Big Data Techniques to Study the Impact of Gender-Based Violence in the Spanish News Media

Hugo J. Bello
University of Valladolid, Soria, Spain
ORCID: 0000-0002-3687-1938

Lourdes Jiménez Navascués
University of Valladolid, Soria, Spain
ORCID: 0000-0002-7801-5398

Nora Palomar-Ciria
Complejo Asistencial de Soria, Soria, Spain
ORCID: 0000-0002-2293-3106

Celia Lozano
AI Department, Bosonit, Logroño, Spain
ORCID: 0000-0001-5830-1154

Elisa Gallego
the Carlos III Health Institute in Madrid, Spain
ORCID: 0000-0002-5323-587X

Abstract: Despite being an underreported topic in the news media, gender-based violence (GBV) undermines the health, dignity, security and autonomy of its victims. Research has studied many of the factors that generate or maintain this kind of violence. However, the influence of the media is still uncertain. This paper used Big Data techniques to explore how GBV is depicted and reported in digital news media. By feeding neural networks with news, the topic information associated with each article can be recovered. Our findings show a relationship between GBV news and public awareness, the effect of well-known GBV cases, and the intrinsic thematic relationship of GBV news with *justice* themes.

Keywords: Big Data; domestic violence; gender-based violence; media; news.

INTRODUCTION

Gender-based violence (GBV) is directly related at an individual based on their biological sex or gender identity (Heise et al., 2002; Violenciagenero (2019) and includes verbal, physical, sexual, and psychological abuse, occurring in either public or private life. All genders globally experience GBV, but women are disproportionately harmed (Russo & Pirlott, 2006; UN Women, 2021; WHO, 2013).

Despite its prevalence, the media largely misreport GBV (Boyle, 2005; Cullen et al., 2019; Moorti, 2002). Recently, the voices of victims, survivors and activists have once again put the spotlight on the issue and raised public awareness by the media through movements such as #MeToo, #TimesUp, #Niunamenos, #NotOneMore, #BalanceTonPorc (Alexander, 2006; Luengo, 2018).

News media are key to understanding how society and the general population react to a topic (Cullen et al., 2019; Dijk, 1995). News can influence public perceptions and consequently, social policies (Carlyle et al., 2008; Cullen et al., 2019; Gillespie et al., 2013; Maydell, 2018), which is the case with GBV with news media being one of the main sources of information. Media may report available help resources added to the piece of news (Comas-d'Argermir, 2015). In Spain, where more than a thousand women have been killed in the last two decades (Menéndez, 2014; Teruelo, 2011), many notorious cases have attracted media attention as never before. For instance, The Manada [wolf pack] case (see News Desk (2019) – a group sexual assault in 2016 that lead to a series of protests across the country. This event (and its corresponding social, juridic and public response) motivated us to study the reporting of GBV news in Spain (LSE, 2020) utilizing Big Data analysis techniques.

Big Data is a novel tool for examining the reporting of GBV in digital media. Moreover, its applications might be used in the future to prevent this kind of violence through the news media. Also, this method could be used for the study and research of other social problems.

THEORETICAL BACKGROUND

Gender-based violence is a pervasive global issue that affects the health, dignity, security, and autonomy of its victims, particularly women and girls. GBV is "violence that is directed against a person on the basis of their gender or sex and includes acts that inflict physical, mental or sexual harm or suffering, threats of such acts, coercion, and other deprivations of liberty" (UN Women, 2012). While GBV can result in physical injuries, psychological trauma, and long-term health consequences (WHO, 2013) of individuals, it also has social and economic costs, including reduced productivity, increased healthcare expenses, and decreased educational attainment (Heise et al., 2019). Research has explored numerous factors to understand the causes and drivers of GBV, including cultural and social norms, economic inequality, and political instability (Heise et al., 2019; Jewkes et al., 2017). However, the role of the media in perpetuating or challenging GBV is still uncertain. Research suggests that media representations of GBV can have significant effects on public attitudes and understanding of the issue (Baker & Ascione, 1995; DeFleur & Ball-Rokeach, 1989). However, the nature

and extent of the influence of media on GBV is not well understood. Research on the media coverage of violence against women has shown that it is often characterized by various patterns of problematic reporting (Boyle, 2005; Cuklanz, 2014). One common issue is the use of victim-blaming language and framing (Moorti, 2002; Wong & Lee, 2018). News media may focus on the victim's behavior or clothing as a cause of the violence, rather than the perpetrator's actions (Howe, 1997). This approach shifts the responsibility for the violence onto the victim and reinforces harmful stereotypes about women's behavior and responsibility for their own safety. Another point is the sensationalizing of violence against women, which can result in graphic and voyeuristic depictions of violence that can further traumatize victims and reinforce harmful stereotypes about women's vulnerability. The use of sensationalist headlines and images can also trivialize the seriousness of the issue and lead to a lack of public engagement (Tranchese & Zollo, 2013). Furthermore, the media often fail to provide context and information about the social and cultural factors that contribute to violence against women. For example, the role of patriarchal beliefs and gender inequality in perpetuating violence against women may be downplayed or ignored (Rollè et al., 2020; Sutherland et al., 2019). In addition, the media may present a narrow and limited view of the aspects that constitute violence against women, with a focus on physical violence rather than other forms of abuse, such as emotional or economic abuse (Easteal et al., 2018). Recent advances in technology and the availability of Big Data have provided an opportunity to explore the manner the media depict and report GBV. By analyzing large datasets of digital news articles, it is possible to gain insights into the ways in which GBV is represented and reported, as well as the factors that shape these representations. This paper aims to use big data techniques to explore the role of the media in perpetuating or challenging GBV. By feeding neural networks with the contents of digital news articles, we recover the topic information associated with each article and analyze the representations of GBV.

MATERIAL AND METHODS

DATASET EXTRACTION PROCESS

We used a Big Data technique called 'web scraping' to extract a great amount of news from the main Spanish online newspapers. This technique requires a group of servers querying massive quantities of data from public pages on the Internet.

We deployed a network of cloud servers and a local server to break up the load of the work. One of the servers stored data and the others searched the Internet for news items of the seven selected Spanish online newspapers— *La Vanguardia,*

El País, El Mundo, ABC, 20minutos, Público and *diario.es*. A local server first trained the neural networks to classify the subjects in the news, and secondly analyzed the full database and by using neural network models, calculated the probability that each news item had of covering the subject of GBV.

This process resulted in a dataset of 784 259 news items from January 2005 to March 2020. Online news media classify every piece of news with tags, which are keywords helping to describe the text and allows search functions to find the tag. Tags are not only descriptive of topics like justice, economy, politics, international, technology and health but also combinations of several data, such as media outlet, title, content, and date.

DATA MINING USING NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a subfield of Artificial Intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. Its goal is to build algorithms capable of understanding the contents of documents, including the contextual nuances of the language within them. The technique can then accurately extract information and insights contained in the documents as well as categorize and organize them.

After the data extraction process using web scraping, we used NLP (Chapman et al., 2011) to analyze local news articles from the selected Spanish online media. Natural language processing consists of a type of data analysis focused on transforming free text (i.e. unstructured data) in documents and databases into normalized, structured data suitable for analysis (Coppersmith et al., 2018; Farzindar & Inkpen, 2017; Mori & Haruno, 2021; Goldberg, 2017). This process enabled us to classify the news. To obtain the main subjects of written media content, we used neural networks (Mikolov, 2017; Goldberg, 2017). Neural networks, with an accuracy of 98%, were also used to obtain news topics related to GBV. This enabled us to formulate three research questions (RQ):

- RQ1—is there a relationship between GBV news topics and public awareness;
- RQ2— do high profile GBV cases in the media affect public awareness,
- RQ3—are there any intrinsic themes in media coverage of GBV.

TOPOLOGICAL DATA ANALYSIS METHODS (MAPPER ALGORITHM)

Topological based Data Analysis (TDA) is an approach to the analysis of datasets using techniques from topology. Extraction of information from datasets that are high-dimensional, incomplete and noisy is generally challenging. The application of TDA provides a general framework to analyze such data in a manner that is insensitive to the particular metric chosen and provides dimensionality reduction and robustness against noise. Beyond this, TDA inherits functoriality,

a fundamental concept of modern mathematics, from its topological nature, which allows it to adapt to new mathematical tools.

One key part of our TDA analysis is the use of the 'Mapper' algorithm (Singh et al., 2007), which focuses on the way that the parts of a system are connected rather than the distance between them. This is useful for analyzing and visualizing data. We used Kepler Mapper, the flexible python implementation of the Mapper algorithm, to study the connectivity of GBV news and other news within a specific period (Veen et al., 2019).

DATA EXTRACTION AND TOPIC CLASSIFICATION

The study used NLP techniques to transform unstructured text into structured data. Also, it allowed topic classification of news based on the list of tags each piece of news contained (such as justice, economy, politics, international, technology, feminism.). However, Spanish media—both analogue and digital—do not commonly use the phrase *la violencia de género'* (GBV) as a tag or topic. To bring light to this issue and, in addition, enhance the description tags, we extracted the subject for each news article and the probability it was related to GBV. To achieve this outcome, we applied a stack protocol that consisted of two consecutive neural network models for NLP: first, by analyzing the content of each text, we extracted the general subject classification. Secondly, we applied a binary GBV classification, which produced a probability rate, where zero means a lack of GBV content in the individual piece of news considering tags such as feminine, domestic violence, female sexual abuse and sexual crimes against women.

ANOMALY DETECTION

In data analysis, anomaly detection (also referred to as *outlier detection*) is the identification of rare items, events or observations which deviate significantly from the majority of the data and do not conform to a well-defined notion of normal behavior. Such observations may arouse suspicions of being generated by another mechanism, or appear inconsistent with the remainder of that set of data.

We used the Prophet forecasting model (a Facebook open software algorithm) (Shen et al., 2020) on our GBV average probability. Prophet is a procedure for analyzing time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. This model works best with time series that have strong seasonal effects and several seasons of historical data.

RESULTS

GENDER-BASED VIOLENCE IN NEWS MEDIA AND PUBLIC AWARENESS—RQ1

To understand the GBV press coverage in more detail, we analyzed the GBV probability over time, i.e. the temporal changes (Tsay, 1989; Mills, 2015). Considering the whole corpus of news, we measured the temporal monthly average of GBV probability. The larger the probability, the greater the chance to be related to GBV in a certain month. Figure 1, Chart (a) shows how GBV probability rises over time, by a factor of 1.8 in the last three years. Remarkably, 20% of the news in 2020 had GBV connotations. We compared such enhancement with public awareness concerning gender movements. Thus, we examined the social perception of GVB using the monthly survey CIS (Centre for Sociological Research, a Spanish public research institute) (Violenciagenero 2019). We applied the same seasonal-trend decomposition model see Figure 1, Chart (b). Comparing Charts (a) and (b), the monthly GBV probability in news and gender violence awareness have the same upward trend. Furthermore, quantitative information about the consistent trend behavior is obtained from the time correlation between average GBV probability against the opinion survey about domestic violence as a function of the displacement of one relative to the other, see Chart (c).

Figure 1. Gender Based Violence (GBV) Time series.
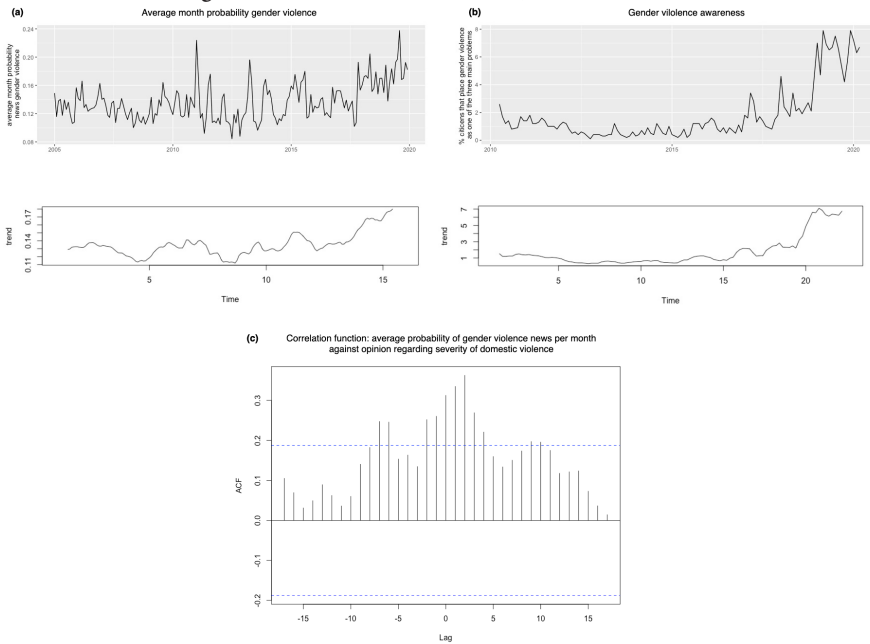
Source: Authors

Key: Chart (a) Monthly GBV probability and its corresponding trend. This time series was obtained by using NPL over each new in the dataset and then grouping by month.

Chart (b) Monthly percent of citizens that place gender violence as one of the three main problems (Centre for Sociological Research data) and its corresponding trend.

Chart (c) Cross-Correlation Function of the Temporal GBV and the Temporal feedback opinion. Cross correlation measures the relationship between two time series, in particular it measures if lags in the first time series can be used to predict future values of the second.

THE IMPACT OF FAMOUS GENDER VIOLENCE CASES—RQ2

Our results show that the GBV monthly probability has increased over the years. A similar trend can be observed in the daily grouped data. Figure 2 shows the temporal trends (green lines) of the GBV data (black symbols). Although it seems that the data fits well in the plot, certain days are out of the range: these are anomalies. Anomalies are understood as values that are sufficiently different or that do not fit the trend of the previous ones and are depicted as red spots.

Figure 2. Prophet forecast model analysis showing Gender Based Violence (GBV) anomalies.
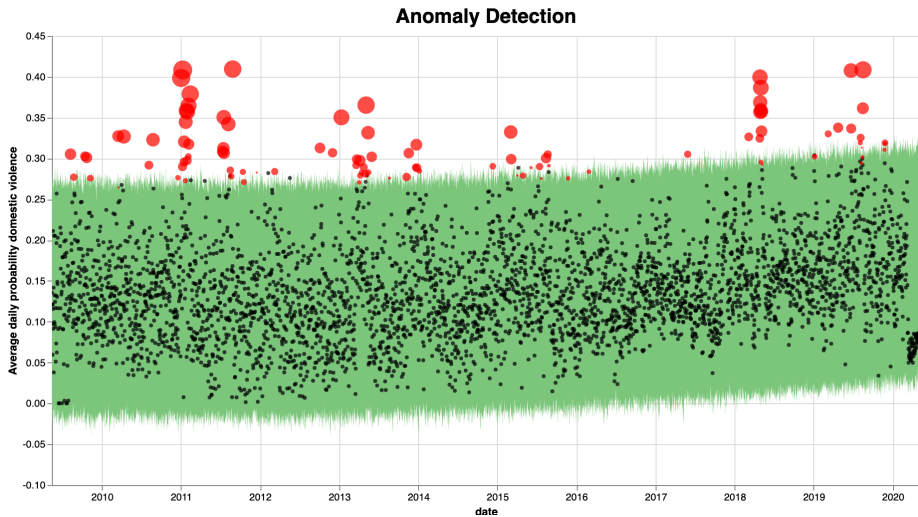
Source: Authors

Key: Prophet forecast model fit of the temporal GBV probability. Solid symbols = GBV data; green lines = the fitting curve; red symbols = the anomalies.

Table 1 shows the list of anomalies detected during 2018 and 2019 by the previous analysis (marked as red points in Figure 2), the corresponding average GBV probabilities, and the event in the news. Note that several of these anomalies correspond to the court's sentence at the end of the 'La Manada [Wolf-pack] rape case' (Aurrekoetxea, 2020; Egea, 2019; News Desk 2019).

Table 1. Gender Based Violence (GBV) anomalies in Spain, 2018–2019, with the corresponding average probabilities and the event.

Date	GBV Average Probability	Event
2019–08–17	0.4085262	News coverage of the murder of a female surgeon in a GBV case. https://www.elmundo.es/madrid/2019/08/17/5d57c84a21efa0ab3b8b45cb.html
2019–06–22	0.4075814	Concerning the Manada (Wolf-pack) case, the Supreme Court of Spain upgraded convictions for sexual abuse to that of continuous sexual assault, and handed down 15-year prison terms.
2018–05–02	0.3866444	News coverage of the case of a GBV victim who was killed in Guadalix. https://elpais.com/politica/2018/02/05/actualidad/1517833337_571770.html
2018–04–29	0.3687508	The five members of the Manada (Wolf-pack) were each condemned to nine years on charges of sexual abuse.
2018–04–28	0.3995969	The five members of the Manada (Wolf-pack) were each condemned to nine years on the charges of sexual abuse.

Source: Authors

To visualize the anomalies and the dates of high GBV probability in more detail, we used a heatmap plot (see Figure 3), where all the dates of 2018, 2019 and 2020 are shown. Lighter colors relate to higher average GBV probabilities. Surprisingly, high probabilities of GBV not only correspond to the anomaly day, but also extended for several days. In addition, the heatmap shows how the years get lighter (in terms of probability of GBV news), which confirms the upward trend observed in the time series (see Figure 1).

Figure 3. Gender Based Violence news heatmap for probabilities each day between 2018 to 2020.
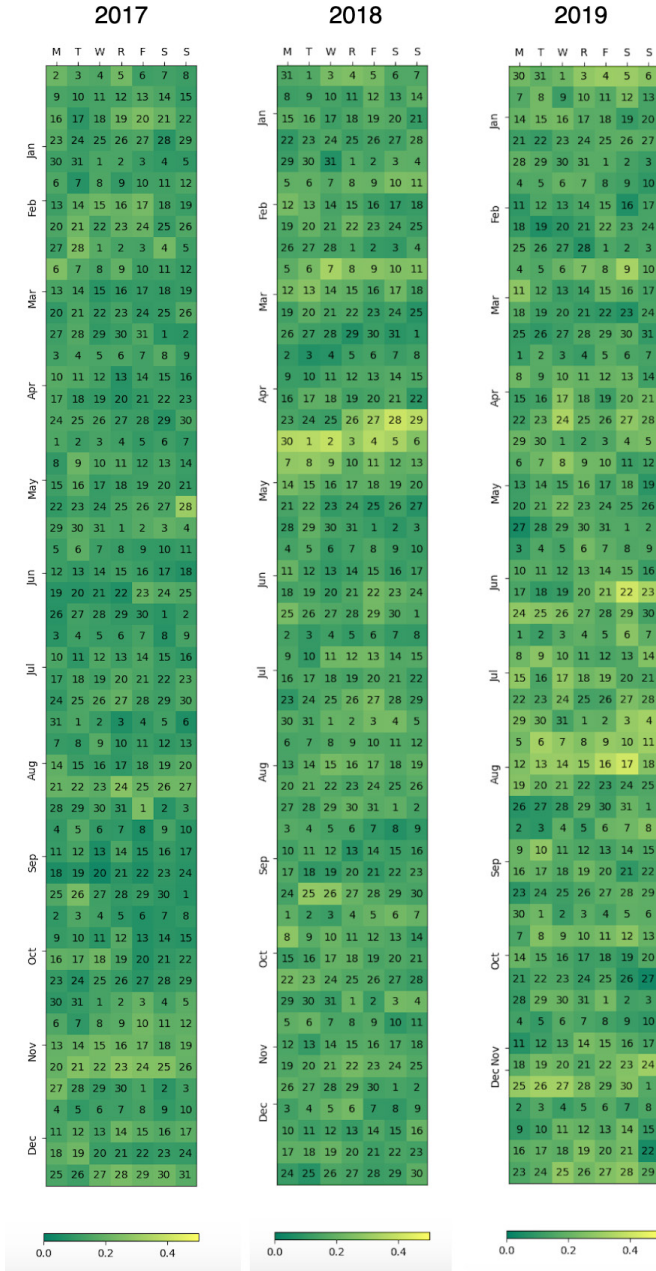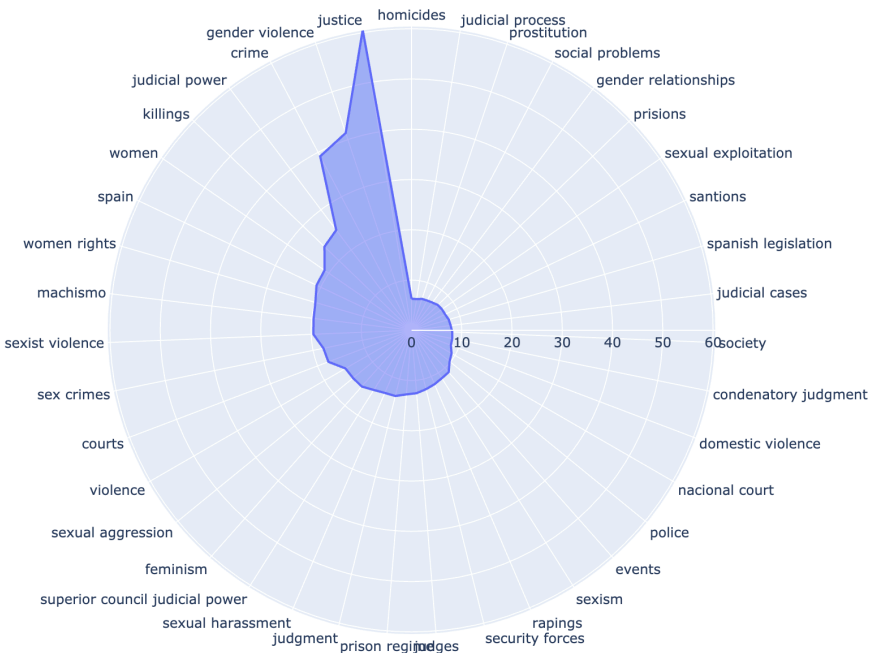

Source: Authors

Key: Colour (= probability) Dark Green (0); Semi-Dark Green (0.2); Semi-Light Green (0.4); Pale Green (0.5). The values depicted were obtained applying Natural Language Process to all the news in the data set and averaging per day.

SUBJECT ANALYSIS OF GENDER-BASED VIOLENCE NEWS

Next, we focused on the intrinsic connection of subjects (or tags) in the news—RQ3. We used the extracted GBV probability to discern, which news were about this issue. We considered a piece of news to mainly cover GBV if the probability returned by the neural network was greater than 0.9999, i.e., if the neural network confirmed it almost without doubt. We thus obtained a set of 5375 news items that covered GBV. Then, we proceeded with a subject classification neural network to tag these news items and to find out their main subjects.

Figure 4. Top topic classification of the Gender Based Violence news by subject frequency (percent).
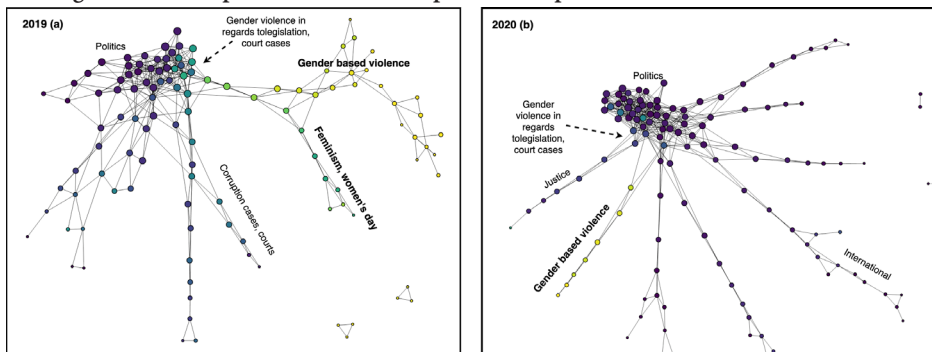

Source: Authors

Figure 4 shows the resulting classification. The most common tags were justice, gender violence and crime. Using TDA, we obtained a mapper algorithm in a diagram (Figure 5) that summarizes the closeness of the news' tags with a geometric interpretation. The resulting graph draws a node for each cluster of news for those close in the context of their themes. The size of the node reflects the quantity of news items they contain. Note that if these nodes overlap, i.e., there are news items belonging to more than one cluster, the nodes will be joined with segments. This allowed us to see the ways that not only subjects change but also news organize themselves in terms of closeness.

In Figure 5, the colour schematic is dark blue (cool-low probability) through green (warm) to yellow (warmer-high probability). The yellow nodes represent news (or groups of news) with higher GBV probability. Each branch represents news' clusters whose subjects evolve similarly. As Figure 5 (2019a) shows, the GBV branch forms at a junction between feminism and politics. These findings fully accord with the central node of the GBV in regard to legislation and court cases. We repeated the analysis for news in the following year, see Figure 5 (2020b), where the GBV branch merges with that of justice.

Figure 5. Interdependence between topics in the Spanish news for 2019a and 2020b.

Source: Authors

DISCUSSION

Gender-based violence (GBV) is a worldwide issue that despite its prevalence has been grossly under-covered in public institutions and also, in the news media (Cullen et al., 2019). Fortunately, awareness of the problem has been growing for years, thanks to social movements and increased media coverage (Luengo, 2018). The media has helped in spreading social awareness about GBV (Luengo, 2018) and transforming a public issue into a social problem (Alexander, 2006). However, GBV depictions continue to be represented wrongly in the news media (Boyle, 2005; Cullen et al., 2019).

In this paper, we presented an analysis of news coverage of GBV using novel Big Data techniques. Our study was limited to Spanish media, but we believe that our methodology has general validity. Our results show that GBV news do not have a specific topic or tag, which contributes to the invisibility of the problem and forces the news to be classified by numerous related tags. However, justice subjects are highly related to GBV news, which can be explained by the famous cases that end up in court or that public attention is fixed on the judicial process.

Additionally, when covering GBV, the media usually present legal communications or discuss protecting the laws.

We found that the media have of late begun to spotlight GBV, the recent feminine movement and gender victims' voices. Our analysis shows an enhancement of the public perception of GBV in recent years resulting from the amount of news about the topic, which echoes well-known events. Moreover, we found that the public opinion lags behind the news media for several months, which means that after a relevant GBV event, people react to it. Indeed, the proposed anomaly detection method showed that a match between the anomaly and the GBV cases confirms this impact. Our results demonstrated that the impact of a GBV event lasts for several days and media has an influential role in raising both social awareness and conscience for the topic.

Our study's strength is the utilization of a GBV probability for each news and how it evolves over a certain period. We mapped how GBV probability appears daily and pinpointed related events that push both the topic's increase in the news and public awareness (UN Women, 2017). Moreover, we used TDA techniques to extract how certain topics (or tags) are related in the news. Our approach shows that GBV tends to relate to the subject of justice and rarely with that of feminism.

Violence against women is a serious social issue and appears in the news media rather frequently. The way that news media portray GBV has significant implications for public perception and policy responses. Research on the media coverage of violence against women has shown that GBV is often characterized by various patterns of problematic reporting. Women's role is usually depicted in the news media as victims (Bleiker & Hutchison, 2019; Busso et al., 2020) and this influences the way GBV is covered. Still, information regarding GBV is biased and the problem it poses is usually silenced by explicit details (Buiten & Salo, 2007; Cuklanz, 2014; ElSherief et al., 2017; Wong & Lee, 2018). Overall, the portrayal of violence against women in news media is a complex issue that requires critical examination and attention. Media outlets should avoid victim-blaming language, but provide context, and accurately reflect the experiences of survivors to raise awareness and promote effective policy responses (Easteal et al., 2015; Sutherland et al., 2019; Wolf, 2018).

Previous research on GBV has utilized Big Data techniques to show some interesting outcomes (Subramani et al., 2018, 2019; Xue et al., 2019). Those works tend to focus on social media, which is an important way of conveying social concern, but none of them approaches this issue in the way, this study did.

Our findings provided a new point of view and representation of the news about GBV that should lead to better reporting. On this line, the creation or improvement of already existing guidelines could be an interesting objective (GenderIT, 2012; UNESCO, 2019). Hence, news media would help raise awareness

on the problem that GBV poses, and thus, promote resources for its prevention. Furthermore, as the media reflects the concerns of our societies, it should help to neutralize GVB and combat it.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request. The source code used in this study is publicly available on GitHub.

REFERENCES

Alexander, J. (2006). *The civil sphere* (Oxford). Oxford University Press.

Aurrekoetxea, M. (2020). San fermines #la manada case [The San Fermin festival—7–14 July— The Wolfpack case]: An exploratory analysis of social support for victims of sexual violence on Twitter. *Computers in Human Behavior*, *108*, 106299. https://doi.org/10.1016/j.chb.2020.106299

Baker, L., & Ascione, F. (1995). Effects of televised violence on aggression. *Journal of Social Issues*, *51*(3), 85–106.

Bleiker, R., & Hutchison, E. (2019). Gender and violence in news media and photography. In *Handbook on Gender and Violence* (pp. 231–247). Edward Elgar Publishing. https://www.elgaronline.com/display/edcoll/9781788114684/9781788114684.00023.xml

Boyle, K. (2005). *Media and violence: Gendering the debates*. Thousand Oaks, Sage.

Buiten, D., & Salo, E. (2007). Silences stifling transformation: Misogyny and gender-based violence in the media. *Agenda*, *21*(71), 115–121. https://doi.org/10.1080/10130950.2007.9674819

Busso, L., Combei, C. R., & Tordini, O. (2020). "Narrating Gender Violence A Corpus-Based Study on the Representation of Gender-Based Violence in Italian Media". In Giusti, G. & Iannàccaro, G. (Eds) *Language, Gender and Hate Speech. A Multidisciplinary Approach*, Fondazione Università Ca' Foscari. https://publications.aston.ac.uk/id/eprint/42283/

Carlyle, K. E., Slater, M. D., & Chakroff, J. L. (2008). Newspaper Coverage of Intimate Partner Violence: Skewing Representations of Risk. *The Journal of Communication*, *58*(1), 168–186. https://doi.org/10.1111/j.1460–2466.2007.00379.x

Chapman, W. W., Nadkarni, P. M., Hirschman, L., D'Avolio, L. W., Savova, G. K., & Uzuner, O. (2011). Overcoming barriers to NLP for clinical text: The role of shared tasks and the need for additional creative solutions. *Journal of the American Medical Informatics Association*, *18*(5), 540–543. https://doi.org/10.1136/amiajnl-2011–000465

Comas-d'Argermir, D. (2015). News of partner feminicides: The shift from private issue to public problem. *European Journal of Communication*, *30*(2), 121–136.

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights*, *10*, 1178222618792860. https://doi.org/10.1177/1178222618792860

Cuklanz, L. M. (2014). Mass media representation gendered violence. In *The Routledge Companion to Media & Gender*. Routledge.

Cullen, P., O'Brien, A., & Corcoran, M. (2019). Reporting on domestic violence in the Irish media: An exploratory study of journalists' perceptions and practices. *Media, Culture & Society*, *41*(6), 774–790. https://doi.org/10.1177/0163443718823141

DeFleur, M., & Ball-Rokeach, S. (1989). *Theories of mass communication* (5th ed.). Longman.

Dijk, T. A. van. (1995). *Power and the news media: Political Communication and Action*. NJ: Hampton Press Cresskill.

Easteal, P., Holland, K., & Judd, K. (2015). Enduring themes and silences in media portrayals of violence against women. *Women's Studies International Forum*, *48*, 103–113. https://doi.org/10.1016/j.wsif.2014.10.015

Easteal, P., Holland, K., Breen, M. D., Vaughan, C., & Sutherland, G. (2018). Australian Media Messages: Critical Discourse Analysis of Two Intimate Homicides Involving Domestic Violence. *Violence Against Women*, 1077801218780364. https://doi.org/10.1177/1077801218780364

Egea, M. A. (2019). Subjetividad y violación social. El caso de la manada [Subjectivity and social violation. The 'wolf pack' case]. *Tropelías: Revista de Teoría de la Literatura y Literatura Comparada*, *31*, Article 31. https://doi.org/10.26754/ojs_tropelias/tropelias.2019313197

ElSherief, M., Belding, E., & Nguyen, D. (2017). #NotOkay: Understanding Gender-Based Violence in Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, *11*(1), Article 1. https://doi.org/10.1609/icwsm.v11i1.14877

Farzindar, A., & Inkpen, D. (2017). *Natural Language Processing for Social Media: Second Edition*. Morgan & Claypool Publishers.

GenderIT (2012). *Learning Resource Kit for Gender-Ethical Journalism and Media House Policy*. GenderIT.Org https://genderit.org/resources/learning-resource-kit-gender-ethical-journalism-and-media-house-policy

Gillespie, L. K., Richards, T. N., Givens, E. M., & Smith, M. D. (2013). Framing deadly domestic violence: Why the media's spin matters in newspaper coverage of femicide. *Violence Against Women*, *19*(2), 222–245. https://doi.org/10.1177/1077801213476457

Goldberg, Y. (2017) *Neural Network Methods in Natural Language Processing.* Morgan and Claypool, Toronto

Heise, L., Ellsberg, M., & Gottmoeller, M. (2002). A global overview of gender-based violence. *International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics*, *78 Suppl 1*, S5–14. https://doi.org/10.1016/S0020–7292(02)00038–3

Heise, L., Greene, M. E., Opper, N., Stavropoulou, M., Harper, C., Nascimento, M., Zewdie, D., & Gender Equality, Norms, and Health Steering Committee. (2019). Gender inequality and restrictive gender norms: Framing the challenges to health. *Lancet (London, England)*, *393*(10189), 2440–2454. https://doi.org/10.1016/S0140–6736(19)30652-X

Howe, A. (1997). "The war against women". Media representations of men's violence against women in Australia. *Violence Against Women*, *3*(1), 59–75. https://doi.org/10.1177/1077801297003001005

Jewkes, R., Dartnall, E., & Sikweyiya, Y. (2017). *Ethical and safety recommendations for research on gender-based violence against women and girls*. World Health Organization.

LSE (2020). *Artificial intelligence could help protect victims of domestic violence*. London School of Economics and Political Science. https://www.lse.ac.uk/News/Latest-news-from-LSE/2020/b-Feb-20/Artificial-intelligence-could-help-protect-victims-of-domestic-violence.aspx

Luengo, M. (2018). Gender violence: The media, civil society, and the struggle for human rights in Argentina. *Media, Culture & Society*, *40*(3), 397–414. https://doi.org/10.1177/0163443717713259

Maydell, E. (2018). 'It just seemed like your normal domestic violence': Ethnic stereotypes in print media coverage of child abuse in New Zealand. *Media, Culture & Society*, *40*(5), 707–724. https://doi.org/10.1177/0163443717737610

Menéndez, M. I. (2014). Retos periodísticos ante la violencia de género. El caso de la prensa local en España [Journalistic challenges in the face of gender violence. The case of the local press in Spain]. *Nueva* Época, *22*, 53–77.

Mikolov, T. (2017). *Neural Networks for Natural Language Processing*. Brno University of Technology, Czechia.

Mills, T. C. (2015). Modelling Stationary Time Series: The ARMA Approach. In T. C. Mills (Ed.), *Time Series Econometrics: A Concise Introduction* (pp. 5–40). Palgrave Macmillan UK. https://doi.org/10.1057/9781137525338_2

Moorti, S. (2002). *The Colour of Rape: Gender and Race in Television's Public Spheres*. State University of New York Press.

Mori, K., & Haruno, M. (2021). Differential ability of network and natural language information on social media to predict interpersonal and mental health traits. *Journal of Personality*, *89*(2), 228–243. https://doi.org/10.1111/jopy.12578

News Desk (2019). '*Wolf Pack' verdict overruled: gang receive 15-year sentences.* https://www.spainenglish.com/2019/06/21/wolf-pack-verdict-15-year-sentences/

Rollè, L., Santoniccolo, F., D'Amico, D., & Trombetta, T. (2020). News Media Representation of Domestic Violence Victims and Perpetrators: Focus on Gender and Sexual Orientation in International Literature. In S. Ramon, M. Lloyd, & B. Penhale (Eds), *Gendered Domestic Violence and Abuse in Popular Culture* (pp. 149–169). Emerald Publishing Limited. https://doi.org/10.1108/978–1–83867–781–720201008

Russo, N. F., & Pirlott, A. (2006). Gender-based violence: Concepts, methods, and findings. *Annals of the New York Academy of Sciences*, *1087*, 178–205. https://doi.org/10.1196/annals.1385.024

Shen, J., Valagolam, D., & McCalla, S. (2020). Prophet forecasting model: A machine learning approach to predict the concentration of air pollutants (PM2.5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea. *PeerJ*, *8*, e9961. https://doi.org/10.7717/peerj.9961

Singh, G., Mémoli, F., & Carlsson, G. (2007). Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In *Eurographics Symposium on Point-Based Graphics 2007* (p. 100). https://doi.org/10.2312/SPBG/SPBG07/091–100

Subramani, S., Michalska, S., Wang, H., Du, J., Zhang, Y., & Shakeel, M. (2019). Deep Learning for Multi-Class Identification from Domestic Violence Online Posts. *IEEE Access*, *7*, 46210–46224. https://doi.org/10.1109/ACCESS.2019.2908827

Subramani, S., Wang, H., Vu, H. Q., & Li, G. (2018). Domestic Violence Crisis Identification from Facebook Posts Based on Deep Learning. *IEEE Access*, 6, 54075–54085. https://doi.org/10.1109/ACCESS.2018.2871446

Sutherland, G., Easteal, P., Holland, K., & Vaughan, C. (2019). Mediated representations of violence against women in the mainstream news in Australia. *BMC Public Health*, *19*(1), 502. https://doi.org/10.1186/s12889–019–6793–2

Teruelo, J. G. F. (2011). Feminicidios de género: Evolución real del fenómeno, el suicidio del agresor y la incidencia del tratamiento mediático [Gender-based femicides: Actual evolution of the phenomenon, the suicide of the aggressor and the incidence of media coverage]. *Revista Española de Investigación Criminológica: REIC*, *9*, 1.

Tranchese, A., & Zollo, S. A. (2013). The Construction of Gender-based Violence in the British Printed and Broadcast Media. *Critical Approaches to Discourse Analysis across Disciplines*, *7*(1), 141–163.

Tsay, R. S. (1989) Identifying Multivariate Time Series Models *Journal of Time Series Analysis—Wiley Online Library* https://onlinelibrary.wiley.com/doi/10.1111/j.1467–9892.1989.tb00034.x

UN Women (2012) *Handbook for National Action Plans on Violence against Women*. UN Women – Headquarters. https://www.unwomen.org/en/digital-library/publications/2012/7/handbook-for-national-action-plans-on-violence-against-women

UN Women (2017) *Media coverage of gender-based violence—Handbook and Training of Trainers*. https://eca.unwomen.org/en/digital-library/publications/2017/09/media-coverage-of-gender-based-violence---handbook-and-training-of-trainers

UN Women—ONU Mujeres (2021). *Violence against women prevalence estimates, 2018. Global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women* https://www.unwomen.org/es/what-we-do/ending-violence-against-women/facts-and-figures

UNESCO (2019) *Reporting on violence against women and girls: A handbook for journalists"* UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000371524

Veen, H. J. van, Saul, N., Eargle, D., & Mangham, S. W. (2019). Kepler Mapper: A flexible Python implementation of the Mapper algorithm. *Journal of Open Source Software*, *4*(42), 1315. https://doi.org/10.21105/joss.01315

Violenciagenero (2019). *Government Delegation against Gender Violence* http://www.violenciagenero.igualdad.gob.es/laDelegacionInforma/home.htm

WHO (2013)| Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. World Health Organization. https://apps.who.int/iris/handle/10665/85239

Wolf, B. (2018). *Gender-based violence in discourse. A comparative study on anti-violence communication initiatives across Europe, in Austria and Spain*. https://doi.org/10.5565/rev/

Wong, J. S., & Lee, C. (2018). Extra! Extra! The Importance of Victim-Offender Relationship in Homicide Newsworthiness. *Journal of Interpersonal Violence*, 886260518789142. https://doi.org/10.1177/0886260518789142

Xue, J., Chen, J., & Gelles, R. (2019). Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter. *Violence and Gender*, *6*(2), 105–114. https://doi.org/10.1089/vio.2017.0066