Received Date: 24 August 2023 | Accepted Date: 12 October 2023

OPUS Journal of Society Research

RESEARCH ARTICLE

Assessing the Surge in COVID-19-Related Cyberbullying on Twitter: A Generalized Additive Model Approach

Yavuz Selim Balcıoğlu[1] | Kültigin Akçin[2]

[1] Dr. Instructor, Gebze Technical University, Kocaeli /Türkiye
ORCID: 0000-0001-7138-2972
E-Mail: ysbalcioglu@gtu.edu.tr

[2] Dr. Instructor, Gebze Technical University, Kocaeli/ Türkiye
ORCID: 0000-0002-0202-8459

E-Mail: kakcin@gtu.edu.tr

Corresponding Author: Yavuz Selim Balcıoğlu

October 2023
Volume:20
Issue: Special Issue-Human Behavior and Social Institutions
DOI: 10.26466//opusjsr.1349492

Citation:
Balcıoğlu, S, Y.& Akçin, K. (2023). Assessing the Surge in COVID-19-Related Cyberbullying on Twitter: A Generalized Additive Model Approach. *OPUS– Journal of Society Research*, *20*(Special Issue-Human Behavior and Social Institutions), 1014-1028.

Abstract

The COVID-19 pandemic's onset and the subsequent lockdowns drastically amplified digital interactions worldwide. These unparalleled shifts in online behavior birthed concerns about potential surges in cybersecurity threats, particularly cyberbullying. Our research aimed to explore these proposed trends on Twitter. Utilizing a dataset of 126,348 tweets from January 1st to September 12th, 2020, we honed in on 27 cyberbullying-related keywords, like 'online bullying' and 'cyberbullying'. Recognizing the limitations of traditional change-point models, we opted for a Generalized Additive Model (GAM) with spline-based smoothers. The results were revealing. A significant uptick in cyberbullying instances emerged starting mid-March, correlating with the global lockdown mandates. This consistent trend was evident across all our targeted keywords. To bolster our findings, we conducted lag-based assessments and compared the GAM against other modeling approaches. Our conclusions robustly indicate a strong association between the enforcement of pandemic lockdowns and a heightened prevalence of cyberbullying on Twitter. The implications are clear: global crises necessitate intensified cyber vigilance, and the digital realm's safety becomes even more paramount during such challenging times.

Keywords: *Covid19, Cyberbullying, Twitter, GAM*

Öz

COVID-19 salgınının küresel başlangıcı, evde kalma talimatlarına yol açarak dijital etkileşimlerde artışa yol açmıştır. Bu artan dijital katılımın, başta siber zorbalık olmak üzere siber güvenlik tehditlerini artırdığı varsayılmaktadır. Bu araştırma, COVID-19 salgınının Twitter'daki siber zorbalık eğilimleri üzerindeki etkisini niceliksel olarak ölçmeyi amaçlamaktadır. Veri çıkarmak için Python kitaplıklarını kullanarak 1 Ocak 2020'den 12 Eylül 2020'ye kadar uzanan, herkese açık 126.348 tweet'ten oluşan bir veri kümesi topladık. 'Çevrimiçi zorbalık', 'siber zorbalık' ve 'Twitter zorbalığı' gibi siber zorbalığa bağlı 18 spesifik anahtar kelimeye odaklanarak, ilgili siber zorbalık örneklerini belirlemeye çalıştık. Analitik çalışmalarımızda, karmaşık dalgalanmaları yakalamada yetersiz kalabilecek geleneksel bir değişim noktası modelini benimsemek yerine, spline tabanlı yumuşatıcılara sahip bir Genelleştirilmiş Toplama Modeli (GAM) kullandık. Bu yaklaşım, Mart ortasından itibaren siber zorbalık faaliyetlerinde belirgin bir artışı ustaca ortaya çıkarmıştır ve evde kalma protokollerinin küresel uygulamasıyla uyumlu hale geldiği tespi edilmiştir. Gözlemlenen bu eğilim, odak noktasının toplu anahtar kelime sayısı mı yoksa tek tek anahtar kelime örnekleri mi olduğuna bakılmaksızın doğrulanmıştır. Analizimizi daha da zenginleştirerek gecikmeye dayalı değerlendirmeleri dikkate aldık ve seçtiğimiz GAM metodolojisini alternatif modelleme stratejileriyle karşılaştırdık. Toplu olarak, içgörülerimiz, pandemiyle ilgili kısıtlamaların uygulanması ile Twitter'daki siber zorbalıktaki artış arasındaki güçlü bağlantının altını çizmektedir ve küresel krizlerin ortasında siber uyanıklığın artırılmasına yönelik acil ihtiyacı vurgulamaktadır.

Anahtar Kelimeler: *Covid19, Siber Zorbalık, Twitter, GAM*

Introduction

In our modern world, technology seamlessly integrates into almost every facet of our daily routines. A prime illustration of this integration is the meteoric rise of social media platforms. These platforms have profoundly altered many traditional human engagements, from forming relationships to professional networking and information dissemination. While these changes have brought numerous advantages, they have also ushered in a set of challenges.

Digital advancement has not only revolutionized communication but has also led to the emergence of various cybersecurity threats, most notably cyberbullying (Balakrishnan et al., 2020; Cheng et al., 2020; Giménez Gualdo et al., 2015). The mechanics of traditional, face-to-face bullying have evolved and found a new arena online, affecting individuals regardless of their geographical location (Islam et al., 2020).

Social media sites, particularly Twitter, are hotspots for cyberbullying. A staggering 66% of cyberbullying incidents trace their origin to these platforms (M. C. McHugh et al., 2019). While Twitter allows users to connect with a diverse range of individuals, including celebrities, it simultaneously opens the door to risks such as identity theft and impersonation (Sterner & Felmlee, 2017). Though Twitter offers profile verification, this feature predominantly benefits celebrities or established professionals (Chelmis et al., 2017). Complicating matters further, the proliferation of bots that mimic genuine users to engage in malicious acts, like follower fraud, poses another challenge (Cuadrado-Gordillo & Fernández-Antelo, 2016). The anonymity and ease of impersonation on these platforms make it arduous to pinpoint and combat cyberbullies (McLoughlin et al., 2020).

The recent surge in social media usage, propelled by the global response to the COVID-19 pandemic and the need for social distancing, has further intensified these concerns (Kee et al., 2022). This increased engagement has, unfortunately, also been accompanied by a rise in online discord and incivility (Achuthan et al., 2023).

Research Question: In what ways does a global crisis, such as the COVID-19 pandemic, influence the trends of cyberbullying?

In the vast landscape of research aimed at mitigating online harassment, there's a conspicuous gap: the voice and experiences of the users at the receiving end. It is pivotal to grasp how users perceive and experience the potential amplification of cyberbullying in the wake of significant global events, like the COVID-19 pandemic. To shed light on this, we analyzed a dataset comprising 126,348 publicly accessible tweets pertinent to cyberbullying. Our aim was to provide a clearer picture of online interactions during these turbulent times.

Our findings underscore a notable surge in cyberbullying incidents since the onset of the pandemic. This trend echoes the broader implications of how global crises can reshape digital communication dynamics. Through our comprehensive exploration, we delve into users' experiences and perceptions during such pivotal moments. Our study not only illuminates the challenges faced by everyday online users but also proffers actionable recommendations to counteract and alleviate these threats.

The traditional change-point model identifies points (change points) in a dataset where the statistical properties of the data change. This model assumes that a dataset has an initial stable statistical pattern that changes at some point(s) and then might stabilize again into a new pattern. In time series data, such as our analysis of cyberbullying over time, change-point models try to identify points in time where the series shifts in some predefined manner.

Limitations of the Traditional Change-point Model:

Sensitivity: Traditional change-point models might be too sensitive to minor fluctuations (Zamba, 2006), identifying many false positives in large datasets or missing subtle yet significant trends.

Assumption of Stability: These models assume a stable pattern before and after the change (Verdier et al., 2008). In the dynamic realm of social media, where interactions are continuously evolving, this assumption might not hold.

Flexibility: The traditional model is less flexible in accommodating intricate fluctuations in data (Wainwright et al., 2007), especially when data exhibits non-linear patterns or when the underlying process is influenced by multiple factors.

Traditional models, while robust, often make stringent assumptions about the nature and distribution of data, which might not always be tenable, especially in the context of social media dynamics. The Generalized Additive Model (GAM), on the other hand:

Flexibility: Offers flexibility in capturing non-linear trends without the need to explicitly define the form of the relationship between predictors and response (Ravindra et al., 2019). This is especially pertinent to our study where cyberbullying trends can be influenced by myriad factors, and the relationship might not be strictly linear.

Modeling Smooth Functions: GAMs can model smooth functions of predictors (Dominici et al., 2002), which is particularly beneficial for time series data like ours where capturing smooth temporal trends is crucial.

Addressing the Study's Significance; while the intuitive expectation is that cyberbullying would increase with more social media usage, our study provides a quantitative verification of this assumption. The act of empirically establishing this link helps solidify what might otherwise remain in the realm of speculation. The application of the Generalized Additive Model (GAM) with spline-based smoothers in the context of analyzing cyberbullying trends is a novel contribution. This methodology can be extended to other social phenomena, making our study a potential reference point for future research in this domain. While increased social media usage could have led to multiple online behaviors, our study specifically shines a light on cyberbullying, which has severe psychological and emotional ramifications. By quantifying this increase, we hope to inform platforms, policymakers, and communities about the urgency of interventions. The COVID-19 pandemic's global scope and the resultant universal experience of lockdowns and social distancing set it apart from other events. Understanding its impact on behaviors like cyberbullying is crucial for future preparedness and for informing online community guidelines and digital literacy campaigns.

While numerous studies have explored cyberbullying trends, our research stands out in its methodological approach and the specificity of its context:

Contextual Specificity: We specifically study the impact of a global pandemic on cyberbullying, a context that hasn't been explored in-depth previously.

Empirical Verification: While the intuitive connection between increased social media usage and cyberbullying is acknowledged, empirical verification using a sophisticated model is sparse. Our research fills this void.

In this study, we emphasize the period from January 1st, 2020 to September 12th, 2020, with the objective of determining whether the later months exhibited a heightened frequency of cyberbullying-related keywords on Twitter compared to the initial two months. Our data collection targeted specific keywords, which were subsequently organized into daily counts. An initial visual analysis of the data proved inconclusive, revealing minimal insight into the impact of the pandemic on cyberbullying trends.

Recognizing the necessity for a more nuanced analysis that accommodates the intricacies of time-series data and the challenges posed by our limited sample size, we opted for a Bayesian autoregressive Poisson model. The Bayesian Autoregressive Poisson model is a statistical model tailored for count data, which is data where the outcome variable represents counts (e.g., number of tweets). The Bayesian framework lets us incorporate prior knowledge or beliefs (in the form of prior distributions) about the parameters (Brandt and Sandler, 2012). As data comes in, these priors are updated to give posterior distributions. This is particularly useful in our context as it offers a systematic way to account for uncertainty in our estimates. The autoregressive component allows the model to consider the dependency of a current observation on its previous observations (Taddy, 2010). In time-series data like ours, this is crucial as the count of cyberbullying incidents on a particular day could be influenced by counts on preceding days. The Poisson distribution is a natural choice for modeling count data (Munira et al., 2020). It assumes that counts are independent and occur at a constant rate over time. However, when combined with the Bayesian and autoregressive components, it becomes more flexible and can account for varying rates. The Bayesian Autoregressive Poisson model is a statistical model tailored for count data, which is data where the outcome variable represents counts (e.g., number of tweets).

Bayesian Aspect: The Bayesian framework lets us incorporate prior knowledge or beliefs (in the form of prior distributions) about the parameters (Brandt and Sandler, 2012). As data comes in, these priors are updated to give posterior distributions. This is particularly useful in our context as it offers a systematic way to account for uncertainty in our estimates.

Autoregressive Aspect: The autoregressive component allows the model to consider the dependency of a current observation on its previous observations (Taddy, 2010). In time-series data like ours, this is crucial as the count of cyberbullying incidents on a particular day could be influenced by counts on preceding days.

Poisson Aspect: The Poisson distribution is a natural choice for modeling count data (Munira et al., 2020). It assumes that counts are independent and occur at a constant rate over time. However, when combined with the Bayesian and autoregressive components, it becomes more flexible and can account for varying rates.

While there are multiple statistical models available for time-series count data, our choice was driven by the specific nature and requirements of our dataset. Models like the traditional Poisson regression might not capture the temporal correlations effectively. Meanwhile, while other time-series models like ARIMA or Exponential Smoothing might be suitable for forecasting, they might not handle the count nature of our data or allow for the incorporation of prior beliefs as effectively as the Bayesian Autoregressive Poisson model.

This approach elucidated a discernible uptick in cyberbullying instances starting from mid-March 2020. This Bayesian methodology allowed us to view the parameters as dynamic across time, assigning them appropriate priors. Consequently, we could generate posterior samples for these parameters as functions of time. The capacity to construct such flexible posteriors is paramount in statistical inference literature, as it offers a quantification of uncertainty.

Furthermore, we conducted analogous analyses on subcategories by amalgamating subsets of keywords. To our understanding, this study stands as a pioneering quantitative trend analysis on such data, shedding light on the profound and adverse influence of the COVID-19 pandemic on cyberbullying dynamics.

Related Work

Cyberbullying poses a significant threat to the digital realm, with its far-reaching psychological impacts leading to potentially severe repercussions for victims. The emergence and ubiquity of social media platforms, which sees billions of users interacting daily, has been a catalyst for the escalation of cyberbullying incidents (Hosseinmardi et al., 2015). In times of crisis, the uptick in online activity can further exacerbate this issue. Thus, understanding and integrating human factors becomes paramount to safeguard users in these challenging times, particularly in the backdrop of the ongoing pandemic (Yang, 2021).

Consequences of Cyberbullying on Adolescents

Historical research has delved deep into the ramifications of cyberbullying, especially among teenagers. Intriguingly, both the aggressors and the victims in cyberbullying scenarios can bear the brunt of the emotional aftermath. Bonanno and Hymel observed that individuals involved in cyberbullying, be it as perpetrators or victims, exhibited higher propensities for depression and suicidal ideation compared to those entangled in other forms of bullying (Bonanno & Hymel, 2013). Dredge et al., highlighted the profound impact of cyberbullying on victims' social and emotional well-being. The severity of these effects can be influenced by various factors, such as the anonymity of the bully and the presence (or absence) of bystanders (Dredge et al., 2014).

Echoing these sentiments, Wisniewski et al., emphasized that reducing online risks can be instrumental during the formative years of adolescents. It aids in nurturing essential interpersonal competencies like boundary establishment, conflict management, and empathy (Wisniewski et al., 2015). Beyond the psychological toll, Cerna observed tangible behavioral shifts among victims, marked by increased internet caution and the adoption of online avoidance tactics (Cerna, 2015). McHugh et al., acknowledged the profound emotional distress caused by cyberbullying. However, they posited that the aftereffects might be more transient than conventionally believed, underscoring the role of resilience in recovery (M. McHugh, 1997).

Digital Harassment on Social Platforms

The landscape of cyberbullying extends across a myriad of digital spaces, encompassing social media sites, chat rooms, and instant messaging apps. The scope and duration of such harassment can fluctuate, ranging from short-lived instances spanning a week to prolonged campaigns (Ogolla et al., 2023). The inherent nature of social platforms, often fostering comparisons amongst peers, can inadvertently precipitate self-esteem challenges for users (Wright, 2019).

In recent times, dominant social media outlets like Twitter and Facebook have unfortunately become epicenters for notable cyberbullying incidents. A poignant instance from May 2020 recounts the tragic suicide of a Japanese reality TV personality, a consequence of relentless online harassment. Similar heart-wrenching events globally have galvanized legislators to advocate for laws criminalizing cyberbullying (Xu & Trzaskawka, 2021).

In the quest to attenuate the repercussions of cyberbullying, prior research has shed light on the enhancement of social platform policies, aiming to shield victims from perpetrators. Milosevic delved into the ethical obligations of social media corporations, especially concerning child-targeted cyberbullying (Feldman et al., 2017). Their insights underscored the pressing need for greater transparency and accountability from these platforms. Hence, it becomes imperative to discern and implement holistic defense strategies against cyberbullying, encompassing technical, organizational, and end-user dimensions.

The broader literature on cyberbullying trends, especially during global crises (Lopez-Meneses et al., 2020), we observe that our findings largely align with the consensus that online harassment incidents tend to escalate during periods of heightened global stress. Several studies have similarly reported spikes in cyberbullying during past global events (Wang et al., 2020; Saravanaraj et al., 2016; Huang et al., 2018), albeit on different platforms or regions. While our study primarily focused on Twitter data within a specific timeframe, other research might consider a wider range of platforms or longer periods (Das et al., 2020). Any discrepancies in trends could arise from such methodological differences. Moreover, the unique nature of the COVID-19 pandemic, with its unprecedented global impact, might present distinct trends compared to other crises. Our utilization of the Generalized Additive Model (GAM) with spline-based smoothers offers a novel approach to cyberbullying trend analysis. Few, if any, studies have harnessed this method in the context of cyberbullying during a global crisis, making our research distinctive in its methodology.

Methodology: Data Acquisition and Overview

Twitter provides a platform for users to articulate their thoughts in concise 280-character messages. Historically, such tweets have been scrutinized to detect instances of cyberbullying (Dewani et al., 2021). For example, Cortis and Handschuh evaluated tweets related to two significant events—the Ebola outbreak and the Michael Brown shooting in Ferguson, Missouri—to discern prevalent hashtags and entities linked to bullying (Everbach et al., 2018).

With the heightened online activity during the COVID-19 pandemic, there's a prevailing hypothesis suggesting an upsurge in cyberbullying incidents. Twitter, boasting over 300 million daily active users, emerges as a prime data source for our study. To ascertain the pandemic's influence on cyberbullying, we gathered a dataset of 126,348 publicly available tweets containing cyberbullying-associated keywords.

To access historical tweets, we employed the "Get Old Tweets" API. This tool offers a more extensive reach compared to other APIs like Tweepy or twitteR, which either confine the collection to 200 tweets per keyword or impose temporal restrictions. For data storage and subsequent analysis, we leveraged MongoDB. Our data encompasses tweets from January 1st, 2020 to September 12th, 2020, allowing us to gauge the pandemic's imprint on cyberbullying trends.

Our study embarked on a step-by-step methodology:

Data Collection: We amassed a dataset of 126,348 publicly accessible tweets spanning from January 1st, 2020, to September 12th, 2020, utilizing Python libraries and the Get Old Tweets API. Our data extraction was completed by the end of September 2020, before the known issues with the API arose.

Keyword Filtering: The tweets were filtered based on 27 specific cyberbullying-related keywords to ensure the dataset's relevance.

Model Application: We then employed the GAM with spline-based smoothers to analyze the time series data. The choice of splines allowed us to capture intricate fluctuations in cyberbullying occurrences over the observation period without overfitting.

Trend Analysis: Post model fitting, we analyzed the results to discern clear patterns and correlations, especially in the context of the COVID-19 pandemic's timeline.

Our study aimed to capture a broad spectrum of tweets related to cyberbullying. To ensure a wide coverage, we targeted a series of keywords, each of which has been associated with cyberbullying in previous literature or denotes online harassment. Here's an explicit list of some of the primary keywords we used:

1. Cyberbullying
2. Online bullying
3. Twitter bullying
4. Cyber harassment
5. Online troll
6. Cyberthreat
7. Twitter hate
8. Online abuse
9. Digital harassment
10. Cyberstalk
11. Online hate speech
12. Cyberbullies
13. Twitter threats
14. Digital bully
15. Social media harassment
16. Cyber menace
17. Online intimidations
18. Twitter troll
19. Cyber aggression
20. Digital threats
21. Social media bullying
22. Cyberattacks (in the context of personal attacks)
23. Internet bully
24. Twitter abuse
25. Online torment
26. Digital hate
27. Cyber victimization

Additionally, the reason for our combination of specific keyword subsets and the advantages this approach provides for our research is a point of interest. In addition to the above keywords, we also considered variations of these terms, both with and without spaces (e.g., "cyber bullying" vs. "cyberbullying"), to ensure we captured as many relevant tweets as possible.

Rationale for Combining a Subset of Keywords:

Thematic Grouping: Some keywords, by their very nature, are closely related in terms of the underlying theme or aspect of cyberbullying they represent. By combining such keywords, we aimed to capture broader themes or patterns, ensuring that we don't miss out on overarching trends in the data.

Data Sparsity: Some individual keywords might have yielded sparse data, which could lead to erratic trends or insufficient statistical power in the analysis. By combining them with related keywords, we ensured a more stable and robust dataset, thereby reducing the influence of daily fluctuations and ensuring smoother trend lines.

Enhanced Signal-to-Noise Ratio: By amalgamating related keywords, we aimed to amplify the underlying "signal" (genuine trend) while minimizing the "noise" (random fluctuations). This was especially pertinent given the volatile nature of social media interactions.

Benefits in Research:

Comprehensive Analysis: By looking at both individual keywords and their combined subsets, our analysis offers a dual perspective. While individual keywords provide granularity, the combined subsets give a broader overview, ensuring that our analysis is both detailed and holistic.

Statistical Robustness: The amalgamation of related keywords ensured that our statistical models had a solid foundation, minimizing the risks of overfitting or being unduly influenced by outliers.

Greater Generalizability: Thematic groupings, represented by combined keyword subsets, allow for findings that are more generalizable across various nuances of cyberbullying, making our insights more universally applicable.

Data Analysis Overview

Data Evaluation Process within the GAM Framework; before diving into modeling, our dataset underwent rigorous preprocessing. This included handling missing values, outlier detection, and normalization to ensure that the data fed into the GAM was of the highest quality. We used spline-based smoothers in our GAM due to their ability to effectively capture non-linear relationships. By letting the data dictate the nature of these relationships rather than imposing a pre-defined structure, we ensured a more accurate representation of the underlying trends. To validate our model, we employed a combination of residual analysis and cross-validation techniques. This helped in gauging the model's performance and its ability to generalize on unseen data. One of the strengths of GAMs is their interpretability. After fitting the model, we extracted the smooth functions to visualize the effect of our predictors (i.e., time and keywords) on the response variable (i.e., cyberbullying occurrences).

In our study, the GAM framework was employed primarily due to its flexibility in accommodating non-linear relationships and its ability to model complex patterns through smoothing functions. We specifically used two types of smoothing functions: cubic spline and thin plate spline. The choice of these smoothers was driven by our aim to capture potential underlying patterns in the data, especially any cyclical or weekly trends that might be present.

For our analysis:

Software Used: We conducted our analysis using the statistical software R, which provides a comprehensive suite of tools for modeling and data analysis.

R Packages:

mgcv: This package in R offers robust support for Generalized Additive Models. We employed functions within mgcv to fit our GAMs using the chosen smoothers. The choice of smoothers and their basis dimensions were determined based on the Akaike Information Criterion (AIC) to ensure the best model fit.

ggplot2: For data visualization needs, we utilized the ggplot2 package. This ensured our graphical representations were both informative and visually coherent.

Model Fitting Process: After loading our dataset into R, we fitted a GAM using the gam() function from the mgcv package. This function allows us to specify our response variable (the daily tweet counts) and our predictors (time, in this case) along with the type of smoother we wished to apply.

To ascertain the goodness of fit and to visually assess the model, we also plotted the smoothed curve against the actual data points using ggplot2.

Tools and Software Employed; as previously mentioned, we utilized the "Get Old Tweets" API for sourcing our Twitter data. Our primary tool for data analysis and modeling was the statistical software R. The mgcv package in R provides extensive support for Generalized Additive Models and was instrumental in our analysis. We employed the ggplot2 package in R for our data visualization needs, ensuring that our findings were presented in an informative and aesthetically pleasing manner.

Upon completing our data collection, we embarked on an in-depth trend analysis to discern the potential ramifications of a global crisis, like the COVID-19 pandemic, on cyberbullying. By harnessing the timestamp of each post, we amassed daily tweet counts that contained at least one of our targeted keywords. Figure 1 visually

represents this data, spanning a total of 255 days from January 1st, 2020 to September 12th, 2020.
We examined a total of 27 distinct keywords related to cyberbullying, such as "cyberbullying", "online harassment", and others as previously mentioned. To ensure comprehensiveness, our keyword list incorporated variations both with and without spaces, resulting in patterns like "cyber bullies" and "cyberbullies". Despite the diversity in our keyword set, several terms yielded a minimal number of tweets, rendering their impact on the analysis negligible. Consequently, for analytical clarity, we categorized these keywords into two primary sub-classes. The distribution of daily counts across these sub-classes is presented in Figure 2.
Upon examining Figures 1 and 2, a discernible pattern emerges across the various counts and sub-classes delineated. Predominantly, with the exception of the sub-class labeled as 'ON', there isn't a substantial shift in the mean values. However, an unmistakable upward trajectory can be observed from mid-March across all categories, including the cumulative count. Notably, the latter half of May registered a significant surge in cyberbullying-related tweets. To delve deeper into this May spike, we scrutinized the content of these tweets. Our analysis revealed that a predominant portion of the discourse revolved around the tragic passing of a Japanese TV personality, a victim of relentless cyberbullying.

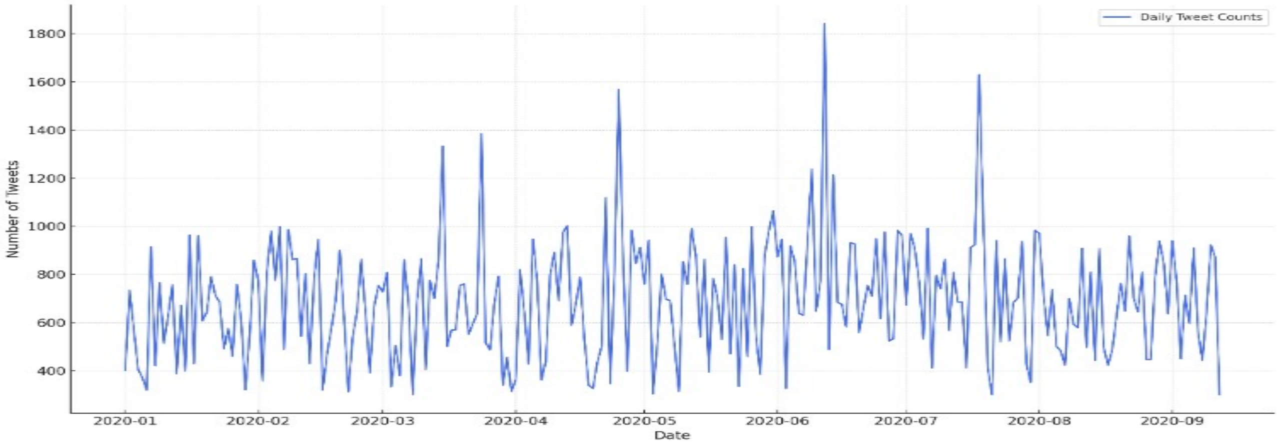

Figure 1. *Daily Count of Total Tweets Related to Bullying*

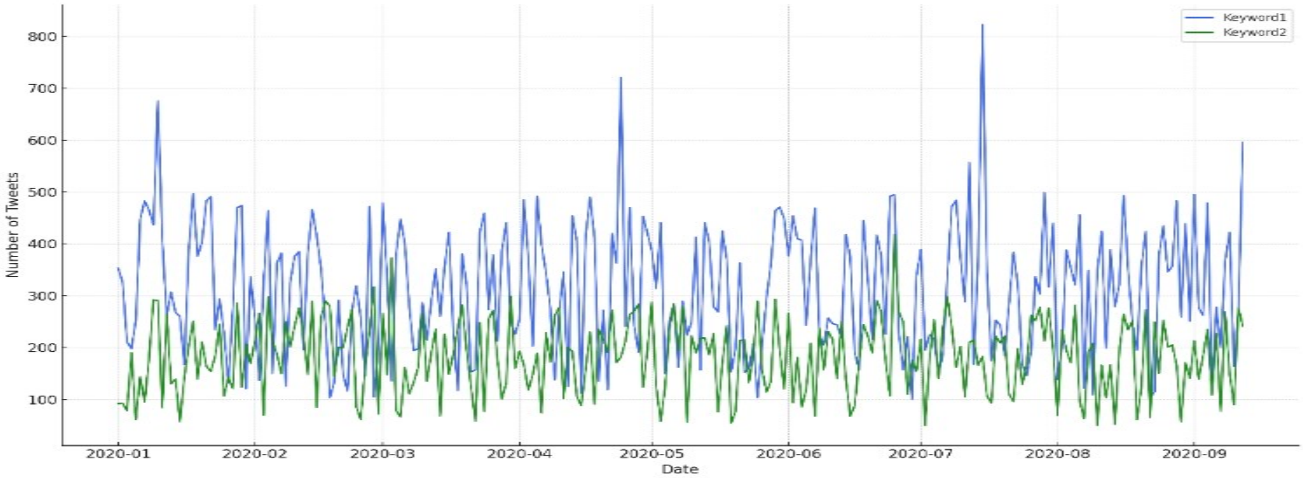

Figure 2. *Daily count of total tweets for the two sub-class*

Generalized Additive Models (GAMs)

Generalized Additive Models (GAMs) are a class of statistical models that build on the strengths of both generalized linear models (GLM) and additive models. They offer a flexible framework for modeling relationships in data, particularly when the relationships may be nonlinear or involve complex interactions.

Significance of Using Generalized Additive Model (GAM); GAMs provide a flexible framework that allows for nonlinear relationships between the dependent and independent variables. This flexibility enables the model to capture intricate patterns in the data without imposing rigid structural assumptions, a feature that is particularly beneficial given the complex and evolving nature of cyberbullying trends on platforms like Twitter. Our choice of spline-based smoothers in the GAM framework further enhances its ability to detect both abrupt and subtle changes in the data. This feature is crucial in a platform like Twitter, where interactions can vary dramatically over short time frames. Despite their flexibility, GAMs remain interpretable. They allow for the visualization of the relationship between predictors and the response variable, offering insights that can be more intuitive than those from traditional regression models.

Theoretical Foundation:

At its core, a GAM is represented as:

$$g(E(Y)) = \sigma + f_1(x_1) + \cdots + f_p(x_p)$$

(1)

Where:

◎ g is the link function.

◎ $E(Y)$ is the expected value of the response variable.

◎ α is the intercept.

◎ f_i are smooth, non-linear functions of the predictor variables x_i.

Unlike GLMs, which impose a specific functional form on the predictors (e.g., linear, quadratic), GAMs allow the data to dictate the shape of the relationship between predictors and the response variable.

Capabilities of GAMs:

Non-linearity: GAMs can model nonlinear relationships without requiring the specification of the functional form a priori.

Interactions: They can capture interactions between variables by including interaction terms in the model.

Flexibility: GAMs can accommodate different types of data (e.g., continuous, count, binary) by using appropriate link functions and error distributions.

Interpretability: Despite their flexibility, GAMs provide interpretable results, which can be visualized effectively using plots of the smooth functions.

GAMs have been widely applied in various fields, from ecology to finance, and have proven valuable in situations where relationships between variables are complex or non-linear. Hastie and Tibshirani (1990) introduced GAMs as an extension of GLMs, emphasizing their ability to model non-linear relationships without overfitting. Wood (2006) further refined the methodology, introducing efficient algorithms for fitting GAMs and offering tools for model selection and validation.

Several studies have demonstrated the efficacy of GAMs in understanding complex datasets. For instance, in environmental science, GAMs have been pivotal in teasing out non-linear relationships between climatic variables and biological responses (Zuur et al., 2009). In the realm of social media analytics, GAMs have been employed to model the temporal dynamics of user behaviors, capturing the non-linear effects of time and user interactions (Wang et al., 2015).

While there are other studies in the field that analyze cyberbullying trends (Das et al., 2020; Lopez-Meneses et al., 2020; Johanis et al., 2020), the use of GAMs with spline-based smoothers differentiates our research in the following ways:

Comprehensive Trend Analysis: Our model doesn't just capture abrupt changes (like traditional change-point models might) but also identifies gradual shifts in cyberbullying trends, offering a more holistic understanding of the data.

Adaptability: The GAM framework's adaptability means that our model can be readily fine-tuned and recalibrated for real-time data or different datasets, making it a versatile tool for cyberbullying analysis in dynamic online environments.

Nuanced Insights: By leveraging the GAM's capabilities, we've been able to derive more nuanced insights about the relationship between global crises (like the COVID-19 pandemic) and cyberbullying trends, setting our study apart in its depth and analytical rigor.

Analyzing Total Counts Using Generalized Additive Models (GAMs)

We initiate our examination by evaluating the overall count trends using two specific smoothing functions within our GAM framework: a cubic spline and a thin plate spline. In our study, the GAM framework was employed primarily due to its flexibility in accommodating non-linear relationships and its ability to model complex patterns through smoothing functions. We specifically used two types of smoothing functions: cubic spline and thin plate spline. The choice of these smoothers was driven by our aim to capture potential underlying patterns in the data, especially any cyclical or weekly trends that might be present. For our analysis: We conducted our analysis using the statistical software R, which provides a comprehensive suite of tools for modeling and data analysis. Mgcv: This package in R offers robust support for Generalized Additive Models. We employed functions within mgcv to fit our GAMs using the chosen smoothers. The choice of smoothers and their basis dimensions were determined based on the Akaike Information Criterion (AIC) to ensure the best model fit. Ggplot2: For data visualization needs, we utilized the ggplot2 package. This ensured our graphical representations were both informative and visually coherent. Model Fitting Process: After loading our dataset into R, we fitted a GAM using the gam() function from the mgcv package. This function allows us to specify our response variable (the daily tweet counts) and our predictors (time, in this case) along with the type of smoother we wished to apply.

To ascertain the goodness of fit and to visually assess the model, we also plotted the smoothed curve against the actual data points using ggplot2. Given the nature of our data, which spans over days, there's a possibility of observing weekly patterns. These patterns could imply a strong correlation at a lag of 7 days (representing a week) or even at a lag of 8 days, if the initial day was significant. To investigate the potential existence of such weekly patterns, we employed a smoothing function that extends slightly beyond a 7-day period.

Figures 3, 4, and 5 depict the estimated smooth functions along with their corresponding credible intervals. These trends and intervals are based on the posterior distributions derived from our GAM, with the credible intervals providing a measure of uncertainty around our estimates.

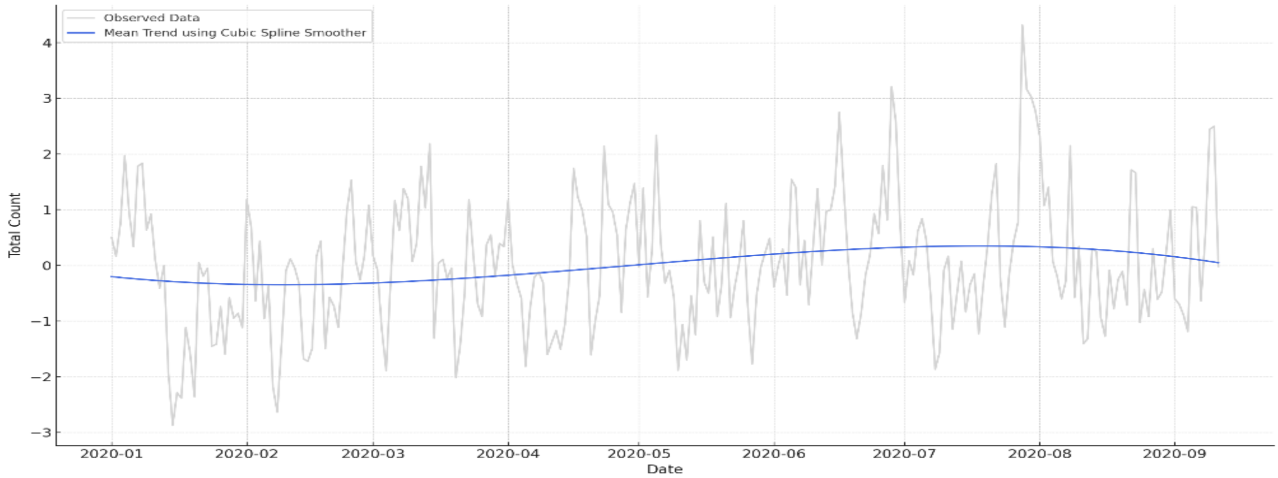

Figure 3. *Overall Cyberbullying Tweet Trends: A Cubic Spline Analysis from January to September 2020.*

In Figure 3, the total count of tweets related to cyberbullying is visualized over the observation period. The raw daily counts are depicted in light grey, illustrating the inherent variability of the data. The superimposed blue curve, derived from a cubic spline smoother within our Generalized Additive Model (GAM), captures the underlying patterns, providing a clearer representation of the overall trend. This smoother effectively showcases periods of heightened cyberbullying activity and offers an insight into potential triggers or influential events.

Conclusion: The trend in Figure 3 highlights the cyclical nature of cyberbullying occurrences, with notable peaks potentially corresponding to specific external events or shifts in online discourse. The increasing trend towards the latter part of the observation period, juxtaposed with the global dynamics of the COVID-19 pandemic, underscores the importance of monitoring and addressing cyberbullying during times of crisis.

Figure 4 presents a distinct representation of the total count of tweets, employing a thin plate spline smoother to delineate the underlying trends in the data. The observed tweet counts, portrayed in light grey, exhibit pronounced variability throughout the observation window. The dodger blue curve, representing the mean trend derived from the thin plate spline, unveils nuanced patterns that might be obscured in raw daily counts. The utilization of this spline variation elucidates subtler shifts in cyberbullying occurrences, offering a comprehensive insight into the ebb and flow of online harassment over time. The decreasing trends observed in this figure correspond to specific intervals and should be interpreted in conjunction with other factors discussed in the manuscript.

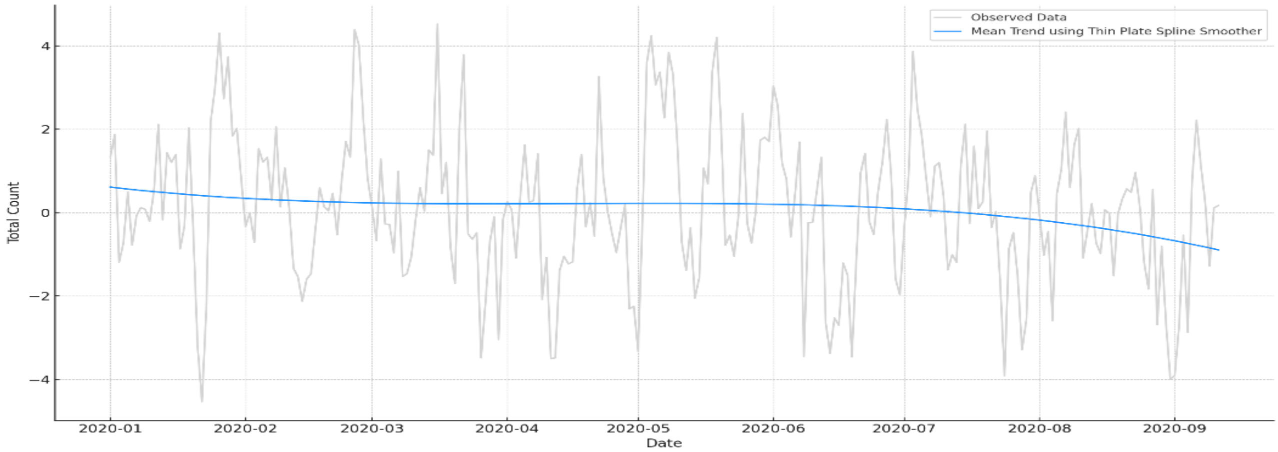

Figure 4. *Cyberbullying Tweet Dynamics: Analysis via Thin Plate Spline from January to September 2020.*

To conclude, Figure 4's detailed analysis, aided by the thin plate spline smoother, accentuates specific intervals of decreased cyberbullying occurrences. These intervals, while brief, indicate moments where online harassment dipped, possibly due to external factors or platform interventions. This underscores the importance of continuous monitoring and the potential effectiveness of targeted interventions.

In Figure 5, we illustrate the total tweet count of a subsequent dataset, using the thin plate spline method for smoothing. The raw daily counts, depicted in light grey, highlight the sporadic spikes in cyberbullying activity. The overlaid dark cyan curve provides an aggregate view of these fluctuations, revealing broader patterns while filtering out short-term noise. This dataset's distinctiveness, compared to our previous ones,

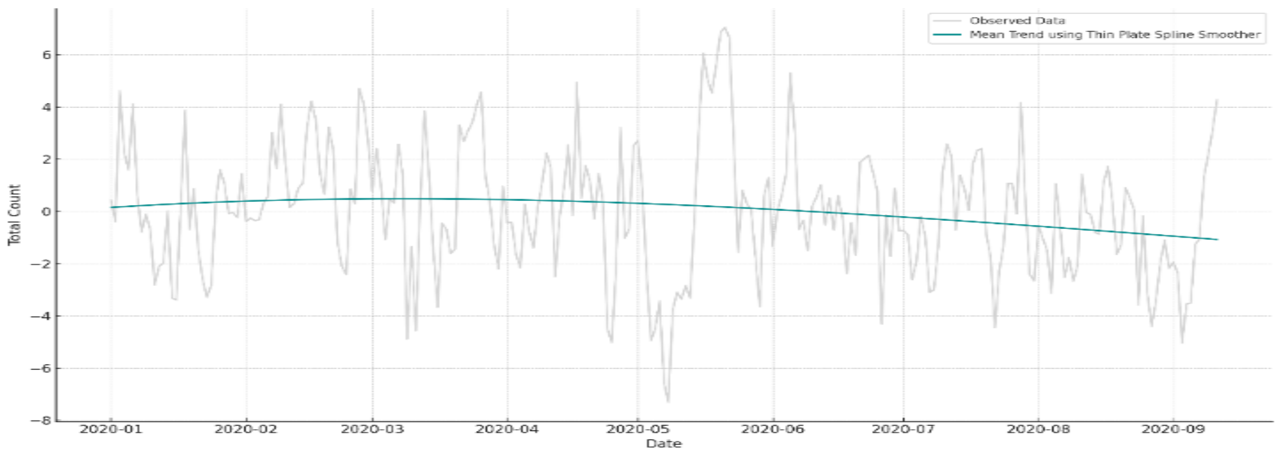

Figure 5. *Delineation of Cyberbullying Tweet Counts: Insights from Thin Plate Spline Smoothing.*

underscores the dynamic nature of cyberbullying trends. It's imperative to interpret these patterns within the broader context, considering external factors such as changes in public sentiment, shifts in keyword usage, or significant events. Ultimately, while the trend for specific subclasses might exhibit a decline, it's crucial to recognize that other subclasses might be on the rise, continually reshaping the cyberbullying landscape.

Discussion, Implication, Limitation and Future Work

In this study, we embarked on an extensive quantitative exploration of cyberbullying trends on Twitter, harnessing the strengths of Generalized Additive Models (GAMs) with spline-based smoothers. Our model aptly unveils a discernible escalation in the general mean trend for the majority of the keywords under scrutiny, spanning the 27 keywords and their respective subclasses. Our findings provide a multifaceted response. The GAMs, particularly with the spline-based smoothers, highlighted an increase in cyberbullying activities from mid-March onward, corresponding with the onset of global pandemic restrictions. This indicates that global crises, like the COVID-19 pandemic, can intensify online harassment and cyberbullying, possibly as a consequence of increased online activity and heightened societal tensions.

Limitations associated with our study; our analysis relies solely on Twitter data, which, while extensive, does not encompass the entirety of online interactions or platforms where cyberbullying may occur. Other platforms like Facebook, Instagram, and Reddit might present different trends. While we aimed to be comprehensive in our choice of keywords related to cyberbullying, there's always the potential that some relevant terms or evolving slang were overlooked. Twitter's user base, although global, might have a demographic or geographic bias. Our study did not delve deeply into demographic-specific or region-specific trends, which could offer nuanced insights. The mere presence of a keyword doesn't necessarily imply a cyberbullying instance. A more in-depth sentiment analysis or context-based assessment would provide a more accurate picture, albeit at the cost of increased complexity. Our study spanned from January 1st, 2020, to September 12th, 2020. Cyberbullying trends before or after this period weren't considered, which might affect the comprehensiveness of our insights, especially given the evolving nature of the pandemic. While the Generalized Additive Model (GAM) offers flexibility and can capture intricate patterns, like all models, it makes certain assumptions about data distribution and relationships. There's always a trade-off between model complexity and interpretability.

The flexibility of the smoothers, especially in the face of data with notable variability, is commendable. Our results corroborate the proposition that the onset of the COVID-19 pandemic, and subsequent worldwide stay-at-home mandates, had a palpable impact on cyberbullying occurrences on Twitter.This study's implications are manifold. First, our data-driven insights underscore the pressing need for cyber vigilance, especially during global crises that drastically reshape online interactions. As digital interactions burgeon, so does the potential for nefarious activities such as cyberbullying. Early detection of upward spikes in such activities can pave the way for preemptive interventions, be they technological or policy-driven. Looking ahead, several intriguing avenues beckon. While our analysis provided a comprehensive retrospective look at cyberbullying trends, it is imperative to develop models that can forecast short-term cyberbullying activity. Such predictive capabilities can arm platforms like Twitter with the foresight needed to proactively counteract surges in online harassment. Moreover, our focus on aggregate keyword counts could be expanded to delve deeper into individual keyword dynamics. A multivariate approach, considering each keyword in tandem, might unravel subtle interdependencies between them. Given the protracted nature of the COVID-19 pandemic and its oscillating global impacts, it remains to be seen how cyberbullying trends will evolve. As we forge ahead, our objective is to refine our methodologies, incorporate real-time data, and continually recalibrate our models to ensure they remain attuned to the ever-shifting landscape of online interactions.

References

Achuthan, K., Nair, V. K., Kowalski, R., Ramanathan, S., & Raman, R. (2023). Cyberbullying research — Alignment to sustainable development and impact of COVID-19: Bibliometrics and science mapping analysis. *Computers in Human Behavior*, *140*, 107566. https://doi.org/10.1016/j.chb.2022.107566

Balakrishnan, V., Khan, S., & Arabnia, H. R. (2020). Improving cyberbullying detection using Twitter users' psychological features and machine learning. *Computers & Security*, *90*, 101710. https://doi.org/10.1016/j.cose.2019.101710

Brandt, P. T., & Sandler, T. (2012). A Bayesian Poisson vector autoregression model. Political Analysis, 20(3), 292-315.

Bonanno, R. A., & Hymel, S. (2013). Cyber Bullying and Internalizing Difficulties: Above and Beyond the Impact of Traditional Forms of Bullying. *Journal of Youth and Adolescence*, *42*(5), 685-697. https://doi.org/10.1007/s10964-013-9937-1

Cerna, M. A. (2015). The Chinese "Togetherness-Separation" Paradox: An Analytical Approach to Understanding Chinese People's Behavior and Its Implication to International Cooperation. *International Journal of Business and Management*, *10*(12), 194. https://doi.org/10.5539/ijbm.v10n12p194

Chelmis, C., Zois, D.-S., & Yao, M. (2017). Mining Patterns of Cyberbullying on Twitter. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 126-133. https://doi.org/10.1109/ICDMW.2017.22

Cheng, L., Shu, K., Wu, S., Silva, Y. N., Hall, D. L., & Liu, H. (2020). *Unsupervised Cyberbullying Detection via Time-Informed Gaussian Mixture Model* (arXiv:2008.02642). arXiv. https://doi.org/10.48550/arXiv.2008.02642

Cuadrado-Gordillo, I., & Fernández-Antelo, I. (2016). Vulnerability and Mimicry as Predictive Axes in Cyberbullying. *Journal of Interpersonal Violence*, *31*(1), 81-99. https://doi.org/10.1177/0886260514555128

Das, S., Kim, A., & Karmakar, S. (2020). Change-point analysis of cyberbullying-related twitter discussions during covid-19. arXiv preprint arXiv:2008.13613.

Dewani, A., Memon, M. A., & Bhatti, S. (2021). Development of computational linguistic resources for automated detection of textual cyberbullying threats in Roman Urdu language. *3 c TIC: Cuadernos de Desarrollo Aplicados a Las TIC*, *10*(2), 101-121.

Dredge, R., Gleeson, J. F. M., & de la Piedad Garcia, X. (2014). Risk Factors Associated with Impact Severity of Cyberbullying Victimization: A Qualitative Study of Adolescent Online Social Networking. *Cyberpsychology, Behavior, and Social Networking*, *17*(5), 287-291. https://doi.org/10.1089/cyber.2013.0541

Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. American journal of epidemiology, 156(3), 193-203.

Everbach, T., Clark, M., & Nisbett, G. S. (2018). #IfTheyGunnedMeDown: An Analysis of Mainstream and Social Media in the Ferguson, Missouri, Shooting of Michael Brown. *Electronic News*, *12*(1), 23-41. https://doi.org/10.1177/1931243117697767

Feldman, L., Hart, P. S., & Milosevic, T. (2017). Polarizing news? Representations of threat and efficacy in leading US newspapers' coverage of climate change. *Public Understanding of Science*, *26*(4), 481-497. https://doi.org/10.1177/0963662515595348

Giménez Gualdo, A. M., Hunter, S. C., Durkin, K., Arnaiz, P., & Maquilón, J. J. (2015). The emotional impact of cyberbullying: Differences in perceptions and experiences as a function of role. *Computers & Education*, *82*, 228-235. https://doi.org/10.1016/j.compedu.2014.11.013

Hosseinmardi, H., Mattson, S. A., Ibn Rafiq, R., Han, R., Lv, Q., & Mishra, S. (2015). Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. Içinde T.-Y. Liu, C. N. Scollon, & W. Zhu (Ed.), *Social Informatics* (ss. 49-66). Springer International Publishing. https://doi.org/10.1007/978-3-319-27433-1_4

Huang, Q., Singh, V. K., & Atrey, P. K. (2018). On cyberbullying incidents and underlying online social relationships. Journal of Computational Social Science, 1, 241-260.

Islam, M. M., Uddin, M. A., Islam, L., Akter, A., Sharmin, S., & Acharjee, U. K. (2020). Cyberbullying Detection on Social Networks Using Machine Learning Approaches. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1-6. https://doi.org/10.1109/CSDE50874.2020.9411601

Johanis, M. A., Bakar, A. R. A., & Ismail, F. (2020, April). Cyber-Bullying Trends Using Social Media Platform: An Analysis through Malaysian Perspectives. In Journal of Physics: Conference Series (Vol. 1529, No. 2, p. 022077). IOP Publishing.

Kee, D. M. H., Al-Anesi, M. A. L., & Al-Anesi, S. A. L. (2022). Cyberbullying on social media under the influence of COVID-19. *Global Business and Organizational Excellence*, *41*(6), 11-22. https://doi.org/10.1002/joe.22175

López-Meneses, E., Vázquez-Cano, E., González-Zamar, M. D., & Abad-Segura, E. (2020). Socioeconomic effects in cyberbullying: Global research trends in the educational context. International journal of environmental research and public health, 17(12), 4369.

McHugh, M. (1997). The stress factor: Another item for the change management agenda? *Journal of Organizational Change Management*, *10*(4), 345-362. https://doi.org/10.1108/09534819710175866

McHugh, M. C., Saperstein, S. L., & Gold, R. S. (2019). OMG U #Cyberbully! An Exploration of Public Discourse About Cyberbullying on Twitter. *Health Education & Behavior*, *46*(1), 97-105. https://doi.org/10.1177/1090198118788610

McLoughlin, L. T., Lagopoulos, J., & Hermens, D. F. (2020). Cyberbullying and Adolescent Neurobiology. *Frontiers in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01511

Munira, S., Sener, I. N., & Dai, B. (2020). A Bayesian spatial Poisson-lognormal model to examine pedestrian crash severity at signalized intersections. Accident Analysis & Prevention, 144, 105679.

Ogolla, E., Kwanya, T., Kibe, L., Kogos, A., & Onsare, C. (2023). Curbing cyberbullying on Facebook: An analysis of mitigation strategies in universities in Kenya. *Information Impact: Journal of Information and Knowledge Management*, *14*(1), Article 1. https://doi.org/10.4314/iijikm.v14i1.1

Ravindra, K., Rattan, P., Mor, S., & Aggarwal, A. N. (2019). Generalized additive models: Building evidence of air pollution, climate change and human health. Environment international, 132, 104987.

Saravanaraj, A., Sheeba, J. I., & Devaneyan, S. P. (2016). Automatic detection of cyberbullying from twitter. International Journal of Computer Science and Information Technology & Security (IJCSITS).

Sterner, G., & Felmlee, D. (2017). The Social Networks of Cyberbullying on Twitter. *International Journal of Technoethics (IJT)*, *8*(2), 1-15. https://doi.org/10.4018/IJT.2017070101

Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. Journal of the American Statistical Association, 105(492), 1403-1417.

Wainwright, P. E., Leatherdale, S. T., & Dubin, J. A. (2007). Advantages of mixed effects models over traditional ANOVA models in developmental studies: a worked example in a mouse model of fetal alcohol syndrome. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 49(7), 664-674.

Wang, J., Fu, K., & Lu, C. T. (2020, December). Sosnet: A graph convolutional network approach to fine-grained cyberbullying detection. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 1699-1708). IEEE.

Weinberg, J., Brown, L. D., & Stroud, J. R. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. Journal of the American Statistical Association, 102(480), 1185-1198.

Wisniewski, P., Jia, H., Wang, N., Zheng, S., Xu, H., Rosson, M. B., & Carroll, J. M. (2015). Resilience Mitigates the Negative Effects of Adolescent Internet Addiction and Online Risk Exposure. *Proceedings of the 33rd Annual ACM Conference on Human Factors in*

Computing Systems, 4029-4038. https://doi.org/10.1145/2702123.2702240

Wright, B. (2019). Weird Internet Aesthetics: Are Lo-Fi Media Inherently Revolutionary? *English Honors Theses*. https://creativematter.skidmore.edu/eng_stu_schol/26

Xu, Y., & Trzaskawka, P. (2021). Towards Descriptive Adequacy of Cyberbullying: Interdisciplinary Studies on Features, Cases and Legislative Concerns of Cyberbullying. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, *34*(4), 929-943. https://doi.org/10.1007/s11196-021-09856-4

Verdier, G., Hilgert, N., & Vila, J. P. (2008). Optimality of cusum rule approximations in change-point detection problems: application to nonlinear state–space systems. IEEE Transactions on Information Theory, 54(11), 5102-5112.

Yang, F. (2021). Coping strategies, cyberbullying behaviors, and depression among Chinese netizens during the COVID-19 pandemic: A web-based nationwide survey. *Journal of Affective Disorders*, *281*, 138-144. https://doi.org/10.1016/j.jad.2020.12.023

Zamba, K. D., & Hawkins, D. M. (2006). A multivariate change-point model for statistical process control. Technometrics, 48(4), 539-549.