



Exploring cyber violence against women and girls in the Philippines through Mining Online News

Explorando la ciberviolencia contra mujeres y niñas en Filipinas a través de Mining Online News

 January-D. Febro-Naga. Associate Professor, Department of Information Technology, MSU-Iligan Institute of Technology (Philippines) (january.febro@g.msuiit.edu.ph) (<https://orcid.org/0000-0002-2977-6377>)

 Mia-Amor-C. Tinam-isan. Associate Professor, Department of Information Technology, MSU-Iligan Institute of Technology (Philippines) (miaamor.catindig@g.msuiit.edu.ph) (<https://orcid.org/0000-0002-5754-2025>)

ABSTRACT

Violence against women and girls (VAWG) is not a recent phenomenon. What is new is the additional increasing threats that millions of women and girls face because of the rapid spread of ICTs and the expansion of social media. Cases of VAWG wherein ICT and social media are used as platforms by cybercriminals can be seen in the news media coverage. This study aims to understand and determine the trend and the state of cyber VAWG to raise awareness through mining online news websites. News articles were scraped from popular news websites between 2015 to 2020. The preprocessed articles (N=3,506) were analyzed by year using the Topic Keyword Model (TKM). It was observed that the cyber VAWG articles topic trends are increasing with most of the articles focusing on the topics "Online sexual exploitation and sexual abuse of children" and "ICT-related violations of privacy". Text mining methods may address the limitations of traditional qualitative approaches. Understanding the cyber VAWG issues by mining news articles is a novel approach that could help create programs and policies to address this societal concern. Additional studies should be conducted related to sentiment analysis of news data to verify and measure the influence of cyber VAWG-related topics.

RESUMEN

La violencia contra las mujeres y las niñas (VCMN) no es un fenómeno nuevo. Lo nuevo son los crecientes peligros a que se enfrentan los millones de mujeres y niñas debido a la difusión de las TIC y redes sociales. Los casos de VCMN donde se utilizan como plataformas las TIC y las redes sociales se encuentran fácilmente por medio de la cobertura mediática. Este estudio tiene como objetivo la comprensión y la definición de la tendencia y el estado de la ciber-VCMN para crear conciencia por medio del análisis de los sitios web de noticias online. Entre 2015 y 2020, se recopilieron artículos de los principales sitios de noticias. Se utilizó el Modelo de palabras claves temáticas para evaluar los artículos preprocesados (N=3.506) por año. Se señaló que la mayoría de los artículos sobre ciber-VCMN se centraban en temas de «Explotación sexual y abuso sexual de niños en línea» y «Violaciones de la privacidad relacionadas con las TIC». El análisis de los textos ayuda a trascender las limitaciones de las metodologías cualitativas tradicionales. Comprender las preocupaciones de la ciber-VCMN mediante la extracción de artículos de noticias podría ayudar a crear iniciativas y políticas para solucionar este problema. Proponemos que se lleve a cabo una investigación utilizando análisis de sentimiento de los datos de noticias para verificar y cuantificar el impacto de los problemas relacionados con la ciber-VCMN.

KEYWORDS | PALABRAS CLAVE

Cyberviolence, news, content analysis, text mining, cyber VAWG, topic modeling.
Ciberviolencia, noticias, análisis de contenido, análisis de textos, ciber-VCMN, modelado de temas.



1. Introduction and state of the art

Because of their gender, millions of women and girls around the world are subjected to deliberate violence and misogyny (Ellsberg et al., 2015; Tandon et al., 2015). It is a global social problem, that has long been overlooked and tolerated (García-Moreno et al., 2015). It occurs in all countries, cutting through borders, races, classes, communities, affecting victims deeply, as well as the people around them and society as a whole (Krantz & Garcia-Moreno, 2005). For that reason, the United Nations has passed resolutions to end Violence Against Women (VAW).

The use of digital technologies can play an important role for women to exercise all human rights, including the right to freedom of opinion and speech, and for them to participate fully, fairly and effectively in political, economic, cultural and social life (Abebe & Jepkiyen, 2016). However, technology has become an unwilling accomplice that enables a gender-based violence called cyberviolence. Cyberviolence is defined as the use of computer systems to cause, facilitate, or threaten violence against individuals that results in, or is likely to result in, physical, sexual, psychological or economic harm or suffering and may include the exploitation of the individual's circumstances, characteristics or vulnerabilities (Council of Europe, 2018: 5). Cyber Violence Against Women and Girls (cyber VAWG) is a type of gender-based violence that occurs online and reinforces physical violence (Van-Der-Wilk, 2018). Although both women and men may be affected by online abuse, women and girls are subjected to more severe forms of cyberviolence. These include online stalking, sextortion, online threats and blackmail, identity theft and online child pornography (including photo and video voyeurism), and "revenge porn" (Malhotra, 2015; Tandon et al., 2015). All these constitute the universe of aggression and violence women face online.

Addressing cyberviolence is important since it is often misunderstood and not considered as serious. It is imperative to note that although cyberviolence may arise online, it usually ends offline and has a detrimental effect for the victims and their families. This poses moral and psychological threats, with victims and survivors suffering anxiety and depression as a result (Nixon, 2014; Saltz et al., 2020). For example, online threats and blackmail, online incitement to suicide, online solicitation of women and girls for sexual purposes can result in self-harm or being subjected to a physical attack by the perpetrator. This can also hinder their economic empowerment and impose direct and indirect costs on individuals and society in the short- and long-term, which may include loss of livelihood (García-Moreno et al., 2015). Thus, it is critical to take action to stop cyber VAWG incidences from happening. Accordingly, there is a global pandemic of cyber VAWG (Tandon, 2015; Web Foundation, 2020) and the number of cyber VAWG cases has increased during COVID-19 (Brudvig et al., 2020; United Nations Women, 2021). The key highlights in the global survey in 180 countries conducted in February 2020 by the World Wide Web Foundation and the World Association of Girl Guides and Girl Scouts (2020) are as follows: 1) 52% of young women and girls have experienced online abuse, including threatening messages, sexual harassment and the sharing of private images without consent; 2) 64% of all respondents know someone who has experienced harassment, abuse or violence; 87% of girls think the problem is getting worse; 3) young people's top concern is the sharing of private images, videos or messages without their consent. Others are concerned about mean and humiliating messages, abusive and threatening language, sexual harassment, and sharing of false content (14%); and 4) 51% of those who have experienced online abuse say it has affected their emotional and/or physical well-being.

The Philippines, like other countries, uses technology in almost all aspects of society and the economy. An estimated 73 million people in the country were already using the Internet as of 2020 (Miniwatts Marketing Group, 2020), with around 72 million active social networking profiles (Statista Research Department, 2021a). It is not a surprise that the cases of cyberviolence in the country are also rising (Gonzales, 2019). According to the Philippine Commission on Human Rights, online sexual harassment, including peer-to-peer cyber violence, is on the rise against women and girls, with victims facing threats of rape, stalking, defamation, and even death (Aguilar, 2020). Moreover, there are studies that have highlighted the relevance of awareness in mitigating different forms of cyberviolence. On the report by the University of New Brunswick (2015), survey participants suggested that teaching people and incorporating in the curriculum strategies to deal with cyberviolence could greatly help to eliminate the problem, while

others shared that being aware of available resources is already a huge influence to eradicate cyberviolence. Further, news media play an important role in raising public awareness, framing public opinions, affecting policy formulation, and acknowledging societal issues (Carll, 2003; Sutherland et al., 2019; Zolnoori et al., 2019). Carll (2003) argued that one of the strongest instruments in combating the endemic problem of violence against women is objective news coverage and information dissemination.

Previous research that focused on news media coverage of violence against women was conducted using traditional content analysis. Over the course of four months in three Australian states, Sutherland et al. (2019) manually retrieved news headlines on violence against women from online news sites using the media monitoring and retrieval service (iSentia). Their study concluded that media reporting is an important indicator of community attitudes and beliefs about violence against women and thus a critical site through which to measure progress towards shifting social norms.

While cyber VAWG is a societal issue and a challenge that must be tackled, it has received little attention. Online news media is one of the platforms where information about cyber VAWG is reposted and one that can be explored through text mining. The increasing amount of text data available from various applications has created a need for advances in algorithmic design (Aggarwal & Zhai, 2012) such as online news articles. Several disciplines have sought to apply text mining to extract useful information and knowledge from huge amounts of text in recent years (Antons et al., 2020; Gupta et al., 2020). Text mining or text analytics is a scientific field that analyzes and processes unstructured data, which accounts for almost 95% of all big data (Gandomi & Haider, 2015). It is an interdisciplinary field encompassing data mining, statistics, artificial intelligence and machine learning, computational linguistics, library and information sciences, and databases (Miner et al., 2012). Text mining methods may address the limitations of traditional qualitative approaches. Qualitative content analysis is expensive, time-consuming, and resource intensive (Zolnoori et al., 2019) it relies on human engagement, affecting the results and limiting the amount of data (Piepenbrink & Gaur, 2017).

In addition, topic modeling is a commonly used text-mining tool for automatically organizing, analyzing, searching, and summarizing large electronic archives to uncover hidden topics and annotate the documents according to the latter. (Cho, 2019).

In this study, the text mining approach, which includes topic modeling, was utilized to efficiently analyze news items and investigate cyber VAWG reports with the aim of understanding and assessing the trend and state of cyber VAWG, as well as to increase awareness by mining news articles. Though several studies have already been conducted in the same context wherein the text mining method was applied to extract information from news articles, these studies tackled different issues. In a study by Zolnoori et al. (2019), they employed state-of-art text mining to conduct sentiment analysis and topic modeling on over 3 million Reuters news articles from 2007 to 2017 to discover coverage, sentiments and focuses for public health concerns based on top keywords from public health scientific publications. Results of the study showed that news coverage for seven public health concerns declined over time, while coverage for "sexual behavior," "pregnancy," and "air pollution" fluctuated during 2007-2017. They concluded that topic modeling represented the media's focus on public health concerns. Hori (2015) utilized the online archives of two newspapers, the Japan News and the International New York Times, to perform an exploratory study by mining for news items on "water" and "society". Here, a clustering technique was applied for dividing a collection of documents into mutually exclusive groups based on related themes.

1.1. Framing cyberviolence

Because of the broad use of ICTs and social media, as well as the ongoing pandemic of VAWG, cyber VAWG has emerged as a growing global issue with serious economic and societal implications (Council of Europe, 2018; Tandon et al., 2015). The concept and types of cyberviolence in this study are adapted from "the mapping study on cyberviolence" conducted by the Council of Europe's Cybercrime Convention Committee (2018). Many of the overlapping examples of cyber-violence types can be seen. Since there is no clear lexicon or typology of crimes categorized as cyber-violence, not all types or instances are similarly serious, and not all of them necessitate a criminal law response. These are briefly discussed in the next section.

1.1.1. Cyber harassment

Perhaps the most prevalent type of cyber-violence is cyber harassment, which involves a persistent and repetitive action, or "storm of abuse" directed at a single person with the intent of causing severe emotional distress and, in some cases, fear of physical harm. In common discourse, cyber harassment may be defined as or associated with "revenge porn" or "sextortion". Cyber harassment encompasses several acts, which include cyberbullying and revenge porn, for example. Cyberbullying can comprise any action by individuals who repeatedly communicate negative or offensive messages via electronic media to harm or discomfort others (Segura et al., 2020). It is more commonly associated with teen victims, whereas "cyberstalking, sextortion, and revenge porn" is more commonly associated with adults or young adults (Patchin & Hinduja, 2020).

Not all types of cyberbullying are inherently violent offenses. Cyberbullying acts include cyberstalking, denigrating, engaging in exclusion or gossip, falsifying an identity to post online material or flaming, impersonating, outing, phishing, sexting, and trickery (Notar et al., 2013; Runcan, 2020). Some of these acts are sometimes more serious than others. They have contributed to sexual exploitation, nonconsensual production, and posting of intimate visual images and coercion that will lead to self-harm and suicide of victims (Myers & Cowie, 2019; Saltz et al., 2020). Revenge porn refers to sexually explicit images that are circulated without the subject's consent. Other terms include "nonconsensual pornography" and "image-based abuse" (Kirchengast & Crofts, 2019). The phenomenon primarily involves a partner spreading the content online to shame or threaten the victim publicly.

1.1.2. ICT-related violations of privacy

Several types of cyberviolence infringe on victims' privacy. This can involve computer intrusions, investigating and distributing private data ("doxing"), or actions like "cyberstalking or sextortion/revenge porn" to procure, steal, expose or exploit intimate data, photo manipulation of data or images, and impersonation.

1.1.2.1. Cyberstalking. Cyberstalking refers to stalking that occurs in an electronic format. With the anonymity, ease, and efficiency of the Internet, cyberstalking can happen in a multitude of ways. Cyberstalkers can use personal information about the victim to threaten intimidation. Cyberstalkers can also send unwanted, repetitious emails or instant messages that may be hostile and threatening in nature. Cyberstalkers can also impersonate their victims online by stealing login information for an email account or social networking page and posting messages on other peers' pages. (Marcum et al., 2014: 48). As per study, cyberstalking by intimate partners is often used as a method of coercion in the context of domestic abuse (Woodlock, 2017). Accordingly, "stalking encompasses a pattern of repeated, intrusive behaviors such as following, harassing, and threatening – that cause fear in victims" (2017: 585).

1.1.2.2. Sextortion. "Sextortion is the threatened dissemination of explicit, intimate, or embarrassing images of a sexual nature without consent, usually for the purpose of procuring additional images, sexual acts, money, or something else" (Patchin & Hinduja, 2020: 1).

Sextortion starts out innocently with a request for explicit videos or images, but soon escalates. Minors are the usual targets as they do not know how to deal with predators who threaten them or pressure them of exposing their explicit images (Hong et al., 2020). According to Howard (2019), this could lead to emotional distress that affects a large number of people, with their explicit images and videos being exposed online after they fail to comply with the predator's demands (Howard, 2019).

1.1.2.3. Online sexual exploitation and sexual abuse of children. Children are the most common victims of cyberviolence, particularly when it comes to online sexual violence (Council of Europe, 2018). While child sexual exploitation and abuse are not recent, ICTs encourage and exacerbate the issue. Other forms include child pornography, child prostitution, and sexual solicitation of children.

There are several ways wherein the Internet can be abused by individuals with a perverted sexual interest in children: a) exchanging child pornography; b) locating potential victims for sexual abuse; c) engaging in inappropriate sexual communication; and d) corresponding with other individuals who have a deviant sexual interest in children (Kloess et al., 2014: 1).

1.1.2.4. ICT-related hate crime. Discrimination based on a victim's perceived personal association may inspire cyberviolence. Race, gender, faith, sexual orientation, and disability are only a few examples of these categories. Hate crime has far-reaching implications for individuals and communities, and it can lead to group disputes and the destabilization of whole communities (Council of Europe, 2018; Iganski & Sweiry, 2018).

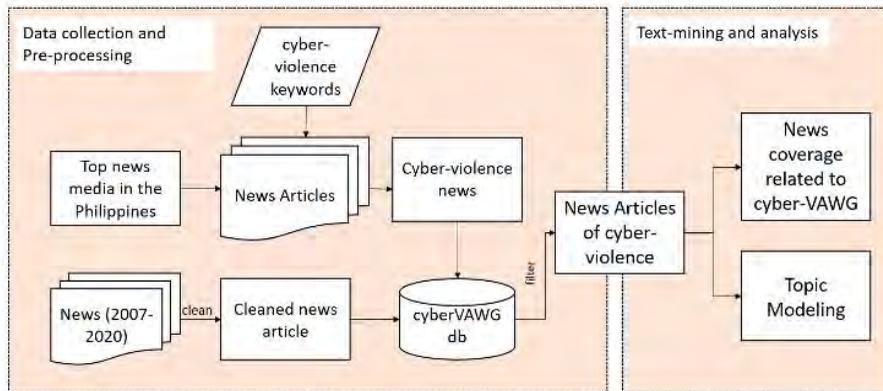
1.1.2.5. ICT-related direct threats of or physical violence. Cyberviolence may also involve explicit threats of violence or actual physical violence. Computer systems may be used in cases of murder, kidnapping, rape, and other acts of sexual assault or extortion. Medical device interference that causes injury or death, as well as cyber-attacks on critical infrastructure, are examples of direct violence (Council of Europe, 2018). Another example is "swatting", which involves deceiving an emergency service by using telephones and, in some cases, computer systems to direct local police to a particular location based on a bogus report.

1.1.2.6. Cybercrime. In light of the above definition, acts of cyberviolence such as unauthorized access to personal data, data destruction, and blocking access to a computer system or data may be categorized as cybercrime.

2. Materials and methods

In this study, text mining methods were employed to efficiently evaluate big data from news media and to extract relevant insights on news coverage of cyber VAWG issues. The methods in this study were based on the study of Zolnoori et al. (2019), which focused on mining news media for understanding health concerns. It consisted of four steps: 1) Identifying cyber-violence-related news; 2) Pre-processing; 3) Identifying the focus of news articles associated with the types of cyber VAWG; and 4) Analyzing news articles trends related to cyber VAWG. Figure 1 below shows the schematic view for mining news sites.

Figure 1. Schematic view for mining news sites



2.1. Identifying Cyber VAWG issues

The types of cyberviolence used to categorize the news were based on the Council of Europe cyberviolence framework (Council of Europe, 2018). As mentioned, cyberviolence can take many forms, including stalking, breach of privacy, sexual assault and exploitation, and bias offenses against social groups or communities.

2.2. Pre-processing news articles

a) Collecting news articles

To assess the coverage of cyber VAWG, reports from top online news sources with a strong media presence and trusted news outlets in the Philippines were analyzed (Statista Research Department, 2021b, 2021c; W3newspapers, 2020). Major online news sites were used as a vital source of our data

considering its reliability and its quality as representative of general opinion (Krawczyk et al., 2021). News articles were collected from Inquirer, GMA, Manila Bulletin, Philippine Star, Philippine Daily Inquirer and Rappler.

Published news articles using cyber VAWG-related terms were scraped through a web crawler developed using DOM (Document Object Model), which navigates through the web pages to download the news articles from an online archive of the mentioned news agencies, to collect the internal hyperlinks and dump them in the database. There were 9,842 news articles specific to cyber violence collected between January 2007 to June 2020.

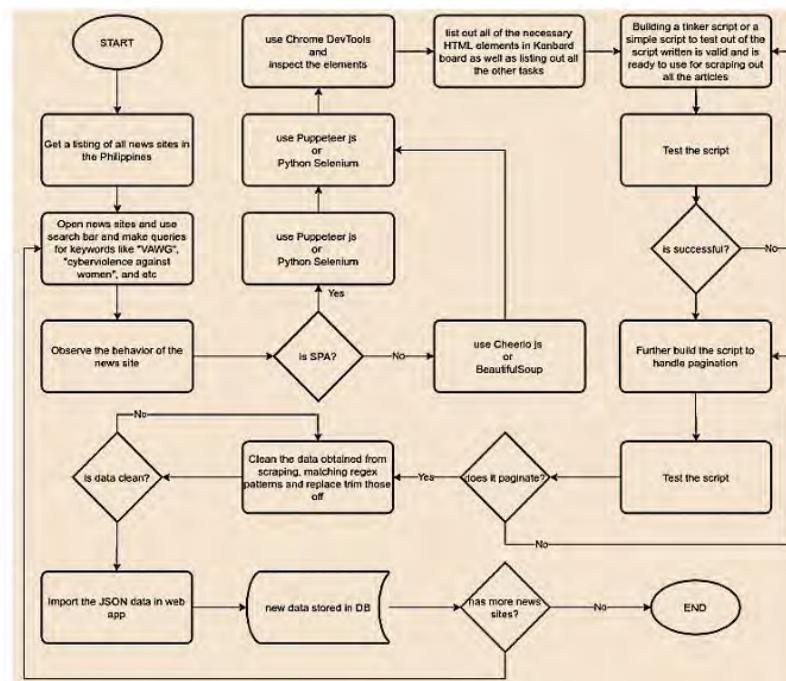
b) Cleaning news articles

Normally, scraped data is noisy. Articles can contain unwanted content or unwanted string characters, and during web scraping, the DOM was relied upon to query specific elements in the web page. A set of specific elements from an article was first located and used to get the data needed from the web page. Regular expression (regex) was used to remove unwanted strings or special characters or to replace them with whitespaces. Regex is an “object that describes a pattern of characters which are used to perform pattern-matching” (Goyvaerts, 2007). For example, characters and tags such as ... or | were removed. In addition, the article keywords were mapped based on the cyber VAWG related terms like, “girl”, “woman”, “arrested”, “nabbed”, “sextortion”, “cybercrime” by creating a regex pattern python script based on Huang (2019).

c) Filtering News Articles on cyber VAWG

A developed python script collected all articles that matched the search keywords related to the cyber violence framework. The filtered article was imported into the cyber VAWG database. The title was carefully read to assess whether it fitted the topic. If not, it was marked as ‘out of scope’ to filter out the article from the data-table. And when it was closer to the topic, the article was read and marked as part of the cyber VAWG scope.

Figure 2. Workflow for mining news sites



The detailed steps on how the data were collected and pre-processed from selected news sites are briefly explained in the steps below. These are applicable to any text-mining related study, and are depicted in the flowchart as shown in Figure 2:

- Step 1: Get a list of all the news sites that have a pagination pattern from their search results.
- Step 2: Use the page's search feature to search for keywords (e.g. VAWG, cyberviolence against women, and cyberviolence) to check if it can handle the query.
- Step 3: Observe the behavior of the news site to know which technology to use. If the website is dynamically rendering HTML elements, either use Puppeteer or Selenium library. If not, use simple libraries like Cheerio js or Beautiful Soup.
- Step 4: Use the Chrome DevTools to inspect the elements of the page, find the necessary elements to get relevant data such as the element containing the rows of news articles from the search results, the title elements, the href attribute to get the URL for that particular content, the element containing date, and the pagination elements to navigate through the other pages.
- Step 5: List out all the elements and write up simple scripts for scraping.
- Step 6: Build and execute a script out of all the limits listed to scrape data from a single page to obtain data.
- Step 7: Write the complete script to export data as a JSON format.
- Step 8: Use regular expressions to clean the data.
- Step 9: Import the JSON file into the web app and parse this so it gets uploaded into the server and stored in the database.
- Step 10: Reiterate.

2.3. Topic modeling

Topic modeling is a technique for extracting a group of words (i.e., topic) from a set of documents that best represents the information in the set. It can be thought of as a methodological approach to derive recurring themes from text corpora that is a subset of text mining (Schmiedel et al., 2018). It is carried out in studies that analyze a variety of content, including newspapers, scientific journals, and social media. In this study, topic modeling was used for content analysis by developing a python script to identify the structure of news articles based on TKM (Topic Keyword Model) package (Schneider, 2018) to identify the hidden topic structure of articles related to the five (5) categories of cyberviolence.

TKM associates a word with a subject if it or its surrounding words have a high topic association score. As a result, the issue to which a word relates is significantly impacted by the words around it. During topic modeling, TKM evaluates the dissimilarity of subjects and only maintains the topics that are significantly different from each other. TKM is able to figure out how many unique subjects to include in a text document. In addition, while inferring word distributions across a subject, TKM distinguishes between a common and a characteristic word for a topic, and it adjusts the association probability (score) of a word with a topic based on its commonness and uniqueness among topics (Zolnoori et al., 2019). The results of the topic modeling were further evaluated by creating an intuitive UI to cluster the data according to category, crime type and location.

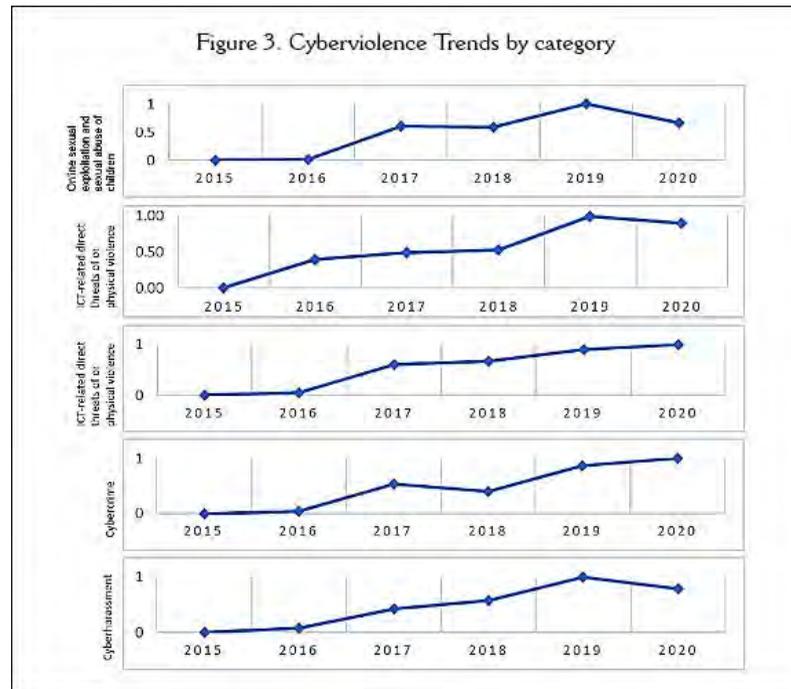
3. Analysis and findings

After excluding repeated articles, culling repeated news and irrelevant entries not related to cyber VAWG, 3,506 news articles were collected from January 1, 2015 to June 6, 2020 as shown in Table 1. These news articles are stored in a database. The result of this study is accessible at <http://app.cybervawgphilippines.co/>.

Year	Number of Articles
2015	464
2016	494
2017	601
2018	608
2019	681
2020	658

TKM was used to classify topics related to cyber VAWG issues from news articles. The news coverage of articles associated with the cyber VAWG category ("Cyber harassment", "ICT-related violations of

privacy”, “Online sexual exploitation and sexual abuse of children”, “ICT-related hate crime”, “ICT-related direct threats or actual violence” and “Cybercrime”) of each year from 2015 to 2020 were calculated to determine the five (5)-year state. Figure 3 shows the number of news articles that were rescaled relative to the highest number on each sub-figure. It showed that the news coverage of Cyber VAWG from the year 2015 up to 2020 were increasing within these years where it reached its highest peak in the year 2019. It can be noted that the data from year 2020 is up to June 6 but coverage was already high.



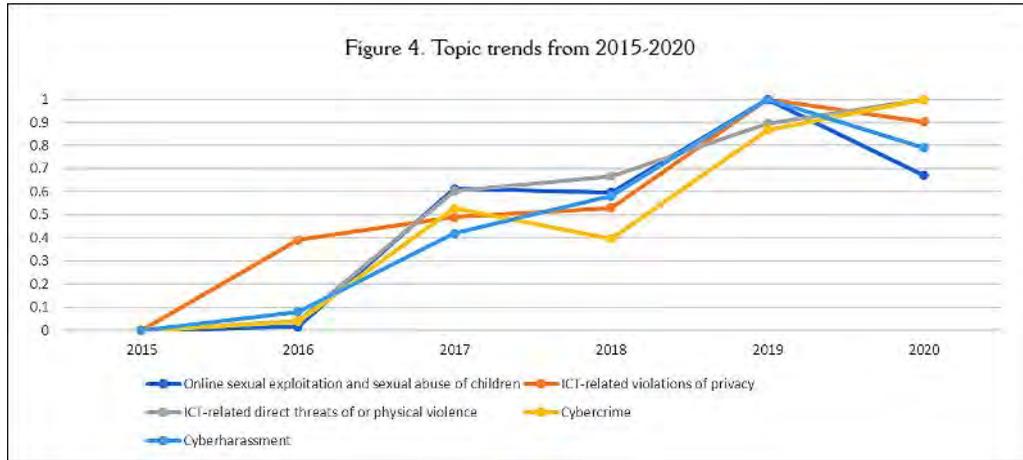
Note. Normalized numbers of news article trends by category.

The period 2015-2020 had five topics on cyber VAWG: T1 “Online sexual exploitation and sexual abuse of children”, T2 “ICT-related violations of privacy”, T3 “Cybercrime”, T4 “ICT-related direct threats of or physical violence” and T5 “Cyber harassment”. Table 2 below shows the result of the topic modeling from 2015-2020.

An examination of the proportions of the topics revealed that most of the articles were focused on topic 1 and topic 2, although all topic trends are increasing. Topic trends are shown in Figure 4 and the most frequent words are shown in Figure 6.

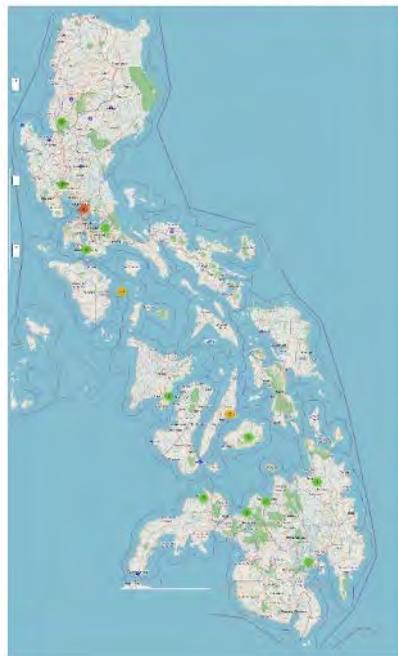
Table 2. Analysis results of topic modeling from 2015 to 2020			
Topic No	Topics	Words	Proportion
T1	Online sexual exploitation and sexual abuse of children	nude videos, prostitution, pornography, corruption, exploitation, solicitation, livestreaming sexual abuse, cybersex, sexual abuse, livestream, sextortion, grooming.	0.21
T2	ICT-related violations of privacy	computer intrusion, taking/sharing/videos /images, data manipulation, sextortion, stalking, doxing, identity theft, impersonation, fake identity, voyeurism, poser, doxing, intrusion, without consent.	0.21
T3	Cybercrime	illegal access, phishing, illegal interception, data interference, forgery, pornography, online scam, hack, libel, computer fraud, system interference, data interference, fraud, hacking/hack.	0.20
T4	ICT-related direct threats of or physical violence	rape, kidnap, murder, sexual violence, torture, extortion, blackmail, swatting, incitement to violence, critical infrastructure, attacks.	0.19
T5	Cyber harassment	coercion, denigration, defamation, reputation damage, cyberbullying, threats of violence, insults, threats, revenge porn, incitement, unwanted post, unwanted sexually explicit emails/online message, hate speech, insult, threaten, sexually harassed, stalk, aggression, impersonation, cyber/online stalking, threats.	0.19

The occurrences of Cyber VAWG in some areas in the Philippines are more frequent in big cities. The hotspots' areas that Cyber VAWG will likely occur are in Metro Manila, Quezon City, and Marikina City. In Visayas region, it is likely to occur in Bacolod City and Cebu areas. In the Mindanao region, it is most likely to occur in Cagayan De Oro City.



Moreover, social networking sites have been used to commit cybercrimes. Facebook is one of these, and it is very popular in the Philippines. Other offenders also use cybersex website platforms.

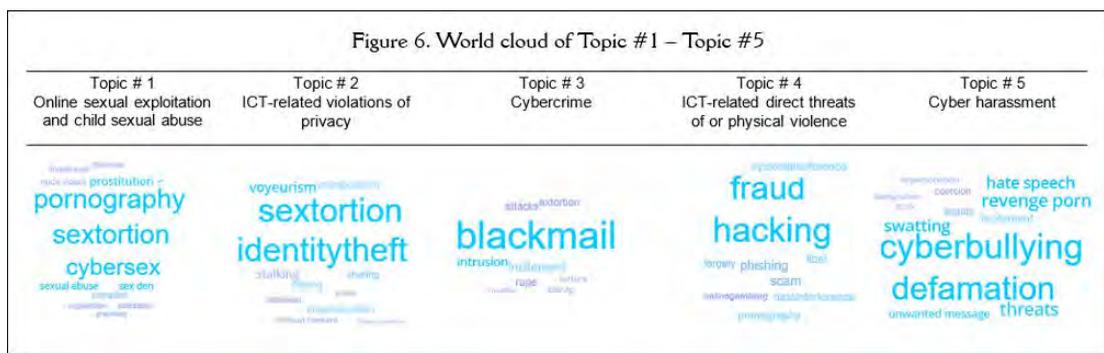
Figure 5. Cyber VAWG hotspots in the Philippines



The most frequent words of Cyber VAWG are shown in Figure 6. These are mostly related to sexual activities such as sexual exploitations, child pornography, sextortion, photo voyeurism, video voyeurism, and other crimes connected to it. The larger-size text in each category reflects heavier weights. In addition, TKM identified 12 topics of the news articles related to Online sexual exploitation and sexual abuse of children. By interpretation of the identified topic keywords, the three (3) meaningful topics of the news

articles on Online sexual exploitation and sexual abuse of children were mostly related to “pornography”, “cybersex”, “sextortion”, prostitution and “solicitation of children for sexual purposes”.

TKM also identified 16 topics of the news articles related to ICT-related violations of privacy. By interpretation of the identified topic keywords, the five (5) meaningful topics of the news articles on ICT-related violations of privacy were mostly related to “identity theft”, “sextortion”, “manipulation”, “doxing, and “impersonation”. TKM identified 16 topics of the news articles related to Cybercrime. By interpretation of the identified topic keywords, the five (5) meaningful topics of the news articles on Cybercrime were mostly related to “fraud”, “hacking”, “phishing” and “forgery”. TKM identified 11 topics of the news articles related to ICT-related direct threats or physical violence. By interpretation of the identified topic keywords, the four (4) meaningful topics of the news articles on ICT-related direct threats were mostly related to “blackmail”, “incitement to violence”, “extortion” and “rape”. TKM identified 26 topics of the news articles related to Cyberharassment. By interpretation of the identified topic keywords, the five (5) meaningful topics of the news articles on Cyber harassment were mostly related to “cyberbullying”, “defamation”, “hate speech” and “revenge porn”.



4. Discussion and conclusions

The convenience in the information flow of online platforms has enhanced civil participation in society. However, one backlash of the comfort offered by social media and internet platforms involves the underlying risks for cyber violations such as the act of cyber VAWG. This is an additional growing threat that millions of women and girls face. It is a social problem and a social issue that needs to be resolved. This study navigated the extent of cyber VAWG and the prevalence of cyber VAWG incidence in the Philippines through mining the online news media sources. Results of the study showed the magnitude of the violence that many women had experienced in the country. Through mining and analyzing all the available data from online news media, the different forms of technology-related-violence that women experience were captured: Online sexual exploitation and sexual abuse of children, ICT-related violations of privacy, cybercrime, ICT-related direct threats of or physical violence, and cyber harassment.

Mutual topics arise from the report mentioned about sextortion, pornography, cybersex, defamation, and blackmail. This result may link to the reports that despite having prostitution as illegal in the country, Philippines is still considered as one of the most popular countries for “sex tourism” (Aguilar, 2019). When it comes to forced labor in the sex industry, 99% of the victims were women and girls, and of these, 21% are children (Aguilar, 2019). These sex providers are considered victims of poverty and social change; thus, it is crucial to address related problems. Oftentimes, victims of cyber VAWG are terrified to share and report their stories and are strained to suffer in silence for fear of reprisal or social stigmatization.

The increasing number of news coverage extracted relating to cyber VAWG is a manifestation of a significant number of cases reported by the mainstream media and the number of cases of cyber VAWG is still increasing yearly despite the convenience of gaining facts and information about VAWG online. The numbers have led to the emergence of cyber VAWG as a continuing problem with potentially significant bearings on victims’ mental health. Reports point to psychological trauma, suicidal ideation and depression, and anxiety due to the fears of shaming, humiliation, harassment, and stigma associated with cyber (sexual) violence (Pashang et al 2018). Big cities also create greater returns of harassment and violence compared

to small cities maybe because perpetrators have a greater density of victims in urban areas. Cyberviolence signifies a daunting challenge to policymakers, law enforcement officials and even to those in the academe. While the Philippines has several legislations in place to protect women, the Commission on Human Rights (CHR) spokesperson pointed out that its implementation remains to be a challenge (Aguilar, 2020). There is a notable absence of public policy to incorporate more preventive programs, and to strengthen institutions and support mechanisms against cyber-VAWG. It is highly suggested to organize awareness programs and attitude-changing educational intervention intended especially for women and girls on how to safeguard their identity, and how to deal with cyberviolence incidents. Empowering women by raising awareness against cyber VAWG in any form, and understanding why such violence happens, may encourage victims to share their stories. Media could also play a vital role to change the stigma raised by society towards victims of cyber VAWG. Through their coverage, they can reach a wider audience and can inform the public of the services available against violence, guarantee fair investigation processes in cases of violence against women and ensure that obligations be translated into policies.

Awareness can change the attitudes and behavior not only by women but also by men who perpetuate or disregard the diverse forms of violence against women and girls. Consequently, it is urgent for policymakers to establish more solid laws that will sanction offenders of cyberviolence. The use of text mining to analyze newspapers can help increase the societal awareness of acceptable and sustainable cyber-VAWG solutions. Moreover, further research is needed using sentiment analysis of news data in order to verify and quantify the impact of cyber-VAWG-related issues.

Author Contribution

Idea, J.D.F.; Literature review (state of the art), J.D.F., M.A.C.T.; Methodology, J.D.F; Data analysis, J.D.F., M.A.C.T.; Results, J.D.F., M.A.C.T.; Discussion and conclusions, M.A.C.T, J.D.F.; Writing (original draft), J.D.F.; Final revisions, J.D.F., M.A.C.T.; Project design and sponsorship, J.D.F., M.A.C.T.

Funding Agency

Gender and Development Center of MSU-Iligan Institute of Technology, SO No. 02831-IIT Series of 2019

References

- Abaido, G.M. (2020). Cyberbullying on social media platforms among university students in the United Arab Emirates. *International Journal of Adolescence and Youth*, 25(1), 407-420. <https://doi.org/10.1080/02673843.2019.1669059>
- Abebe, O.J., & Jekpieny, A. (2016). Accelerating action to end violence against women and girls in Africa. *SRRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2797611>
- Aggarwal, C.C., & Zhai, C. (2012). An introduction to text mining. In C. C. Aggarwal, & C. Zhai (Eds.), *Mining text data* (pp. 1-10). Springer. https://doi.org/10.1007/978-1-4614-3223-4_1
- Aguilar, K. (2020). *Online sexual harassment on the rise, says CHR*. Inquirer. <https://bit.ly/3BpBSxw>
- Aguilar, M.G. (2019). Sex tourism in the Philippines: A basis for planning and policy making and amendments. *Journal of Tourism & Hospitality*, 8(5). <https://bit.ly/3y8k6NA>
- Antons, D., Grunwald, E., Cichy, P., & Salge, T. (2020). The application of text mining methods in innovation research: current state, evolution patterns, and development priorities. *R & D Management*, 50, 329-351. <https://doi.org/10.1111/radm.12408>
- Blanchfield, L. (2011). *United Nations system efforts to address violence against women*. Congressional Research Service. <https://bit.ly/3kz7Kdc>
- Brudvig, I., Chair, C., & Van-Der-Wilk, A. (2020). *Covid-19 and increasing domestic violence against women: The pandemic of online gender-based violence*. Word Wide Web Foundation. <https://bit.ly/36ELUgp>
- Carll, E.K. (2003). News portrayal of violence and women: implications for public policy. *American Behavioral Scientist*, 46(12), 1601-1610. <https://doi.org/10.1177/0002764203254616>
- Cho, H.W. (2019). Topic modeling. *Osong Public Health and Research Perspectives*, 10, 115-116. <https://doi.org/10.24171/j.phrp.2019.10.3.01>
- Council of Europe (Ed.) (2018). *Mapping study on cyberviolence*. <https://bit.ly/3hMAoFY>
- Ellsberg, M., Arango, D.J., Morton, M., Gennari, F., Kiplesund, S., Contreras, M., & Watts, C. (2015). Prevention of violence against women and girls: What does the evidence say? *The Lancet*, 385, 61703-61710. [https://doi.org/10.1016/S0140-6736\(14](https://doi.org/10.1016/S0140-6736(14)
- European Union Agency for Fundamental Rights (Ed.) (2010). *Violence against women: An EU-wide survey*. FRA - European Union Agency for Fundamental Rights. <https://bit.ly/2Uf3pkV>
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- García-Moreno, C., Zimmerman, C., Morris-Gehring, A., Heise, L., Amin, A., Abrahams, N., Montoya, O., Bhatte-Deosthali, P.,

- Kilonzo, N., & Watts, C. (2015). Addressing violence against women: A call to action. *The Lancet*, 385(9978), 61830-61834. [https://doi.org/10.1016/S0140-6736\(14\)61830-4](https://doi.org/10.1016/S0140-6736(14)61830-4)
- Gonzales, C. (2019). *Cybercrime on the rise over the last 6 years*. Inquirer. <https://bit.ly/3b6vv7h>
- Goyvaerts, J. (2007). *Regular expressions: The complete tutorial*. Jan Goyvaerts. <https://bit.ly/3wMQRhD>
- Gupta, A., Dengre, V., Kheruwala, H.A., & Sha, M. (2020). Comprehensive review of text-mining applications in finance. *Financial Innovation*, 6(39). <https://doi.org/10.1186/s40854-020-00205-1>
- Hong, S., Lu, N., Wu, D., Jimenez, D.E., & Milanaik, R.L. (2020). Digital sextortion: Internet predators and pediatric interventions. *Current Opinion in Pediatrics*, 32(1), 192-197. <https://doi.org/10.1097/MOP.0000000000000854>
- Hori, S. (2015). An exploratory analysis of the text mining of news articles about "water and society". *WIT Transactions on The Built Environment*, 168, 501-508. <https://doi.org/10.2495/SD150441>
- Howard, T. (2019). *Sextortion: Psychological effects experienced and seeking help and reporting among emerging adults*. [Doctoral Dissertation, Walden University]. Walden University Scholar Works. <https://bit.ly/3xNzDlv>
- Huang, M. (2019). *News mining*. [Source Code]. Github. <https://bit.ly/3wScXPQ>
- Iganski, P., & Sweiry, A. (2018). Hate crime. In A. Treviño (Ed.), *The Cambridge Handbook of Social Problems* (pp. 399-410). Cambridge University Press. <https://doi.org/10.1017/9781108550710.024>
- Kirchengast, T., & Crofts, T. (2019). The legal and policy contexts of 'revenge porn' criminalisation: The need for multiple approaches. *Oxford University Commonwealth Law Journal*, 19(1), 1-29. <https://doi.org/10.1080/14729342.2019.1580518>
- Kloess, J.A., Beech, A.R., & Harkins, L. (2014). Online child sexual exploitation: Prevalence, process, and offender characteristics. *Trauma, Violence, and Abuse*, 15, 126-139. <https://doi.org/10.1177/1524838013511543>
- Krantz, G., & Garcia-Moreno, C. (2005). Violence against women. *Journal of Epidemiology and Community Health*, 59(10), 818-821. <https://doi.org/10.1136/jech.2004.022756>
- Krawczyk, K., Chelkowski, T., Laydon, D., Mishra, S., Xifara, D., Gibert, B., Flaxman, S., Mellan, T., Schwämmle, V., Röttger, R., Hadsund, J., & Bhatt, S. (2021). Quantifying online news media coverage of the COVID-19 pandemic: Text mining study and resource. *J Med Internet Res*, 23(6), e28253. <https://doi.org/10.2196/28253>
- Malhotra, N. (2015). *Good questions on technology-related violence*. Association of Progressive Communications. <https://bit.ly/36HEqZN>
- Marcum, C.D., Higgins, G.E., & Ricketts, M.L. (2014). Juveniles and cyber stalking in the United States: An analysis of theoretical predictors of patterns of online perpetration. *International Journal of Cyber Criminology*, 8(1), 47-56. <https://bit.ly/2UpET0b>
- Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., & Nisbet, R. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Elsevier. <https://bit.ly/3BrK0O3>
- Miniwatts Marketing Group (Ed.) (2021). *Internet usage statistics the Internet big picture*. Internet World Stats. <https://bit.ly/3hMefYk>
- Myers, C.A., & Cowie, H. (2019). Cyberbullying across the lifespan of education: Issues and interventions from school to university. *International Journal of Environmental Research and Public Health*, 16(7). <https://doi.org/10.3390/ijerph16071217>
- Nixon, C. (2014). Current perspectives: The impact of cyberbullying on adolescent health. *Adolescent Health, Medicine and Therapeutics*, 5, 143-158. <https://doi.org/10.2147/AHMT.S36456>
- Notar, C.E., Padgett, S., & Roden, J. (2013). Cyberbullying: Resources for intervention and prevention. *Universal Journal of Educational Research*, 1(3), 133-145. <https://doi.org/10.13189/ujer.2013.010301>
- Pashang, S., Clarke, J., Khanlou, N., & Degendorfer, K. (2018). Redefining cyber sexual violence against emerging young women: toward conceptual clarity. In S. Pashang, N. Khanlou, & J. Clarke (Eds.), *Today's youth and mental health. Advances in Mental Health and Addiction* (pp. 77-97). Springer. https://doi.org/10.1007/978-3-319-64838-5_5
- Patchin, J.W., & Hinduja, S. (2020). Sextortion among adolescents: Results from a national survey of. *Sexual Abuse*, 32(1), 30-54. <https://doi.org/10.1177/1079063218800469>
- Piepenbrink, A., & Gaur, A.S. (2017). Topic models as a novel approach to identify themes in content analysis: the example of organizational research methods. In *Academy of Management Proceeding* (pp. 11335-11335). Academy of Management. <https://doi.org/10.5465/AMBPP.2017.141>
- Runcan, R. (2020). Conflict solution in cyberbullying. *Revista de Asisten Sociala*, 2, 45-57. <https://bit.ly/2UZnnje>
- Saltz, S.B., Rozon, M., Pogge, D.L., & Harvey, P.D. (2020). Cyberbullying and its relationship to current symptoms and history of early life trauma: A study of adolescents in an acute inpatient psychiatric unit. *The Journal of Clinical Psychiatry*, (8), 81-81. <https://doi.org/10.4088/JCP.18m12170>
- Schmiedel, T., Müller, O., & Brocke, J.V. (2018). Topic modeling as a strategy of inquiry in organizational research. *Organizational Research Methods*, 22(4), 941-968. <https://doi.org/10.1177/1094428118773858>
- Schneider, J. (2018). *TKM Package*. [Source Code]. Github. <https://bit.ly/2O9sdXY>
- Segura, L., Estévez, J.F., & Estévez, E. (2020). Empathy and emotional intelligence in adolescent cyber aggressors and cyber victims. *International Journal of Environmental Research and Public Health*, (13), 17-17. <https://doi.org/10.3390/ijerph17134681>
- Statista Research Department (Ed.) (2021a). *Number of social network users in the Philippines from 2017 to 2019 with a forecast until 2025*. Statista. <https://bit.ly/3rBYkPE>
- Statista Research Department (Ed.) (2021b). *Trust in news outlets in the Philippines in 2021, by brand*. Statista. <https://bit.ly/3zkhBla>
- Statista Research Department (Ed.) (2021c). *Leading news websites Philippines 2019 by daily pageviews per visitor*. Statista. <https://bit.ly/3kziPLi>

- Sutherland, G., Eastal, P., Holland, K., & Vaughan, C. (2019). Mediated representations of violence against women in the mainstream news in Australia. *BMC Public Health*, 502, 19. <https://doi.org/10.1186/s12889-019-6793-2>
- Tandon, N. (2015). *Cyber violence against women and girls*. ITU News. <https://bit.ly/36GMLgr>
- Tandon, N., & Pritchard, S. (2015). *Cyber violence against women and girls: A world-wide wake-up call*. UN Broadband Commission for Digital Development. <https://bit.ly/3kyivfK>
- United Nations Women (Ed.) (2021). *The shadow pandemic: Violence against woman during COVID-19*. UN Women. <https://bit.ly/3xPml81>
- University of New Brunswick (Ed.) (2015). *Behind the screen: assessing needs for the prevention and elimination of cyber violence against young women in New Brunswick*. University of New Brunswick. <https://bit.ly/3etWI51>
- Van-Der-Wilk, A. (2018). *Cyber violence and hate speech online against women*. The European Parliament's Committee on Women's Rights and Gender Equality. <https://bit.ly/3rhQvhQ>
- W3newspapers (Ed.) (2020). *Philippine newspapers and news sites*. <https://bit.ly/2URILXQ>
- Web Foundation (Ed.) (2020). *There's a pandemic of online violence against women and girls*. World Wide Web Foundation. <https://bit.ly/3BakGw2>
- Woodlock, D. (2017). The abuse of technology in domestic violence and stalking. *Violence Against Women*, 23(5), 584-602. <https://doi.org/10.1177/1077801216646277>
- World Wide Web Foundation & World Association of Girl Guides and Girl Scouts (Ed.) (2020). *Survey - Young people's experience of online harassment*. <https://bit.ly/3kpkKxX>
- Zolnoori, M., Huang, M., Patten, C.A., Balls-Berry, J., Goudarzvand, S., Brockman, T., Sagheb, E., & Yao, L. (2019). Mining news media for understanding public health concerns. *Journal of Clinical and Translational Science*, 5(1). <https://doi.org/10.1017/cts.2019.434>