic situation [15].

2.1. Impacts of Cyberbullying

Studies in the past have explored the effect of cyberbullying on targets, specifically on teenagers; sometimes cyberaggressors and cybervictims can both be impacted by such abuses. Bonanno and Hymel found that both victims and perpetrators of cyberbullying were more likely to develop depression and suicidal thoughts compared to those involved in other types of bullying [16]. Dredge et al. describeded the critical effects of cyberbullying as it related to the targets social and emotional lives, with the impacts severity relying on different factors. Such factors include the perpetrators anonymity and whether or not bystanders were present [17]. Similarly, Wisniewski et al. noted that lower online risk can help in the developmental stages of the teens, while developing and enhancing crucial interpersonal skills, such as boundary setting, conflict resolution, and empathy [18]. In addition to the mental impact, Šléglová and Cerna found that cyberbullying led to behavioral changes, with victims displaying more cautious browsing habits and avoidance strategies [19]. McHugh et al. noted the negative emotions victims of cyberbullying experience, though they also found that the impact may be more short-term than previously thought, emphasizing the importance of resilience [20].

2.2. Social Media Bullying

Cyberbullying takes place across various online platforms. such platforms include social networks, chat rooms, and mobile messaging applications. Regardless of geographic proximity; such bullying varies in the amount of time it occurs, from as short as a week to much longer [21]. Due to social networking platforms often being used to compare oneself to others, they can lead to the development of self-esteem issues [22]. Recently, major social media platforms, including Twitter and Facebook, have been the location for several impactful cyberbullying events. In May 2020, a Japanese reality TV star took her own life after being subject to abuse on social media [3] Similar incidents across the world have led lawmakers to pass legislation that would make cyberbullying criminal [23].

To help mitigate cyberbullying's effects, prior studies have focused on the improvement of social media policies to protect victim's from their abusers. Milosevic looked into social media companies

[3] https://www.bbc.com/news/world-asia-52782235

responsibility regarding cyberbullying amongst children [24]. They mention concerns on the transparency and accountability of these platforms in addressing and mitigating such issues. Thus, it is critical to understand the various defensive mechanisms against cyberbullying from the technical, organizational, and user perspectives.

2.3. Cyberbullying Trend Analysis

Studies which involve cyberbullying trend analysis are important for understanding digital users are impacted by events.By conducting four separate surveys across 17 high schools, Schneider et al. found an increase in the overall rate of cyberbullying from 2006 to 2012 [25]. Snell and Englander conducted a survey to study the experiences of the cyberbullying victims. They found that it is more likely for females to be both victims and perpetrators when they are involved in cyberbullying, indicating that gender plays an important factor in reducing online bullying [26]. Mangaonkar et al. analyzed tweets utilizing a distributed design to detect cyberbullying in as it occurs [27].

On Twitter users can express themselves using 280 character long 'tweets;' , such messages have been analyzed in prior studies for cyberbullying [28, 29]. In terms of trend analysis of cyberbullying related tweets as an impact of some real-world events Cortis and Handschuh analyzed bullying tweets in the context of two trending events (the Ebola outbreak and shooting of Michael Brown in Ferguson, Missouri) [30]. They found some common hashtags which were utlizied in our study to collect and analyze the data. Through these discussions they attempted to identify cyberbullies, but the impact of such crisis situations regarding cyberbullying was not studied. Due to individuals' online digital presence recently increasing, some people may assume that the pandemic situation caused by COVID-19 has resulted in an increase of cyberbullying attacks. Hence, the goal of this study was to find conclusive evidence that either supports or contradicts this hypothesis.

2.4. Cyberbullying- User Studies

For the improvement of anti-harassment measures, prior research has investigated cyberbullies motivations. Lee and Kim conducted an interview of 110 subjects to better understand what drives social media users to post malicious or benevolent comments [31]. They found that malicious comments are often made to "resolve feelings of dissatisfaction" and identified anonymity and lack of real-world consequences as primary drivers for online hostility. Whittaker and Kowalski discovered that cyber aggression was very prevalent in online comment sections and forum replies, even more so than on Facebook, reiterating how important anonymity is to cyberbullies [32]. Lai and Tsai collected 224 survey responses and found that users who were either victims of or witnesses to previous cyberbullying campaigns were more likely to become perpetrators themselves [33]. Also noteworthy, respondents lacked confidence in the reporting systems of social networking platforms, with only 36.36% of them saying they believed the site had taken appropriate measures to help them.

Likewise, cyberbullying has also been studied from the victims' perspective. Whittaker and Kowalski found that the categorization of an incident of cyberbullying partly depends on the relationship between the victim and the perpetrator; aggressive comments directed at peers were seen as less acceptable than those directed at random people [32]. DiFranzo et al. studied the effect of improving site interface design on bystander intervention, which is critical to reducing and deescalating cyberbullying [34]. McNally et al. found that children acknowledged their online safety needs and were accepting of certain parental controls on their mobile applications [35]; along the same lines, Singh et al. found that students hold application designers responsible for the consequences of abuse conducted on their platform [36].

These prior studies helped us build a concrete foundation to understand cyberbullying both from the perpetrators' and victims' perspectives. It is also critical to understand how users discuss cyberbullying on social media at a large scale. Thus, we collected publicly available tweets about cyberbullying. Additionally, crisis situations can lead to harmful impacts; thus, analyzing the impact of cyberbullying was a goal for this study.

3. Data collection and summary

Twitter allows users to express themselves in 280 character 'tweets (Given our analysis and paper writing time till October 2020);' in prior studies such messages have been analyzed for cyberbullying [28, 29]. Cortis and Handschuh analyzed bullying tweets in the context of two trending events (the Ebola outbreak and shooting of Michael Brown in Ferguson, Missouri) and identified commonly used hashtags and named entitites in bullying tweets [30]. Due to an increase in individuals' online digital presence, assumptions have been made that the pandemic situation from COVID-19 has resulted in an increase cyberbullying related attacks. Comprised of over 300 million active daily users, Twitter [4] is the optimal data source for our research [5] in order to assess COVID-19's impact on cyberbullying, we comprised a total of $454,046$ public tweets that mentioned cyberbullying related keywords.

The data was obtained utilizing Get Old Tweets API [6], which enabled us access to older tweets. We

[4] https://twitter.com/login

[5] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

[6] https://github.com/Jefferson-Henrique/GetOldTweets-python

implemented an API for webcrawling and obtaining more data than what Tweepy [7] or twitteR [8] provides, which limits you to 200 tweets per keyword or limits the data collection time as well. For the storage and analysis of the data we used MongoDB. We collected our data from tweets made between January 1st 2020–June 7th 2020 to evaluate the impact of COVID-19 on these reported tweets.

The following key terms were used when we were conducting our search: *Internet bullying, Internet bully, Internet bullies, online abuse, online harassment, online shaming, online stalking, cyberbullying, social media bullying, stop cyberbullying, cyberbully, cyberbullies, FB bullying, FB cyberbullying, FB harassment, FB victim, Facebook bullying, Facebook cyberbullying, Facebook victim, Facebook harassment, Twitter bullying, Twitter cyberbullying, Twitter harassment, Twitter victim, Insta bullying, Insta cyberbullying, Insta harassment, Insta victim*. We only collected direct tweets and removed any retweets or duplicate tweets.

Some examples of the collected tweets include:

> Dear X[Twitter handle of a Twitter User] Not everything that happens in A[country name] is political. Y[Name of a Person] is a brilliant young woman who made the right decisions that helped to stop the spread of Covid 19. STOP cyber-bullying her.
>
> Online predators target kids in COVID19 lockdown http://www.dailytelegraph.com.au/ coronavirus/news-story eSafety ”recorded a 40% increase in reports over the past 3 weeks ... image-based abuse have increased by about 86% cyberbullying has risen 21%” - says @eSafetyOffice Investigations Manager,Toby Dagg
>
> Increased free time, greater online access and anxiety over covid-19 anxiety is leading to more cases online bullying globally. How can cyber-bullying be contained and what do you do if you get bullied?...
>
> I hope B[country name] stop bullying here. Z[Twitter Handle of a Twitter User] You guys are doing cyberbullying. This can be a crime. W[Name of an Individual] is a victim of COVID19 virus, too. You guys do not know exactly where he was COVID19 infected. Everything is based on guess only.

3.1. Summary

After collecting all of our data, we performed a comprehensive trend analysis based around evaluating the impact a crisis situation, such as the COVID-19 pandemic, can have on cyberbullying. Using the posts timestamp, we gathered a daily count of tweets which included, at minimum, one of our keywords. Figure 1 showcases the daily count for the 159 days beginning at 01^{st} January, 2020 and concluding at 07^{th} June, 2020.

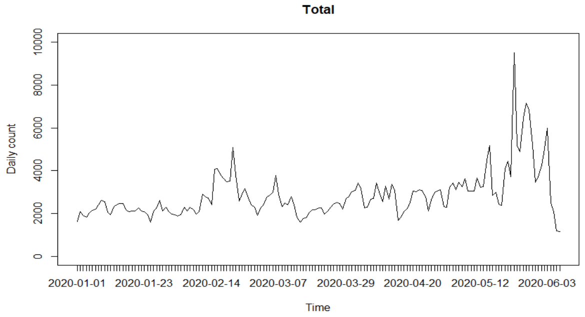


Figure 1. Daily count of total tweets related to bullying

There were 46 different keywords including, cyberbullying, online harassment, etc (mentioned above). In addition to these keywords, we also included keywords with and without spaces thus making 27 different types of keywords, for example cyber bullies/cyberbullies. Some of the types of keywords had fewer tweets with negligible impact on the analysis. Thus, we broadly divide them into 3 sub-classes. The daily count distribution for these sub-classes is shown in Figure 2.

Name	Contains	#keywords	#tweets
CY	”Cyber”	7	235,542
ON	”Online”	6	96,629
TW	”Twitter”	3	96,147

Table 1. Description of 3 sub-classes

3.2. Some naive visual analysis for a change-point

Looking at Figure 1 and 2, we see a clear pattern relating to all the counts and sub-classes presented here. Overall, excluding sub-class ‘ON’ , there does not appear to be a significant change in the mean except the slight upward trend beginning in mid-March seen in all categories including the total. In the second half of may, We noticed a considerable spike in tweets related to cyberbullying. Here to explore the rise in the tweets about cyberbullying in May, we looked intot the tweets which were shared and we found that mostly people discussed about the unfortunate demise of a Japanese TV star who was impacted by cyberbullying [9].

[7] https://www.tweepy.org/

[8] https://cran.r-project.org/web/packages/twitteR/README.html

[9] https://www.japantimes.co.jp/news/2020/05/30/national/media-national/online-harrassment/

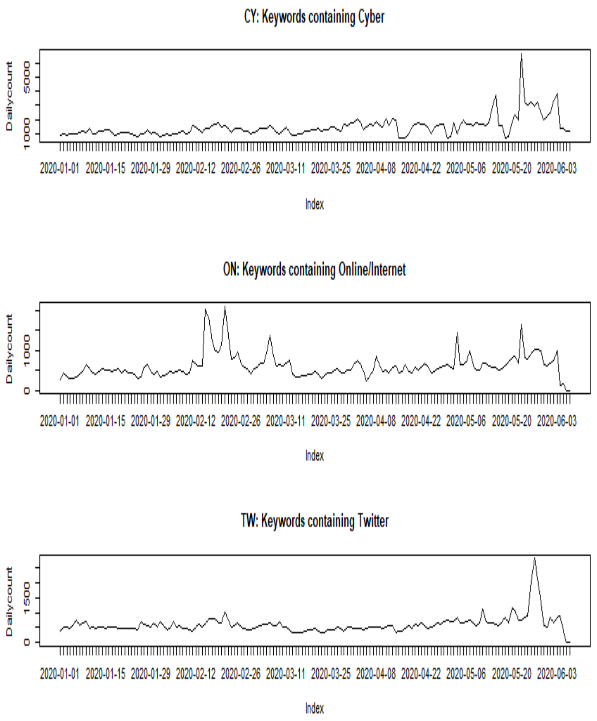


Figure 2. Daily count of total tweets for the three sub-class

> The horror of this Cyber Bullying Young Hana Kimura is without a question a tradegy. My thoughts and prayers are with Her and her family. As well as everyone who knew her personally.

Moreover for the class ‘ON’ , one can observe a significant spike in the later half of February, and overall the mean had an upward trend. It is possible that this may not be a result of the pandemic. Note that excluding the spike in the second half of May, there is not an abrupt break caused by COVID-19.

We decided from the visual summary that such a simple model is simply not enough to capture the complicated dynamics of how the daily counts were affected. In the next section we discuss some of the key points this simple model might be missing and introduce a novel Bayesian perspective.

4. A Comprehensive Bayesian Model

We start with the following set of motivations for choosing a Bayesian Model for our statistical analysis of the dataset:

4.1. Motivation

Small sample size: As is well-known in the frequentist time-varying literature, one needs a large sample (Sample size at least 500) to estimate (using proper kernels) any time-trend with precision. Such a short-coming can be overcome by a Bayesian method where one can compensate the lack of sample size by simulating from the posteriors.

Poisson count time series evolves differently: Also note that the daily number of occurrences is a count series, but unfortunately the traditional change-point analysis often assumes normality (Normal distribution). Another advantage of using Poisson random variable is it can model the mean and variance through a single parameter.

Abrupt changepoint models are insufficient: Before proposing a smoothly -varying coefficient model, we refer to a prior preliminary work [13] by the authors where they address a simple change-point model. However, since the crowd on Twitter is heterogeneous, one can expect the change might not be abrupt which can explain why just from the daily count summaries of the total tweets or the sub-classes do not reveal any abrupt change in either mean or variance in general. A more meaningful model could be where the parameters change smoothly over time and we can estimate these parameters as a function of time and see whether the trend is increasing due to COVID-19 or not.

Dependence cannot be ignored: Note that daily count of number of ’bad’ tweets is a time-series and the total count, th subseries and the individual counts all show significant autocorrelation. Any visual analysis of change-point would heavily disregard the inherent dependence assumption that is present in a time-series. These counts depend heavily on current trend and are expected to show strong correlation with recent pasts. We decide to furnish this through the auocorrelation plots in the following Figure 3.

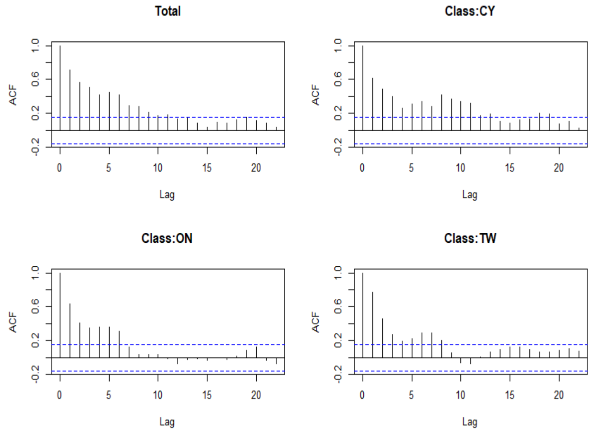


Figure 3. ACF plots of Daily count of total tweets for the three sub-classes

Under such heavy dependence for the total count and the three sub-series as mentioned above, one needs to take the dependence of count per tweet given the sub-series into account. Otherwise any analysis, be it abrupt change point or a smooth time-varying parameter model will not be justified.

In order to address these issues we discuss the following time-varying Bayesian autoregressive count (TVBARC) model from [12].

4.2. Bayesian Model

The possibly non-stationary (over time) nature of the data (See Figure 1), motivates us to propose a time-varying version of the linear Poisson autoregressive model [37, 38]. Since we are collecting the number of tweets here, we preferred using a count random variable like Poisson distribution. The autoregression part of our model is motivated from the fact that these counts are a time-series that, as shown in Figure 3, shows significant correlation. Note that time-constant models assume the parameters do not change over time, whereas our modelling is providing a non-stationary model that maintains the same data generating process while allowing the parameters to vary over time. For a comprehensive overview on usefulness and applicability of time-varying models see [39]. The conditional distribution for a count-valued time-series X_t given $\mathcal{F}_{t-1} = \{X_i : i \leq (t-1)\}$ is,

$$X_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \tag{1}$$

$$\text{where } \lambda_t = \mu(t/T) + \sum_{i=1}^{p} a_i(t/T)X_{t-i} \tag{2}$$

In the above section, $\mu(t/T)$ stands for the overall mean trend at time t whereas $a_i(t/T)$ denotes the effect of i-th lag at time t. The rescaling of time-scale from t to t/T is standard in in-filled asymptotics literature for meaningful modelling. Due to the Poisson link in (2), both conditional mean and conditional variance depend on prior observations. The conditional expectation of X_t in the above model (2) is $E(X_t|\mathcal{F}_{t-1}) = \mu(t/T) + \sum_{i=1}^{p} a_i(t/T)X_{t-i}$, which is positive-valued. Additionally, we imposed the following constraints on parameter space for the time-varying parameters,

$$\mathcal{P}_1 = \{\mu, a_i : \mu(x) > 0, 0 \leq a_i(x) \leq 1, \tag{3}$$

$$\sup_x \sum_k a_k(x) < 1\}.$$

When $p = 0$, our proposed model reduces to a routinely used nonparametric independent Poisson regression model as in [40]. The $\mu(\cdot)$ function corresponds to the general mean trend at time t and $a_p(\cdot)$, the $p - th$ order autoregressive (AR hereafter) coefficient function denotes how the observation at time t is affected by a past observation at lag p. The strong correlation pattern in Figure 3 shows we should opt for a $p > 0$.

4.3. Posterior Computation

In a time-varying Bayesian world the coefficients are random variable functions and thus we put priors on the unknown functions $\mu(\cdot)$ and $a_i(\cdot)$'s such that they are supported in the restriction $\mathcal{P}_1$.

The complete likelihood L corresponding to (2) is given by

$$L_1 \propto \exp\Bigg(\sum_{t=p}^{T}\big[-\{\mu(t/T) + \sum_{i=1}^{p} a_i(t/T)X_{t-i}\} \tag{4}$$

$$+ X_t \log\big\{\mu(t/T) + \sum_{i=1}^{p} a_i(t/T)X_{t-i}\}\big] \tag{5}$$

$$- \sum_{j=1}^{K_1} \beta_j^2/(2c_2) - \sum_{l=0}^{p} \delta_l^2/(2c_1)\Bigg)\mathbf{1}_{0\leq\theta_{ij}\leq 1}, \tag{6}$$

where

$$\mu(x) = \sum_{j=1}^{K_1} \exp(\beta_j)B_j(x), \tag{7}$$

$$a_i(x) = \sum_{j=1}^{K_2} \theta_{ij}M_iB_j(x), 0 \leq \theta_{ij} \leq 1, \tag{8}$$

$$M_i = \frac{\exp(\delta_i)}{\sum_{k=0}^{p}\exp(\delta_k)}, \quad i = 1, \ldots, p, \tag{9}$$

$$\delta_l \sim N(0, c_1), \text{ for } 0 \leq l \leq p, \tag{10}$$

$$\beta_j \sim N(0, c_2) \text{ for } 1 \leq j \leq K_1, \tag{11}$$

$$\theta_{ij} \sim U(0, 1) \text{ for } 1 \leq i \leq p, 1 \leq j \leq K_2. \tag{12}$$

Here B_j's are the B-spline basis functions and δ_j's are unbounded.

The prior distributions on these functions are induced through basis expansions in B-splines with suitable constraints on the coefficients to impose the shape constraints as in $\mathcal{P}$. The verification of the fact that priors induced by above construction are $\mathcal{P}$-supported is straight-forward. In above construction, $\sum_{j=0}^{P} M_j = 1$. Thus $\sum_{j=1}^{P} M_j \leq 1$. Since $0 \leq \theta_{ij} \leq 1$, $\sup_x a_i(x) \leq M_i$. Thus $\sup_x \sum_{i=1}^{P} a_i(x) \leq \sum_{i=1}^{P} M_i \leq 1$. We have $\sum_{j=1}^{P} M_j \leq 1$ if and only if $\delta_0 = -\infty$, which has the probability zero. On the other hand, we also have $\mu(\cdot) \geq 0$ as we have $\exp(\beta_j) \geq 0$. Thus, the induced priors, described above are well supported in $\mathcal{P}$.

This enables us to develop an efficient MCMC algorithm to sample the parameter β, θ and δ from the above likelihood. See [12] for the computation of the partial derivatives and [41] for a theoretical justification of why such methods of posterior computation can guarantee contraction to the true parametric function.

5. Analysis of daily count data

5.1. Analyzing total counts using our Poisson count model

We first analyze the total count trends through two different choices of lag p: an AR(1) and an AR(10) model. Often there could be weekly patterns which could mean high correlation at lag 7 and also lag 8 if lag 1 was significant. To see if there is really a weekly pattern we decided to take a lag that is slightly higher than 7. The trend functions with their corresponding credible intervals (we omit the credible intervals for 10 AR coefficients for clarity) are shown in Figure 4, 5 and 6. The trend and the credible intervals are mean and quantiles of 20000 posterior MCMC samples after 10000 burn-in.

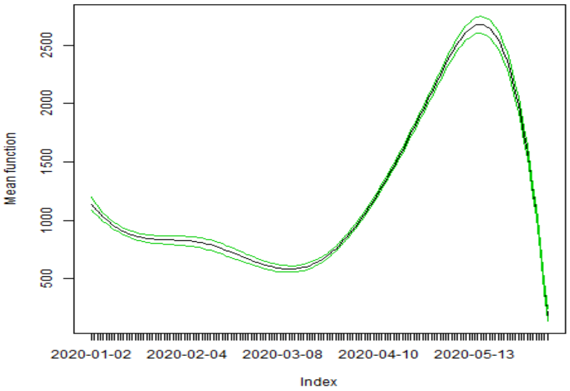


Figure 4. Total count: Mean trend of AR(1) model

When we increase the number of lags to 10, we no longer report the credible intervals as to keep the pictorial descriptions interpretable. We also provide a more fine-grained lag analysis, as suggested by a reviewer to inspect the impact of lag 2-5 in Figure 7. From the above figures we summarize the findings as follows:

- **Mean trend Total count increases:** The mean trend on a contrary to the naive visual image in Figure 1 shows significant increase from Mid-March. This shows our sophisticated model is able to find some significant results that are also more interpretable on the light on COVID-19 disease spread.
- **Mean trend Subseries increases:** All the subseries also show significant increase around the same time, i.e. mid-March which says that the increase for daily total count is not due to random fluctuation. We also analyzed a few individual series and observed the same pattern

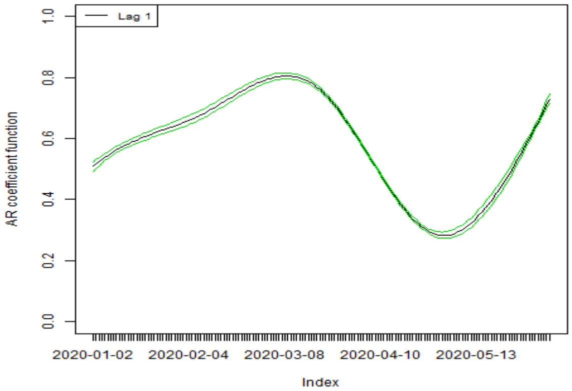


Figure 5. Total count: AR(1) trend of AR(1) model

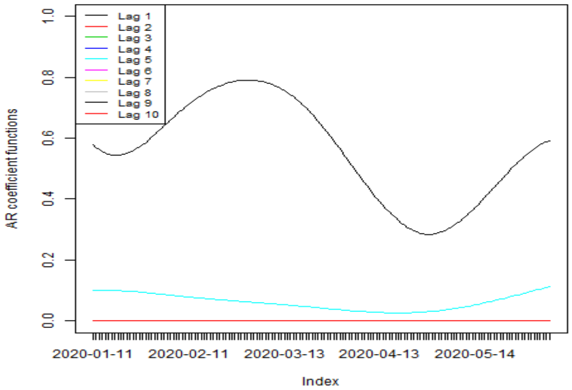


Figure 6. Total count: AR trends of AR(10) model

- **Mean trend similarity with COVID-19 numbers:** From the analysis, we can notice how similar the mean trend are with how number of confirmed cases COVID-19 has progressed in the month of January to June 2020. See Figure 8.
- **AR(1) trend for these series:** For the AR(1) model, it increased upto March and then slowly decreased, it was again reaching a peak around the end of May.
- **AR(1) model is sufficient:** The AR(p) models are comparable and usually only the first lag accounts for most of the correlation. Even with a more fine-grained analysis for smaller lags, the results are very similar. This similarity for different AR(p) model bolsters our hypothesis that

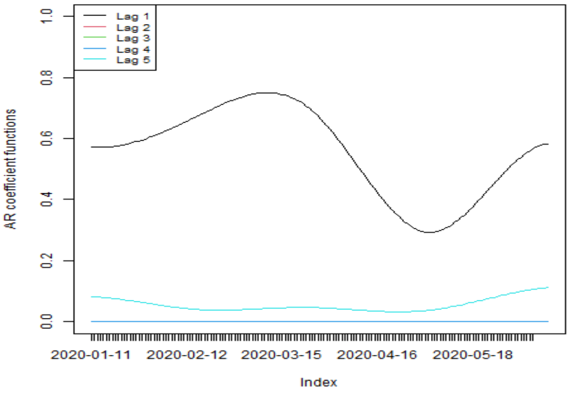
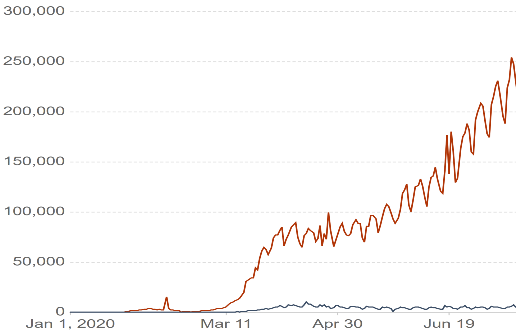


Figure 7. Total count: AR trends of AR(5) model



Source: European CDC – Situation Update Worldwide – Last updated 4 October, 10:35

Figure 8. COVID-19 daily confirmed cases

probably an AR(1) modelling is sufficient.

- **Credible intervals are narrow:** The 95% credible intervals provided are very narrow and thus gives us significant confidence about the true trends being of a similar nature.

Next we present the analysis for the three sub-classes in Figure 9. We just restrict ourselves here for the AR(1) model for the sake of conciseness. We see the mean trends show similar increasing trend except for sub-class TW. Also note that the credible intervals for ON and TW are wider which can be explained due to smaller number of observations in these. We also performed higher lags for each of these subclasses but none of the lags except lag 1 show significant difference.

5.2. Analyzing total counts assuming normal distribution

A reviewer asked whether our inference changes if we use Normal random variable to model the daily count data. To answer this question we used an AR(1)-ARCH(1) model on the demeaned data since our time series counts showed significant variation in both mean and variance over a window of size 30. We assume the following structure

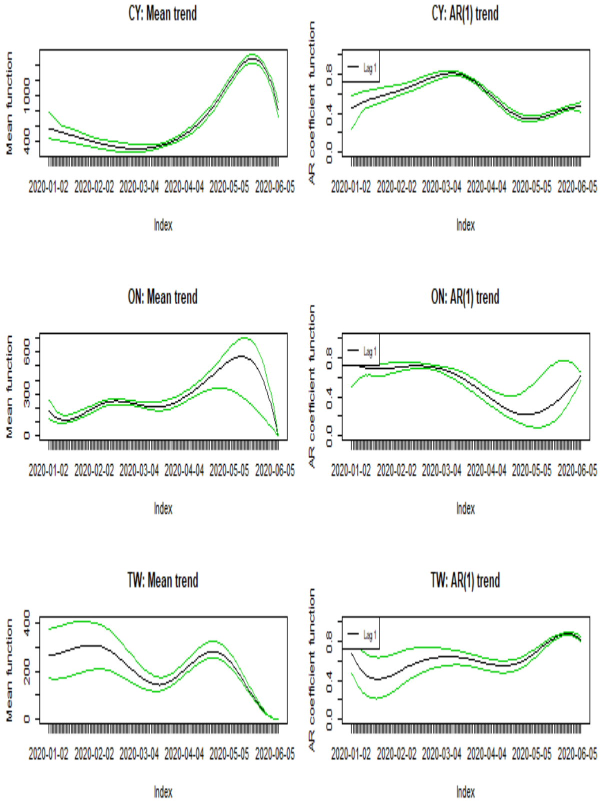


Figure 9. Sub-classes: Mean and AR(1) trends

$$X_i - \mu_i | X_{i-1} \sim N(a_{\mu,1}(i/n)X_{i-1}, \qquad (13)$$

$$\mu_(i/n) + a_1(i/n)X_{i-1}^2), \qquad (14)$$

where μ_i is a localized mean estimate at the i-th time-point respecting the time-variation. Our Bayesian modelling for this part is similar to what is done in [42] and we skip notational details for conciseness. We provide the plot of $a_{\mu,1}(\cdot), \mu(\cdot)$ and $a_1(\cdot)$ in Figure 10. One can see that both the variance trends also increase around the end of April which is when the virus spread globally to almost all countries whereas the mean AR(1) function remained pretty steady and slowed down after May. Understanding that the Poisson random variable has the same mean and variance, these results somewhat agree with what we obtained through Poisson modelling.

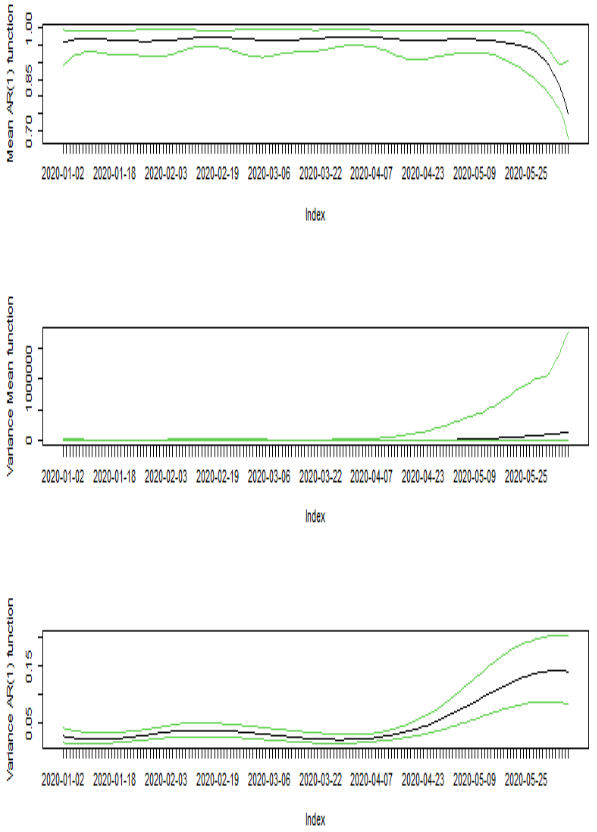


Figure 10. AR(1)-ARCH(1) model $a_{\mu,1}, \mu, a_1$ trends

6. Discussion, Implication, and Future Work

In this paper, we performed a thorough statistical evaluation including Bayesian and Poisson distribution analysis on the daily count of cyberbullying occurrences and its dominant classes mentioned above. Our model perfectly exhibits that there is a considerable increase in the general mean trend for a majority of these keywords from the subclasses mentioned above and the total keywords comprehending 27 keywords.

The most significant AR correlation can be seen at lag 1 where it evolved approximately 0.4-0.6. To the best of our knowledge a similar time-trend analysis is not done in the cybersecurity literature. Another significant advantage of our model is its adaptability in even small sample sizes. This allows us to quickly re-evaluate these trends and see whether a sudden change has occurred over the past few days or weeks, and prepare with defensive mechanisms if necessary. Since COVID-19 is a pandemic which has been spreading in multiple phases, it will be of the utmost importance to detect such increases as soon as they occur like we noticed a rise during March 2020 as well. Note that this analysis was to estimate what has already happened and the impact of COVID-19 on the reported incidents of cyberbullying through user-generated tweets.

We wish to discuss short term forecasting of these daily counts in our future work. Also note that here we mostly addressed cumulative counts of multiple keywords. In our future work we wish to properly model this through multivariate Poisson data with time-varying parameters, and we also wish to analyze each keyword simultaneously. Also, since COVID-19 is not diminishing anytime soon it will be interesting to see how these trends pan out over time and how we can adopt some suitable defensive mechanism against this if it continues to rise.

7. Acknowledgement

We would like to thank Umang Mehta for his help with the data collection, Andrew Kim for the initial editing, and John Jones Jr. for proofreading and copy editing the final version of the paper. We would also like to acknowledge the support of Secure and Privacy Research in New-Age Technology (SPRINT) Lab, University of Denver; and Human and Technical Security (HATS) Lab, Indiana University, and University of Florida. Any opinions, findings, and conclusions or recommendations expressed in this material are solely those of the author(s).

References

[1] M. L. Ybarra and K. J. Mitchell, “Online aggressor/targets, aggressors, and targets: A comparison of associated youth characteristics,” *Journal of child Psychology and Psychiatry*, vol. 45, no. 7, pp. 1308–1316, 2004.

[2] M. L. Ybarra, K. J. Mitchell, J. Wolak, and D. Finkelhor, “Examining characteristics and associated distress related to internet harassment: Findings from the second youth internet safety survey,” *Pediatrics*, vol. 118, no. 4, pp. e1169–e1177, 2006.

[3] M. L. Ybarra, M. Diener-West, and P. J. Leaf, “Examining the overlap in internet harassment and school bullying: Implications for school intervention,” *Journal of Adolescent Health*, vol. 41, no. 6, pp. S42–S50, 2007.

[4] R. Slonje and P. K. Smith, “Cyberbullying: Another main type of bullying?,” *Scandinavian Journal of Psychology*, vol. 49, no. 2, pp. 147–154, 2008.

[5] S. Das, J. Goard, and D. Murray, “How celebrities feed tweeples with personal and promotional tweets: Celebrity twitter use and audience engagement,” in *Proceedings of the 8th International Conference on Social Media & Society*, pp. 1–5, 2017.

[6] A. Tsoutsanis, “Tackling twitter and facebook fakes: Id theft in social media,” *World Data Protection Report*, vol. 12, no. 4, pp. 1–3, 2012.

[7] A. M. Meligy, H. M. Ibrahim, and M. F. Torky, “Identity verification mechanism for detecting fake profiles in online social networks,” *International Journal*

of Communication Networks and Information Security (IJCNIS), vol. 9, no. 1, pp. 31–39, 2017.

[8] O. Goga, G. Venkatadri, and K. P. Gummadi, "The doppelgänger bot attack: Exploring identity impersonation in online social networks," in *Proceedings of the 2015 Internet Measurement Conference*, pp. 141–153, 2015.

[9] M. A. Al-garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the twitter network," *Computers in Human Behavior*, vol. 63, pp. 433–443, 2016.

[10] B. K. Wiederhold, "Social media use during social distancing," 2020.

[11] B. Kim, "Effects of social grooming on incivility in covid-19," *Cyberpsychology, Behavior, and Social Networking*, 2020.

[12] A. Roy and S. Karmakar, "Bayesian semiparametric time varying model for count data to study the spread of the covid-19 cases," *arXiv preprint arXiv:2004.02281*, 2020.

[13] S. Das, A. Kim, and S. Karmakar, "Change-point analysis of cyberbullying-related twitter discussions during covid-19," *arXiv preprint arXiv:2008.13613*, 2020.

[14] S. Karmakar and S. Das, "Evaluating the impact of covid-19 on cyberbullying through bayesian trend analysis," in *Proceedings of The European Interdisciplinary Cybersecurity Conference (EICC) co-located with European Cyber Week*, 2020.

[15] A. Depoux, S. Martin, E. Karafillakis, R. Preet, A. Wilder-Smith, and H. Larson, "The pandemic of social media panic travels faster than the covid-19 outbreak," 2020.

[16] R. A. Bonanno and S. Hymel, "Cyber bullying and internalizing difficulties: Above and beyond the impact of traditional forms of bullying," *Journal of Youth and Adolescence*, vol. 42, no. 5, pp. 685–697, 2013.

[17] R. Dredge, J. Gleeson, and X. De la Piedad Garcia, "Cyberbullying in social networking sites: An adolescent victims perspective," *Computers in Human Behavior*, vol. 36, pp. 13–20, 2014.

[18] P. Wisniewski, H. Xu, M. B. Rosson, D. F. Perkins, and J. M. Carroll, "Dear diary: Teens reflect on their weekly online risk experiences," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3919–3930, 2016.

[19] V. Šléglová and A. Cerna, "Cyberbullying in adolescent victims: Perception and coping," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 5, no. 2, 2011.

[20] B. C. McHugh, P. J. Wisniewski, M. B. Rosson, H. Xu, and J. M. Carroll, "Most teens bounce back: Using diary methods to examine how quickly teens recover from episodic online risk exposure," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, pp. 1–19, 2017.

[21] P. K. Smith, J. Mahdavi, M. Carvalho, and N. Tippett, "An investigation into cyberbullying, its forms, awareness and impact, and the relationship between age and gender in cyberbullying," *Research Brief No. RBX03-06. London: DfES*, 2006.

[22] E. A. Vogel, J. P. Rose, L. R. Roberts, and K. Eckles, "Social comparison, social media, and self-esteem.," *Psychology of Popular Media Culture*, vol. 3, no. 4, p. 206, 2014.

[23] S. Hinduja and J. W. Patchin, "Cyberbullying legislation and case law," *Implications for School Policy and Practice*, vol. 5, p. 2016, 2015.

[24] T. Milosevic, "Social media companies' cyberbullying policies," *International Journal of Communication*, vol. 10, p. 22, 2016.

[25] S. Kessel Schneider, L. O'Donnell, and E. Smith, "Trends in cyberbullying and school bullying victimization in a regional census of high school students, 2006-2012," *Journal of School Health*, vol. 85, no. 9, pp. 611–620, 2015.

[26] P. A. Snell and E. Englander, "Cyberbullying victimization and behaviors among girls: Applying research findings in the field," *Journal of Social Sciences*, 2010.

[27] A. Mangaonkar, A. Hayrapetian, and R. Raje, "Collaborative detection of cyberbullying behavior in twitter data," in *2015 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 611–616, IEEE, 2015.

[28] S. Alim, "Analysis of tweets related to cyberbullying: exploring information diffusion and advice available for cyberbullying victims," *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, vol. 5, no. 4, pp. 31–52, 2015.

[29] H. Nurrahmi and D. Nurjanah, "Indonesian twitter cyberbullying detection using text classification and user credibility," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, pp. 543–548, IEEE, 2018.

[30] K. Cortis and S. Handschuh, "Analysis of cyberbullying tweets in trending world events," in *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, pp. 1–8, 2015.

[31] S.-H. Lee and H.-W. Kim, "Why people post benevolent and malicious comments online," *Communications of the ACM*, vol. 58, no. 11, pp. 74–79, 2015.

[32] E. Whittaker and R. M. Kowalski, "Cyberbullying via social media," *Journal of School Violence*, vol. 14, no. 1, pp. 11–29, 2015.

[33] C.-Y. Lai and C.-H. Tsai, "Cyberbullying in the social networking sites: An online disinhibition effect perspective," in *Proceedings of the 3rd Multidisciplinary International Social Networks Conference*, pp. 1–6, 2016.

[34] D. DiFranzo, S. H. Taylor, F. Kazerooni, O. D. Wherry, and N. N. Bazarova, "Upstanding by design: Bystander intervention in cyberbullying," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12, 2018.

[35] B. McNally, P. Kumar, C. Hordatt, M. L. Mauriello, S. Naik, L. Norooz, A. Shorter, E. Golub, and A. Druin, "Co-designing mobile online safety applications with children," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2018.

[36] V. K. Singh, M. L. Radford, Q. Huang, and S. Furrer, “”they basically like destroyed the school one day” on newer app features and cyberbullying in schools,” in *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 1210–1216, 2017.

[37] S. L. Zeger, “A regression model for time series of counts,” *Biometrika*, vol. 75, no. 4, pp. 621–629, 1988.

[38] P. T. Brandt and J. T. Williams, “A linear poisson autoregressive model: The poisson ar (p) model,” *Political Analysis*, vol. 9, no. 2, pp. 164–184, 2001.

[39] S. Karmakar, S. Richter, and W. B. Wu, “Simultaneous inference for time-varying models,” *In revision,* `https://sayarkarmakar.github.io/publications/sayar1.pdf`, 2020+.

[40] W. Shen and S. Ghosal, “Adaptive bayesian procedures using random series priors,” *Scandinavian Journal of Statistics*, vol. 42, no. 4, pp. 1194–1213, 2015.

[41] A. Roy and S. Karmakar, “Time-varying auto-regressive models for count time-series,” *arXiv preprint arXiv:2009.07634*, 2020.

[42] S. Karmakar and A. Roy, “Bayesian modelling of time-varying conditional heteroscedasticity,” *arXiv preprint arXiv:2009.06007*, 2020.